electronics

Multi-Sensory Interaction for Blind and Visually Impaired People

Edited by

Jun Dong Cho

Printed Edition of the Special Issue Published in *Electronics*

www.mdpi.com/journal/electronics

MDPI

Multi-Sensory Interaction for Blind and Visually Impaired People

Multi-Sensory Interaction for Blind and Visually Impaired People

Editor

Jun Dong Cho

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editor
Jun Dong Cho
Department of Electronic and Electrical Engineering,
Sungkyunkwan University
Korea

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Electronics* (ISSN 2079-9292) (available at: https://www.mdpi.com/journal/electronics/special_issues/Human_Centered_Intelligence).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-3058-1 (Hbk)
ISBN 978-3-0365-3059-8 (PDF)

Contents

About the Editor

Jun Dong Cho was born in Seoul, Korea, in 1957. He received his B.S. degree in Electronic Engineering, Sungkyunkwan Univ., in Seoul, Korea, in 1980, an M.S. degree from Polytechnic University (NYU-Poly), Brooklyn, NY, in 1989, and a Ph.D. degree from Northwestern University, Evanston, IL, in 1993, both in computer science. He was a CAD Engineer at Samsung Electronics, Co., Ltd. (1983–1987,1993–1995). He received the Best Paper Award at the 1993 Design Automation Conference. He has been an IEEE Senior member since 1996. He held visiting scholarships at IBM T.J. Watson Research Center (NY, USA) and received the Invention Achievement Award, IBM, USA (2001). In September 2013, he founded "H-Lab." and the department of Human ICT (Information and Cognition Technology) Convergence (humanict.skku.edu) and served as a department chair. He is now in several faculty positions at Sungkyunkwan University: Professor of the Department of Electronic and Electrical Engineering (Director of H-Lab. http://humanict.skku.edu), Adj. Professor at the Department of Digital Healthcare, Samsung Advanced Institute for Health Sciences & Technology, and Adj. Professor at the Department of Sports Interaction Science, College of Sports. He published about 400 papers in the area of embedded digital design automation, human–computer interaction, and user experience of smart health care devices. He received the Korean Minister of Knowledge Economy Award for Improvement of semiconductor design performance for ICT and healthcare (2012); Korean Minister of Science, ICT and Future Planning Award for Best Convergence Research Best Practice (2014); Korean Minister of Trade, Industry and Energy Award—Contribution to industrial development through creative industry convergence (2017). His recent works (blindtouch.org) for helping people with visual impairment to appreciate visual artworks were published online in the Special Issue in 'Multi-sensory Interaction', Electronics. His current research topics are Embodied, Playful, Tangible, Wearable Interaction.

Preface to "Multi-Sensory Interaction for Blind and Visually Impaired People"

Multi-sensory integration is an essential part of information processing, by which various forms of sensory information such as sight, hearing, and touch are combined into a single experience. Information is typically integrated across sensory modalities when the sensory inputs share certain common features. Though many studies introduce the use of other modalities of sensation, such as haptic, sound, and scent, for the user interface to act as a supplement for the absence of vision, they are still not close to what vision is to the people. Contemporary art has also been influenced by this trend, and the number of artists interested in creating novel multi-sensory works of art has increased substantially. As a result, the opportunities for visually impaired people to experience artworks in different ways are also expanding. Despite all of this, the research focusing on multimodal systems for experiencing visual arts is not extensive, and user tests comparing different modalities and senses, particularly in the field of art, are insufficient.

"Starry Night" is a masterpiece by Vincent van Gogh, the master of post-Impressionist art, depicting a beautiful night scene. Van Gogh's night sky, drawn by tracing memories of the past, is filled with 11 large and small stars, adding to the sense of mystery. We can meet a slightly different "Starry Night". It is a three-dimensional picture with embossing added using 3D printing. When touched, a sound effect is produced. Small stars make small sounds and big stars make loud sounds, completing a rich three-dimensional effect. It has been implemented so that even the visually impaired can enjoy "Starry Night" vividly. It has a different meaning because it allows the general public to experience a multifaceted art experience that allows them to appreciate works using their senses of sight, touch, smell, and hearing.

We often see color as a concept. Color is expressed through various sensory organs such as scent, sound, temperature, and touch. This book will be dedicated to introducing a set of assistive Internet of Things (IoT) tools with multisensory information to help visually impaired people appreciate works of art. These tools are used to represent visual elements (form, distance, location, and color) by applying touch, sound, and smell. These research results are convergence studies of art, science, technology, and the humanities that include various academic fields such as human–computer interface technology, electronic engineering, media art, and sound design technology.

This book conveyed the visual elements of artwork to the visually impaired through various sensory elements to open a new perspective for appreciating visual artwork. In addition, the technique of expressing a color code by integrating patterns, temperatures, scents, music, and vibrations was explored, and future research topics were presented. A holistic experience using multi-sensory interaction acquired by people with visual impairment was provided to convey the meaning and contents of the work through rich multi-sensory appreciation. A method that allows people with visual impairments to engage in artwork using a variety of senses, including touch, temperature, tactile pattern, and sound, helps them to appreciate artwork on a deeper level than can be achieved with hearing or touch alone. The development of such art appreciation aids for the visually impaired will ultimately improve their cultural enjoyment and strengthen their access to culture and the arts. The development of this new concept ultimately expands opportunities for the non-visually impaired as well as the visually impaired to enjoy works of art and breaks down the boundaries between the disabled and the non-disabled in the field of culture and arts through continuous efforts to enhance accessibility. In addition, the developed multi-sensory expression and delivery tool can be used as an educational tool to increase product and artwork accessibility and usability through multi-modal interaction. Training the multi-sensory experiences introduced in this book may lead to more vivid visual imageries or seeing with the mind's eye.

First of all, I would also like to express my sincere gratitude to the authors who, through their serious and noble research, have made this book bright. Additionally, I would like to express my sincere thanks to the editors and reviewers who reviewed the papers. Without their precious time and effort, this book would not have existed.

Jun Dong Cho
Editor

electronics

Editorial

Multi-Sensory Interaction for Blind and Visually Impaired People

Jun-Dong Cho

Department of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon 16419, Korea; jdcho@skku.edu

Citation: Cho, J.-D. Multi-Sensory Interaction for Blind and Visually Impaired People. *Electronics* **2021**, *10*, 3170. https://doi.org/10.3390/electronics10243170

Received: 15 December 2021
Accepted: 17 December 2021
Published: 20 December 2021

1. Introduction

Multi-sensory interaction aids learning, inclusion, and collaboration because it accommodates the diverse cognitive and perceptual needs. Multi-sensory integration is an essential part of information processing, by which various forms of sensory information such as sight, hearing, touch, and proprioception (also called kinesthesia, the sense of self-movement and body position) are combined into a single experience. Information is typically integrated across sensory modalities when the sensory inputs share certain common features. Cross-modality refers to the interaction between two different sensory channels. Cross-modal correspondence is defined as the surprising associations that people experience between seemingly unrelated features, attributes, or dimensions of experience in different sensory modalities. For visually impaired people, conventional human–computer interaction devices are inconvenient, as these devices rely heavily on visual information. Though many studies introduce the use of other modalities of sensation like haptic, sound, and scent for the user interface to act as a supplement for the absence of vision, they are still not close to what vision is to the people. The topics of interest include but are not limited to the following:

1. Universal access in human–computer interaction;
2. Haptic interfaces for accessibility;
3. Tactile artworks and interactions;
4. Flexible haptic displays;
5. Ambient assistive intelligence;
6. Human-centered user accessibility for people with visual impairments;
7. Assistive technology;
8. Multi-sensory color coding.

This Special Issue contains 10 research papers [1–10] and two review papers [11,12].

2. Research Papers

The development of assistive technologies for art appreciation for visually impaired people can enhance their cultural and perceptual appreciation. These opportunities result in better comprehension and accessibility at museums and exhibitions and in everyday life. These multi-sensory interactions may also offer enhanced usability and understanding and promote educational tools aiding synesthetic capabilities to promote creative thinking. Color associations with aspects such as symbolism, culture, and preferences play an influential role, demanding the promotion of color comprehension in the daily lives of virtually impaired people as well. Despite the availability of tactile graphics and audio guides, the visually impaired still face challenges in experiencing and understanding visual artworks.

Visually impaired people can take advantage of multimodal systems in which visual information is communicated through different modes of interaction and types of feedback. Among the possible interaction modes, thermal interaction in the context of assistive devices for visually impaired people lacks research despite its potential.

Bartolomé et al. [1] proposed a temperature depth mapping algorithm and a thermal display system to convey the depth and color depth of artworks' features in the context of tactile exploration by visually impaired people. Tests were performed both during the mapping algorithm's design and after developing a tactile temperature prototype artwork model to assess the potentials of thermal interaction for recognizing depth and color depth in tactile art appreciation. These tests showed both an existing correlation between depth and temperature and that the mapping based on that correlation was appropriate for conveying depth during tactile artwork exploration. Throughout this work, two types of tests were performed to assess a method to convey depth information by using temperature cues. The first test showed that warm and cold temperatures could be used as cues to communicate to the user how near or far the features of an artwork are. Based on these results, a complete thermal display prototype was designed and developed. Similarly, a relief artwork was designed and installed on top of the prototype, which was used for performing the final test. This final test's results proved that thermal interaction is a proper way of conveying the depth information of the artwork to visually impaired people. This is an addition to the current technologies which, as demonstrated above, used to communicate the depth of the features of an artwork either by using audio or by adding depth into a tactile model by extruding the features. The addition of thermal interaction as a way of communicating depth can open the door to many new ways of experiencing art for visually impaired people. Moreover, the developed thermal display system can be used for adding thermal interaction to any type of paper-based relief artwork not only by using the thermal cues as a substitute for depth, but also by giving them another role, such as expressing color warmness and color coolness or for making hot objects (such as the sun) warm and cold objects (such as water) cold.

There are many ways in which this work can be continued or improved on, such as the following:

(1) Increasing the number of peltiers, adding the possibility of creating more complex temperature regions on the artwork;
(2) Finding a way to make the system smaller and more portable;
(3) Changing the use of temperature cues from depth representation to an artwork featuring semantic mapping, such as making the water feel cold. A necessary addition for this semantic mapping would be to be able to make the prototype work with 2.5D relief artworks, which present depth by extruding the features in the z direction and not only with tactile paper artworks. In this way, visually impaired people could be aware of depth through tactile exploration while also feeling the temperature of the different artwork features while exploring. For that, a method to make the peltier temperature reach all the way to the surface of the 2.5D relief model should be found.

The recent development of color coding in tactile pictograms helps visually impaired people to appreciate the visual arts. The auditory sense, in conjunction with (or possibly as an alternative to) the tactile sense, would allow people with visual impairments to perceive colors in a way that would be difficult to achieve with just a tactile stimulus. Sound color coding [2] can replicate three characteristics of colors (i.e., hue, chroma, and value) by matching them with three characteristics of sound (i.e., timbre, intensity, and pitch). This paper examines the relationships between sound (melody) and color and provides color coding with sound for the hue, chroma, and value for deepening their relationship with visual art. Two sets of proposed methods for coding colors with sound use melodies to improve upon the current method in use by adding more colors (18 colors in 6 hues). User experience and identification tests were conducted with 12 visually impaired and 8 sighted adults, and the results suggest that the sound color coding was helpful for the participants.

Despite the use of tactile graphics and audio guides, blind and visually impaired people still face challenges in experiencing and understanding visual artworks independently at art exhibitions. Art museums and other art places are increasingly exploring the use of interactive guides to make their collections more accessible. In [3], Quero et al. presented an interactive multimodal guide prototype that used audio and tactile modalities to improve the autonomous access to information and experience of visual artworks. The prototype was composed of a touch-sensitive 2.5D artwork relief model that could be freely explored by touch. Users could access localized verbal descriptions and audio by performing touch gestures on the surface while listening along to themed background music. Eighteen participants evaluated and compared the multimodal and tactile graphic accessible exhibits. The results from a usability survey indicated that the presented multimodal approach was simple, easy to use, and improved confidence and independence when exploring visual artworks.

Feedback collected during the multiple exhibition points in new directions for this work. The presented interactive multimodal guide is sometimes used as a collaboration tool to socially interact with art. Moreover, the presented prototype was designed for use in an exhibition environment. Art educators at schools have expressed their interest in using the guide as an educational tool in class. The current prototypes only make use of tactile and audio modalities. As a future work, new experiences with other modalities such as smell need to be explored for how they might improve visual artwork exploration.

Tactile perception enables people with visual impairments to engage with artworks and real-life objects at a deeper abstraction level. The development of tactile and multisensory assistive technologies has expanded their opportunities to appreciate visual arts. Tactile color pictograms [4] using tactile sensing attain sensational conception along with the other physical properties of artwork, such as contour, size, texture, geometry, and orientation. Although several tactile color patterns [4] have been developed, they are limited to fixed tactile color interpretation, which requires outgoing resources. Jabbar et al. [5] have proposed the design for ColorWatch, integrating colors from Goethe's color triangle and the Munsell color system with an analog wristwatch, allowing spatial color-to-tactile interpretation. They developed a tactile interface based on the proposed concept design under the considerations of people with visual impairments with tactile actuation, color perception, and learnability. The proposed interface automatically translates reference colors into spatial tactile patterns like a timepiece wristwatch. A range of achromatic colors and six prominent basic colors with three levels of chroma and values are considered for the cross-modular association. In addition, a simplified and affordable tactile color watch design has been proposed. This scheme enables people with visual impairments to explore artwork or real-life object colors by identifying the reference colors through a color sensor and translating them to the tactile interface. They have associated achromatic and monochromatic colors with chroma and value levels to the cross-modular tactile interface. The tactile interface manifests angular positions of tactile patterns. These patterns can be transformed automatically, corresponding to the reference color. The arrangement of the tactile pattern is based on intuitive learning, which is translated through the analog wristwatch's tactile interface. This integrated approach offers ease of learnability to provide the essence of particularly emotional or psychological states. The color identification tests using this scheme on the developed prototype exhibit good recognition accuracy. The workload assessment and usability evaluation for people with visual impairments demonstrated promising results. Usability tests based on a system usability scale and workload assessment by NASA-TLX tests suggest that the proposed ColorWatch system can help people with visual impairments in color identification and reduce a factor that hinders their museums' accessibility and real-life color perception. The function of ColorWatch may be expanded to represent a color gamut of 42 colors with 12 color hues, based on the RYK color wheel originally described by Issac Newton. The 6 additional color hues can be represented at uniform 30° angular distances, alternating between existing chromatic color hues.

Playing board games is important for people with a visual impairments, as it promotes interactive socialization and communication skills. However, some board games are not accessible to them at present. In [6], Miyakawa et al. proposed an auditory card game system that presents a card's contents with auditory stimuli to all players, geared toward playing equally with others regardless of whether they have a visual impairment or not, as one of the solutions to make board games accessible. This proposal contributes significantly to expanding the range of inclusive board games for the visually impaired. The purpose of this paper is to determine whether the game allows for fair competition for people with visual impairments and to clarify the effects of the valuable parameters of the system on the players. The effectiveness of the proposed system was verified by having the experiment's participants play "Auditory *Uta-Karuta*". The results suggest that the proposed system has the potential for an accessible board game design regardless of visual impairment. In the following experiment, the impact of each valuable parameter of the system on the player's perception of the board games is investigated to clarify the appropriate audio cue design method. The results of this experiment will greatly assist in designing an appropriate board game using the proposed system.

Although this proposal cannot improve the accessibility of all board games, it contributes to expanding the range of inclusive board games for the visually impaired. Furthermore, clarifying the impact of the element of the system on the players will greatly assist in designing an appropriate board game using the proposed system.

Contemporary art is evolving beyond simply looking at works, and the development of various sensory technologies has had a great influence on culture and art. Accordingly, opportunities for the visually impaired to appreciate visual artworks through various senses such as auditory and tactile senses are expanding. However, insufficient sound expression and a lack of portability make it less understandable and accessible. Lee et al. [7] attempted to convey a color and depth coding scheme for the visually impaired based on alternative sensory modalities, such as hearing (by encoding the color and depth information with 3D sounds for audio description) and touch (to be used for interface-triggering information such as color and depth). The proposed color coding scheme represents light, saturated, and dark colors for red, orange, yellow, yellow-green, green, blue-green, blue, and purple. The paper's proposed system can be used for both mobile platforms and 2.5D (relief) models. They presented a methodology of 3D sound color coding using HRTF. The color hue is represented by the sound simulation of the position of the color wheel, and the lightness of the color is reflected by using the pitch of the sound. The correlation between a sound's loudness and depth was found through experiments on the correlation between the sound variables and depth, and the correlation was used to represent depth by changing the sound's loudness and increasing the reverberation in addition to the original sound codes. Additionally, an identification test and system usability test were conducted in this study. A total of 97.88% of the identification test results showed that the system had excellent recognition. The results of the NASA TLX test and user experience test also showed the good usability of the system. Experiments with visually impaired subjects will be implemented in future studies. This is a new attempt to express color. Although there are many ways to use sound to express color, there are few ways to use changes in a sound's position to express color accurately. The variable of sound position is very common and familiar to the visually impaired. The use of this method also opens a new direction in the way that art can be experienced by the visually impaired. However, there is still room for improvement in this method. Further refinements will increase the accuracy and usability. Future improvements in sound processing will also make recognition easier. Neither sighted people nor people with visual impairments had experienced the proposed 3D sound coding colors before. Therefore, it was judged that there were no significant differences in the perception ratings between the sighted and visually impaired test people. However, future extended testing will be necessary to analyze the differences in the speed of perception between those two groups.

Visually impaired visitors experience many limitations when visiting museum exhibits, such as a lack of cognitive and sensory access to exhibits or replicas. Contemporary art is evolving in the direction of appreciation beyond simply looking at works, and the development of various sensory technologies has had a great influence on culture and art. Thus, opportunities for people with visual impairments to appreciate visual artworks through various senses such as hearing, touch, and smell are expanding. However, it is uncommon to provide a multi-sensory interactive interface for color recognition, such as integrating patterns, sounds, temperatures, and scents. Cho at al. [8] attempted to convey color cognition to the visually impaired, taking advantage of multi-sensory color coding. In previous works, musical melodies with different combinations of pitch, timbre, velocity, and tempo were used to distinguish vivid (i.e., saturated), light, and dark colors. However, it was rather difficult to distinguish among warm, cool, light, and dark colors using only sound cues. The method of using poems together when appreciating works has the advantage of enhancing the expressiveness of one work. Although several researchers have previously performed art–poetry matching, to the best of the author's knowledge, there is no art–poetry matching for conveying different color dimensions such as warmth, coolness, lightness, and darkness. Moreover, there are no previous works that suggest a method of providing poems suitable for the various senses (sight, touch, hearing, and smell) of the artwork. These motivated the authors to develop a systematic algorithm to automate the generation of poetry which could be applied consistently to artworks, especially to help visually impaired users perceive colors in the artworks. Therefore, this paper aims to build a multi-sensory color coding system by combining sound and poem such that the poem leads to representing more color dimensions, such as including warm and cool colors for red, orange, yellow, green, blue, and purple. To this end, an implicit association test was performed to identify the most suitable poem among the candidate poems to represent colors in artwork by finding the common semantic directivity between the given candidate poem with voice modulation and the artwork in terms of light, dark, and warm color dimensions. Finally, a system usability test confirmed that the poem will be an effective supplement for distinguishing between vivid, light, and dark colors with different color appearance dimensions, such as warm and cold colors.

Kim et al. [9] attempted to improve the user experience when appreciating visual artworks with soundscape music chosen by a deep neural network based on weakly supervised learning. The baseline was improved by different perspectives, such as modeling methods, learning methods, and domain adaptation methods. They also proposed a multi-faceted approach to measuring ambiguous concepts, such as the subjective fitness, implicit senses, immersion, and availability. They showed improvements in the appreciation experience, such as the metaphorical and psychological transferability, time distortion, and cognitive absorption, with in-depth experiments involving 70 participants. The proposed method can also help spread soundscape-based media art by supporting traditional soundscape design. Furthermore, the proposed method will help people with visual impairments to appreciate artworks through its application to a multi-modal media art guide platform. However, their research suffered from three major limitations. First, this was not state-of-the-art performance from the perspective of the deep neural network. In particular, models with low accuracy from the feature extraction perspective were selected for mutual learning. Therefore, the development of models with improved audio feature representation is required. Second, the author's current database comprised only 2000 music items. In addition, these 2000 music pieces were limited from the genre perspective. A wide spectrum of databases containing music from various cultures and times should be used in future studies. Third, they conducted experiments using 70 individuals, but each experiment was conducted on a group of only 10 people. Therefore, the author's experimental results cannot be generalized. Large-scale experiments need to be conducted to generalize these experimental results.

For years, the HCI community's research has been focused on the hearing and sight senses. However, in recent times, there has been an increased interest in using other types of senses, such as smell or touch. Moreover, this has been accompanied by growing research related to sensory substitution techniques and multi-sensory systems. Similarly, contemporary art has also been influenced by this trend, and the number of artists interested in creating novel multi-sensory works of art has increased substantially. As a result, the opportunities for visually impaired people to experience artworks in different ways are also expanding. Despite all of this, the research focusing on multimodal systems for experiencing visual arts is not extensive, and user tests comparing different modalities and senses, particularly in the field of art, are insufficient. Baltimore et al. [10] attempted to design multi-sensory mapping to convey color to visually impaired people by employing musical sounds and temperature cues. Through user tests and surveys with a total of 18 participants, a multi-sensory system was properly designed to allow the user to distinguish and experience a total of 24 colors (6 hues and 4 different color dimensions of dark, bright, warm, and cold). The authors attempted to find out the best way to mix previous unisensory temperature colors with sound color codes to reach a satisfactory multi-sensory cross-modal code. In addition, a semantic study of the musical sound cues and temperatures from those methods were acquired. The tests consisted of several semantic correlational adjective-based surveys for comparing the different modalities to find out the best way to express colors through musical sounds and temperature cues based on previously well-established sound color and temperature color coding algorithms. In addition, the resulting final algorithm was also tested with 12 more users. The results showed that the musical sounds and temperatures could be used as a substitute for color hues and color dimensions. Additionally, the data from the results guided us in designing the optimum multi-sensory temperature sound method based on those musical sounds and temperatures. This work can encourage researchers to consider thermal and sound multi-sensory interaction both as a substitute for color and to improve accessibility for visually impaired people in visual artworks and color experience.

The development of assistive technologies is improving the independent access of blind and visually impaired people to visual artworks through non-visual channels. Current single-modality tactile and auditory approaches to communicate color contents must compromise between conveying a broad color palette, ease of learning, and suffer from limited expressiveness. Quero [11] proposed a multi-sensory color code system that uses sounds and scents to represent up to 30 colors. It uses melodies to express each color's hue and scents to express the saturation, lightness (light or dark), and temperature (warm or cool) color dimensions.

In collaboration with 18 participants, the color identification rate was evaluated when using the multi-sensory approach. Seven (39%) of the participants improved their identification rate, five (28%) remained the same, and six (33%) performed worse when compared with an audio-only color code alternative. The multi-sensory color code system improved the convenience and confidence of the participants. The scent selection and pairing were made through a semantic correspondence survey in collaboration with 18 participants. The color code system was evaluated to determine if using a multi-sensory approach eased the effort used to recognize the encoded colors and help improve the color identification compared with the commonly used unisensory method. The results from the evaluation suggest that the multi-sensory approach did improve color identification, but not for everyone. The cause of this seemed to be the extra cognitive effort and sensory overload experienced by some. In addition, the authors integrated the color code into a sensory substitution device prototype to determine if the color code could be more suitable and expressive when exploring visual art color content compared with a tactile graphics alternative. The results from this evaluation indicate that the multi-sensory-based prototype is more convenient and improves the confidence for visual artwork color content exploration. Moreover, the results also suggest suitability for artwork exploration, since the multi-sensory stimuli improve the experience and reaction to the artwork. In the future, the authors would like to expand the color code to include more color audio–scent pairs to study the applicability across different styles of visual artworks. In this work, two non-visual sensory reproductions of artworks were used to evaluate the proposed color coding system. As for future work, experiments can be carried out using more diverse styles of non-visual sensory reproductions to further support the results proposed in this paper. In addition, it is relevant to explore the effect of the semantic incoherence that could happen due to color code usage and its influence on the artwork experience and interpretation. While experiencing color is just a fraction of the art appreciation process, this work might contribute toward designing and studying accessible art appreciation frameworks for all.

3. Review Papers

Visually impaired visitors experience many limitations when visiting museum exhibits, such as a lack of cognitive and sensory access to exhibits or replicas. Contemporary art is evolving in the direction of appreciation beyond simply looking at works, and the development of various sensory technologies has had a great influence on culture and art. Thus, opportunities for people with visual impairments to appreciate visual artworks through various senses such as hearing, touch, and smell are expanding. However, it is uncommon to provide an interactive interface for color recognition, such as applying patterns, sounds, temperatures, or scents. Cho [12] conveyed the visual elements of the work to the visually impaired through various sensory elements. In addition, to open a new perspective on appreciation of the works, the technique of expressing a color coded by integrating patterns, temperatures, scents, music, and vibrations was explored, and future research topics were presented.

In this review, a holistic experience using synesthesia acquired by people with visual impairment was provided to convey the meaning and contents of the work through rich multi-sensory appreciation. In addition, pictograms, temperatures, scents, music, and new forms incorporating them were explored to find a new way of conveying colors in artworks to the visually impaired. A method that allows people with visual impairments to engage in artwork using a variety of senses, including touch and sound, helps them to appreciate artwork at a deeper level than can be achieved with hearing or touch alone. The development of such art appreciation aids for the visually impaired will ultimately improve their cultural enjoyment and strengthen their access to culture and the arts. The development of this new concept of post-visual art appreciation aids ultimately expands opportunities for the non-visually impaired as well as the visually impaired to enjoy works of art at the level of weak synesthesia and breaks down the boundaries between the disabled and the non-disabled in the field of culture and arts through continuous efforts to enhance accessibility. In addition, the developed multi-sensory expression and delivery tool can be used as an educational tool to increase product and artwork accessibility and usability through multi-modal interaction. Training the multi-sensory experiences introduced in this paper may lead to more vivid visual imageries or seeing with the mind's eye.

Several studies have been conducted to improve the accessibility of images using touchscreen devices for screen reader users. Oh et al. [13] conducted a systematic review of 33 papers to get a holistic understanding of existing approaches and to suggest a research road map given identified gaps. As a result, the authors identified the types of images, visual information, input device, and feedback modalities that were studied for improving image accessibility using touchscreen devices. The findings also revealed that there is little research on how the generation of image-related information can be automated. Moreover, the involvement of screen reader users is mostly limited to evaluations, while input from target users during the design process is particularly important for the development of assistive technologies. Then, two recent studies on the accessibility of artwork and comics (AccessArt and AccessComics) are presented. Based on the identified key challenges, the authors suggested a research agenda for improving image accessibility for screen reader users. To have complete understanding of the existing approaches and identify challenges to be solved as the next step, a systematic review of 33 papers was conducted on touchscreen-based image accessibility for screen reader users. The results revealed that image types other than maps, graphs, and geometric shapes such as artwork and comics were rarely studied. Furthermore, the authors found that only about one third of the papers provided audio and haptic multi-modal feedback. Moreover, ways to collect image descriptions were out of the scope of interest for most studies, suggesting that automatic retrievals of image-related information are one of the bottlenecks for making images accessible on a large scale. Finally, while most previous studies did not involve people who were blind or had low vision during the system design process, future studies should consider inviting target users early in advance and reflecting on their comments when making design decisions.

4. Conclusions and Future Works

Synesthesia appears in all forms of art and provides a multisensory form of knowledge and communication. It is not subordinated but can expand the aesthetic through science and technology. Science and technology could thus function as a true multidisciplinary fusion project that expands the practical possibilities of theory through art. Synesthesia is divided into strong synesthesia and weak synesthesia. Strong synesthesia is characterized by a vivid image in one sensory modality in response to the stimulation of another sense. Weak synesthesia, on the other hand, is characterized by cross-sensory correspondences expressed through language or by perceptual similarities or interactions. Weak synesthesia is common, easily identified, remembered, and can be manifested by learning. Therefore, weak synesthesia could be a new educational method using multisensory techniques. Synesthetic experience is the result of a unified sense of mind; therefore, all experiences are synesthetic to some extent. The most prevalent form of synesthesia is the conversion of sound into color. In art, synesthesia and metaphor are combined. Through art, the co-sensory experience became communicative. The origin of the co-sensory experience can be also found in painting, poetry, and music (visual, literary, and musical). To some extent, all forms of art are co-sensory. The core of an artwork is its spirit, but grasping that spirit requires a medium which can be perceived not only by the one sense intended but also through various senses. In other words, the human brain creates an image by integrating multiple nonvisual senses and using a matching process with previously stored images to find and store new things through association. So-called intuition thus appears mostly in synesthesia. To understand reality as much as possible, it is necessary to experience reality in as many forms as possible. Thus, synesthesia offers a richer reality experience than the separate senses, and that can generate unusually strong memories. An intensive review on multi-sensory experiences and color recognition in visual arts appreciation in persons with visual impairment and promising future works can be found in Cho's work [12].

Funding: This research was funded by the Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea, grant number 2018M3C1B6061353.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Iranzo Bartolomé, J.; Cho, J.-D.; Cavazos Quero, L.; Jo, S.; Cho, G. Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People. *Electronics* **2020**, *9*, 1939. [CrossRef]
2. Cho, J.-D.; Jeong, J.; Kim, J.H.; Lee, H. Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments. *Electronics* **2020**, *9*, 1981. [CrossRef]
3. Cavazos Quero, L.; Iranzo Bartolomé, J.; Cho, J. Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics. *Electronics* **2021**, *10*, 297. [CrossRef]
4. Cho, J.D.; Quero, L.C.; Bartolomé, J.I.; Lee, D.W.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2021**, *46*, 103–116. [CrossRef]
5. Jabbar, M.S.; Lee, C.-H.; Cho, J.-D. ColorWatch: Color Perceptual Spatial Tactile Interface for People with Visual Impairments. *Electronics* **2021**, *10*, 596. [CrossRef]
6. Miyakawa, H.; Kuratomo, N.; Salih, H.E.B.; Zempo, K. Auditory *Uta-Karuta*: Development and Evaluation of an Accessible Card Game System Using Audible Cards for the Visually Impaired. *Electronics* **2021**, *10*, 750. [CrossRef]
7. Lee, Y.; Lee, C.-H.; Cho, J.-D. 3D Sound Coding Color for the Visually Impaired. *Electronics* **2021**, *10*, 1037. [CrossRef]
8. Cho, J.-D.; Lee, Y. ColorPoetry: Multi-Sensory Experience of Color with Poetry in Visual Arts Appreciation of Persons with Visual Impairment. *Electronics* **2021**, *10*, 1064. [CrossRef]
9. Kim, Y.; Jeong, H.; Cho, J.-D.; Shin, J. Construction of a Soundscape-Based Media Art Exhibition to Improve User Appreciation Experience by Using Deep Neural Networks. *Electronics* **2021**, *10*, 1170. [CrossRef]
10. Bartolome, J.I.; Cho, G.; Cho, J.-D. Multi-Sensory Color Expression with Sound and Temperature in Visual Arts Appreciation for People with Visual Impairment. *Electronics* **2021**, *10*, 1336. [CrossRef]

11. Cavazos Quero, L.; Lee, C.-H.; Cho, J.-D. Multi-Sensory Color Code Based on Sound and Scent for Visual Art Appreciation. *Electronics* **2021**, *10*, 1696. [CrossRef]
12. Cho, J.-D. A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment. *Electronics* **2021**, *10*, 470. [CrossRef]
13. Oh, U.; Joh, H.; Lee, Y. Image Accessibility for Screen Reader Users: A Systematic Review and a Road Map. *Electronics* **2021**, *10*, 953. [CrossRef]

electronics

MDPI

Article

Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People

Jorge Iranzo Bartolomé [1,*], Jun Dong Cho [2], Luis Cavazos Quero [1], Sunggi Jo [2] and Gilsang Cho [1]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Korea; luis@skku.edu (L.C.Q.); Jus5041@skku.edu (G.C.)

[2] Department of Human ICT Convergence, Sungkyunkwan University, Suwon 16419, Korea; jdcho@skku.edu (J.D.C.); whtjdrl410@skku.edu (S.J.)

* Correspondence: jorgedavid@skku.edu

Received: 6 October 2020; Accepted: 11 November 2020; Published: 17 November 2020

Abstract: Visually impaired people can take advantage of multimodal systems in which visual information is communicated through different modes of interaction and types of feedback. Among the possible interaction modes, thermal interaction in the context of assistive devices for visually impaired people lacks research in spite of its potential. In this paper, we propose a temperature-depth mapping algorithm and a thermal display system to convey depth and depth-color of artworks' features in the context of tactile exploration by visually impaired people. Tests with a total of 18 sighted users and six visually impaired users were performed both during the mapping algorithm design and after developing a tactile temperature prototype artwork model to assess the potentials of thermal interaction for recognizing depth and color-depth in tactile art appreciation. These tests showed both an existing correlation between depth and temperature and that the mapping based on that correlation is appropriate for conveying depth during artwork tactile exploration.

Keywords: visually impaired people; accessibility; art appreciation; color; temperature-depth coding; thermal interaction

1. Introduction

1.1. Introduction to Our Work

Multimodality, or combining several modes in order to communicate information, is an important technique in the field of HCI (Human Computer Interaction). Since each type of interface and interaction has its own strengths and weaknesses, the combination of many modes of communication results in a more efficient user-machine communication compared to the one accomplished through a system that is based on only one type of user interaction. As a result, researchers have extensively studied different types of interaction and sensing modalities that could be used by combining several of them together in assistive technology solutions for the VIP (Visually Impaired People) [1]. However, even though research related to cross-modal associations based on haptic interfaces exists, research about thermal cues as a possible sensing modality for assistive technologies for VIP is quite small, especially when compared to other haptic interfaces such as vibrotactile actuators.

In this work, the correlation between temperature and depth was studied and evaluated with real users in order to find out a proper mapping between temperature and depth. After that, the algorithm was applied to two different contexts: for conveying depth and for conveying color-based depth (effect known as chromostereopsis [2]). Also, a physical prototype was designed, implemented, and evaluated always in the context of tactile artwork exploration for VIP. The system contains an array of petlier devices in which the artwork can be installed. The user can explore the artwork with the hands while feeling the temperature of the different features of the artwork as a way of conveying the depth levels. All information related to the prototype and the mapping algorithm will be given in the corresponding sections.

This work is the continuation of a series of works whose main goal is the development of a multi-modal tactile artwork system that can help blind people appreciate bi-dimensional pieces of art. The main system, presented in [3,4], consisted of a 3D-printed *2.5D* relief replica of an artwork installed on top of the multimodal guide system. In those works, the *2.5D* relief model is a replica of the bi-dimensional artwork but with a z dimensional depth added to the different features. The user's finger positions were recognized in real time through the use of conductive paint and real-time audio feedback was triggered by double or triple tapping on the different features of the artwork. Double tapping activated explanatory audio feedback and triple tapping activated a sound effect feedback related to what is touched. A user voice interface was also added to facilitate the control of the system by giving the users the chance of using their voice as input. Feedback from the visually impaired users during the tests encouraged us to add thermal interaction to the system, first for representing colors, as can be seen in [5]. There, a double petlier device was used to communicate a total of 54 different colors to the visually impaired user through temperature cues. Lastly, exploring the possibilities of thermal interaction in the context of artwork tactile exploration led us to the current work, in which temperature cues were used as a way of conveying depth levels of the different features of the artwork.

To sum up, in this work thermal interaction, a way of mapping it to depth, and the applications of that thermal-depth mapping will be explored with the goal of assigning temperature to objects in the painting that enables conveying depth and/or color-based depth. Tests were performed to a total of 18 sighted users and six visually impaired users both during the mapping algorithm design and after developing a tactile temperature prototype artwork model to assess the potentials of thermal interaction for recognizing depth and color-depth in tactile art appreciation. These tests consisted of both prototype testing and interviews with the users. The tests showed both an existing correlation between depth and temperature and that the mapping based on that correlation successes on conveying depth in a new way during artwork tactile exploration. Also, it proved to be a promising technique for improving visually impaired people's artwork exploration assistive devices. Particularly, eight users assessed the similarity between the depth feeling created by the temperature tactile perception and the one which arose from the visual perception. The results showed that the temperature-depth mapping algorithm was able to successfully translate the visual depth feeling into temperature cues. All these results and tests will be shown in the corresponding section.

1.2. Background and Related Work

1.2.1. Thermal Perception and Thermal Interaction

Humans are able to detect temperature and temperature variations thanks to the thermoreceptors located in the dermal and epidermal skin layers. These thermoreceptors located in the skin code the relative changes and the absolute temperature and send the information to the brain. Thermoreceptors can be of two types: high-threshold receptors and low-threshold receptors. Low-threshold receptors are activated by temperatures that fall within the range of 15 °C and 45 °C. On the contrary, high-threshold detectors are activated by temperatures falling outside of that range. In general, within the 15–45 °C range, the activation of low-threshold receptors are not accompanied by pain. On the contrary, any temperature outside of that range can be painful and temperatures below 0 °C or above 50 °C can even cause tissue damage [6].

The non-painful temperature range has been used extensively for adding thermal feedback to a large variety of applications. For example, in [7] the authors designed structured thermal cues for conveying icons when using the phone. They created both thermal icons and intramodal tactile icons that mixed both thermal cues and vibrotactile cues. The users were successfully able to identify most of the icons correctly. Another interesting application of thermal interaction is the one found in [8].

There, thermal cues are used for car driving assistance. Thermal cues are provided in both sides of the steering wheel indicating the driver which way to turn next. These works are some examples of all the potential possibilities of thermal interactions in the field of HCI. Similarly, our work contemplates possible new applications of thermal interaction but in the context of tactile artwork exploration by VIP.

1.2.2. Thermal Interaction for Assistive Devices

There has been some research about thermal interaction in the context of mobile device applications and even some design guidelines about it [9]. However, the research about thermal interactive assistive devices for the VIP is not extensive and only a few examples exist. In [10], a system for VIP to feel a virtual sun while exploring virtual environments is presented. The virtual sun is produced by means of a device consisting of twelve infrared lamps. Also, in [5] a thermal interactive assistive device was used to aid VIP know the color of the different features of an artwork. In spite of the fact that this works also explore thermal interaction in the context of artistic or recreational activities, there are no other common traits with our present work. The present research focuses on the existing correlation between depth and temperature and some of the potential applications which arise from it for aiding visually impaired users explore tactile artworks. Semantic applications of temperature (such as recreating a virtual sun) or temperature-color mapping are not part of this work's scope, which instead focuses on temperature-depth correlation, possible mappings, and derived applications.

1.2.3. Thermal Interaction for Artwork Exploration

Thermal interaction has already been used in the context of artwork exploration for the VIP. Nevertheless, the common way of applying it has been using temperature as a way of conveying color, since previous research has suggested the existence of a color-temperature association given by the warm-cool spectrum, based in the amount of red and blue. This led research to investigate about the possible mapping between temperature and color, some of them with an art exploration application in mind, such as in [5,10]. While both of those works focused on the design of a tactile-thermal display for haptic exploration of paintings with temperature conveying color, in the case of [10], this was done only as a concept design, without implementation. On the other hand, in [5] a whole tactile-thermal display was prototyped and user tests, as well as interviews, were performed. The results showed not only that the users were able to recognize colors by feeling temperature, but also that they enjoyed the thermal interaction and that the use of thermal interaction in the context of tactile artworks was promising. Our present work follows the same path and tries to extend the possible uses of thermal interaction for tactile artwork exploration. However, this work does not focus on conveying color but, instead, on conveying the depth of the different features of the artwork.

1.2.4. Assistive Devices for Communicating Depth

There are many ways of communicating depth, or how near or far an object is, to the VIP, both in a general context and also in the particular context of artwork tactile exploration. However, to our knowledge, none of them has used temperature interaction as a way of communicating depth. In non-artistic contexts, VIP are usually communicated how far or near objects are when using navigation assistive device. For example, in [11] an obstacle detection system was implemented in a navigation assistive device. The system used ultrasound to detect the nearest obstacle via stereoscopic sonar system. Once the obstacle was detected, a vibrotactile feedback was used to inform the blind person about the obstacle's location.

Similarly, in artistic contexts there are several methods for communicating depth levels to VIP. We can find all of them at museums in what are called 'guides'. There are several types of guides, but all of them share the same goal: giving information about the artwork to the visually impaired user. This information might or might not include depth, but in general it does include it. The three types of guides are: audio guides, relief guides, and volumetric guides. Audio guides are the most common method for providing descriptions of the artwork to VIP. They provide a verbal description of the different features of the artwork. Most of the time, the relative distance and depth of the different features from the artwork are also explained with words.

Relief guides are a type of tactile guides that allow the user to comprehend the visual information by means of the sense of touch. In the case of relief guides, the artwork is translated into an embossed picture or relieve image. Sometimes, only the contour of the different features is salient, so the user can only feel those contours. In those types of relief guides, the depth information is not given to the user in a tactile way. However, in many instances, all the features of the artwork are extruded and have a third dimension added to them, which allows the user to feel the depth of the different artwork features with the fingers. This type of relief guides can be called *2.5D* guide. An example of a *2.5D* image artwork model can be seen in [12].

The other type of tactile guides are volumetric guides. These are completely three-dimensional volumetric works that the visually impaired user is able to explore by touching. They are usually the types of guides that gives more information through tactile interaction. However, if the original work is bi-dimensional, it is usually hard or not possible to transform it into a volumetric form. As a result, most of the research related to bi-dimensional art for visually impaired people takes advantage of audio and relief guides, rather than volumetric guides.

In the case of this work, the copy of the artwork is a relief model based on tactile paper embossing technology. The result is a relief guide where only the contours of the features are embossed. This type of thin tactile paper allow us to add thermal interaction so the user can feel the temperature in the fingers while exploring the relief model. This temperature is what communicates depth to the user by way of a novel and intuitive temperature-depth mapping algorithm whose foundation will be explained in the next section.

1.2.5. Temperature-Depth Cross-Modality

Some studies have researched color-concept or color-emotion association. For example, some previous research examined existing beliefs, either subconscious or conscious, about color through color-emotion association by means of an adjective list [13]. Their test consisted of a list of 30 adjectives, randomly ordered, so the users could associate any of the adjectives with any of the colors that were given to them.

While some temperature-concept association research exists, there has not been similar adjective list approaches for inferring the emotional and conceptual responses of the users towards temperature. The existing research focuses on cases such as exploring heat as an expression medium when focusing on interpersonal communication [14] or mapping temperature to the dimensional models of emotion by ratings along valence and arousal dimensions, based in Russell's circumflex model [15].

However, even with these few pieces of research and approaches about temperature and the conceptual and emotional association that arises from it, there are some results that give a hint about the fact that cold temperatures and warm temperatures are interpreted subjectively as a feeling of remoteness or nearness, respectively. In [16], tests for figuring out subjective interpretations of thermal stimuli in three different scenarios (social media activity, colleague's presence, and the extent of use of digital content) were performed. The results showed both that there was a strong degree of agreement among participants about what temperature conveyed and that warm temperatures conveyed the presence of a colleague while cool temperatures conveyed absence.

In this work we continue researching in that direction by performing some tests to find out whether users find any correlation between temperature and depth (near–far). These tests, which were performed before the temperature-depth mapping design, are shown in the following section. The results were the basis for the final temperature-depth mapping algorithm.

2. Materials and Methods

2.1. Temperature-Depth Correlation

2.1.1. Temperature Range Used

Before going forward, it is important to define the range of temperatures that will be used from now on, both for the tests and for the design of the algorithms. Our algorithms are designed to help visually impaired people being aware of depth in a more interactive way through thermal interaction. Therefore, the temperature range to use during the tests and for defining the algorithm needs to be fixed from the start. As stated above, the temperature pain threshold is [15, 45 °C] although some research suggest it might be [11, 45 °C] [17], so the temperature range needs to be within those extreme limits. In [5], it was experimentally validated that the temperature range of [14, 38 °C] was convenient and comfortable for the visually impaired users, so it is also the range used in this work.

2.1.2. Temperature-Depth Correlation Test

As has been shown in the background section, temperature-depth correlation has not been extensively researched and only a rough relationship between temperature and the presence or absence of a person has been found. These results give us a glimpse about the possibility of the concepts near/far being somehow correlated to warm/cold temperatures. As a result, our team felt encouraged to figure out whether the hypothesis of depth and distance (near/far) being correlated to temperature might make sense or not.

The test was based, as in [13], on an adjective list for the users to select associations with the warm and cold temperatures. The list had similar adjectives to the ones found in there, with the exception of two added adjectives: the adjectives 'near' and 'far'. The list was made so every adjective had another one with the opposite meaning, creating a list of adjective pairs. The main purpose for having so many adjectives when the research was focused on the depth-temperature relationship was to make the user select adjectives without having any clue that the near/far dichotomy was the one the test was focused on. Two different adjectives lists were prepared: one with all the adjectives ordered randomly (Table 1) and another one with the adjectives ordered in pairs (Table 2). Each one of the lists was used at a different stage during the test, which will be explain next.

The test was performed twice, first with a total of ten users, five women and five men. The users had an average age of 24 years and were volunteer college students. Each test lasted around 25 min. Then, for verifying whether the findings could be applied to visually impaired people, the same test was carried out again with six visually impaired users. The users were an average age of 17 years and all of them were students from Chungju Sungmo School, a Korean school for visually impaired people. Two of them were nine-year-old kids and all of them where totally blind from birth. Each test lasted around 30 min. The temperature actuators consisted on two petlier devices, a fan, and a heat-sink, which were controlled by an Arduino Mega board. Both temperature actuators can be seen in Figure 1. The test was performed according to the following four steps:

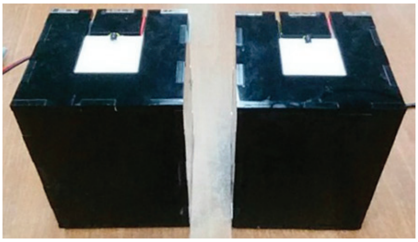


Figure 1. Thermal devices.

Step 1

The tester was given an explanation about the procedure of the test and its purpose. However, the testers were told that the main purpose was to find out which concepts were correlated to temperature, without giving any kind of stress or importance to the near/far concepts as to not make the users tempted to select those options on purpose.

Step 2

After the explanation, the user was asked to touch two different petliers, one which was at a temperature of 15 °C and another one which was at a temperature of 38 °C. After having felt both petliers for as long as desired, they were asked to touch only one of them and to select any number of adjectives that seemed to be related to or conveyed by the temperature they were feeling. Then, the same process was done with the remaining petlier. For each one of the users, the order of the petliers were changed. In other words, if one user started describing the adjectives that suited the warm temperature, then the next user was asked to select the adjectives which were correlated to the cool temperature first.

Step 3

Finally, the same process was repeated. However, this time the adjectives were purposely ordered in pairs, with the option of 'Not Applicable' added to each pair for the cases when the user felt none of both concepts were related to the temperature. This last step was performed for two main reasons: First, the freedom given to the user during the previous step of choosing any adjective from a randomly ordered long list could cause the user to not consider all of them seriously, but rather to just skim through some of them without paying too much attention. On the contrary, the ordered list would force them, at least once more, to go over the near/far adjective dichotomy and consider it in relationship to the felt temperature. Secondly, forcing the user to consider all adjectives and its correlation to the temperatures twice could aid in making the user be more aware of the reason for his/her choices.

Table 1. Complete list of adjectives from the test, ordered randomly. Here the user would choose as many adjectives as desired if he/she felt the experienced temperature was somehow correlated to them.

Vivid	Modern
Mysterious	Near
Classical	Sad
Happy	Confidence
Tiring	Dynamic
Depressive	Fearful
Cheerful	Far
Simple	Boring

Table 2. Complete list of adjectives from the test, ordered in pairs. Also, the "Not Applicable" option was added for each pair. The users were asked to choose one option per row.

Vivid	Boring	Not applicable
Sad	Cheerful	Not applicable
Classical	Modern	Not applicable
Happy	Depressive	Not applicable
Tiring	Dynamic	Not applicable
Confidence	Fearful	Not applicable
Near	Far	Not applicable
Simple	Mysterious	Not applicable

Step 4

Lastly, the testers were asked about the reasons why they chose (or did not choose) the near and far adjectives. Basically, they were asked to justify their answers in order to find out which was the reasoning behind their choice. Also, a brief five-minute conversation about it took place.

2.1.3. Results from Temperature-Depth Correlation Test with the Sighted Users

The results of the tests for the case of the ten sighted users for each one of the two stages can be seen in Figures 2 and 3. The data related only to the near and far adjectives can be seen more clearly in Tables 3 and 4, where the number of people that selected each option is indicated next to the percentage in relation to the total number of sighted participants.

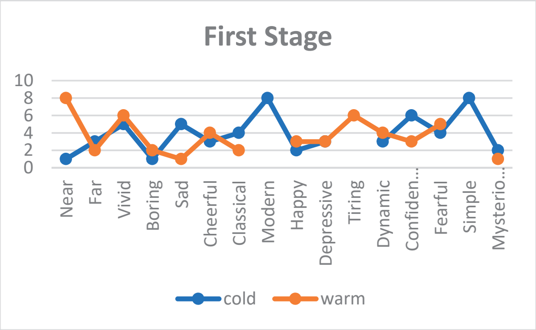


Figure 2. Results during the first stage when users were given the adjective list shown in Table 1.

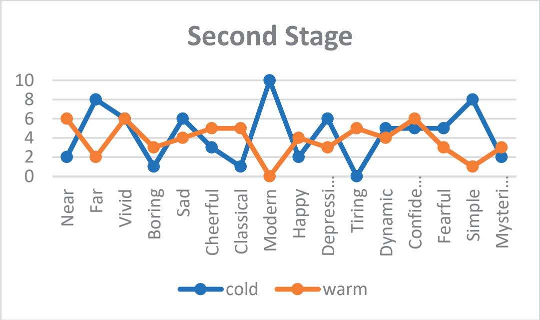


Figure 3. Results during the second stage when users were given the adjective list shown in Table 2.

Table 3. Results during first stage only for the "near" and "far" adjectives.

First Stage	Near	Far
Warm	8 (80%)	2 (20%)
Cold	1 (10%)	3 (30%)

Table 4. Results during second stage only for the "near" and "far" adjectives.

Second Stage	Near	Far
Warm	6 (60%)	2 (20%)
Cold	2 (20%)	8 (80%)

2.1.4. Results from Temperature-Depth Correlation Test with the Visually Impaired Users

The results of the tests for the case of the six visually impaired users for each one of the two stages can be seen in Figures 4 and 5. The data related only to the near and far adjectives can be seen more clearly in Tables 5 and 6.

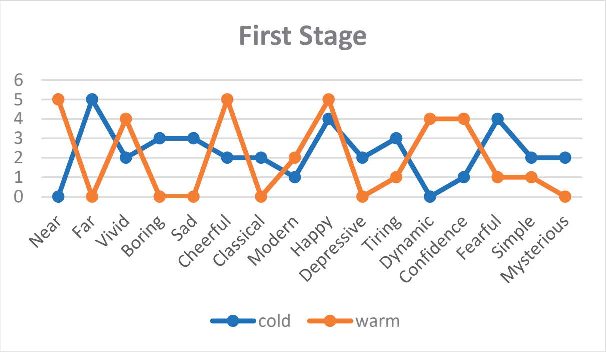


Figure 4. Results during the first stage when users were given the adjective list shown in Table 1.

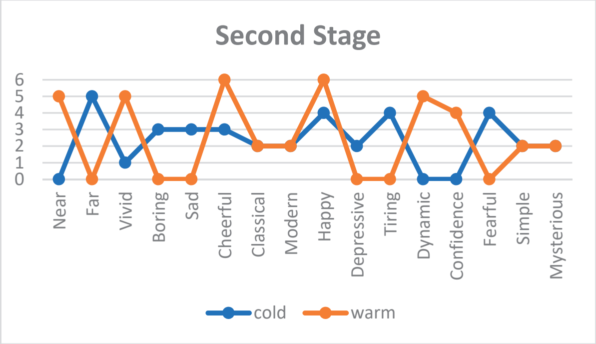


Figure 5. Results during the second stage when users were given the adjective list shown in Table 2.

Table 5. Results during first stage only for the "near" and "far" adjectives.

First Stage	Near	Far
Warm	5 (83.3%)	0
Cold	0	5 (83.3%)

Table 6. Results during second stage only for the "near" and "far" adjectives.

Second Stage	Near	Far
Warm	5 (83.3%)	0
Cold	0	5 (83.3%)

2.1.5. Temperature-Depth Correlation Test Results

In general, a correlation between warm and near, and cold and far, can be clearly seen in both stages with both groups. In particular, it is really interesting to see that, in the case of the VIP group, both stages of the test gave quite similar results (totally similar as far as the near/far adjectives are concerned), which indicates that the visually impaired people were probably putting attention and effort during the test since they are more used to, and find more meaning, in trying out different types of interactions. On the contrary, the sighted people seemed to show less interest during the test, which might be the reason why the answers from stage 1 and stage 2 were somehow different. Nevertheless, even in the case of the sighted people, at some point of the test, eight of the ten users agreed that the conceptual dichotomies warm—near and cold—far were correlated. This fact was proven during the test with VIP, where 83% of the users (a total of five out of six) linked in both stages the warm temperature to the concept of being near something, and the cold temperature to the concept of being far. The exception was one of the nine-year-old young kids, who, after hesitating, decided to select the "not applicable" answer, arguing that even though there was some correlation felt, it was not a clear and transparent feeling. This is really interesting since it might not be a matter of chance that the user who had more troubles correlating temperature to the near/far concept was a child. The process which defines temperature-depth association (or temperature and its association to any other concepts whatsoever) is an interesting issue, which should also be taken into account in the future.

2.1.6. Temperature-Depth Correlation Test Interviews

To make sure what their answers and thoughts were, we asked participants the reasons for their choice and what the feeling or thought was that made them choose that particular adjective. The interviews were useful both for finding out the real thoughts of the users and the feeling and thinking behind their answers. In general, every answer considering a correlation between the warm temperature and the concept of "near" were related to the feeling of warmness we feel when we are surrounded by people, when someone is near us. For example, some of the testers said:

> "I chose near because I remembered how I feel warm and nice when I am close to people".

> "I chose near because I felt a warm feeling like in a warm atmosphere with people coming towards me".

However, there were also two users who believed the concept of "far" was better suited to the warm temperature, and they also had their own reasons to feel that. One of the most interesting comments was the following:

> "I chose far for the warm temperature because I felt a warm hazy feeling like that of smoke, like distant far away memories".

Regarding the far–cold correlation, it is interesting that most of the users that chose far for the cold temperature seemed to agree on the reasons. Eight users answered something related to feeling cold because of being lonely and away from people. Some of the exact words were:

> "The cold temperature reminded me of winter and of feeling lonely, away from everything"

> "The cold temperature made me feel sad and dead, so I felt far away from others and life"

However, there were also answers that justified the opposite: that the cold temperature was a reminder of something being near:

> "The cold temperature was felt fast, immediately, like a knife. So it reminded me of something that is really near and true, something that I can easily feel".

In spite of this differences, it is important to note that most users justified the near–warm and far–cold correlation by sharing similar ideas, feelings and conceptualizations, which in turn, proved that the correlation between farness and cold temperatures, and nearness and warm temperatures can be useful for designing a temperature-depth mapping.

2.2. Temperature-Depth Mapping

The existing correlation between warm and near, and cold and far, was utilized for conveying depth and distance by means of temperature cues. The main idea is simple: the nearer an object is to the user, the warmer the temperature cue. Similarly, the farther an object from the user, the cooler the temperature cue for that object. This idea can be applied in many ways, but we decided to follow a simple mapping method which is explained next, in four steps:

First step

The temperature range is selected. This can be selected freely (as long as it falls within the comfortable temperature range stated above). In general, the visual perceived distance between the extreme depth levels will be assessed and the extreme temperatures selected accordingly (higher temperature difference for higher distances). However, if the simpler algorithm is to be used, then the extremes depth levels will always be linked to the extreme temperatures of 14 and 38 °C, regardless of their perceived relative distance.

Second step

The total number of perceived depth levels are counted. For example, in the case of an image with two objects, one in front of the other, there are two depth levels: front and back.

Third step

The temperature is equally divided in as many temperatures as needed for assigning a temperature to each depth level. The highest temperature and the lowest one are usually assigned to the nearest and the farthest depth level respectively but the use of other initial temperatures might also be possible if is it is considered more appropriate by the designer.

Fourth step

Optionally, if the difference between the temperatures of two consecutive depth levels is less than 3 °C, some of the consecutive depth levels are clustered together. In other words, some objects from different depth levels are put into a similar intermediate depth level. This helps the user recognizing the different depth levels better by decreasing the total different temperatures to be felt and recognized.

Following these four steps creates a simple mapping that does not consider the absolute distance between depth levels (except for the nearest and farthest depth levels, whose relatively absolute distance is considered for choosing the initial temperature range), but only the number of depth levels and the order in which they approach the user. More complex mappings could be designed, such as a mapping which took into consideration the absolute depth levels of all the features. However, this simple mapping is enough for conveying the different depth levels through temperature cues to give an idea to the VIP of where the different features of the artwork are placed according to depth.

Next, this temperature-depth mapping will be applied and examples of its use will be given in two different types of applications: for representing depth of the different objects of a bi-dimensional artwork and for representing color-based depth of a bi-dimensional image (an effect called chromostereopsis).

2.2.1. Application 1: Artwork Depth

The temperature-depth mapping can be used to convey through temperature the different depth levels of the objects of an artwork. In that way, the visually impaired user can sense more deeply the depth presented in a painting.

In this case, different temperatures for the objects that are at different levels of depth are assigned by following the method presented above.

Before applying the method, first, some techniques used by artists for creating the illusion of depth in visual arts will be contemplated and the method applied to those simple cases. After that, some examples of the temperature-depth mapping method presented above will be applied to two real famous artworks.

Illusion of depth in 2D visual arts

In 2D visual arts, there are many ways of creating the illusion of depth, such as:

- Overlapping and layering
- Size and placement and perspective
- Shading
- Texture and detail
- Color, hue, and value

The most relevant ones are layering and overlapping, shading, and size, placement and perspective. However, color, hue, and value can also contribute to create strong feelings of depth, an effect called chromostereopsis [2]. This effect will be explored separately in the following section. First, shading and size, placement and perspective techniques will be explained (since layering and overlapping is a really intuitive and common technique, no explanation will be given).

Shading

Volumetric objects always create shade when being hit by a source of light. As a result, in 2D visual arts, the use of light and shade is one of the methods for creating the illusion of depth. Figure 6 shows an example of an effect called "the crater illusion" [2] in which the central square seems to be in front of the background (right image) or behind the background (left image) depending on the position of highlighted or shadowed edges. The results of applying the depth-temperature algorithm in this case can be seen in Table 7, for the left side of the figure, and in Table 8 for the right side of the figure.

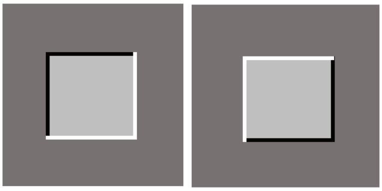


Figure 6. The crater illusion.

Table 7. Temperature-depth mapping of right side of Figure 6. The central square seems much nearer than the external square, so they are given both extreme temperatures of 38 and 14 °C, respectively.

Color	Layer	Depth Temp. (°C)
Central Square	1	38
Background	2	14

Table 8. Temperature-depth mapping of left side of Figure 6. The external square seems much nearer than the central square so they are given both extreme temperatures of 38 and 14 °C, respectively.

Color	Layer	Depth Temp. (°C)
Central Square	2	14
Background	1	38

Size, placement, and perspective

Vertical placement: we perceive objects that are placed lower in the image as closer to us, and objects that are placed higher as being further away. A really clear example of this will be seen later when we apply the temperature-depth mapping algorithm to the artwork "*Starry Night*" by Vincent Van Gogh.

Diagonal perspective: we perceive diagonal lines as receding into the distance. As shown in Figure 7 and Table 9, the red-colored square seems to recede due to the diagonal perspective.

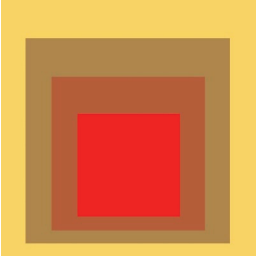


Figure 7. An example of diagonal perspective: Josef Albers, Homage to the Square [18].

Table 9. Temperature-depth mapping for Figure 7. The temperature range selected was [26–38 °C], since the external square and the central one seem to be quite far from each other, but not that much far away. The range was then divided by four and each temperature assigned to each one of the depth levels. In this case, basically, the user would feel the temperature decreasing as he/she approached the central square.

Color	**Layer**	**Depth Temp. (°C)**
Saturated yellow	1	38
Dull yellow	2	34
Dull red	3	30
Saturated red	4	26

Examples

In this painting, by Matisse (Figure 8), there are seven different elements, which can be seen in Table 8. The dancers have been numbered to aid identification. By looking at the drawing, sighted people can generally agree on five different levels of depth, these levels of depth have been linked to the different depth layers from 1 to 5, with 1 being the nearest layer and 5 being the farthest depth layer from the viewers' location as a reference. Nearer depth layered elements need to have higher temperatures so temperatures are assigned to the extreme layers (38 °C for the Dancer 5 °C and 14 °C for Sky) and then that temperature range is divided by five (since we have five depth layers). The reason for choosing 38 °C and 14 °C was because Sky and Dancer 5 are visually really far away from each other, so the temperature for those extreme depth layers were chosen in a way that the temperature difference was maximum: the lowest and highest temperatures from the defined [14 °C, 38 °C] temperature range in which we are working. The resulting temperatures are linked to their respective layers and they can all be seen in Table 10.

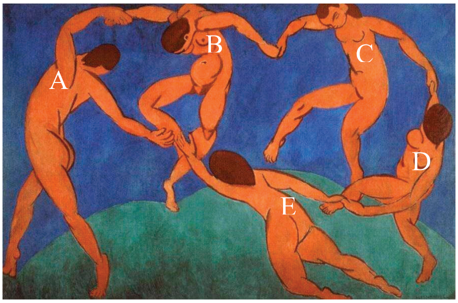


Figure 8. *Dance II* by Henri Matisse, 1910 (Hermitage Museum, Saint Petersburg, Russia).

Table 10. Temperature-depth mapping of Matisse's "*The Dance*".

Element	Layer	Depth Temp. (°C)
Dancer A	2	32
Dancer B	3	26
Dancer C	3	26
Dancer D	2	33
Dancer E	1	38
Soil	4	20
Sky	5	14

In the case of "*Starry Night*" (Figure 9) by Van Gogh, there are many depth layers, as can be seen in Table 11. As before, these depth layers are selected in a visual way, by contemplating the artwork and choosing the main depth levels defined by the features. In our case, nine depth layers were defined. However, since such a high number of layers would force the user to feel and differentiate many different temperatures, we decided to simplify the number of depth layers. For that, elements were visually and conceptually grouped to check whether some of the elements from different depth levels could be grouped under a common depth level. As a result, the starts and the moon, and the mountain and the forest (both pairs of elements having both conceptual common traits and being visually near to each other) were grouped together in two common depth levels. This can be seen in Table 11 where Forest and Mountains share depth layer number 3, and Stars and Moon share the layer number 4. In this way, the number of temperatures is less and the user can identify the different temperatures and depth layers in an easier way. Therefore, even though technically the forest and the mountains are not in the same depth level, we can simplify it to aid identification. In general, this technique should be performed when the temperature difference between layers becomes less than 3 °C.

Figure 9. *The Starry Night* by Vincent Van Gogh, 1889 (Museum of Modern Art, New York City).

Table 11. Temperature-depth mapping of Van Gogh's "*Starry Night*".

Element	Layer	Depth Temp. (°C)
Tree	1	38
Village	2	32
Forest	3	26
Mountains	3	26
Stars	4	20
Moon	4	20
Sky	5	14

2.2.2. Application 2: Chromostereopsis

Another possible application of the temperature-depth algorithm is for conveying the effect of chromostereopsis through temperature. Chromostereopsis [2] is the effect produced by colors on a flat two-dimensional surface by which each color seems to be located in different depth planes, in spite of the two-dimensionality of the image [19]. It is important not to mistake this effect with the association made by artists between red colors and blue colors as advancing and receding colors, since that idea might be based on the brightness produced by atmospheric haze, which is associated with distance, rather than with the effect of chromostereopsis [20]. Chromostereopsis is produced by an effect called chromatic aberration, which is the result from the differential refraction of light depending on its wavelength, causing some light rays to converge before others in the eye and/or to be located on non-corresponding locations of the two eyes during binocular viewing.

Next, an exploratory analysis of the main features that make a color seem farther or nearer will be given, followed by a simple algorithm for conveying chromostereoptic depth by means of temperature. However, first, a brief explanation about colors needs to be given.

Color

The spectrum of color is a continuous one for which there has been several representation models [21]. One of the earliest models is called the Munsell color model, which organized the color perception into a color cylindrical space with three dimensions: hue, chroma (or saturation), and value (or lightness), as can be seen in Figure 10.

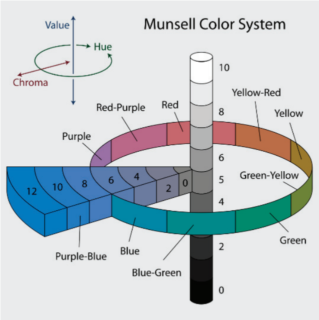


Figure 10. Munsell color system [22].

Hue refers to the color itself. Luminance means the brightness of a color. The higher the luminance, the closer it is to white, and the lower the luminance, the closer it is to black. Saturation is the vividness (clearness) of a color. As an example, 27 colors of varied hue, luminance, and saturation can be seen in Figure 11.

Figure 11. Colors of varied hue, luminance, and saturation [23].

Chromostereopsis

The chromostereoptic effect is complex and its effects can vary due to many different reasons. Nevertheless, for simple images and in a dark background, red objects tend to appear closer to the observer than blue objects, as can be seen in Figure 12. There, the red and blue stripes will seem to be in separated depth levels for most observers, with the red being apparently nearer to the observer.

Figure 12. Red and blue stripes on top of a dark background. The red stripes tend to be seen as being nearer by most people.

This can be extrapolated to warm and cool colors, since, in general (and always when in a black background) warm colors come forward and cool colors retreat [16,19]. However, in [20] researches have also proved that, when the background is white, the effect is reversed, and the warm color seems to be further away than the cool one, as can be seen in Figure 13, which sets the same blue and red colored image to both a black and white background.

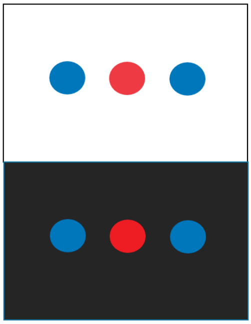


Figure 13. A white background inverts the effect of chromostereopsis. In the top image, most people would see the red color as receding into the distance and the blue nearer to the viewer, while in the bottom image it will mostly be seen as nearer than the blue color.

An algorithm for conveying the chromostereoptic effect in simple images through temperature cues will be presented. However, first it is necessary to find out why some of these colors seem to recede or advance when in company with other colors. Even though warm colors tend to be felt nearer and cool colors tend to be felt further away, that is not always the effect produced. It seems that the most important features for this chromostereoptic effect in simple images are luminance and saturation.

Luminance

In [20], it was observed that one of the reasons why some colors seemed nearer than others was luminance difference, with bright objects appearing closer than dim ones. This would support the claim that "warm colors tend to advance and cool colors recede" since warm colors tend to be brighter than cool colors. In the following Figure 14, and Tables 12 and 13, the same color can be seen next to each other with different luminance levels. In both cases, the high luminance version of the color seems to be nearer than the dark color.

Table 12. Temperature-depth mapping of the left side of Figure 14. Even though both colors were visually similar and the depth difference conveyed was not too large, the extreme temperatures of 14°C and 38°C were selected but since there are only two depth levels, other temperature difference could have been used, but with the brighter color being warmer.

Color	Layer	Depth Temp. (°C)
Light red	1	38
Dark red	2	14

Table 13. Temperature-depth mapping of right side of Figure 14.

Color	Layer	Depth Temp. (°C)
Dark blue	2	14
Light blue	1	38

Figure 14. Higher luminance colors tend to be seen as been nearer than low luminance ones.

Saturation

In [24], patches of colored paper against a black background were shown to a total of 17 subjects. In general, they seemed to agree that a desaturation of a color made its depth effect be diminished. This can be seen in Figure 15, where two colors appear at different levels of saturation. In both cases the muted color (the one that is less clear) seems to be farther away. As before, the range of temperature between both colors can be selected freely after assessing the visual depth contrast (similar to applying the first step of the mapping method stated above) as long as the saturated color is the warmest. We chose again the extreme temperatures, as can be seen in Tables 14 and 15.

Table 14. Temperature-depth mapping of left side of Figure 15.

Color	Layer	Depth Temp. (°C)
Saturated red	2	38
Muted red	1	14

Table 15. Temperature-depth mapping of right side of Figure 15.

Color	Layer	Depth Temp. (°C)
Saturated blue	2	38
Muted blue	1	14

Figure 15. A saturated color will appear to advance where muted color will appear to recede.

Chromostereoptic Temperature-Depth Algorithm

Considering all these features, an algorithm for representing simple chromostereoptic effects with temperature was designed. The algorithm consists of several steps:

First step

First, the background needs to be chosen (sometimes there is no background or the background color can be simplified and not used, especially when it is a color that is not black nor white). If there is a clear background and it is black or white, it will influence the direction in which the chromostereoptic effect is produced so it is important to take it into consideration.

Second step

For each color that is not the background, the saturation and luminance level needs to be calculated, summed up, and halved. So, for each color, a value representing its level of saturation and luminance is acquired.

Third step

The defined temperature range (which is again selected freely through visual relative distance assessment, like was explained in the first version of the temperature-depth mapping method above) is divided by the total number of colors.

Fourth step

Each one of the temperatures is then assigned to each color in order, according to their luminance-saturation level value. If there is no background or there is a background and it is black, higher luminance-saturation values correspond to higher temperatures; if the background is white, lower luminance-saturation values correspond to lower temperatures.

The algorithm is presented in a more formal and concise way here:

(1) For each color (except background) => find saturation and luminance level
(2) For each color (except background) => (saturation + luminance)/2
(3) Order colors by its luminance-saturation value from highest to lowest into a vector V
(4) If white background: reverse V.
(5) Select temperature range and divide it by number of colors.
(6) For each color in V, assign the temperatures in order from highest to lowest.

As an example, the temperatures of the different colors of two artworks will be shown next.

Examples

In Figures 16 and 17, two artworks of the artist called Mark Rothko can be seen. The temperatures were chosen by following the chromostereopsis-temperature algorithm presented above. For calculating the saturation and luminance level of each color, an app called "Visual Color Picker 2.6", created by NOVOSIB software co., was used. The saturation (S), lumination (L), and the hexadecimal color code are shown in both Tables 16 and 17, where the final depth temperatures of each color is also given. The temperature range selected was the 14 °C and 38 °C. As commented before, this is the temperature range to select when the simplest temperature-depth mapping is desired, one in which the extreme depth levels are always mapped to the extreme temperatures, which users can feel without pain. Similarly, the number of depth layers in each image are three, since those are the number of clearly differentiated colors which contribute to the chromostereoptic effect.

Figure 16. *No 1 (Royal Red and Blue)* by Mark Rothko, 1954 [25]. © Mark Rothko.

Figure 17. *No5/No22* by Mark Rothko, 1950 [26]. © 1998 Kate Rothko Prizel & Christopher Rothko/Artists Rights Society (ARS), New York.

Table 16. Temperature-depth mapping of Figure 14.

Color	Layer	Depth Temp. (°C)
Saturated red (S = 92, L = 92) 🟥 #EB3812	1	38 °C
Cool red (S = 80, L = 87) 🟥 #DC2B51	2	26 °C
Saturated blue (S = 82, L = 76) 🟦 #2373C3	3	14 °C

Table 17. Temperature-depth mapping of Figure 15.

Color	Layer	Depth Temp. (°C)
Saturated yellow (S = 95, L = 79) 🟨 #CA9C09	1	38 °C
Saturated red (S = 73, L = 61) 🟥 #9E332A	3	14 °C
Saturate orange (S = 81, L = 76) 🟧 #C47C24	2	26 °C

2.3. Thermal Display System Prototype

We designed and developed a thermal display system prototype where the artworks can be installed to feel the different temperatures while exploring them. The system consists of an array of petliers, each one with its own heat-sink and fan, which are driven by a dual H-bridge board controlled through an Arduino Mega microcontroller. The peltier element is a device that releases heat through one side while absorbing heat through the other side when an electric current goes through it. The direction of the current determines which side heats up and which one cools down.

Each petlier from the setup was able to adjust their temperature from as low as 13 °C to as high as 40 °C. The array of petliers can be seen in Figure 18. The artwork is placed on top of the petlier array. The artwork is printed on Thermal Foamed Capsule Paper by means of a braille printer called TactPlus by Kanematsu USA Inc. Tactplus is a printer that uses thermal technology. By heating the paper, braille and graphics can be easily made. As a result, the user can explore the artwork with the hands while feeling the different temperatures. However, for the different temperatures to be felt more clearly, a thin copper layer was placed between the petliers and the artwork. The schematic of this set up can be seen in Figure 19.

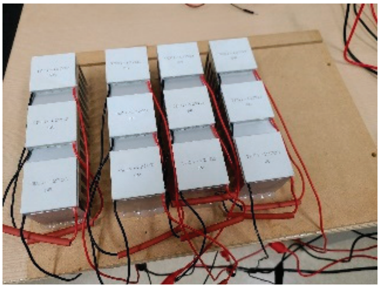


Figure 18. Peltier temperature sensor and controller.

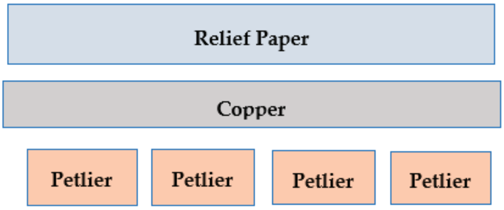


Figure 19. Squematic set up of the prototype, with the relief paper on top of the copper layer and the array of petliers below it.

Each petlier is set to the desired temperature that the visually impaired user feels with the fingers while exploring that zone of the thermal interactive relief artwork. Mark Rothko's work was used as the artwork model due to the strong chromostereoptic effect present there and the simplicity of the shapes. However, under the limitation of the array of twelve petliers, any artwork could be installed and the petliers used both for representing either real depth or chromostereoptic depth. An image of the final prototype can be seen in Figure 20.

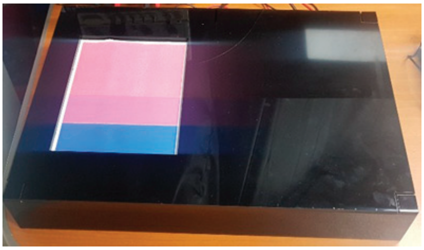


Figure 20. Completed thermal display prototype with the artwork installed on top of the petlier array.

2.4. *Mark Rothko's Artwork Experiment*

To verify the accuracy of the proposed algorithm, a final test was performed. The test was announced through the university bulletin board and, in total, eight college students agreed to form part of the experiment. All of them claimed to be interested in arts and technology. The average age was 26.6 years.

The purpose of this experiment was to test whether visual color depth appearing in Rothko's work could be transmitted through tactile sense by means of temperature cues, which resulted from the designed temperature-depth mapping algorithm. For this, the relative distance between two objects or colors was defined in a scale from 0 to 5, with 0 meaning no relative distancing at all, and 5 meaning a really strong distance variation between objects. Negative numbers meant the second object was felt farther away than the first. The intention of this test was not notified to the participants until the test was over since the goal was to capture natural feelings through their vision and touch. Therefore, the only information given to the users was the context in which the test was placed (improving art exploration experience for the visually impaired people) and the scale system for defining relative distances that was going to be used during the test.

The test sessions included an explanation of the test and its procedure, visual exploration of the artwork with depth degree scale questionnaire, and tactile exploration of the artwork with temperature conveyed depth degree scale questionnaire. Test duration was about 20 min per person. The testing procedure was the following: (1) Introducing the context of the research, explaining about visually impaired people, art exploration, and thermal cues as a way of presenting different features such as color and depth; (2) Introduction of the petlier display prototype and the artwork installed on it. Also, explanation about depth, color-depth and temperature as a way of conveying nearness and farness; (3) The tester was asked to assess and give a scale degree of the relative distance between color 1 and 2, and color 3 and 2 of the Rothko artwork that can be seen in Figure 21. This assessment was done both visually and though tactile exploration with temperature feedback. The order was reverted for each person, so if one user started exploring the artwork and giving a scale to the relative depth visually, then the next user would start assessing the depth by touching the artwork. While touching the artwork, the petliers under it controlled the temperature of each color.

Figure 21. Rothko's artwork (*No 1 (Royal Red and Blue)*) used during the experiment and temperatures associated to each color (as calculated through the color-based temperature-depth mapping in Section 2).

3. Results

The results of the test can be seen in Table 18. V (A ⇒ B) and T (A ⇒ B) indicate the degree of feeling of the difference in depth between colors A and B perceived through visual and tactile sense, respectively. After having all the results, the visual and tactile feeling of the difference in depth between colors were subtracted as a way to measure the difference of depth feeling between the visual and the thermal cues. As an example for understanding the table, the columns related to (3 ⇒ 2) will be explained.

Table 18. Results of the final test.

User	V(1 ⇒ 2)	V(3 ⇒ 2)	T(1 ⇒ 2)	T(3 ⇒ 2)	V(1 ⇒2)–T(1 ⇒ 2)	V(3 ⇒ 2)–T(3 ⇒ 2)
P1	−3	+3	−2	+2	−1	+1
P2	+2	+3	−3	+2	+5	+1
P3	−3	+3	−2	+2	−1	+1
P4	−3	+2	−3	+2	0	0
P5	−3	−3	−2	+1	−1	−4
P6	+1	+2	−2	+2	+3	0
P7	+3	+3	−4	+3	+7	0
P8	−2	+2	−3	+2	+1	0
Avg.	−1	1.9	−2.6	2	1.63	−0.13

V(3 ⇒ 2) is the visual different of depth created when looking at color 2 after color 3. Because of the chromostereoptic effect, most people find the red to be nearer than the blue and, as a result, the effect is that of looking at a color that is nearer than the first one. This effect is related to the positive scale from [0, +5], with a higher number if the depth difference between both colors is felt higher. Similarly, T(3 ⇒ 2) is the depth sensation created by touching the color that is at 26 °C after the one that is at 14 °C, also graded in the same manner (considering that, as has been proved before, most people feel warmer temperatures as that of something being nearer to us). As a result, we can compare V(3 ⇒ 2) to T(3 ⇒ 2) to assess how similarly or differently the thermal cue allows the user to be aware of the depth difference between two objects or colors compared to the depth difference acquired by the visual sense.

As can be seen in the results, most of the times the difference between the visual and thermal tactile depth feeling was less than one whole scale degree. However, there seem to be some extreme differences in some cases, usually given by the fact that the chromostereoptic effect is not totally universal and a few people seem to perceive the depth levels differently. Such is the case with the participant number seven, who felt that, visually, color 1 was actually farther away than color 2 by a considerable distance.

However, even with this little differences, the average difference between visual and thermal tactile depth cues was less than two whole scale degrees for the comparison between color 1 and color 2, and only −0.1 degree of difference between color 3 and color 2. These are promising results that show that the temperature mapping is a proper translation for depth. Nevertheless, some statistical data analysis can also be performed for narrowing down the confidence interval of the mean difference between visual and temperature cue-based depths. However, since the sample data is not too large, any outlier should be properly spotted and deleted from the sample. For that, first, the median and quartiles of both columns are calculated as in equations 1 and 2. For ease of reading, the column representing V(1 ⇒2)–T(1 ⇒ 2) will be called A, and the column representing the value V(3 ⇒ 2)–T(3 ⇒ 2) will be called B.

$$Q_{A1} = \frac{-1-1}{2} = -0.5 \; Q_{A2} = \frac{0+1}{2} = 0.5 \; Q_{A3} = \frac{5+3}{2} = 4 \tag{1}$$

$$Q_{B1} = \frac{0+0}{2} = 0 \; Q_{B2} = \frac{0+0}{2} = 0 \; Q_{B3} = \frac{1+1}{2} = 1 \tag{2}$$

The interquartile range is then calculated:

$$IQR_A = Q_{A3} - Q_{A1} = 4.5 \; IQR_B = Q_{B3} - Q_{B1} = 1 \tag{3}$$

Any value that is below the first quartile or above the third one by an amount of $1.5IQR$ would be considered an outlier. However, in this case there is no outlier so all data needs to be used for the statistical analysis. For calculating the confidence interval of both column A and column B values, the t-distribution is used since not the mean nor the standard deviation of the population are known. Also, other distributions, such as the normal distribution, give better results only when the number of samples exceeds 30.

First, the mean and the standard deviation of columns A and B can be calculated as in (4), where μ is the mean, σ is the standard deviation, N is the number of values, and x_i is each individual value.

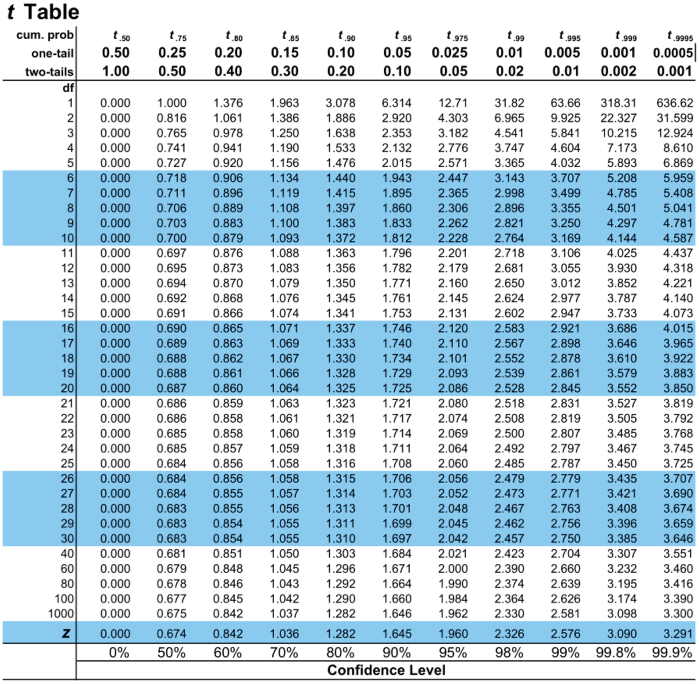
$$\mu = \sqrt{\frac{\Sigma(x_i)}{N}} \; \sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}} \tag{4}$$

The results after applying those formulas are: $\mu_A = 1.63$, $\mu_B = -0.13$, $\sigma_A = 3.07$, $\sigma_B = 1.64$. The confidence interval for the population mean following a t-distribution can be calculated as in (5), where t_{n-1}. is the cumulative probability of the t-distribution given a degree of freedom and confidence level, and n is the degrees of freedom calculated as $N - 1$.

$$\mu \pm t_{n-1} \frac{\sigma}{\sqrt{n}} \tag{5}$$

As can be seen in Figure 22, in this case, for a confidence level of 95%, the value of t_{n-1} is 2.365. As a result, the population average of column A and column B falls, with a 95% of confidence, within the range that can be seen in (6). It can be observed that in the case of column B, the difference between the visual and temperature-based depth assessment would not be larger than a scale degree and a half (of the 5-point scale degree that has been defined above). In the case of column A, there is more uncertainty due to the small sample size, but the result is still promising and encouraging for considering temperature-depth mapping and its based temperature interaction for artwork exploration as an interesting option for future assistive devices for the VIP.

$$\mu_A = [-0.9366, \; 4.1966] \; \mu_B = [-1.5011, \; 1.2411] \tag{6}$$



t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Figure 22. T-distribution table indicating the cumulative probabilities depending on the confidence level. (https://www.tdistributiontable.com).

In conclusion, the accuracy of the proposed algorithm was verified through the experimental process and the temperature cues were found to be a promising way for conveying the chromostereoptic depth of the artwork. The next step was to install the prototype at an exhibition hall in Chungju Sungmo school for the visually impaired in Korea (Figure 23) to assess the responses of the visually impaired users. The reaction of the users was observed and visitors were briefly interviewed for finding out their impressions about the prototype.

Figure 23. Temperature-depth system prototype being exhibited together with other art assistive devices for the visually impaired at the Chungju Sungmo School for visually impaired people, in Korea. The temperature prototypes, of which there are two, are placed in the right side.

Comments from the visually impaired users at exhibition hall

Visually impaired and sighted visitors were briefly interviewed after using the temperature-depth art prototype exhibited in the Chungju Sungmo School for the blind of Korea. They were asked to comment about their feelings and thoughts regarding the temperature interface and its use for representing depth. In general the responses were positive, with some people stating that it was really interesting to explore the artwork in different ways for which they had not been able to do before. Also, some sighted school teachers pointed about the fact that this kind of new interactions keeps some of the visually impaired people interested in learning since most of the books and tools they use to learn are only braille books or audio recordings. Most of the VIP agreed that temperature was an intuitive way of describing depth because of the correlation between warm and near, and cold and far that we stated above. Also, some of them added, especially the children, that the temperature gave a gamification-like feeling to the prototype, making the artwork exploration more enjoyable and engaging. This statement about the gamification making art exploration engaging and interesting seems to be directly correlated to the teachers' comments about VIP getting a lot of benefit from new and unusual ways of interaction for keeping up interest by trying out new ways of learning. There were also some complaints, particularly about the fact that in the boundaries of the objects, the temperature from adjacent objects would mix a little bit and the distinction between temperatures was not very clear at those points. Also, some of them suggested that the same system could also be used for adding temperature to some hot or cold objects, such as the sun or water, instead of for representing depth, which could be done instead by adding a third dimension to the tactile model.

4. Conclusions

Throughout this work, two types of tests were performed with sighted and visually impaired users to assess a method to convey depth information by using temperature cues. The first test showed that warm and cold temperatures can be used as cues to communicate to the user how near or far the features of an artwork are. Based on these results, a complete thermal display prototype was designed and developed. Similarly, a relief artwork was designed and installed on top of the prototype, which was used for performing the final test. This final test's results proved that thermal interaction is a proper way of conveying depth information of the artwork to the VIP. This is an addition to the current technologies which, as was seen above, used to communicate the depth of the features of an artwork either by using audio, or by adding depth into a tactile model by extruding the features. The addition of thermal interaction as a way of communicating depth can open the door to many new ways of experiencing art for VIP. Moreover, the developed thermal display system can be used for adding thermal interaction to any type of paper-based relief artwork, not only by using the thermal cues as a substitute for depth, but also by giving them another role, such as expressing color warmness and color coolness, or for making hot objects (such as the sun) warm and cold objects (such as the water) cold.

Future Work

There are many ways in which this work could be continued or improved on, such as:

1. Increasing the number of petliers adding the possibility of creating more complex temperature regions on the artwork;
2. Finding a way to make the system smaller and more portable;
3. Changing the use of temperature cues from depth representation to an artwork feature semantic mapping, such as making the water feel cold.
4. A necessary addition for this semantic mapping would be to be able to make the prototype work with *2.5D* relief artworks, which present depth by extruding the features in the z direction, and not only with tactile paper artworks. In that way the visually impaired people could be aware of depth through tactile exploration while also feeling the temperature of the different

artwork features while exploring. For that, a method to make the petlier temperature reach all the way to the surface of the *2.5D* relief model should be found.

Last Words

In this work, a temperature-depth cross modal mapping for conveying depth in the context of tactile artwork exploration for visually impaired people was designed. The mapping was based in a conceptual and intuitive correlation between temperature and depth. In addition, the developed mapping was applied to two different types of contexts, a complete prototype for adding temperature interaction to paper relief artworks designed and developed, and that same prototype was tested both with sighted and visually impaired users to assess its functionality and the temperature-depth mapping algorithm performance. The results showed that a relationship between depth and temperature exists and that, on the basis of that relationship, depth of artwork features can be transmitted successfully through temperature cues during tactile exploration of an artwork. We hope this work can encourage researchers to consider thermal interaction both as a substitute for depth and as a viable way to improve accessibility for visually impaired people in tactile artwork exploration contexts.

Author Contributions: Conceptualization, J.D.C.; data curation, S.J.; formal analysis, J.I.B.; funding acquisition, J.D.C.; methodology, J.I.B. and L.C.Q.; project administration, J.D.C.; software, J.I.B. and G.C.; supervision, J.D.C.; validation, J.I.B., S.J., and G.C.; writing—original draft, J.I.B. and J.D.C.; writing—review & editing, J.I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Science Technology and Humanity Converging Research Program of National Research Foundation of Korea (2018M3C1B6061353).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bhowmick, A.; Hazarika, S.M. An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends. *J. Multimodal User Interfaces* **2017**, *11*, 149–172. [CrossRef]
2. Kitaoka, A. Chromostereopsis. In *Encyclopedia of Color Science and Technology*; Luo, M.R., Ed.; Springer: New York, NY, USA, 2016.
3. Quero, L.C.; Bartolomé, J.I.; Lee, S.; Han, E.; Kim, S.; Cho, J. An Interactive Multimodal Guide to Improve Art Accessibility for Blind People. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; pp. 346–348. [CrossRef]
4. Bartolome, J.I.; Quero, L.C.; Kim, S.; Um, M.-Y.; Cho, J. Exploring Art with a Voice Controlled Multimodal Guide for Blind People. In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, Tempe, AZ, USA, 17–20 March 2019; pp. 383–390. [CrossRef]
5. Bartolome, J.D.I.; Quero, L.C.; Cho, J.; Jo, S. Exploring Thermal Interaction for Visual Art Color Appreciation for the Visually Impaired People, In Proceedings of the International Conference on Electronics, Information, and Communication (ICEIC). Barcelona, Spain, 19–22 January 2020; pp. 1–5. [CrossRef]
6. Eliav, E.; Gracely, R.H. Chapter 3—Measuring and Assessing Pain. In *Orofacial Pain and Headache*; Sharav, Y., Benoliel, R., Eds.; Mosby: Maryland Heights, MO, USA, 2008; pp. 45–56. ISBN 9780723434122.
7. Wilson, G.; Brewster, S.; Halvey, M.; Hughes, S. Thermal Icons: Evaluating Structured Thermal Feedback for Mobile Interaction. In Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services, San Francisco, CA, USA, 29 September 2012; pp. 309–312. [CrossRef]
8. Di Campli, S.V.P.; Brewster, S.; Pollick, F.; White, S. Thermal in-car interaction for navigation. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, Scotland, 13–17 November 2017; pp. 510–511. [CrossRef]
9. Graham, W.; Martin, H.; Stephen, B.; Stephen, H. Some Like It Hot: Thermal Feedback for Mobile Devices. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 2555–2564. [CrossRef]
10. Lécuyer, A.; Mobuchon, P.; Mégard, C.; Perret, J.; Andriot, C.; Colinot, J.P. HOMERE: A Multimodal System for Visually Impaired People to Explore Virtual Environments. In Proceedings of the IEEE Virtual Reality Conference 2003 (VR 2003), Los Angeles, CA, USA, 22–26 March 2003; pp. 251–258. [CrossRef]

11. Bousbia-Salah, M.; Redjati, A.; Fezari, M.; Bettayeb, M. An Ultrasonic Navigation System for Blind People. In Proceedings of the 2007 IEEE International Conference on Signal Processing and Communications, Dubai, UAE, 24–27 November 2007; pp. 1003–1006. [CrossRef]
12. Andreas, R.; Werner, P.; Helena, G.C.; Joanna, W.; Svenja, S.; Christian, L.; Laura, L.; Maria, S.; Anton, F.; Stefan, F. Pictures in your mind: Using interactive gesture-controlled reliefs to explore Art. *ACM Trans. Access. Comput.* **2018**, *11*, 1–39. [CrossRef]
13. Banu, M. Color-emotion associations and color preferences: A case study for residences. *Color Res. Appl.* **2007**, *32*, 144–150. [CrossRef]
14. Lee, W.; Lim, Y.K. Explorative Research on the heat as an expression medium: Focused on interpersonal communication. *Pers. Ubiquitous Comput.* **2012**, *16*, 1039–1049. [CrossRef]
15. Wilson, G.; Dobrev, D.; Brewster, S.A. Hot Under the Collar: Mapping Thermal Feedback to Dimensional Models of Emotion. In *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 4838–4849. [CrossRef]
16. Graham, W.; Gavin, D.; Stephen, B. In the Heat of the Moment: Subjective Interpretations of Thermal Feedback During Interaction. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015. [CrossRef]
17. Stevens, J.C. Thermal Sensibility. In *The Psychology of Touch*; Heller, M.A., Schiff, W., Eds.; Erlbaum: Mahwah, NJ, USA, 1991; pp. 61–90.
18. FriedeWie/CC0. Available online: https://commons.wikimedia.org/wiki/File:Grafik_nach_Josef_Albers.jpg (accessed on 20 October 2020).
19. Vos, J.J. Depth in colour, a history of a chapter in physiologie optique amusante. *Clin. Exp. Optom.* **2008**, *91*, 139–147. [CrossRef] [PubMed]
20. Peter, T.; Keith, M.; Robert, S. Chromostereopsis: A multicomponent depth effect? *Displays* **1993**, *14*, 227–234. [CrossRef]
21. Noor, I.; Mokhtar, H.; Zaman, K.R.; Pramod, M. Understanding Color Models: A Review. *ARPN J. Sci. Technol.* **2012**, *2*, 265–275.
22. Available online: https://commons.wikimedia.org/wiki/File:Munsell-system.svg (accessed on 15 September 2020).
23. Available online: https://commons.wikimedia.org/wiki/File:HLSColorSpace.png (accessed on 8 November 2020).
24. Kishto, N.B. The colour stereoscopic effect. *Vis. Res.* **1965**, *5*, 313–329. [CrossRef]
25. Available online: https://www.wikiart.org/en/mark-rothko/no-1-royal-red-and-blue-1954 (accessed on 2 July 2020).
26. Available online: https://www.wikiart.org/en/mark-rothko/no-5-no-22 (accessed on 20 June 2020).

electronics

MDPI

Article

Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments

Jun Dong Cho [1,2,*], Jaeho Jeong [3], Ji Hye Kim [4] and Hoonsuk Lee [2]

[1] Department of Human Information and Cognition Technology Convergence, SungkyunKwan University, Suwon, Gyeonggi-do 16419, Korea
[2] Center for Human Information and Cognition Technology Convergence, SungkyunKwan University, Suwon, Gyeonggi-do 16419, Korea; hoonsuk@skku.edu
[3] Sound Plan, Seoul 06974, Korea; dotoricps@gmail.com
[4] Gangbuk Braille Library, Seoul 06974, Korea; weione@nate.com
* Correspondence: jdcho@skku.edu; Tel.: +82-10-4332-7127

Received: 8 October 2020; Accepted: 19 November 2020; Published: 23 November 2020

Abstract: The recent development of color coding in tactile pictograms helps people with visual impairments (PVI) appreciate the visual arts. The auditory sense, in conjunction with (or possibly as an alternative to) the tactile sense, would allow PVI to perceive colors in a way that would be difficult to achieve with just a tactile stimulus. Sound coding colors (SCCs) can replicate three characteristics of colors, i.e., hue, chroma, and value, by matching them with three characteristics of sound, i.e., timbre, intensity, and pitch. This paper examines relationships between sound (melody) and color mediated by tactile pattern color coding and provides sound coding for hue, chroma, and value to help PVI deepen their relationship with visual art. Our two proposed SCC sets use melody to improve upon most SCC sets currently in use by adding more colors (18 colors in 6 hues). User experience and identification tests were conducted with 12 visually impaired and 8 sighted adults, and the results suggest that the SCC sets were helpful for the participants.

Keywords: user experience; visually impaired; color sound coding; accessibility; art appreciation

---

1. Introduction

Artists believe that art is central to human life. Unfortunately, many people with visual impairments (PVI) do not have access to the world's visual culture or the opportunity to experience the life-enhancing power of visual art. PVI must have access to the world's visual culture if they are to participate fully in their communities and the world at large. Such access will improve the quality of their lives and help them gain skills crucial to their education and employment opportunities [1]. Findings from participant observations in touch tours for blind and visually impaired people at the Metropolitan Museum of Art were discussed by Hayhoe [2]. Touch Graphics Inc. exhibits a reproduced painting called "Talking Tactile" at the San Diego Museum of Art [3]. Recently, "BlindTouch" [4,5] has been introduced to enhance PVI's artwork experience by reproducing painting masterpieces as 3D-printed models that contain touch recognition sensors (based on conductive painting) and provide relevant audio descriptions and sound effects (such as the sound of the wind or a flying butterfly) while the user explores the different features of the artwork with their fingers. The touch interface was evaluated by PVI and found to be effective. The audio feedback can be activated by tapping the fingers. However, PVI still find it difficult to effectively experience color when appreciating a museum's artwork. Tactile color pictograms (TCPs), which are embossed on a surface that PVI can touch to perceive color information, are used to allow easy accessibility to color information. TCPs have two advantages as an assistive tool when used in conjunction with audio descriptions. First, they allow immediate access to color information through color patterns, just as a sighted person sees colors immediately. Second, embossing a TCP directly on a piece of artwork reproduction surface allows PVI to grasp color and color-related information (e.g., shape, size, brightness, and position) through tactile interaction [6].

However, for color recognition, tactile patterns coding colors (TPCs) alone might not provide a good user experience for PVI (especially those with congenital blindness or complex disabilities) because tactile interactions tend to be slow. Congenitally blind people understand colors through physical and abstract associations. Color audition means the reaction of feeling color in one sound. Gauguin tried to pursue a symbol of innerness as a color. He said, “Color which, like music, is a matter of vibrations, reaches what is most general and therefore most indefinable in nature: its inner power” [7].

Therefore, it might be desirable to provide color images for PVI that convert the color being touched into a corresponding musical sound, such as orchestral or classical music, that codes the color. In other words, to provide better color perception for PVI, a multisensory user experience that combines tactile and musical stimuli might be effective. A combination of TPCs and sound coding colors (SCCs) could also enable PVI to interpret the overall color composition of a piece of artwork. Therefore, we here explore the association of sound tones (timbre, pitch, and intensity) with the color properties (hue, value, and chroma) of the Munsell color system. Based on the observations and series of user tests performed in this research, we here introduce two SCC sets that represent 18 colors in 6 hues.

1.1. Review of the Color System

The Munsell color system is a color space that specifies colors based on the three properties of hue (basic color), value (lightness), and chroma (color intensity) [8]. In this system, the higher the lightness value, the closer the color is to white, and the lower the value, the closer it is to black. Chroma is the vividness (clearness) of a color. Colors with the highest chroma are pure colors (saturated colors), meaning that they are not mixed with other colors, and colors with the lowest chroma are achromatic (white, gray, black). Palmer et al. described 37 colors that reflect the color preferences of US college students [9]. Munsell’s hue/value/chroma data from those 37 colors can be found in [9]. The SCC sets proposed in this paper can express six unique hues (red (R), orange (O), yellow (Y), green (G), blue (B), and purple (P)) using three color dimensions—saturated (S), light (L), and dark (D)—for each hue, as shown in Figure 1.

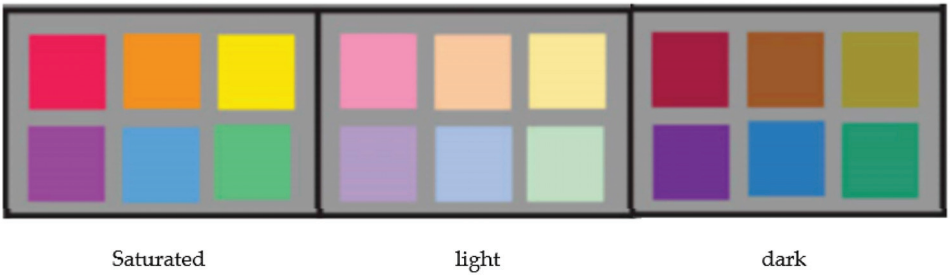


Figure 1. Red (R), orange (O), yellow (Y), green (G), blue (B), and purple (P) using three color dimensions—saturated (S), light (L), and dark (D).

For example, in Figure 2, the color marked “Light” has a value of seven and chroma of eight. The color marked “Dark” has a value of three and a chroma of eight. The color marked “Saturated” has a value of five and a chroma of 15. The colors with the lowest chroma (0) are achromatic.

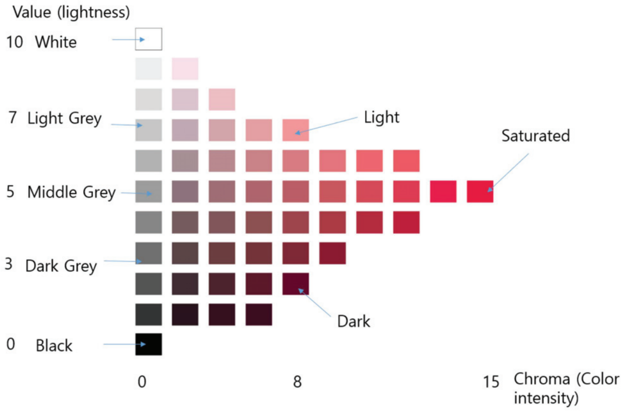


Figure 2. Saturated (S), light (L), and dark (D) for red in [9].

1.2. Review of the Sound Representations of Colors

The perception of a sound has several aspects:

Pitch: the frequency of a sound (high or low).
Tone: the timbre of a pitch.
Key: is it a major or minor scale?
Timbre: the characteristic sound of an instrument (e.g., a flute and an oboe playing the same note sound differently).
Chords: two or more notes played at the same time.
Melody: sequentially arranged notes.
Volume: loudness.
Velocity (the musical term): **the force with which a note is played, and it is vitally important in making MIDI performances sound human.**

Synesthesia, a mixing of the senses, occurs in some individuals. People with strong synesthesia, synesthetes, experience a perception in one sense when a stimulus for another sense is presented (e.g., seeing a color when hearing a sound). This is what Martino and Marks [10] call "strong synesthesia" as opposed to "weak synesthesia," which describes the simple cross-sensory correspondences that most people experience, e.g., being able to match a tone of a certain pitch to a light of a certain brightness [11]. One of the most common types of synesthesia is seeing color while listening to music or musical notes (e.g., [12–15]). Gaboski and Odebert [12] showed some tendency to associate color with short musical selections. Slow music is associated with blue, high tones are light, and low tones are dark. Fast and cheerful notes are reminiscent of bright and intense warm colors.

As a synesthete, Kandinsky saw colors corresponding to the sounds of different instruments and used those correspondences in his paintings [13]. He attributed musicality to his paintings and said that painting could generate energy like music, "drawing is rhyming the form with color and showing the moving power through color." He believed that just as musicians express their feelings in musical forms such as rhythm, timbre, and melody, artists can express their inner experiences of fear, sadness, and joy using various arrangements of color and form. He associated colors and forms with specific emotions, connected them with elements of music, and then sought to write those subjective and emotional experiences into logical and universal laws. "Yellow is stimulus, red is energy, blue is infinite sensibility, and green is calm," said Kandinsky.

People without synesthesia also report correspondences between music and color. For example, [14,16] found some evidence that music–color correspondences result from an emotional link. Brighter colors such as yellow, red, green, and blue were usually assigned to happy songs, and gray was usually assigned to sad songs [17]. A study by Palmer [9] also suggested that fast notes in a major key are yellow or orange and slow notes in a major key are blue and gray. He found cross-modal correlations based on the associative sensibility between color and music with the help of music that involves emotion.

Newton's *Opticks* [18] showed that the colors of the spectrum and the pitches of musical scales are similar (for example, "red" and "C"; "green" and "A*b*"). Maryon [19] also explored the similarity between the ratio of each tone to the wavelength of each color to connect them. This method of associating the pitch frequency of the scale with color can be a way of substituting colors and notes for one another [20]. However, the various sensibilities that can be obtained through color are limited by simply substituting colors into the musical scale. Alber Lavigna [21] found that the technique of a composer in organizing an orchestra seems very similar to the technique of a painter applying colors. In other words, a musician's palette is a list of orchestral instruments.

A comprehensive survey of associations between color and sound can be found in [22], including how different color properties such as value and hue are mapped onto acoustic properties such as pitch and loudness. Using an implicit associations test, those researchers [22] confirmed the following cross-modal correspondences between visual and acoustic features. Pitch was associated with color lightness, whereas loudness mapped onto greater visual saliency. The associations between vowels and colors are mediated by differences in the overall balance of low- and high-frequency energy in the spectrum rather than by vowel identity as such. The hue of colors with the same luminance and saturation was not associated with any of the tested acoustic features, except for a weak preference to match higher pitch with blue (vs. yellow). In other research, high loudness was associated with orange/yellow rather than blue, and high pitch was associated with yellow rather than blue [23].

Chroma has a relationship with sound intensity [23,24]. When the intensity of a sound is strong and loud, its color is close, intense, and deep. However, when the sound intensity is weak, the color feels pale, faint, and far away. Higher value is associated with higher pitch [17,25]. Children of all ages and adults matched pitch to value and loudness to chroma. The value (i.e., lightness) is high and heavy dependent on the light and dark levels of the color. Using the same concept in music, sound is divided into light and heavy feelings according to the high and low octaves of a scale. When the intensity of the sound is strong, the color sensed is close and sharp, whereas when the intensity of the sound is weak, the color becomes distant and muted [26].

Another way to match color and sound is to associate an instrument's tone with color, as in Kandinsky [14]. A low-pitched cello has a low-brightness dark blue color; a violin or trumpet-like instrument with a sharp tone feels red or yellow; and a high-pitched flute feels like a bright and saturated sky blue. As shown in Table 1, it is possible to compare the sensibility felt in each instrument tone with the sensibility felt in color.

In "SeeColor" [27], when you touch a relief-shaped, embossed outline area, the color associated with that area is transmitted to an instrument's sound.

In color sonification [28], the color hue, chroma, and value attributes of pixels are mapped onto sound parameters as the f0 (which is related to a sound's perceived pitch), the spectral envelope of the sound (which influences the perception of timbre), and the intensity (which is related to the sound's perceived loudness), respectively. The sound output of a hue value is thus a sinusoid whose frequency depends on that value. The lower frequency colors, such as red and orange, give sensations of strength and power, so they relate well with higher pitched sounds.

The future work from [28] is to deal with more complex sounds and rhythmic patterns. However, there was no assessment provided from these methods to confirm the sound-code literacy of PVI.

In this paper, based on the aforementioned observations, we introduce two SCC sets, VIVALDI and CLASSIC, produced with rhythmic instrumental sounds of classical melody. Table 1 shows the previously proposed color–instrument matchings and the two SCC sets proposed by us.

Table 1. Existing SCCs with instruments and ours.

Colors	Kandinsky, 1912 [14]	Lavignac,1895 [29]	Deville, 2009 [27]	Ours: VIVLADI	Ours: CLASSIC
R	Tuba + Trumpet (warm light red) Violin (cool light red) Cello-bass (cool red) Tuba-bass (warm red)	Cornet (Red) Clarinet (Red brown:) Trumpet (Purple red)	Oboe	Violin + Cello	Violin
O	Viola	Trumpet+Trombone (Crimson with orange)	Violin	Clarinet + Bassoon	Viola
Y	Trumpet (light yellow)	French horn	Violin pizzicato	Trumpet + Trombone	Trumpet
G	Violin-bass	Oboe	Flute	Guitar	Oboe
B	Flute (Light blue) Cello (Dark blue) Contrabass (Thick blue) Bass of pipe organ (Deep blue)	Flute (Azure blue) Violin (Blue)	Piano	Piano	Cello
P	English Horn	English Horn	Saxophone	Organ	Organ

2. Materials, Methods

2.1. Chord Coding Colors (CCC)

The purpose of this study was to create sounds by which the colors in an image are expressed. Our first SCC set is a so-called chord coding colors (CCC) set as shown in Table 2. Each color's chroma and value are represented by a unique musical chord (e.g., the G chord is made up of the notes G, B, and D. The Gm chord is comprised of G, Bb, and D) that sounds as a single note. In the CCC set, the colors red, orange, yellow, blue, green, indigo, and purple are represented by the instruments violin, viola, trumpet, oboe, cello, horn, and saxophone, respectively. The color–instrument assignment mostly stems from Kandinsky in Table 1. High chroma (saturated) maps to a sound intensity denoted by "ˆ" (velocity = 127), and the medium chroma light and dark map to sound intensity (velocity = 60). A high value (light) maps to a high pitch, denoted by "H" (C4 to C6), and a low value (dark) maps to a low pitch, denoted by "L" (C2 to C4). The codes between colors are all the same, so there is no difficulty to understanding saturated, light, and dark.

Table 2. Chord coding colors (CCC).

Hue	Instrument	Saturated	Light	Dark
Red	Violin	$\text{G}^{\wedge\text{M}}$ S.wav	GM7^{H} L.wav	Gm^{L} D.wav
Orange	Viola	$\text{A}^{\wedge\text{M}}$ S.wav	AM7^{H} L.wav	Am^{L} D.wav
Yellow	Trumpet	$\text{B}^{\wedge\text{M}}$ S.wav	BM7^{H} L.wav	Bm^{L} D.wav

Table 2. *Cont.*

Hue	Instrument	Saturated	Light	Dark
Green	Oboe	♪ C$^{\wedge\mathrm{M}}$ S.wav	♪ CM7$^{\mathrm{H}}$ L.wav	♪ Cm$^{\mathrm{L}}$ D.wav
Blue	Cello	♪ D$^{\wedge\mathrm{M}}$ S.wav	♪ DM7$^{\mathrm{H}}$ L.wav	♪ Dm$^{\mathrm{L}}$ D.wav
Indigo	Horn	♪ E$^{\wedge\mathrm{M}}$ S.wav	♪ EM7$^{\mathrm{H}}$ L.wav	♪ Em$^{\mathrm{L}}$ D.wav
Purple	Saxophone	♪ F$^{\wedge\mathrm{M}}$ S.wav	♪ FM7$^{\mathrm{H}}$ L.wav	♪ Fm$^{\mathrm{L}}$ D.wav

To listen to a wav file marked with ♪ (In Supplementary Materials), left-click the wav audio file and drag it to the computer screen; then right-click to enter the program menu and launch Windows Media Player.

2.2. *VIVALDI SCC*

The user feedback on the initial pilot test motivated us to develop our two melody-based SCC sets (Tables 3–5). Our VIVALDI SCC was inspired by Vivaldi's The Four Seasons, which musically represents elements of the four seasons of the year. We extracted some of Vivaldi's melodies from spring, autumn, and summer that matched the characteristics of saturated, light, and dark colors, respectively. The most important part in matching instruments and colors is avoiding similar tones so that the sounds (colors) are clearly distinguishable. Therefore, to express each color, we used two strings, two wind instruments, and two percussion instruments for easy identification. In addition, the composition of the excerpted part was changed from original sound, and the velocity and pitch were adjusted to clarify the distinction between value and chroma.

2.2.1. Hue

It seems to be a cross-cultural, innate phenomenon that humans do not arbitrarily attach sounds to shapes. The classic example of this is the kiki/bouba effect: when subjects are asked to pick a word that most corresponds to a particular shape, the word kiki is invariably attached to jagged shapes, and bouba to smooth, rounded shapes [30]. One study [31] found that smooth musical timbres were associated with bouba, as were the colors blue, green, and light gray, whereas harsh timbres and the colors red, yellow, and dark gray were associated with kiki. A rounded shape has been associated with sounds from a piano, along with the colors blue and green [31]. In other words, the piano has a rounded tone that corresponds with bouba, so it goes well with green and blue rather than red or yellow. That correspondence is also consistent with Parise's findings that sine waves (soft sounds) are associated with a rounded shape and square waves are associated with sharp angular shapes [32].

Therefore, to represent our six hues, we chose: red, strings (violin + cello); orange, guitar; yellow, brass (trumpet + trombone); green, woodwinds (clarinet + bassoon); blue, piano; and purple, organ.

Red, a warm color that gives the feeling of a hot temperature, is represented by stringed instruments with a passionate tone (violin + cello). Brass instruments convey energy, as if a bright light were expanding, and thus represent yellow (trumpet + trombone). Orange, which is a mixture of yellow and red, is represented by the warm and energetic acoustic guitar. Green, which gives the eyes a comfortable and psychologically stable feeling, is represented by woodwinds with a soft and non-irritating tone (clarinet + bassoon). Blue, a cold color, is represented by the piano, which produces a solid and dense tone with a refreshing feel. Purple, a combination of warm red and cold blue, is represented by the pipe organ, which combines a keyboard with brass tubes.

To express the low, medium, and high values of each color using three pitch levels, a wide range of pitches is required to distinguish between saturated, light, and dark colors. As some instruments such as the violin, trumpet, and clarinet have limited pitch ranges, instruments with similar timbres and different pitch ranges need to be mixed.

Table 3. VIVALDI SCC set with wav sound sources.

Ours: VIVALDI	S	L	D
Red–(Violin + Cello)	WAV S.wav	WAV L.wav	WAV D.wav
Orange–Guitar	WAV S.wav	WAV L.wav	WAV D.wav
Yellow–(Trumpet + Trombone)	WAV S.wav	WAV L.wav	WAV D.wav
Green–(Clarinet + Basoon)	WAV S.wav	WAV L.wav	WAV D.wav
Blue–Piano	WAV S.wav	WAV L.wav	WAV D.wav
Purple–Organ	WAV S.wav	WAV L.wav	WAV D.wav

To listen to a wav file marked with WAV (In Supplementary Materials), left-click the wav audio file and drag it to the computer screen; then right-click to enter the program menu and launch Windows Media Player.

We rearranged the overlapping notes from the original score from Vivaldi's The Four Seasons in Figure 3 with the one shown in Figure 4 such that the harmony can be maintained sufficiently with using the least number of overlapping notes. The reason to use the least number of overlapping notes while maintaining harmony is to highlight the melody. In the two scores shown in Figures 4 and 5, the number of notes differs, and the second score is divided into two parts: high-end and low-end. The high-end instruments, such as the violin, trumpet, and clarinet, use two to three tracks, and the low-end instruments, such as the cello, trombone, and bassoon, use two tracks. The melody is placed on the first track and is played only by the high-end instruments. The low-end instruments play only chords.

In the case of dark colors, playing too many notes at a low volume could make it difficult to hear the notes clearly, so one of the high-frequency chords was excluded for simplicity. Instruments such as the guitar, piano, and organ have wide ranges to express all the required melodies and chords, and thus can represent both high and low values on their own.

Here, the melodies and chords of VIVALDI SCC set (for red, orange, yellow, green) were recorded in a soundproof studio with the collaboration of a sound designer, a composer, and performers. The blue (piano) and purple (organ) sound codes were produced by MIDI.

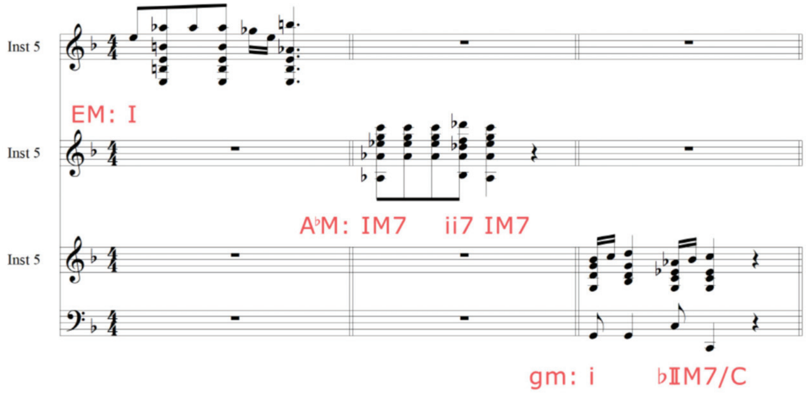



Figure 3. Original track scores for sound coding saturated (left), light (middle), and dark (right) colors from Vivaldi's The Four Seasons.

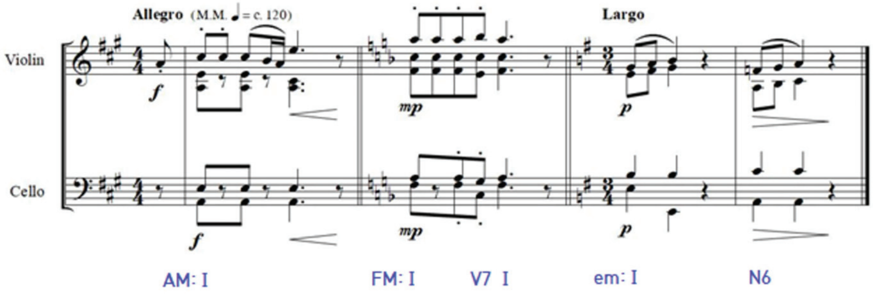


Figure 4. Modified scores for sound coding saturated (left), light (middle), and dark (right) colors played with a violin (same for trumpet and clarinet).

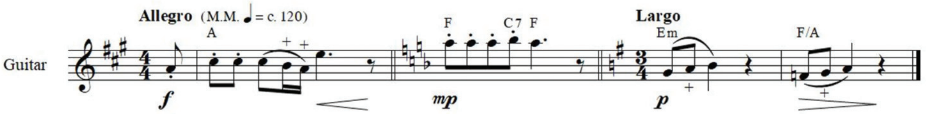



Figure 5. Modified scores for sound coding saturated (left), light (middle), and dark (right) colors played with a guitar (same for piano and organ).

2.2.2. Value and Chroma

The expressions of value and chroma are based on each season's theme sound. Table 2 shows VIVALDI SCC containing with wav sound source for each color. "Saturated" is from spring, "Light" is from autumn, and "Dark" is from summer. In Vivaldi, each season has entry words. Those for spring are: "Spring has come. Little birds say hello to spring"; those for autumn are: "Villagers rejoice and celebrate the joy of harvesting by dancing and singing"; and those for summer are: "In the season when the sun is strong, men and flocks languish, and the trees and grass are exhausted from the heat." We chose rhythmic melodies to represent saturated, light, and dark while considering those entry words.

High pitched stimuli are generally matched with white and light, highly saturated colors, whereas stimuli from the lower octaves tend to be paired with dark colors [33]. Chroma is expressed through the melody's velocity [16,23–25]; at "saturated," it is expressed as a velocity value of 120 or similar.

"Saturated" uses the melody excerpted from spring, and it was judged to be suitable for high chroma by using an A major chord at A4 (middle pitch) to express high chroma. "Light" uses the main theme of autumn, with a high-pitched A5 sound that is played in quick succession and a light atmosphere using F major at A5 (high pitch) to highlight the high-brightness feature. "Dark" uses a slow melody excerpted from summer made dull by progression through E minor at G4 (low pitch) to express a low value. All instrument groups were transposed (compared to the original sound source) to enable actual performance without losing the pitch characteristics of saturated, light, and dark.

2.3. CLASSIC SCC V1

The user feedback on the initial pilot test also motivated us to develop the second melody-based SCC set called "CLASSIC SCC V1" in Table 4 CLASSIC SCC V1 has extracted melodies that match the characteristics of each color from classical music sources which also express the characteristics of the instrument well. The tone, pitch, and intensity of the first note of each melody extracted from a classical music source was chosen to match the characteristics of hue, value, and saturation. The numbers in the table refer to the start time and end time of the extracted sound source. We chose classical music for two reasons. **(1)** In order for the PVI to image the relative position and harmony of the various colors in a piece of artwork, as it will be explained in detail later in Section 2.3, it is necessary to combine the SSCs that express the colors of the picture to produce a well-finished song. Therefore, it is necessary to use sound sources of similar genres as SCCs. **(2)** The word viridi is meaningless to anyone who does not know Latin, but those who have studied Latin immediately think of green. In a similar way, VIVALDI and CLASSIC SCC are color codes expressed by sound, and visually impaired people learn these codes so that they can understand the color composition of the whole picture in a few minutes (e.g., it takes 3 min and 29 s in Table 4). Otherwise, it takes a lot of effort and time to read the tactile color pattern, as in [6], even with all 10 fingers, just like reading Braille.

Table 4. CLASSIC SCC V1.

	S	L	D
R-Violin. High frequency string instrument Vivaldi: The Four Seasons, Violin concertos https://www.youtube.com/watch?v=H49yX03tuHY	MP3 R-S.mp3	MP3 R-L.mp3	MP3 R-D.mp3
O-Viola. Medium frequency string instrument Schubert: Arpeggione Sonata https://www.youtube.com/watch?v=S0YLqYI6$\times$1A	MP3 O-S.mp3	MP3 O-L.mp3	MP3 O-D.mp3
Y-Trumpet. Brass instrument Haydn: Trumpet Concerto https://www.youtube.com/watch?v=NHjgSiTBddM	MP3 Y-S.mp3	MP3 Y-L.mp3	MP3 Y-D.mp3
G-Oboe. Woodwind instrument with reed Rossini: Variations for Oboe https://www.youtube.com/watch?v=obHxRCc7Y-A	MP3 G-S.mp3	MP3 G-L.mp3	MP3 G-D.mp3
B-Cello. Low frequency string instrument Saint-Saëns: Cello Concerto https://www.youtube.com/watch?v=TJVGB6Bf3uE&t=820s	MP3 B-S.mp3	MP3 B-L.mp3	MP3 B-D.mp3
P-Organ. Keyboard instrument with simultaneous expression Mozart: Eine Kleine Nachtmusik https://www.youtube.com/watch?v=U9BO1dazswE	MP3 P-S.mp3	MP3 P-L.mp3	MP3 P-D.mp3

To listen to a wav file marked with MP3 (In Supplementary Materials), left-click the wav audio file and drag it to the computer screen; then right-click to enter the program menu and launch Windows Media Player.

2.3.1. Hue

The sound sources for the CLASSIC SCC set are solo performances by each chosen instrument that well express the characteristics and performant aspects of the instruments. Musical instruments expressing each color are thus classified and designated to make the colors easily distinguishable from one another. We chose the following instruments to represent the colors (hues) in the CLASSIC SCC.

Red, a warm color that gives a feeling of hot temperature, is a violin that plays a passionate and strong melody. A trumpet plays a melody in the high-frequency range with energy, as if bright light were expanding, to represent yellow. Orange is a viola playing a warm yet energetic melody. Green, which makes the eyes feel comfortable and psychologically stable, is a fresh tonal oboe that plays a soft melody. Blue, a cold color, is a cello that plays a low yet calm melody. Purple, a combination of warm red and cold blue, is a pipe organ that plays a splendid yet solemn melody.

2.3.2. Value and Chroma

In the tactile color pattern CELESTIAL [6], the light red pattern consists of three dots, the medium bright one consists of two dots, and the dark one consists of one dot. The light is symbolized as dots, and the more dots there are, the brighter it is. In a similar way, beat-heavy rhythms in SCC have brighter (i.e., higher value) colors. In addition, “saturated” conveys visual glare and a feeling of being close to me with an intense and clear melody in mid-tone. “Light” uses relatively high and fast notes to convey a light particle feel. “Dark” conveys the feeling of separation. Table 4 shows CLASSIC SCC V1 containing with wav sound source for each color.

2.4. Creating Artwork Music Using CLASSIC SCC V2

CLASSIC SCC V2 (Table 5), a modified version of CLASSIC SCC V1 (Table 4), was used to create music that expresses the composition of the overall color of the work. Thus, each color in the artwork is converted into sound, and then the converted sounds are combined into one piece of music.

Table 5. CLASSIC SCC V2 with wav sources.

Instrument Range	CLASSIC	S	L	D
Red-Violin	Tchaikovsky: Violin Concerto in D	WAV R-S.wav	WAV R-L.wav	WAV R-D.wav
Orange-Viola	Stamitz: Viola Concerto in D	WAV O-S.wav	WAV O-L.wav	WAV O-D.wav
Yellow-Trumpet	Haydn: Trumpet Concerto in E flat	WAV Y-S.wav	WAV Y-L.wav	WAV Y-D.wav
Green-Oboe	Rossini Variations for oboe	WAV G-S.wav	WAV G-L.wav	WAV G-D.wav
Blue-Cello	Bach: Cello Suite No.1 in G	WAV B-S.wav	WAV B-L.wav	WAV B-D.wav
Purple-Pipe Organ	Mozart—Eine Kleine Nachtmusik	WAV P-S.wav	WAV P-L.wav	WAV P-D.wav

To listen to a wav file marked with WAV (In Supplementary Materials), left-click the wav audio file and drag it to the computer screen; then right-click to enter the program menu and launch Windows Media Player.

Since the classic music used is different for each sound coding color, the composition will be different. Thus, to combine each sound coding color into one piece of music, it is necessary to unify the composition. The composition of all colors is unified with the key of F. F is suitable for the ranges of the instruments used. The cello's two bars are in the same time as the oboe's and the trumpet's single bar. Thus, the tempo was adjusted as follows. Trumpet (yellow) and oboe (green) were allegro with tempos of 120 to 140 bpm, and adjusted to 78 bpm. The cello (blue) had a tempo between 150 and 170 bpm, and was adjusted to 156 bpm, which is twice as much as 78 bpm.

Rather than reproducing each piece of music in a row, it takes a characteristic melody from the sound code and makes them harmonize to form the music, so that color information and artistry can be saved at the same time. The sound strength (velocity) was adjusted to 120 for "saturated" and 50 for both "light" and "dark" so that they could be clearly distinguished.

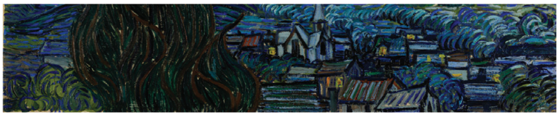
For example, for Gogh's starry night, a blue cello plays the bass and continues to play, and a yellow trumpet and a green oboe play the upper notes for the main melody.

Artwork music provides information about the approximate color placement and arrangement throughout a work of art before it separately detects the color of each local image within the piece. To provide sounds that correspond to the positions of colors on an image, each piece of artwork is divided into three or four tracks, and the color flow within each track is expressed as a continuous sound code. It is very important to identify the position of color on an image to form a standardized shape with a track of the same height. As shown in Table 6, we used the CLASSIC SCC to produce sheet music for Vincent van Gogh's The Starry Night. We decomposed the artwork space into four tracks (rows) of equal height, analyzed the prominent color characteristics of each part, converted them using the SCC set, and recombined them into music with a total length of 3 min and 29 s.

The first track in Table 6 has a star in the blue sky and the moon on the far right. Thus, the cello (blue) plays a bass line, and the trumpet (yellow) and oboe (green) play the main melody. To combine each SCC into a single piece of music, consistency of composition is necessary. The composition of all colors was consistent in the key of F, and the tempo was set to 80/40 bpm. Yellow and green were adjusted to allegro, with a tempo of 120 to 140 bpm. The cello (blue) was set to vivace, between 150 and 170 bpm. All the sounds are consistent with a 4/4 beat.

As shown in Table 6, we used five colors: saturated and dark blue, saturated and bright yellow, and dark green. Thus, the blue night sky on the top track is played by the cello; the bright yellow stars and the moon are played on trumpets; and the dark green cypress tree visible vertically in the second to fourth tracks is played on oboes. As blue, green, and yellow are brought together, a feast of colors unfolds in the form of low and high notes combined to create beautiful music. The wind wriggling in the night sky is a cello with the shining stars in between. The soft glow that follows the wind is a trumpet, and the tranquil village in the bottom track uses a cello and an oboe to create the melodies given to each instrument. Accordingly, the blue color that dominates the overall hue of the picture appears as a cello that plays the bass of the entire song, and while moving the gaze to the right, as if reading a sheet of music from the top left; the visual feeling conveyed by each element that catches the eye is conveyed by the trumpet and oboe. You can feel it aurally through the melody shared between two instruments. A cymbal sounds at the end of every track to indicate which track is playing. This method not only allows the listener to analyze the individual elements that make up a piece of artwork, but also exhibits the completeness of the piece of art as a combination of various visual elements. It is significant in that it allows the overall feeling of a piece of artwork to be conveyed intuitively while retaining the image.

Table 6. Music composition with CLASSIC SCC for Vincent van Gogh's The Starry Night, 1889 (Museum of Modern Art, New York).

Partition	**Color and Music Composition** Starry Night-.wav
Night sky with stars and moon	 Blue/S/Cello (Bach: Cello Suite No.1 in G) Yellow/S,L/Trumpet (Haydn: Trumpet Concerto In E-Flat Major, Hob.VIIe:1—I. Allegro)
Night sky and whirlwind	 Blue/S/Cello (Bach: Cello Suite No.1 in G) Green/D/Oboe (Rossini Variations for oboe) Yellow/S/Trumpet (Haydn: Trumpet Concerto In E-Flat Major, Hob.VIIe:1—I. Allegro)
Cypress trees and whirlwinds	 Blue/S/Cello (Bach: Cello Suite No.1 in G) Green/D/Oboe (Rossini Variations for oboe) Yellow/S,L/Trumpet (Haydn: Trumpet Concerto In E-Flat Major, Hob.VIIe:1—I. Allegro)
Cypress trees and cold land	 Blue/S,D/Cello (Bach: Cello Suite No.1 in G) Green/D/Oboe (Rossini Variations for oboe)

To listen to a wav file marked with (In Supplementary Materials), left-click the wav audio file and drag it to the computer screen; then right-click to enter the program menu and launch Windows Media Player.

3. User Tests

This user test aimed to improve the recognition rate of the proposed SCC by evaluating and analyzing the cognitive accuracy, intuition, and texture of the sound.

3.1. Chord Coding Colors (CCC)

Fifteen college students took part in the exam, which began after participants spent 15 min learning each color corresponding to the sound of an instrument. Two 7" 2-Way Active Studio Monitor speakers were used to play MIDI sound clips by installing RME TotalMix FX v1.50 to control the hardware mixers and effects on an RME audio interface. Each MIDI sound clip was played to the participants, who were asked to identify the color corresponding to each clip. As the participants listened to the sound clips in random order, they were asked to report on the color hue and color attributes (saturated, light, dark) they associated with each sound. The stringed instruments, such as violin, cello, and viola, were difficult to distinguish from one another in the CCC SCC. Likewise, participants found it difficult to distinguish between the horn and oboe, and the oboe and trumpet.

Moreover, low frequency violins and high frequency cellos were difficult to distinguish. On the other hand, the match between yellow and trumpet was easy to remember. The violin (red) was associated with roughness, and the cello (blue) was deemed the softest. The primary color identification rate for each color was as follows: red (violin): 93%; blue (cello): 80%; yellow (trumpet): 93%. However, it was difficult to distinguish the secondary colors: orange (viola) = 53%; purple (horn) = 47%; green (oboe) = 47%.

The subjective feedback from test participants was as follows:

It would be better to express color with a melody.
It would be better to use a clearly distinguished instrument rather than a similar instrument such as the oboe and horn.
Mechanically regenerated sound is not natural.
If you only hear one sound, it seems difficult to distinguish colors. Listening and comparing multiple sounds at the same time could find a connection.

Thanks to this feedback, we decided that it would be more advantageous to distinguish musical instrument color matching by expressing a melody with a lot of musical personality than to distinguish only by chords.

3.2. CLASSIC SCC V1 vs. VIVLADI SCC

Participants were recruited into two groups: PVI musicians and sighted non-musicians. Eight sighted non-musicians participated in the test. The average age of the sighted participants was 22 years (range 20–25 years). The PVI musicians all had congenital blindness with no color experience. The musicians had all been involved in college-level music studies or had a significant amount of personal education. The non-musicians had not been involved in college-level music research and had no significant amount of personal education.

First, nine congenitally blind musician participants attended the test for CLASSIC SCC V1 and VIVLADI SCC. The average age of the participants was 26 years (range 22–30 years). The participants were given a one-hour tutorial (introduction to color basics and an explanation of how we made the VIVALDI and CLASSIC SCC sets). The sound source and user testimonial questions to be used in the experiment were distributed just before the experiment began. After orientation, participants were presented with sound clips corresponding to 18 colors in each SCC in random order. After listening to each color sound code, the participants evaluated their user experiences. Figure 6 shows the user experience evaluation scores that blind musicians gave the CLASSIC SSC V1 and VIVALDI SCC V1. The overall average scores for the two SCC sets were 3.09 (77%) and 3.47 (87%), respectively. As a result of the paired t-test between CLASSIC SCC v1 and VIVALDI SCC, “ease of use” ($t = -4.0$, $p = 0.04$) and “texture”($t = -2.63$, $p = 0.03$), were statistically significant, but “usefulness” ($t = -1.51$, $p = 0.169$), “ease of learning” ($t = -2.00$, $p = 0.081$), and “satisfaction” ($t = -1.51$, $p = 0.169$) were statistically insignificant. As a result of the analysis, both SCC sets received good scores, but it was found that VIVLADI SCC should improve “sound texture” and CLASSIC SCC should further improve “ease of use.” We shall show the improved versions like CLASSIC SCC V2 and VIVLADI SCC V2 in the next sections.

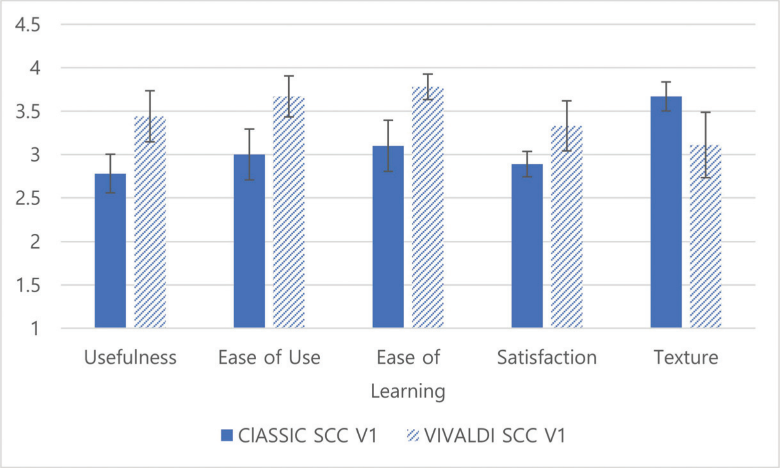


Figure 6. User experience evaluation: strongly disagree (one point)–strongly agree (four points).

3.3. VIVALDI SCC V1 vs. VIVALDI SCC V2

Table 7 lists the participants' feedback after reviewing the two versions of VIVLADI SCCs. The user experience scores from the eight sighted participants were high (80%) for the VIVLADI SCC V1. As mentioned above, recall that scores from the nine congenitally blind musician participants were even higher (84%) for the VLIVLADI SCC V1. The sound texture of the VIVALDI SCC V1 is annoying because it was created in MIDI. This motivated us to enhance the quality of sound by recording the SCCs in a soundproof studio with the collaboration of a sound designer, composer, and performers. As seen from Table 7, the negative feedback, "It is difficult to distinguish due to differences between instruments," became positive: "It was easy to distinguish and fun because only the instruments were different with the same melody."

Table 7. User feedback for VIVALDI SCC V1 and VIVALDI SCC V2.

VIVALDI SCC V1 Evaluated from Eight Sighted Non-Musician Students. (before Recording Real Instruments Playing)	**VIVALDI SCC V2 (Table 5) Evaluated from Nine Congenitally Blind Musician Participants. (after Recording Real Instruments Playing)**
The same melody is repeated, a little annoying. *I think it is less useful than CLASSIC SCC.* *It seems to be very helpful to the blind.* *The music itself seems to create a good synergy with the image.* *It was amazing because the characteristics of the color and the one of the instruments matched well.* *It is easy to listen continuously with the same melody.* *The feeling of each of Vivaldi's [The Four Seasons] is well expressed in color, making it easy to remember.*	*The melody was suitable to distinguish.* *The distinction was made well because only the instrument was different for the melody.* *Especially difficult to distinguish saturation.* *The sound source is concise and easy to understand.* *It was easy to distinguish and fun because only the instruments were different, but with the same melody.* *I heard impressively.*

The PVI with musical experience may be more sensitive to frequency differences and differences in timbre between two signals and may have better short-term auditory memory than blind people with no musical experience [34]. Also, compared with sighted subjects, early blind subjects showed advantages in auditory spectral and temporal resolution, while late blind subjects showed an advantage in temporal resolution [35].

In the negative evaluations other than the positive user feedback in Table 7, non-musicians who evaluated VIVALDI SCC V1 said, "There are many unfamiliar instruments"; "There are instruments that are difficult to discern"; and "Color transmission is difficult." Other said, "It's difficult without appreciating art at the same time"; "In most cases, the sound mood doesn't match the chosen color." However, congenitally blind musician participants who participated in the VIVALDI SCC V2 evaluation provided all positive feedback.

The participant who provided negative feedback thought that in the case of a congenitally blind person without color experience, the color name would have a completely different meaning from what is visible, so it is sufficient to convert the image itself into a melody directly without converting it to a melody corresponding to the color of the image. This method has the advantage of enhancing the expressive power of one work, but also has the disadvantage that it cannot be applied consistently to other works.

Moreover, representing the image itself in sound is a much more difficult and complex process than using color names. The feeling that comes from an image or color is very subjective, so deciding which image to express with some sound is very difficult. Therefore, the current method of linking the characteristics of an instrument to the characteristics of color is objective and easy to access. Just as sighted people learn color names, blind people need to learn not only color names, but also musical codes that correspond to colors.

3.4. CLASSIC SCC V1 vs. CLASSIC SCC V2

Table 8 lists the user feedback for the CLASSIC SCC V2 after it had been updated to reflect the user feedback from the test results of CLASSIC SCC V1. User experience scores from the eight sighted participants averaged 79% for the CLASSIC SCC V2. As mentioned above, recall that scores from the nine congenitally blind musician participants were even higher (84%) for the CLASSIC SCC V1.

Table 8. User feedback for CLASSIC SCC V1 and CLASSIC SCC V2.

CLASSIC SCC V1 (Table 4) Evaluated from Nine Congenitally Blind Musician Participants	**CLASSIC SCC V2 (Table 8) Evaluated from Eight Sighted Students. (after Updating to Reflect the Feedback from V1)**
The distinction between the pitch range and volume is good. *Compared to VIVLADI SCC, it was difficult to distinguish S, L, and D due to the difference in pitch."* *The color of the instrument is not clear because there is accompaniment by various instruments.* *"Dark" is well distinguished because he has a clear musical personality.* *Sometimes it is difficult to distinguish between S and L.* *L is particularly well described, but S and D are difficult to distinguish due to the influence of melodies and instruments.*	*It is useful for expressing paintings, as it seems to express colors well.* *It seems to be very helpful to the blind.* *Most of the music was familiar, so it was easy to use and explain.* *It is more comfortable than VIVALDI SCC, so it is okay to listen for a long time.* *If you learn through enough practice, you will be able to remember it well.* *I want to recommend it because the sound is more abundant.* *If the length of each segment is the shorter, I'll be able to better remember and distinguish S, L, D.*

In addition to the feedback on CLASSIC SCC V1 in Table 8, we got the following: "It is difficult to change the pitch range for each instrument because a specific pitch range is not established"; "Due to the variety of instrumental accompaniment, the color of the instrument is not clear." To solve this problem as a whole, we replaced the red, orange, and blue sound sources of CLASSIC SCC V1, Table 4, with the sound sources of CLASSIC SCC V2, Table 5.

There was negative feedback about the CLASSIC SCC V2: "The shorter the length of the sound segment representing each color, the better you can remember and distinguish S, L and D." That is true, but in this case, when creating music that represents the color composition of the entire artwork, we should make sure that each segment is not too short in length to seamlessly connect the sound segments corresponding to each color.

3.5. VIVLADI SCC V2 vs. CLASSIC SCC V2

Three male participants (26 years old on average) with congenital blindness and no experience of color attended the test, in which we wanted to learn how easily the sounds of the VIVALDI SCC V2 and CLASSIC SCC V2 could be identified. Among them, two participants were musicians who played the trombone and guitar, respectively, in a chamber orchestra. After a 30 min tutorial orientation for both SCCs, the subjects were asked to listen to the sound sources threre times to familiarize themselves with the VIVALDI SCC V2. They were then asked to listen to the VIVALDI SCC V2's 18 color sound codes in a random order, and they were asked to choose one of three color attributes (hue, saturation, or light and dark) that corresponded to each code. Next, the same procedure was used to test the CLASSIC SCC V2. Both SCCs scored 100% on the identification test.

3.6. Vincent van Gogh's The Starry Night with CLASSIC SCC V2

The user experience score evaluated from eight sighted students for artwork music composition of "Vincent van Gogh's The Starry Night" applied with CLASSIC SCC V2 was 83.5%.

The feedback on the composition of Vincent van Gogh's The Starry Night has all been very positive as follows:

The purpose is clearly necessary and useful in expressing color vision through hearing. I really want to recommend it.
It takes a lot of imagination.
It is easy to remember because it can remind us of a rough expression of color in your head. If the rules of the code are maintained but composed of various melodies, the beauty of the work will be conveyed well. It also shows variety.
It is clear and simple, but even looking at one piece of artwork, there seem to be many people making different music with similar pitches.
The division of each part of the artwork was expressed with cymbals, and it was good to appreciate.
When it was arranged with the image of Van Gogh, it was clearer because it was suitable music.
I am satisfied because it goes well with art, and I like music.

4. Conclusions

In this paper, we presented a sound coding color (SCC) design methodology to ascertain whether a person with a visual impairment (PVI) can interact with color through sound without a complex learning process. Although several researchers have done color–sound matchings previously, to the best of our knowledge, no one has experimented with whether PVI can distinguish the color–sound matchings effectively. In this paper, a series of user tests was performed on increasingly refined SCCs developed through user feedback. During the test, the user experience evaluation rate from nine blind musicians for the CLASSIC SSC V1 and VIVALDI SCC V1 was 77% and 84%, respectively. The user experience scores from eight sighted participants were 79% and 80% for CLASSIC SCC V2 and VIVLADI SCC V1, respectively. The music composed to convey the color arrangement of Van Gogh's starry night using CLASSIC SCC V2 received a user experience rating of 83.5%. Finally, after training three congenitally blind adults in both VIVALDI SCC V2 and CLASSIC SCC for about one hour, the recognition rate for both CLASSIC SCC V2 and VIVLADI SCC V2 was 100%. Therefore, the CLASSIC and VIVALDI SCCs helped participants appreciate the overall color harmony of artwork. In [34], in pitch perception, the congenitally visually impaired group was found to be superior to the acquired visually impaired and sighted groups. The "time change analysis ability" of sound was also found to be superior in the congenitally visually impaired group and the acquired visually impaired group than in the sighted group. In [35], early blindness was linked to enhanced perception of the auditory world, including pitch perception. We shall leave extra experiments on more PVI who are not musicians as future work, along with an experiment to find out the differences in perception among various levels of visually impaired and sighted people regarding our proposed SCC sets. Additionally, as a future study, we shall explore the simultaneous cognition ability of PVI when dealing with both color–temperature code [36] and the color–sound code presented in this paper.

Supplementary Materials: The following are available online at http://www.mdpi.com/2079-9292/9/11/1981/s1.

Author Contributions: Conceptualization, J.D.C. and J.J.; methodology, J.D.C., J.J., H.L. and J.H.K.; validation, J.D.C., H.L. and J.H.K.; formal analysis, J.D.C.; investigation, J.D.C., J.J. and J.H.K.; resources, J.D.C., J.J. and J.H.K.; data curation, J.D.C., J.J. and J.H.K.; writing—original draft preparation, J.D.C. and J.J.; writing—review and editing, J.D.C.; visualization, J.D.C.; supervision, J.D.C.; project administration and funding acquisition, J.D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea, grant number 2018M3C1B6061353.

Acknowledgments: We would like to thank all volunteers for their participation and the reviewers for insights and suggestions.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Axel, E.S.; Levent, S.L. *Art beyond sight: Guidelines for making tactile diagrams & standard tactile patterns, lines, and icons. Art Education for the Blind*; AFB Press: Arlington, VA, USA, 2003.
2. Hayhoe, S. *Blind Visitor Experiences at Art Museums*; Rowman & Littlefield: New York, NY, USA, 2017.
3. San Diego Museum of Art Talking Tactile Exhibit Panel. Available online: http://touchgraphics.com/portfolio/sdma-exhibit-panel/ (accessed on 22 November 2020).
4. Bartolomé, J.I.; Quero, L.C.; Kim, S.; Um, M.-Y.; Cho, J. Exploring art with a voice controlled multimodal guide for blind people. In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, Tempe, AZ, USA, 17–20 March 2019; ACM: New York, NY, USA, 2019; pp. 383–390. [CrossRef]
5. Quero, L.C.; Bartolomé, J.I.; Lee, S.; Han, E.; Kim, S.; Cho, J. An interactive multimodal guide to improve art accessibility for blind people. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; ACM: New York, NY, USA, 2018; pp. 346–348. [CrossRef]
6. Cho, J.D.; Bartolomé, J.I.; Quero, L.C.; Lee, D.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2020**. [CrossRef]
7. Color Quoted by Paul Gauguin. Available online: https://www.sensationalcolor.com/gauguin-quotes-color/ (accessed on 22 November 2020).
8. Munsell Color. Available online: http://munsell.com (accessed on 22 November 2020).
9. Palmer, S.E.; Schloss, K.B. An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8877–8882. [CrossRef] [PubMed]
10. Martino, G.; Marks, L.E. Synesthesia: Strong and weak. *Curr. Dir. Psychol. Sci.* **2001**, *10*, 61–65. [CrossRef]
11. Marks, L.E. On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychol. Bull.* **1975**, *82*, 303–331. [CrossRef] [PubMed]
12. Odebert, H.S.; Karwoski, T.F.; Eckerson, A.B. Studies in synesthetic thinking: I. Musical and verbal associations of color and mood. *J. Gen. Psychol.* **1942**, *26*, 153–173. [CrossRef]
13. Karwoski, T.F.; Odebert, H.S. Color-music. In *Psychological Monographs: General and Applied*; The Pshychological Review Company: Columbus, OH, USA, 1938.
14. Kandinsky, V. *Concerning the Spiritual in Art*; Dover Publications: Mineola, NY, USA, 1977.
15. Haack, P.A.; Radocy, R.E. A case study of a chromesthetic. *J. Res. Music Educ.* **1981**, *29*, 85–90. [CrossRef]
16. Barbiere, J.M.; Vidal, A.; Zellner, D.A. The color of music: Correspondence through emotion. *Empir. Stud. Arts* **2007**, *25*, 193–208. [CrossRef]
17. Jonas, C.; Spiller, M.J.; Hibbard, P. Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychon. Bull. Rev.* **2017**, *24*, 1104–1112. [CrossRef] [PubMed]
18. Newton, I. *Opticks or a Treatise of the Reflections, Refractions, Inflections & Colors of Light*; Dover Publications, Inc.: New York, NY, USA, 1952.
19. Maryon, E. *MARCOTONE The Science of Tone-Color*; Birchard Hayes & Company: Boston, MA, USA, 1924.

20. Peacock, K. Synesthetic perception: Alexander Scriabin's color hearing. *Music Percept.* **1985**, *2*. [CrossRef]
21. Lavignac, A. *Music and Musicians*; Henry Holt and Company: New York, NY, USA, 1903.
22. Anikin, A.; Johansson, N. Implicit associations between individual properties of color and sound. *Atten. Percept. Psychophys.* **2019**, *81*. [CrossRef] [PubMed]
23. Hamilton-Fletcher, G.; Witzel, C.; Reby, D.; Ward, J. Sound properties associated with equiluminant colors. *Multisens. Res.* **2017**, *30*, 337–362. [CrossRef] [PubMed]
24. Giannakis, K. Sound Mosaics: A Graphical User Interface for Sound Synthesis Based on Audio-Visual Associations. Ph.D. Thesis, Middlesex University, London, UK, 2001.
25. Cogan, R.D. *Sonic Design: The Nature of Sound and Music*; Prentice Hall: Upper Saddle River, NJ, USA, 1976.
26. Marks, L.E.; Hammeal, R.J.; Bornstein, H.M. Perceiving similarity and comprehending metaphor. *Monogr. Soc. Res. Child Dev.* **1987**, *52*, 1–102. [CrossRef] [PubMed]
27. Deville, B.; Bologna, G.; Vinckenbosch, M.; Pun, T. See color: Seeing colors with an orchestra. In *Human Machine Interaction*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 251–279.
28. Cavaco, S.; Henriquesbc, J.T.; Mengucci, M.; Correia, N.; Medeiros, F. Color sonification for the visually impaired. *Procedia Technol.* **2013**, *9*, 1048–1057. [CrossRef]
29. Donnell-Kotrozo, C. Intersensory perception of music: Color me trombone. *Music Educ. J.* **1978**, *65*, 32–37. [CrossRef]
30. Köhler, W. Gestalt psychology. In *An Introduction to New Concepts in Modern Psychology*; Liveright Publishing Corporation: New York, NY, USA, 1947.
31. Adeli, M.; Rouat, J.; Molotchnikoff, S. Audiovisual correspondence between musical timbre and visual shapes. *Front. Hum. Neurosci.* **2014**, *8*, 252. [CrossRef] [PubMed]
32. Parise, C.V.; Spence, C. Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Exp. Brain Res.* **2012**, *220*, 319–333. [CrossRef] [PubMed]
33. Reuter, C.; Kruchten, S. Colors and Timbres—Consistency and Tendencies of Color-Timbre Mappings in non-synesthetic Individuals. In Proceedings of the "Musik im audiovisuellen Kontext"—34. Jahrestagung 2018 der Deutschen Gesellschaft für Musikpsychologie (DGM), Universität Gießen, Giessen, Germany, 7–9 September 2017.
34. Bogusz-Witczak, E.; Skrodzka, E.; Turkowska, H. Influence of musical experience of blind and visually impaired young persons on performance in selected auditory tasks. *Arch. Acoust.* **2015**, *40*, 337–349. [CrossRef]
35. Shim, H.J.; Go, G.; Lee, H.; Choi, S.W.; Won, J.H. Influence of visual deprivation on auditory spectral resolution, temporal resolution, and speech perception. *Front. in Neurosci.* **2019**, *13*. [CrossRef] [PubMed]
36. Iranzo Bartolomé, J.; Cho, J.D.; Cavazos Quero, L.; Jo, S.; Cho, G. Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People. *Electronics* **2020**, *9*.

electronics

Article

Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics

Luis Cavazos Quero [1], Jorge Iranzo Bartolomé [1] and Jundong Cho [2,*]

[1] Deptartment of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Gyeonggi-do 16419, Korea; luis@skku.edu (L.C.Q.); jorgedavid@skku.edu (J.I.B.)
[2] Department of Human ICT Convergence, Sungkyunkwan University, Suwon, Gyeonggi-do 16419, Korea
* Correspondence: jdcho@skku.edu

Abstract: Despite the use of tactile graphics and audio guides, blind and visually impaired people still face challenges to experience and understand visual artworks independently at art exhibitions. Art museums and other art places are increasingly exploring the use of interactive guides to make their collections more accessible. In this work, we describe our approach to an interactive multimodal guide prototype that uses audio and tactile modalities to improve the autonomous access to information and experience of visual artworks. The prototype is composed of a touch-sensitive 2.5D artwork relief model that can be freely explored by touch. Users can access localized verbal descriptions and audio by performing touch gestures on the surface while listening to themed background music along. We present the design requirements derived from a formative study realized with the help of eight blind and visually impaired participants, art museum and gallery staff, and artists. We extended the formative study by organizing two accessible art exhibitions. There, eighteen participants evaluated and compared multimodal and tactile graphic accessible exhibits. Results from a usability survey indicate that our multimodal approach is simple, easy to use, and improves confidence and independence when exploring visual artworks.

Keywords: accessibility technology; multimodal interaction; auditory interface; touch interface; vision impairment

Citation: Cavazos Quero, L.; Iranzo Bartolomé, J.; Cho, J. Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics. *Electronics* **2021**, *10*, 297. https://doi.org/10.3390/electronics10030297

Received: 12 November 2020
Accepted: 19 January 2021
Published: 26 January 2021

1. Introduction

Museums have traditionally employed several methods to make their collections more accessible in support of the participation of blind and visually impaired people in arts and culture and to comply with laws [1,2] that protect the right to access art. For example, some leading art institutions [3–5] offer accessible "touch tours" and workshops similar to Art Beyond Sight [6] and the Mind's Eye Program [7] where participants can experience art by touching some of the collection artworks while listening to tailored audio descriptions given by the staff. Two additional methods to support access are descriptive audio guides and accessible Braille leaflets of the artworks that may include embossed tactile graphic diagrams. Unfortunately, these methods have limitations. Accessible tours and workshops are available only on specific dates, schedules, and often must be reserved in advance. Moreover, they fail to support independent visits, exploration, and the artworks prepared for touch exploration are not the most prominent collection pieces due to the risk of damage [8]. Audio descriptions and accessible leaflets fail to convey much of the spatial information in the artwork. The latter also requires Braille proficiency, which remains low even in developed countries (about 5% in the UK [9] and less than 10% in the USA [10]).

Nowadays, the development and display of relief models of artworks made using low-cost digital fabrication techniques such as 3D printing are becoming an alternative for improving the accessibility to art. Several art institutions like the Prado Museum [11] and The Andy Warhol Museum [12], among others, have pioneered the use of this alternative in their exhibitions. Compared to tactile graphic diagrams, they offer advantages like improved volume shape, depth, and more diverse texture representation. However, without any verbal descriptions, they might still be challenging to understand. Interactive multimodal guides (IMGs) combine modalities such as audio, tact, smell, flavor, or others to convey and communicate information. Doing so mitigates the individual modalities' shortcomings and complements their strengths.

In this work, we describe our approach to the design, implementation, and evaluation of an interactive multimodal guide for blind and visually impaired people that uses localized on-demand audio descriptions and tactile relief models to improve the independent access and understanding of visual artworks.

Motivation and Objective

Several challenges prevent the adoption of interactive multimodal guides at art museums and galleries. One of them is the preservation efforts and prioritization of the primacy of vision to experience the art pieces [13]. Also, making and exhibiting models based on artists' works may lead to ownership, copyright infringement, and artistic integrity arguments [14]. Furthermore, determining effective methods for accessible art representation is challenging. Motivated by these challenges, our objective was to develop an interactive multimodal guide and study its feasibility to improve accessible art representation compared to tactile graphics. Our main contributions are:

1. A formative study performed with the help of eight blind and visually impaired participants, art museum and gallery staff, and two artists to understand the different needs of these stakeholders and the current state of the accessibility tools available to experience visual artworks.
2. A low-cost alternative implementation of an interactive multimodal guide that enables blind and visually impaired people without previous training to independently access and experience visual artworks.
3. In collaboration with an accessible art gallery and a school for blind and visually impaired people, we performed two art exhibitions using the proposed guide. Within those exhibitions, we performed a survey with eighteen blind and visually impaired participants to compare the proposed interactive guide and a tactile graphics alternative.

2. Related Work

2.1. Tactile Graphics

Tactile graphics (TG) are made using raised lines and textures to convey drawings and images by touch. They are frequently used by blind and visually impaired people because the tactile modality is the best for their graphical image comprehension [15]. Their use is recommended where spatial relationships among the graph's objects are important [16], such as simple graphs, diagrams, and drawings. Unfortunately, they are ineffective to express visual information of complex images [17,18], such as those present in many visual artworks. For this case, adding Braille labels is of limited use due to the large space needed by the Braille characters to be legible. Moreover, including labels within the artwork area obstructs exploration. Advances in low-cost prototyping and 3D printing technologies bring the potential to tackle the complexity of expressing complex images without exploration obstruction by adding interactivity to tactile graphics.

2.2. Interactive Tactile Graphics and 3D Models

In the last decades, researchers have explored the improvement of tactile graphics accessibility by adding interactivity through diverse technologies. Some of the improvements are better content exploration [18], learning facilitation [19], and expansion of the amount of information provided without over-complications [20]. Table A1 summarizes several of these projects and their interaction technologies. Three early works are NOMAD [21], The Talking Tablet [22], and IVEO [23], all of which function by placing a tactile graphic on a high-resolution touch-sensitive pad that detects user touch gestures that trigger audio descriptions. This method provides independent and detailed access to graphic elements, and since it does not rely on Braille, the possible audience is broader. Taylor et al. [24] and LucentMaps [25] make use of the touch screens in mobile devices to detect user–touch interactions in a portable way. They attach 3D printed tactile overlays of city maps to the device screen. Taylor et al. [24] 3D print sections of the overlay using conductive filament to provide interaction points on discrete sections of the map. LucentMaps instead uses translucent filament for their overlays coupled with a mobile application that visually highlights sections of the overlay using the device screen. MapSense [26] also uses a touchscreen to identify user touch gestures and conductive tangible tokens placed on the surface. The tangibles are additionally infused with smell and taste to foster reflective learning and memorization. Using touch-sensitive surfaces to detect user input and trigger audio feedback increases the amount of information communicated to the user. However, this approach is limited to thin overlays. Otherwise, the system can't recognize the touch gestures.

An alternative approach is using cameras to track either the content or the user's hands. CamIO [27], Tactile Graphics with a Voice [28,29], and The Tactile Graphics Helper [30] are examples of projects using this approach. The Tactile Graphics with Voice projects work by using a mobile or wearable device's camera to identify QR codes printed along a tactile graphic. Then, the system tracks the user's hand to trigger localized verbal descriptions. CamIO and The Tactile Graphics Helper use mounted cameras that identify the content using image processing algorithms, instead of using QR codes or visual markers. With the exception of CamIO, the previous projects focus on adding interactivity to 2D tactile graphics, and mainly propose their use for STEM (science, technology, engineering and mathematics) education and orientation and mobility improvement. Both approaches are effective to improve the amount of information and the comprehension of the spatial arrangement of images. However, to facilitate comprehension, they abstract the complexity of images to contour lines, which hinders the aesthetic aspect and exploration of artwork images.

3D printing opens up the possibility to create low-cost reliefs and 3D models of objects with added expressive volume. Holloway et al. [31] propose a touch interactive prototype that uses 3D printed volumetric representations of map models embedded with discrete capacitive touch points that users can touch to trigger audio descriptions. This approach improved the short term recollection and the understanding of the relative height among the map elements. Other studies focused on symbolic representation on 3D maps models, like Holloway et al. [32] and Gual et al. [33,34], they report improvements in terms of accuracy, efficiency, and memorability compared to two-dimensional symbols. Alternative methods to add interactivity involve using other type of devices. For example, pen-shaped devices like The Talking Tactile Pen [35] or wearables like the ring-shaped Tooteko [36]. In this approach, the user must hold or wear the device, which can detect sensors embedded in the tactile graphic or models on approximation.

2.3. Interactive Multimodal Guides for Blind and Visually Impaired People

The body of work on interactive multimodal guides focused on artwork exploration is limited, as seen in Table A1. However, there are several related works. The American Foundation for the Blind offers guidelines and resources for the use of tactile graphics for the specific case of artworks [37]. Cho et al. [38] present a novel tactile color pictogram system to communicate the color information of visual artworks. Volpe et al. [39] explore the semi-automatic generation of 3D models from digital images of paintings, and classifies four classes of 3D models (tactile outline, textured tactile, flat-layered bas-relief, and bas-relief) for visual artwork representation. After an evaluation with fourteen blind participants, the results indicate that audio guides are still required to make the models understandable. Holloway et al. [14] evaluated three techniques for visual artwork representation: tactile graphic, 3D print (sculpture model), and laser cut. Notably, 3D print and laser cut are preferred by most participants to explore visual artworks. Hinton [40] describes the use of tactile graphics of visual artworks made using thermoforming intended to be explored along with tape recordings. Blind study participants reported that the approach helped them understand the space and perspective of the artworks and found the approach fun, interesting, informative, and even stimulating to their creative efforts.

There are the few projects that add interactivity to visual artwork representations and museum objects. Anagnostakis et al. [41] use proximity and touch sensors to provide audio guidance through a mobile device of museum exhibits. Vaz et al. [42] developed an accessible geological sample exhibitor that reproduces audio descriptions of the samples when picked up. The on-site use evaluation revealed that blind and visually impaired people felt more motivated and improved their mental conceptualization. Leporini et al. [43] explore the use of a three-dimensional archeological map and fascade models to communicate historical, practical, and architectural information on demand, using 3D printed buttons with success to provide autonomous and satisfying exploration. Reichinger et al. [44–46] introduce the concept of a gesture-controlled interactive audio guide for visual artworks that uses depth-sensing cameras to sense the location and gestures of the user's hands during tactile exploration of a bas-relief artwork model. The guide provides location-dependent audio descriptions based on the user's hand position and gestures.

We designed and implemented an interactive multimodal guide prototype based on the needs found through our preliminary study described in Section 3.1 and inspired mainly in the related works Holloway et al. [31] and Reichinger et al. [44]. Table 1 compares the main technical differences between the related works and our approach. Besides these differences, this work introduces a comparison between our approach and using traditional tactile graphics to measure potential improvements of the multimodal approach.

Table 1. Features of the proposed interactive multimodal guide and selected related works.

Author	Description
Halloway et al. [31]	- Sensing technology: Capacitive sensor board connected to discrete copper interaction points placed on the surface of the model. - Input: Double tap and long tap gestures on the surface. - Tactile presentation: Tactile 3D map model. - Output: Audio Descriptions. - Objective: Improve Mobility and Orientation.
Reichinger et al. [44–46]	- Sensing technology: Color and depth mounted camera. - Input: Tap gestures on the surface and hand gestures above the surface. - Tactile presentation: Tactile bas-relif model. - Output: Audio Descriptions. - Objective: Improve visual artwork exploration.
Cavazos et al. *	- Sensing technology: Capacitive sensor connected to conductive ink-based sensors embedded under the surface of the model. - Input: Double tap and triple tap gestures on the surface. - Tactile presentation: Tactile bas-relief model. - Output: Audio Descriptions, Sound effects, and Background music - Objective: Improve visual artwork exploration.

* This work.

3. Materials and Methods

3.1. Formative Study

To better understand the current state of the accessibility tools available to experience visual artworks and to explore the requirements for the use of interactive multimodal guides, we conducted a formative study with blind and visually impaired participants, art museums and gallery staff, and artists.

3.1.1. Accessible Visual Artworks for Blind and Visually Impaired People

The formative study focused on the current access to visual artworks through tactile graphics and other means with eight blind and visually impaired participants, with an average age of 29.13 (standard deviation of 7.7). Other characteristics of the participants are described in Table 2. Of the eight participants in the study, three (37.5%) are male, and five (62.5%) are female. While five (62.50%) of the participants attend university studies, three (37.5%) of them work. All the participants gave signed informed consent based on the procedures approved by the Sungkyunkwan University Institutional Review Board.

Table 2. Characteristics of blind and visually impaired participants in our formative study.

Participant	Sex	Age	Occupation	Sight
FP1	Female	24	University student	Total vision loss
FP2	Male	40	Worker	Near vision loss
FP3	Female	42	Worker	Total vision loss
FP4	Female	30	Worker	Profound vision loss
FP5	Male	27	University student	Near vision loss
FP6	Male	24	University student	Total vision loss
FP7	Female	23	University student	Total vision loss
FP8	Female	23	University student	Total vision loss

We followed a semi-structured interview focused on the access and availability of tactile materials at museums, galleries, and through their education. Moreover, we inquired about their experience when using tactile graphics and interactive guides, if any. While all the participants stated having experience using tactile graphics, most of the encounters with this type of materials were limited to educational materials and tactile books during their early education or related to STEM subjects and maps. Four participants stated having experience with tactile graphics related to visual artworks. All the participants that said having experience with tactile graphics in the art fields had access to them during their primary and secondary studies. Only two mentioned having experienced them during a visit to a museum or gallery. All of the participants expressed having visited a museum or art gallery; they reported that the most common accessible tools during their visit were guided tours and the use of audio guides. Seven of the participants mentioned that they were accompanied by someone (relatives or friends) during their visits. They added that they mostly relied on that person's comments and help to use the audio guide during their visit to experience the artworks.

Regarding their experience exploring tactile graphics, the participants mentioned that they are convenient to understand simple diagrams of mathematical concepts or simple graphics in educational fields, learning language characters, and storybooks. Mixed results were reported in their use for tactile maps. Three participants considered tactile graphics easy to understand, while five found them over-complicated or not very useful. However, all of the participants with previous experience with tactile graphics of visual artworks stated dissatisfaction due to their limitations. In particular, one participant commented: *"FP2: Using the tactile graphics is a hit and miss. If the contents are simple and separated is easy to get an idea of what the picture looks like, but often there are so many shapes and textures that is difficult to imagine what the picture looks like, it becomes hard, like thinking about math, art is not supposed to be like that."* This reflects the known problem of producing tactile graphics of complex images, which is usually dealt with by simplifying and abstracting the objects in the image. However, this approach often doesn't solve the problem in the case of tactile artworks. *"FP3: So much detail is lost when touching a tactile graphic. Even if I can find and feel the silhouette of a person or their face, I cannot know if the person in the painting is smiling or crying, and that's what people usually talk about."* Another problem is the challenge to represent perspective and volume. *"FP3: When exploring a tactile graphic everything is on the same level, there's no depth like in the real world. If it's a landscape, I don't know what is in front* *and what's on the back. Even something simple like a ball, I only feel a circle, and many things can be a circle. I'm told that in the painting you can know it's a ball because of the color and shadows, but I just feel a circle."* Despite the shortcomings, the participants expressed the need for tactile graphics and desired for them to be available for more artworks and more locations. *"FP5: Even when they are not perfect (tactile graphics), they are still useful to know what is where in the painting, I still can be in the conversation. I just hope they were available in more places and for all the works."*

3.1.2. Accessible Visual Artwork at Art Museums and Galleries

Some of the participants in the formative study mentioned the shortage in the availability of tactile graphics or other accessibility tools in their visits to art museums and galleries. We met with a couple of administrators and curators at a national art museum, a private art gallery, and an accessible gallery at a social welfare center for blind and visually impaired people, to shed some light on their approach and efforts towards the accessibility to their collection. At the national art museum, they described several of their initiatives towards accessibility. Their current effort is mostly directed to accessible tours. Besides the tours and available audio guides, some of their exhibitions are made accessible through 3D-printed models that can be explored by touch. However, this tool is not always available, and it is used mostly for large modern art installations. The private gallery just offered guided tours by its staff. There were two main concerns. First, any accessible tool or display must be unobtrusive. One of the concerns was that any display co-located with the artwork can become a distraction and deviate the attention from the artwork. The second concern is about the contents. The administrators commented that presenting the artwork through a different medium than the one used by the artist could have implications in the message and intention that the artist wanted to express. Because of this, the use of accessible exhibits is more often available for modern artworks, where the artist can provide guidelines or collaborate in the development of the exhibits or even make their artworks considering accessibility needs.

3.1.3. Accessible Visual Artwork and Artists

We interviewed two artists separately to inquire about the use of accessibility tools and other mediums to experience their art. To generate richer insights, we provided one tactile graphic representation of a painting and discussed it with them. Both artists agreed on the importance of making visual art more accessible to blind and visually impaired people and that it may require the introduction of other tools or mediums. To this end, they strongly suggested collaboration with the author or experts when possible, noting that while the artist may not be an expert on the added medium, it can provide feedback to improve it. One of the artists expressed his concern regarding tactile graphics *"Artist 1: I believe too much emphasis is placed on what is in the painting and not the painting itself. Yes, the recognition of shapes, objects, colors, and elements is relevant, but I dare to say it is not the most important aspect. Viewers should not be passive, just saying to them 'this is this' or 'this means this' is a failure. The goal of my art is to cause a reaction when someone sees it, they (viewers) should think, they should react. That's what experiencing art is."* We believe that this is a very relevant point, since most of the research literature is centered in the improvement of recognition of the objects in the painting, but there is almost no improvement related to the reaction and interpretation studies when using accessible artwork guides.

3.1.4. Design Requirements

Based on the feedback obtained during the formative study, we identified the following design requirements to develop our interactive multimodal guide. Independent exploration is the most important need derived from the formative study. It is largely derived from two factors, adequate access to the artwork and the information presentation method to facilitate understanding and experience. To improve it, the IMG should tackle the following:

1. Simple to learn and use. The guide should offer a low entry barrier to the user. It should avoid the need for Braille literacy for operation and exploration to improve the access for blind and visually impaired people without or limited Braille literacy. It should avoid, as much as possible, the need for training or previous experience for its operation. For example, using a limited set of intuitive and well-known interaction gestures and interfaces to avoid cognitive load.
2. Self-contained. The guide should avoid requiring blind and visually impaired people to carry external devices or install software on their own. Blind and visually impaired visitors often already carry several items such as a personal bag, white cane, leaflets, and audio guides. External devices add to their carrying load, add the need to check-in and out the device, as well as to learn the device operation and interface.
3. Facilitate access to information. Exploring the artworks by touch is essential to understand the spatial arrangement of the artwork. The design of the model should be simple and abstract enough for easy comprehension, while avoiding oversimplification. Audio descriptions should be detailed but not long. Users should be able to skip them if desired.
4. Promote active engagement. The IMG should promote active user engagement by facilitating exploration rather than just providing information. As much as possible, the guide should encourage critical thinking, reflection, and emotional responses.
5. Unobtrusive and versatile. The guide should avoid being obtrusive to the original artwork within an art museum and gallery environment such that it can be colocated and avoid user isolation. The IMG should be able to support different artwork styles, sizes, and shapes.

3.2. Interactive Multimodal Guide (IMG)

Based on the design requirements that we identified from the formative study to address the limitations of tactile graphics and audio guides, we decided to develop an interactive multimodal guide. Our IMG will use a combination of tactile and audio modalities to communicate information and promote the exploration of visual artworks such as paintings. The tactile modality is covered by employing a 2.5-dimensional bas-relief model representation of the visual artwork. This model is accessible by touch and will convey the spatial and composition information of the artwork and will be the primary input interface of the IMG. The audio modality will be delivered through speakers or headphones and will include: narrations, sounds, and background music to convey iconographic and iconological information. The following subsections will cover the implementation of the several components of our proposed IMG.

3.2.1. 2.5D Relief Model

Users of the IMG can touch the 2.5-dimensional model to get an idea of the objects, textures, and their locations in the artwork. The main difference between a tactile graphic and a 2.5D model is that the latter can provide depth perception by giving volume to the objects in the model. There are several techniques to extract the topographical information from artworks like paintings to make a 2.5D model. Three of them are 3D laser triangulation, structured light 3D scanning, and focus variation microscopy [47]. The advantage of these techniques is that they are highly automated and provide close to exact information to reproduce the artwork's surface. Blind and visually impaired people using a model designed using these techniques can perceive the direction of the strokes made by the artist, but often cannot recognize the objects. Only artworks made with simple strokes or rich in textures like splatter, impasto, or sgraffito are good candidates to be experienced with models designed using these techniques. Instead, we decided to use a semi-automated hybrid approach combining a technique known as shape from shading (SFS) [48]. SFS only requires a single image of the painting to generate the depth information to create a 2.5-dimensional model [49]. We chose this technique for three reasons: First, we do not need to have direct access to the artwork. Only a high-resolution image of the artwork is required to generate the depth information. Second, the process is automated and does not need specialized equipment like stereo cameras. Third, the output of the process is a greyscale height-map image that can be easily modified with any image editing software for corrections, or like in our case, to abstract, simplify or accentuate features and objects on the image. The process to design a 2.5-dimensional relief model to use with our IMG is graphically described in Figure 1 and is as follows:

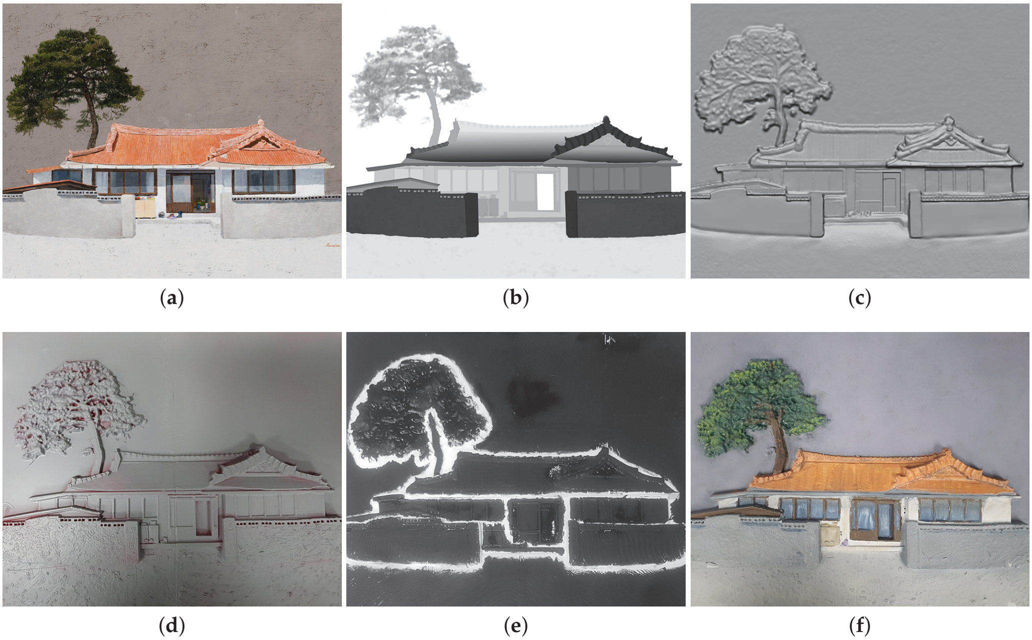

(**a**) (**b**) (**c**)
(**d**) (**e**) (**f**)

Figure 1. Touch sensitive 2.5D relief model fabrication process. (**a**) Original image; (**b**) Grey scale height-map; (**c**) 2.5D digital model; (**d**) 2.5D printed model; (**e**) Conductive paint coat (**f**) Completed 2.5D relief model.

1. A high resolution picture of the visual artwork is taken or obtained. Figure 1a.
2. The picture is processed using the SFS based methodology proposed in Furferi et al. [50] to obtain a grayscale height-map. Figure 1b.
3. The height-map image is modified using a digital image software to correct, modify, abstract, simplify or accentuate features and objects in the painting to improve their legibility and recognition by blind and visually impaired people.
4. A three-dimensional model is generated from the original picture and the height-map image using the ’Embossing Tool’ in the ZW3D 3D drawing software. Figure 1c.

Once the digital model of the relief model is ready, there are several methods to produce it. We chose to 3D print it using a fused filament fabrication 3D printer due to the variety and low cost of the materials, as well as the popularity and production services available (Figure 1d). It is also possible to 3D print the model using other 3D printing methods, as long as the material is non-conductive. Such methods are selective laser sintering (SLS) or stereolithography (SLA), which offer improved printing resolution at a cost trade-off. Another alternative is to use a CNC mill to carve the model out of a solid block of material.

The relief model is the primary input interface of our IMG. The touch interactivity on the relief surface is implemented by treating the surface with conductive paint. Conductive paints are electrical conductive solutions composed of dissolved or suspended pigments and conductive materials such as silver, copper, or graphite. We chose to use a water-based conductive paint that uses carbon and graphite for their conductive properties because of its easy to use, safe, and low cost nature. For our IMG, we used electric paint by bare conductive, but there are other suppliers in the market, as well as online guides to self-produce it.

Once the relief model has been 3D printed, making touch-sensitive areas is a simple procedure that only requires painting the areas that must be sensitive using conductive paint. The only requirement is to be careful to paint each touch-sensitive area isolated from the others, as seen in Figure 1e. If two treated areas with conductive paint overlap, they will act as one. The conductive paint dries at room temperature and does not require any special post-processing. One limitation of this method is that while extending or adding zones to the relief model is as simple as painting more areas or extending the existing ones, reducing or splitting existing ones is a more complicated process that involves scrapping or dissolving the paint. Therefore, it is recommended to plan the location and shape of the touch-sensitive areas. Each sensitive area must be connected with a thin conductive thread or wire to the circuit board. To this end, holes can be included in the model design before production or be made using a thin drill. Once the process is complete, the relief model can be sealed using a varnish or coating, preventing smudging and acting like a protective layer. It is possible to add subsequent layers of paint to produce a range of more aesthetic finishes, like a single color finish, a colored reproduction (Figure 1f), or different color palette combinations to improve visibility.

3.2.2. Control Board

The control board is the processing center of the IMG. It receives the touch sensor input from the 2.5D relief model described in Section 3.2.1 and peripherals, processes the signals, and provides audio output feedback. The control board is primarily composed of three components: An Arduino Uno microcontroller (Arduino, Somerville, MA, USA), a WAV Trigger polyphonic audio player board (SparkFun Electronics, Boulder, CO, USA) and an MPR121 proximity capacitive touch sensor controller (Adafruit Industries, New York, NY, USA). The wire leads from each of the touch sensitive areas of the relief model connect to one of the electrode inputs of the MPR121 integrated circuit. The MPR121 processes the capacitance of each of the touch areas in the relief model, which changes when the users touch the area, and it communicates touch and release events to the microcontroller through an I2C interface. One MPR121 integrated circuit is limited to 12 electrodes. It can only handle input for up to 12 touch areas. While this was enough for our prototypes, if more touch areas are required, up to four MPR121 can be connected by configuring different I2C addresses for a total of 48 touch areas. If more areas are needed, an I2C multiplexer, such as the TCA9548A (Adafruit Industries, New York, NY, USA), can be used to extend the number of supported touch zones. The microcontroller acts as the orchestrator of the control board. It receives input signals from the MPR121 and its general purpose input/output ports, processes them, and depending on the current state of execution, issues commands through its UART port to the audio board to trigger audio feedback. The WAV Trigger polyphonic audio player is a board that can play and mix up to 14 audio tracks at the same time and outputs the amplified audio through a mini-plug speaker connector. The audio files are read from an SD card and should be stored using WAV format.

3.2.3. External Hardware

Besides the relief model and the control board, the IMG is composed of an enclosure display. The enclosure was designed for different exploration scenarios. For example, for our preliminary test, a portable box-shaped enclosure is made of laser-cut acrylic. The box itself acts as an exhibit, the relief model is on its top surface, and the control board and electronics are in its interior. Headphones or speakers are connected to listen to the audio, and there is a button that the user can push to start using the IMG prototype. This prototype is meant to be placed on a desk to be used in a seated position during the early preliminary tests to make its use more comfortable for longer periods. For the IMG evaluation, we designed an exhibition display made of plywood for standing up use, as this is the more frequently used display arrangement in art museums and galleries. This version includes three physical buttons with labels in Braille to listen to use instructions, general information of the artwork, and to change the speed of the audio. Headphones are on the right side of the display. Depending on the size of the relief model or the floor space of the gallery, it might be difficult to explore the relief model if it is displayed horizontally or at a near angle, so a full-size vertical display was also developed, as seen in Figure 2c.

(**a**) (**b**) (**c**)

Figure 2. Interactive Multimodal Guide prototypes. (**a**) Portable IMG prototype; (**b**) Standing exhibition IMG; (**c**) Vertical exhibition IMG.

3.2.4. Interaction Design

Since there is no standard for interactive relief interfaces, and users are likely to lack previous experience with them, it is important to carefully design the interaction so that using the IMG is intuitive and easy to learn. A session with the IMG starts with the user already located in front of the display. The first task is to wear the exhibit's headphones. The exhibition stand only has a label in Braille inviting the user to wear the headphones and indicating their location. This is a barrier for blind and visually impaired people with limited Braille literacy. While it is possible to trigger a speaker to inform the user about the location of the headphones using a proximity sensor, from our user test experience, just verbally informing the user one time and maintaining consistency on the location is enough for users to find and wear the headphones independently across different exhibition stands. In our prototypes, we maintained consistency, by placing the headphones hanging on a hook at the right side of the exhibition display.

The interactive session with the IMG starts when the user either touches anywhere on the relief or presses the “Instructions” physical button on the surface to the right of the artwork relief model. At the beginning of the session, the user listens to a short instruction recording that suggests exploring the relief using both hands. Then, it instructs the user to double-tap to hear more localized detailed information about any point of interest in the relief model or triple tap to listen to localized sounds. The recording also introduces the functionality of the other two physical buttons on the surface. The “General description” button provides general information about the artwork. The “Audio Speed” button changes the speed of the audio narrations. The “Instruction” and “General description” narrations can be interrupted any time another button is pressed or by double or triple tapping on the relief model. This is intended to give freedom to the user to skip the narration if desired.

3.2.5. Information Hierarchy

To provide intuitive artwork information access, we divided the information into two layers:

1. General information: Refers to the general information of the artwork such as name, author, short visual description, and any information that is not already present in the artwork or related to information that can be accessed in a single point of interest.
2. Localized information is information related to a specific point of interest in the artwork such as the object name, detailed description, color, meaning, and their relationship with neighboring points of interest and their sound, among others.

The general information narration of the artwork is accessed only through the physical button on the IMG. Localized information is accessed by double or triple tapping on any of the points of interest in the relief model. Mapping the localized description to the point of interest being touched helps the user to relate what is touched (location, shape, and texture) to what is heard (localized information narration or sound). Sound design plays an important role in the IMG to communicate non-textual information. In collaboration with a music expert, background music was composed for each of the artworks to reflect the artwork’s general mood. This track is reproduced through the entire exploration session. Sound effects representing the objects in each of the points of interest are reproduced on demand. The objective of these sounds is to facilitate the formation of a mental image of the artwork, using familiar sounds instead of images like sighted people would do.

3.3. Evaluation

3.3.1. Accessible Exhibitions Using IMG and Tactile Graphics

We expanded our formative study to receive feedback on our interactive multimodal guide prototype and compare it with a tactile graphics approach as a reference.

3.3.2. Participants

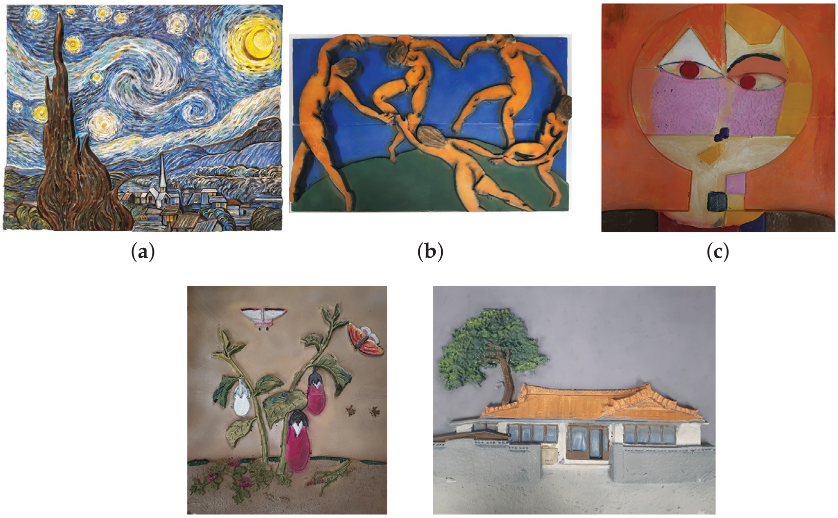
We recruited eighteen participants for the study and divided them into two groups. We held the study with the first group of seven participants at an accessible gallery at a social welfare center for blind and visually impaired people. At a later date, we performed the study with the second group of eleven participants at a school for blind and visually impaired people. Participant age ranged from 15 to 52, with an average of 27.7 years. All of the participants had previous experience using tactile graphics and stated having an interest in arts. None of the participants took part in the formative study. Other characteristics of the participants are described in Table 3. All the participants or their legal guardians gave signed informed consent based on the procedures approved by the Sungkyunkwan University Institutional Review Board.

Table 3. Characteristics of participants in our Standard Usability Scale evaluation study.

Participant	Sex	Age	Occupation	Sight
EP1	Female	16	High school student	Total vision loss
EP2	Female	16	High school student	Near vision loss
EP3	Female	19	High school student	Profound vision loss
EP4	Male	15	High school student	Total vision loss
EP5	Male	15	High school student	Total vision loss
EP6	Male	18	High school student	Total vision loss
EP7	Female	19	High school student	Profound vision loss
EP8	Female	16	High school student	Total vision loss
EP9	Male	17	High school student	Near vision loss
EP10	Male	18	High school student	Profound vision loss
EP11	Female	15	High school student	Total vision loss
EP12	Female	39	Worker	Total vision loss
EP13	Male	38	Worker	Total vision loss
EP14	Female	43	Worker	Total vision loss
EP15	Male	52	None	Near vision loss
EP16	Male	50	Worker	Near vision loss
EP17	Female	47	Housewife	Near vision loss
EP18	Female	45	Worker	Total vision loss

3.3.3. Materials and Apparatus

Two sets of test materials were prepared for the usability study. The IMG set is composed of five standing exhibition IMG prototypes similar to Figure 2b. Each prototype exhibits the 2.5D relief model of a distinct artwork from the selection in Figure 3. Since the participants may not had recently experienced visual artworks through tactile graphics, the second set of materials consisted of tactile graphics reproductions of the same artworks and was produced by a designer with extensive experience in the production of tactile graphics and reading materials for blind and visually impaired people. Descriptions of the artworks in Braille were provided side by side with the tactile graphics.

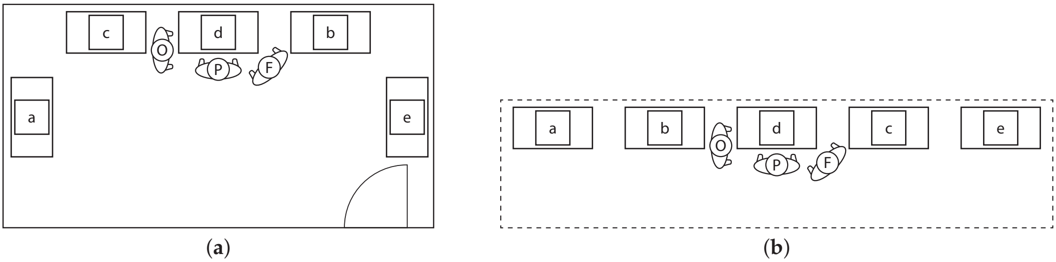


(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 3. Usability study IMG artwork models. (**a**) The Starry Night—Vincent van Gogh; (**b**) Dance—Henri Matisse; (**c**) Senecio—Paul Klee; (**d**) Flowers and Insects—Sin Saimdang; (**e**) Hyunsook’s House—Kim Yong-il.

3.3.4. Methodology

The first group study was held at an accessible gallery located in a social welfare center for blind and visually impaired people. The gallery has a permanent accessible exhibition, and we were able to install our test materials and perform our study in a temporal gallery next to the main gallery, and arranged them as shown in Figure 4a. The second group study was held at a school for blind and visually impaired people. The materials were installed in the main hall of the school, as shown in Figure 4b. The study was performed in the absence of other people.



(**a**) (**b**)

Figure 4. Usability study setup (P = Participant; F = Facilitator; O = Observer) (**a**) Accessible gallery setup; (**b**) School for blind and visually impaired people setup.

It began with a short introduction of our team and an interview with the participant to learn about their personal information, level of vision, interests in arts, and their experience at art museums and galleries. Participants were told that they would be experiencing visual artworks through different mediums and would be asked about their experience. A 2 × 2 Latin square test design was used to counterbalance the medium (tactile graphic or IMG) and presentation order, so that the participants would experience both mediums. The artwork selection was random among the five artworks prepared, and the participants responded to a standard usability scale survey immediately after each of the first two interactions with the exhibits. After the survey and a questionnaire, they could freely explore the rest of the exhibits. To replicate the experience that they would face at an art gallery, no training on how to use the exhibits was given to the participants. Only the location of the headphones in the IMG exhibit was communicated. Participants were able to freely explore the artwork exhibit for about ten minutes, after which, they completed the survey and moved to the next exhibit.

4. Results and Discussion

4.1. General Impressions

All the participants received the interactive multimodal guide and tactile exhibits well. The first impression of the IMG was much more exciting for the participants. They expressed surprise since, for most, it was the first time to use such a system, while reading tactile graphs was something they had already experienced. They eagerly expressed their desire to use both tactile graphics and IMG frequently at art galleries, and museums (Table 4-S1) and even demanded it, with expressions such as *"EP13: I can't understand why these (tactile graphics) are not available everywhere for every single artwork."*.

The IMG was considered extremely easy to use (Table 4-S3), mostly for two reasons; because it requires almost no effort to start using it, *"EP7: With this exhibit (IMG) you can feel the artwork from the beginning, you touch it, and it automatically starts telling things to you."*, and because it is easier to access and confirm information about their point of interest in the artwork directly *"EP11: I think one of the advantages is that with the speaking model (IMG), I can check what I'm touching by tapping two times right there, it is immediate. With the other one (tactile graphics), I need to go and read the Braille and come back, and sometimes I get lost in the graphic or with the Braille."* Having to switch between the Braille annotations, texture legends, and the tactile graphics was perceived as the largest factor to perceive the tactile graphics as unnecessarily complex (Table 4-S2).

Table 4. Tactile Graphics and Interactive Mulitmodal Guide Exhibits Standard Usability Scale report.

1	2	3	4	5	M	SD
S1. I think that I would like to use this system frequently.						
		2	8	8	4.33	0.69
			9	9	4.50	0.51
S2. I found the system unnecessarily complex.						
4	5	4	5		2.56	1.15
11	4	2		1	1.67	1.08
S3. I thought the system was easy to use.						
		5	8	5	4.00	0.77
			6	12	4.67	0.49
S4. I think that I would need the support of a technical person to be able to use this system.						
2	2	5	7	2	3.28	1.18
9	8			1	1.67	0.97
S5. I found the various functions in this system were well integrated.						
		2	6	10	4.44	0.70
			4	14	4.78	0.43
S6. I thought there was too much inconsistency in this system.						
4	6	5		3	2.56	1.34
2	7	3	3	3	2.89	1.32
S7. I would imagine that most people would learn to use this system very quickly.						
	2	3	9	4	3.83	0.92
		1	5	12	4.61	0.61
S8. I found the system very cumbersome to use.						
4	5	4	5		2.56	1.15
12	5	1			1.39	0.61
S9. I felt very confident using the system.						
		7	9	2	3.72	0.67
			3	15	4.83	0.38
S10. I needed to learn a lot of things before I could get going with this system.						
1	3	5	7	2	3.33	1.08
11	5	2			1.50	0.71
⬛	Tactile Graphics					
🟦	Interactive Multimodal Guide					

SUS score range from 1 (“Strongly disagree”) to 5 (“Strongly agree”).

Participants found the functions of both approaches well integrated (Table 4-S5). Participants were already used to exploring tactile graphics accompanied by Braille annotations. The simple touch interface on the artwork relief of the IMG coupled with the localized audio descriptions was well received. The participants expressed that hearing the localized audio while touching the 3D model area helped them to create a better spatial image of the shape and location of the object to the canvas. A couple of participants perceived background music.

One of them reported two effects; the first was that it made them think about the atmosphere of the scene in the artwork and the second was that it made her wonder about the time and circumstances that the artwork was made. *“EP8: When I heard the Korean traditional background music of the painting (Figure 3d) I could feel the solemnity of the painting and I wondered if the painter felt that way when making the painting”*.

All the participants expressed feeling very confident when using the IMG (Table 4-S10) because they could always revisit the points of interest quickly and trigger the audio descriptions or sounds to confirm the object that they are touching. For the tactile graphics, the opinion was divided between participants that felt very confident and those that didn’t because of the uncertainty of not being sure that they were correctly identifying the point of interest.

In general, the IMG was less cumbersome to use compared to the tactile graphics exhibit (Table 4-S8). Participants stated the following reasons: the difficulty of Braille, *"reading Braille is more difficult than listening to a conversation"*, the cognitive load of switching between the tactile graphic and Braille annotations: *"touching the object and getting its information is much better than having to read through Braille text and tactile graphics."* which adds up with each session: *"after trying several tactile graphics and Braille notes I felt more tired."*.

4.2. Interaction

One of our design requirements was to make interaction with the IMG as simple to learn and use as possible. Requiring to remember the location and use of buttons as well as gestures or commands can be burdensome for most people since it will be the first time that they use a device. Moreover, many users often skip instructions, even if they are short. Because of this, the IMG only has three user interactions, pressing buttons with a single-use, and double and triple tapping on the relief model to access localized information and audio. A simple interaction interface has its benefits. It makes the system easy to learn to use (Table 4-S7) and avoiding the feeling of the burden that can come when facing a new device (Table 4-S10) as evidenced by one participant's response, *"EP8: With the talking exhibit you don't need to know anything, you just stand there, touch something, and it starts talking to you about the picture."* By keeping consistency throughout the IMGs, once a user knows how to use one IMG, it knows how to use the rest. Unfortunately, tactile graphics have drawbacks. Experience goes a long way to read tactile graphics proficiently, and every time the user faces a new tactile graphic, it will need to learn the meaning of the texture and line styles to recognize their meaning. As expressed by one of the participants, *"EP15: You need to know Braille to read the tactile drawings with Braille and that takes time and effort."* Moreover, the lack of Braille proficiency affects the experience across all the exhibits, since the burden is on the user.

The participants reported a higher degree of inconsistency (Table 4-S6) for the IMG. Upon further investigation, we found out that it was due to a failure in some of the IMG prototypes to register some touch gestures correctly, causing the wrong audio feedback to trigger or not at all. Similarly, at the exhibition, not all the interactive zones in some of the artworks had ambient sound audio feedback, causing some users to believe that the system was malfunctioning or that their gesture was not recognized when they tapped the area and audio was not reproduced. Audio feedback should be added to the interaction zones that lack audio tracks, like empty or background space, to manage user expectations. Non-obtrusive audio or vibrotactile feedback could be added to help the user become aware that their input is sensing.

5. Conclusions and Future Work

In this work, we have presented the development of an interactive multimodal guide for improving the independent access and understanding of visual artworks. The IMG design was developed following the needs uncovered through a formative study in collaboration with people with vision impairments, art museums and gallery staff, and artists. Through an evaluation with eighteen participants, results demonstrate that the multimodal approach coupled with a simple to learn interaction interface is more effective in comparison to tactile graphics guides in providing independent access across a diverse style of artworks. Feedback collected during the multiple exhibition points in new directions for our work. As seen in Figure 5b, the IMG is sometimes used as a collaboration tool to socially interact with art. We would like to explore this possibility, as this could alleviate the perceived burden that some participants expressed when going to the art gallery with an acquaintance. Moreover, our current prototype was designed for use in an exhibition environment. Art educators at schools have expressed their interest in using the guide as an educational tool in class. To this end, more research is needed to explore the difference in audio description content and delivery methods to provide tailored information, while making it manageable for users with different content needs. The current prototypes only make use of tactile and audio modalities. We look forward to develop new experiences with other modalities such as smell, and explore how they might improve visual artworks exploration.


(**a**) (**b**)

Figure 5. Exhibition visitors using the interactive multimodal guide. (**a**) Stand alone use (**b**) Social interaction.

Author Contributions: Conceptualization, methodology, formal analysis, data curation, writing—original draft preparation, visualization, and project administration, L.C.Q.; Software, J.I.B. and L.C.Q.; Investigation, L.C.Q., J.I.B. and J.C.; Resources, writing—review and editing, supervision, funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2018 Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea grant number 2018M3C1B6061353.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Sungkyunkwan University (SKKU202011005-UE003 2020-11-05).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available within the article.

Acknowledgments: We would like to thank all volunteers for their participation and the reviewers for their insights and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IMG	Interactive Multimodal Guide
TG	Tactile Graphics
STEM	Science, Technology, Engineering and Mathematics

Appendix A

Table A1. Interactive tactile graphics and multimodal guide projects.

Author - Name	Input	Output	Focus
Parkes [21] NOMAD	- Touch (Surface)	- Tactile overlay- Verbal descriptions	- Mathematics, Geometry, Geography, and Biology Education- Orientation & Mobility
Landau et al. [22] The Talking Tablet	- Touch (Surface)	- Tactile overlay- Verbal descriptions	- Mathematics, Geometry, Geography, and Biology Education- Orientation & Mobility
Gardner et al. [23] IVEO	- Touch (Surface)	- Tactile overlay- Verbal descriptions	- Education & Scientific Diagrams
Taylor et al. [24]	- Touch (Touchscreen)	- Tactile overlay- Verbal descriptions	- Orientation & Mobility
Gotzelmann et al. [25] LucentMaps	- Touch (Touchscreen)- Voice	- Tactile overlay- Visual agumentation	- Orientation & Mobility
Brule et al. [26] MapSense	- Touch (Touchscreen)- Tokens (Capacitive)	- Tactile overlay- Smell and taste infused tangible tokens- Verbal descriptions	- Geography Education- Map Exploration- Orientation & Mobility
Shen et al. [27] CamIO	- Touch (Mounted camera)	- Tactile graph- Tactile 3D Map-Tactile ObjectVerbal descriptions	- Access to 3D objects- Map Exploration- Access to appliances- Access to documents
Baker et al. [28]Tactile Graphics with a Voice	- Touch (Mobile Camera)	- Tactile graph- Verbal descriptions	- STEM Education
Baker et al. [29] Tactile Graphics with a Voice	- Touch (Wearable Camera)- Voice	- Tactile graph- Verbal descriptions	- STEM Education
Fusco et al. [30] The Tactile Graphics Helper	- Touch (Mobile Camera)- Voice	- Tactile graph- Verbal descriptions	- STEM Education- Map Exploration
Holloway et al. [31]	- Touch (Embedded capacitive sensors)	- Tactile 3D Map- Verbal descriptions	- Orientation & Mobility
Vaz et al. [42]	- Touch (Embedded capacitive sensors)	- Tactile Objects- Verbal descriptions- Visual augmentation	- Museum Object Exploration
Anagnostakis et al. [41]	- Touch (PIR and touch sensors)	- Tactile Objects- Verbal descriptions	- Museum Object Exploration
Leporini et al. [43]	- Touch (Physical buttons)	- Tactile 3D Map & Model- Verbal descriptions	- Archeological site exploration- Artwork exploration
Reichinger et al. [44–46]	- Touch (Camera)- Hand gestures (Camera)	- Tactile 3D Artwork Model- Verbal descriptions	- Artwork exploration
Landau et al. [35] The Talking Tactile Pen	- Touch (Pen device)	- Tactile graph- Verbal descriptions	- STEM Education- Map Exploration- Games
D'Agnano et al. [36] Tooteko	- Touch (Ring NFC reader)	- Tactile 3D model- Verbal descriptions	- Archelological site exploration- Artwork exploration

References

1. Maintaining Accessibility in Museums. Available online: https://www.ada.gov/business/museum_access.htm (accessed on 9 November 2020).
2. United Nations Department of Economic and Social Affairs (UN-DESA). Convention on the Rights of Persons with Disabilities (CRPD). Available online: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html (accessed on 9 November 2020).
3. The Guggenheim Museums and Foundation. Mind's Eye Programs. Available online: https://www.guggenheim.org/event/event_series/minds-eye (accessed on 9 November 2020).
4. The Metropolitan Museum of Art. For Visitors Who Are Blind or Partially Sighted. Available online: https://www.metmuseum.org/events/programs/access/visitors-who-are-blind-or-partially-sighted. (accessed on 9 November 2020).
5. Museum of Moden Art. Accessibility. Available online: https://www.moma.org/visit/accessibility/index#individuals-who-are-blind-or-have-low-vision. (accessed on 9 November 2020).
6. Hoyt, B. Emphasizing Observation in a Gallery Program for Blind and Low-Vision Visitors: Art Beyond Sight at the Museum of Fine Arts, Houston. Available online: https://dsq-sds.org/article/view/3737/3278 (accessed on 9 November 2020).
7. Krantz, G. Leveling the Participatory Field: the Mind's Eye Program at the Guggenheim Museum. Available online: https://dsq-sds.org/article/view/3738/3277 (accessed on 9 November 2020).
8. Mesquita, S.; Carneiro, M.J. Accessibility of European museums to visitors with visual impairments. *Disabil. Soc.* **2016**, *31*, 373–388. [CrossRef]
9. Phillips, A.; Beesley, L. Braille Profiling Project. 2011. Available online: https://www.rnib.org.uk/sites/default/files/braille_profiling.doc (accessed on 9 November 2020).
10. The Braille Literacy Crisis in America. Available online: https://www.nfb.org//images/nfb/documents/pdf/braille_literacy_report_web.pdf. (accessed on 9 November 2020).
11. Museo del Prado. Touching the Prado. Available online: https://www.museodelprado.es/en/whats-on/exhibition/touching-the-prado/0d94a9bf-07d7-491a-866a-37976169f929 (accessed on 9 November 2020).
12. The Andy Warhol Museum. Available online: https://www.warhol.org/accessibility-accommodations/ (accessed on 9 November 2020).
13. Candlin, F. The dubious inheritance of touch: Art history and museum access. *J. Vis. Culture* **2006**, *5*, 137–154. [CrossRef]
14. Holloway, L.; Marriott, K.; Butler, M.; Borning, A. Making Sense of Art: Access for Gallery Visitors with Vision Impairments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–12
15. Schiff, W.; Foulke, E. *Tactual Perception: A Sourcebook*; Cambridge University Press: Cambridge, UK, 1982.
16. BANA. Guidelines and Standards for Tactile Graphics. Available online: http://www.brailleauthority.org/tg/web-manual/index.html (accessed on 9 November 2020).
17. Sinha, P.; Kalia, A.A. Tactile picture recognition: Errors are in shape acquisition or object matching? *See. Perc.* **2012**, *25*, 287–302. [CrossRef] [PubMed]
18. Brock, A.M.; Truillet, P.; Oriola, B.; Picard, D.; Jouffrais, C. Interactivity improves usability of geographic maps for visually impaired people. *Hum. Comput. Interact.* **2015**, *30*, 156–194. [CrossRef]
19. Landau, S.; Wells, L. Merging tactile sensory input and audio data by means of the Talking Tactile Tablet. In Proceedings of the EuroHaptics, Dublin, Ireland, 6–9 July 2003; pp. 414–418.
20. Tatham, A.F. The design of tactile maps: Theoretical and practical considerations. In *Mapping the Nations: Proceedings, Volume 1*; Rybaczuk, K., Blakemore, M.J., Eds.; International Cartographic Association: Bern, Switzerland, 1991.
21. Parkes, D. "NOMAD": An audio-tactile tool for the acquisition, use and management of spatially distributed information by partially sighted and blind persons. In Proceedings of the Second International Conference on Maps and Graphics for Visually Disabled People, Nottingham, UK, 20 April 1988; pp. 24–29.
22. Landau, S.; Holborow, R.; Jane, E. Use of the talking tactile tablet for delivery of standardized tests. In Proceedings of the CSUN 2004, Los Angeles, CA, USA, 15–20 March 2004.
23. Gardner, J.A.; Bulatov, V. Scientific Diagrams Made Easy with IVEOTM. In *Computers Helping People with Special Needs*; Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I., Eds.; Springer: Berlin, Germany, 2006; pp. 1243–1250.
24. Taylor, B.; Dey, A.; Siewiorek, D.; Smailagic, A. Customizable 3D printed tactile maps as interactive overlays. In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, Reno, NV, USA, 23–26 October 2016; pp. 71–79.
25. Götzelmann, T. Visually Augmented Audio-Tactile Graphics for Visually Impaired People. *ACM Trans. Access. Comput.* **2018**, *11*, 1–31. [CrossRef]
26. Brule, E.; Bailly, G.; Brock, A.; Valentin, F.; Denis, G.; Jouffrais, C. MapSense: multi-sensory interactive maps for children living with visual impairments. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 445–457.
27. Shen, H.; Edwards, O.; Miele, J.; Coughlan, J.M. *CamIO: A 3D Computer Vision System Enabling Audio/Haptic Interaction with Physical Objects by Blind Users*; Association for Computing Machinery: New York, NY, USA, 2013. [CrossRef]

28. Baker, C.M.; Milne, L.R.; Scofield, J.; Bennett, C.L.; Ladner, R.E. Tactile graphics with a voice: using QR codes to access text in tactile graphics. In Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility, Rochester, NY, USA, 20–22 October 2014; pp. 75–82.
29. Baker, C.M.; Milne, L.R.; Drapeau, R.; Scofield, J.; Bennett, C.L.; Ladner, R.E. Tactile graphics with a voice. *ACM Trans. Access. Comput. (TACCESS)* **2016**, *8*, 1–22. [CrossRef]
30. Fusco, G.; Morash, V.S. The tactile graphics helper: Providing audio clarification for tactile graphics using machine vision. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, Lisbon, Portugal, 26–28 October 2015; pp. 97–106.
31. Holloway, L.; Marriott, K.; Butler, M. Accessible Maps for the Blind: Comparing 3D Printed Models with Tactile Graphics. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–13. [CrossRef]
32. Holloway, L.; Marriott, K.; Butler, M.; Reinders, S. 3D Printed Maps and Icons for Inclusion: Testing in the Wild by People who are Blind or have Low Vision. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility, Pittsburgh, PA, USA, 28–30 October 2019; pp. 183–195.
33. Gual, J.; Puyuelo, M.; Lloveras, J. Three-dimensional tactile symbols produced by 3D Printing: Improving the process of memorizing a tactile map key. *Br. J. Vis. Impair.* **2014**, *32*, 263–278. [CrossRef]
34. Gual, J.; Puyuelo, M.; Lloveras, J. The effect of volumetric (3D) tactile symbols within inclusive tactile maps. *Appl. Ergon.* **2015**, *48*, 1–10. [CrossRef] [PubMed]
35. Touch Graphics. TTPen Stem Binder. 2015. Available online: http://touchgraphics.com/portfolio/ttpen-stem-binder/ (accessed on 9 November 2020).
36. D'Agnano, F.; Balletti, C.; Guerra, F.; Vernier, P. Tooteko: A case study of augmented reality for an accessible cultural heritage. Digitization, 3D printing and sensors for an audio-tactile experience. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *40*, 207–213. [CrossRef]
37. Axel, E.S.; Levent, N.S. *Art Beyond Sight: A Resource Guide to Art, Creativity, and Visual Impairment*; American Foundation for the Blind: Arlington, VA, USA, 2003.
38. Cho, J.D.; Quero, L.C.; Bartolomé, J.I.; Lee, D.W.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2021**, *46*, 103–116. [CrossRef]
39. Volpe, Y.; Furferi, R.; Governi, L.; Tennirelli, G. Computer-based methodologies for semi-automatic 3D model generation from paintings. *Int. J. Comp. Aided Eng. Technol.* **2014**, *6*, 88–112. [CrossRef]
40. Hinton, R. Use of tactile pictures to communicate the work of visual artists to blind people. *J. Vis. Impair. Blind.* **1991**, *85*, 174–175. [CrossRef]
41. Anagnostakis, G.; Antoniou, M.; Kardamitsi, E.; Sachinidis, T.; Koutsabasis, P.; Stavrakis, M.; Vosinakis, S.; Zissis, D. Accessible museum collections for the visually impaired: combining tactile exploration, audio descriptions and mobile gestures. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Florence, Italy, 6–9 September 2016; pp. 1021–1025.
42. Vaz, R.; Fernandes, P.O.; Cecília, A. Designing an Interactive Exhibitor for Assisting Blind and Visually Impaired Visitors in Tactile Exploration of Original Museum Pieces. *Proc. Comp. Sci.* **2018**, *138*, 561–570. [CrossRef]
43. Leporini, B.; Rossetti, V.; Furfari, F.; Pelagatti, S.; Quarta, A. Design Guidelines for an Interactive 3D Model as a Supporting Tool for Exploring a Cultural Site by Visually Impaired and Sighted People. *ACM Trans. Access. Comp. (TACCESS)* **2020**, *13*, 1–39. [CrossRef]
44. Reichinger, A.; Fuhrmann, A.; Maierhofer, S.; Purgathofer, W. A Concept for Re-useable Interactive Tactile Reliefs. In *Computers Helping People with Special Needs*; Miesenberger, K., Bühler, C., Penaz, P., Eds.; Springer: Cham, Switzerland, 2016; pp. 108–115.
45. Reichinger, A.; Fuhrmann, A.; Maierhofer, S.; Purgathofer, W. Gesture-Based Interactive Audio Guide on Tactile Reliefs. In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, Reno, NV, USA, 24–25 October 2016; pp. 91–100. [CrossRef]
46. Reichinger, A.; Carrizosa, H.G.; Wood, J.; Schröder, S.; Löw, C.; Luidolt, L.R.; Schimkowitsch, M.; Fuhrmann, A.; Maierhofer, S.; Purgathofer, W. Pictures in Your Mind: Using Interactive Gesture-Controlled Reliefs to Explore Art. *ACM Trans. Access. Comput.* **2018**, *11*, 1–39. [CrossRef]
47. Elkhuizen, W.S.; Callewaert, T.W.; Leonhardt, E.; Vandivere, A.; Song, Y.; Pont, S.C.; Geraedts, J.M.; Dik, J. Comparison of three 3D scanning techniques for paintings, as applied to Vermeer's 'Girl with a Pearl Earring'. *Herit. Sci.* **2019**, *7*, 89. [CrossRef]
48. Zhang, R.; Tsai, P.S.; Cryer, J.E.; Shah, M. Shape-from-shading: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 690–706. [CrossRef]
49. Prados, E.; Faugeras, O. Shape From Shading. In *Handbook of Mathematical Models in Computer Vision*; Paragios, N., Chen, Y., Faugeras, O., Eds.; Springer: Boston, MA, USA, 2006; pp. 375–388. [CrossRef]
50. Furferi, R.; Governi, L.; Volpe, Y.; Puggelli, L.; Vanni, N.; Carfagni, M. From 2D to 2.5D i.e. from painting to tactile model. *Graph. Model.* **2014**, *76*, 706–723. [CrossRef]

electronics

Article

ColorWatch: Color Perceptual Spatial Tactile Interface for People with Visual Impairments

Muhammad Shahid Jabbar [1], Chung-Heon Lee [1] and Jun Dong Cho [1,2,*]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Gyeonggi-do, Korea; eeshahid@skku.edu (M.S.J.); yi940715@g.skku.edu (C.-H.L.)
[2] Department of Human ICT Convergence, Sungkyunkwan University, Suwon 16419, Gyeonggi-do, Korea
* Correspondence: jdcho@skku.edu

Abstract: Tactile perception enables people with visual impairments (PVI) to engage with artworks and real-life objects at a deeper abstraction level. The development of tactile and multi-sensory assistive technologies has expanded their opportunities to appreciate visual arts. We have developed a tactile interface based on the proposed concept design under considerations of PVI tactile actuation, color perception, and learnability. The proposed interface automatically translates reference colors into spatial tactile patterns. A range of achromatic colors and six prominent basic colors with three levels of chroma and values are considered for the cross-modular association. In addition, an analog tactile color watch design has been proposed. This scheme enables PVI to explore artwork or real-life object color by identifying the reference colors through a color sensor and translating them to the tactile interface. The color identification tests using this scheme on the developed prototype exhibit good recognition accuracy. The workload assessment and usability evaluation for PVI demonstrate promising results. This suggest that the proposed scheme is appropriate for tactile color exploration.

Keywords: color identification; tactile perception; cross modular association; universal design; accessibility; people with visual impairment

Citation: Jabbar, M.S.; Lee, C.-H.; Cho, J.D. ColorWatch: Color Perceptual Spatial Tactile Interface for People with Visual Impairments. *Electronics* **2021**, *10*, 596. https://doi.org/10.3390/electronics10050596

Academic Editor: Calogero Maria Oddo

Received: 3 February 2021
Accepted: 28 February 2021
Published: 4 March 2021

1. Introduction

People with visual impairments (PVI) can form concepts regarding real-world object properties, including abstract conceptualization and relative differences between colors [1–3]. Symbolism, preferences, terminology, and other daily life concepts associate with colors by engaging psychological and aesthetic stimuli. These color associations serve a significant role in culture, faith, art, commercial branding, and everyday lifestyle. Therefore, comprehensive color information is necessary for PVI to think, consider, make actions, and cause reactions more prudently [4].

Color recognition through non-visual stimuli has been an active research area for PVI [5]. The color-sound cross modular associations convey visual color information through auditory senses. These associations code color characteristics of hue, chroma, and value through a combination of music instrument, music tone, pitch variations, etc. The auditory codings might be used for conveying color information independently, or in conjunction with tactile sense. Cho et al. [6] have studied these relationships and proposed two sound coding color melodies. They considered the tone, intensity, and pitch of melody sounds extracted from classic music to express the brightness and saturation of colors. The sound code system represented 18 chromatic and five achromatic colors with using classical music sounds played on different instruments. While using sound to depict color, tapping a relief-shaped embossed outline area transforms the color of that area into the sound of an orchestra instrument. Furthermore, the overall color composition of Van Gogh's "The Starry Night" was expressed as a single piece of music that accounted for color using the tone, key, tempo, and pitch of the instruments. The shape can be distinguished by touching it with a hand, but the overall color composition can be conveyed as a single piece of music, thereby reducing the effort required to recognize color from that needed to touch each pattern one by one. Gilbert et al. confirmed that humans have an association mechanism that unconsciously associates a specific scent with a specific color through olfactory sense [7]. Temperature differences have also been studied to convey color information to PVI. However, as the temperature is a single modality, merely one color characteristic can be coded through temperature levels. Bartolomé et al. [8] proposed thermal stimuli to code depth of colors in visual artworks. Cavadini et al. proposed an automatic color perception system, which provides haptic feedback at three points on the palm corresponding to Red, Green, and Blue channel values of detected colors through color sensors [9]. This method uses a wearable glove, and wearing the glove can hinder everyday tasks for PVI. In addition, the RGB color system uses a combination of red, green, and blue channel numerical values to represent any arbitrary color, which does not match the human perception of colors. Researchers reflect that spatial perception, tactile perception, and linguistics engage the brain's visual cortex. In addition, not having to process the visual stimuli makes PVIs' performance superior for tactile tasks [10], which suggests the preference of tactile stimuli for PVI perception. The cognitive aspects of multi-sensory activation and recognition by the sense of touch are covered by Lawrence et al. [11].

1.1. Tactile Representation of Colors

Tactile color patterns (TCP) are embossed surface patterns for conveying color information through touch to PVI. TCPs might be helpful as they can be used in conjunction with other tactile modalities, like contour information or object boundaries in the artwork. They offer immediate color perception by tapping onto the artwork relief pattern, unlike audio description which needs to be triggered. TCPs as an assistive tool for visual aspects of artwork can be supplemented to the audio description; this helps shorten the audio description and improve localization of artworks' objects information. Ramsamy-Iranah et al. [12] designed color symbols for children. The design process for the symbols was influenced by the children's prior knowledge of shapes and links to their surroundings. For example, a small square box was associated with dark blue reflecting the blue square soap, a circle represented red as it was associated with the 'bindi' [13]. Yellow was represented by small dots reflecting the pollen of flowers. Since orange is a mixture of yellow and red, circles of smaller dimension were used to represent orange. Horizontal lines represented purple and curved lines were associated with green representative of bendable grass stems. Shin et al. [4] coded nine colors (pink, red, orange, yellow, green, blue, navy, purple, brown, and achromatic) using a grating orientation (a regularly spaced collection of identical, parallel, elongated elements). The texture stimuli for color were structured by matching variations of orientation to hue, width of the line to chroma, and the interval between the lines to value. The eight chromatic colors were divided into 20º angles and achromatic at 90º. Each color has nine levels of value and of chroma. Levels 1–4 used a different grating interval to represent value, levels 6–9 used a different grating width to represent chroma, and level 5 represented the pure hue. In the survey on whether or not 3D printed colors were distinguished by texture, color identification tests were performed on five visually impaired people to distinguish the direction, width, and spacing of the proposed color patterns. Adjacent color hues in this scheme are oriented at an angular distance of 20°, but research suggests that tactile accuracy for grating orientation is significantly distinguishable for 30° or 45° angle [14,15]. The Munsell color system based TCP schemes by using ideographic characters were proposed by Cho et al. [16]. They employed an experimental investigation and adaption based approach for representing wide color gamut of 29 and 53 colors shades in the basic and extended versions for three TCP schemes, respectively. Taras et al. [17] presented a color code created for viewing on braille devices. The primary colors, red, blue, and yellow, were each coded by two dots. Mixed colors, for example violet, green, orange, and brown, were coded as combinations of dots representing the primary colors. In addition, the light and dark shades were added by using the 2nd and 3rd dots in the left column of the Braille cell. In addition, color can be represented by using the spatial color wheel that expresses the angular orientation. This is another way of color information transmission. Our proposed system codes a similar number of colors compared to other color patterns. Moreover, the color wheel depicts essentially the visible spectrum of colors enclosed by a circle, and is a useful tool for describing what happens when you mix paints together, complementary color relationships, and adjacent colors.

It can be difficult for individuals with visual impairments to fully participate in the visual arts due to the lack of inclusion and assistive technologies. Their participation in the visual culture of the world and visual art is important as the inclusion opportunities improve their life quality and help them gain skills crucial to their education and employment opportunities [18]. The visual centricity of exhibitions and museums has typically been a barrier regarding visit and appreciation of PVIs [19]. However, as “The event happens as a question mark before happening as a question (Lyotard, 1989: 197),” many of these institutions are now focusing on accessibility to enhance PVIs’ experiences by incorporating assistive technologies, contextual information delivery, and multisensory experiences [20].

The TCP schemes in recent literature are mainly focused on tactile color translation for artwork, whereas the significance of colors in daily life imparts the need to convey comprehensive color information to PVI for them to participate in society more prudently. Most of these schemes present static interpretation of colors, which causes the need to arise to develop assistive technologies that can dynamically translate detected colors from real-life objects or artwork for PVI. Moreover, the usability of these tactile patterns is limited, as learnability of these tactile patterns may be required for individual TCP associations. In this paper, we propose a tool for people with visual impairment (i.e., congenital blind people who have not experienced color and the acquired blind people whose color has disappeared from memory) that can intuitively recognize and understand the three elements of color, that is, color hue, lightness, and saturation, taking advantage of the timepiece watch design.

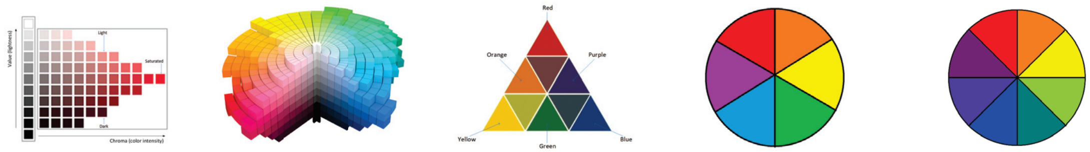
1.2. Timepiece Watches for PVI

There are many wearable wristwatches for PVI in the context of timepiece operation. Nevertheless, the common way of applying it has been using the braille interface for conveying numeric braille patterns similar to the digital watches for sighted people, since interactable hands of the traditional analog watch are fragile and prone to damage by heavy touch. The electromagnetic solenoid based cost-effective electronic braille display was proposed by Adnan et al. [21]. Tyler [22] used the 4-dot condensed braille code to introduce a design scheme of wearable timepiece wristwatch for PVI. The mechanical design was adapted to represent numeric braille cells for time and date, with time/date adjustment and alarm options. Dot watch [23] is a commercially designed smartwatch based on a similar concept which uses braille as its interface using braille coding. In addition to presenting time information by tactile stimuli, the Dot watch can be connected to a smartphone via Bluetooth to perform few basic tasks such as caller identification, and either pick or decline the call. Another design scheme consisting of a disk, a plate, an actuator, and a plurality of four pins mounted to a slide within the four respective holes was proposed for the braille timepiece wristwatch by Anderson et al. [24]. A haptic feedback scheme for a digital smartwatch display was proposed by Twyman et al. [25], which engaged side-mounted piezoelectric actuators to cause glass screen vibrations via ultrasonic frequencies. The braille literacy of PVI is on the decline, and it is projected to decrease in upcoming years, with the widespread use of smart devices. Velázquez [26] provided a comprehensive study on workload, learnability, design concept understanding, and latest advancements in wearable assistive devices for blind people and recognized low user acceptance for these devices. This led our research to investigate the durable design of the analog watch and its possible mapping with color perception for the PVI.

1.3. Review of the Color Systems

The Munsell color system arranges colors by accounting for the human visual response into systematic color space [27]. It considers hue, chroma, and lightness as three properties of color for color space organization as shown in Figure 1a. The hues or basic colors are arranged in a circular manner placed apart at each horizontal circle Figure 1b. Chroma or saturation of color is measured by the distance from the center of the circle to the edge. High chroma implies clearness of color or pure colors, while low chroma implies less saturated color with the lowest chroma colors being achromatic. The lightness or value of colors varies vertically where white is represented by the highest lightness and black holds the lowest lightness value.

The Goethe's color triangle is an excellent model for color relationships and the relative differences between them, such as additive color mixing and complementary colors [28]. These differences can be interactively simulated. This model arranges colors as primary, secondary, and tertiary colors. The three vertices of triangle are associated by primary colors of red, blue, and yellow. The secondary colors orange, green, and purple are obtained by mixing primary colors on either side of them (Figure 1c). The tertiary colors are obtained by mixing primary colors adjacent to them. Goethe arranged the primary, secondary, and tertiary colors based on physical grounds, as well as for their emotional content linked with them. An American educator Josef Albers extended Goethe's work for studying and teaching colors through an experimental way [29]. The particular arrangement of Goethe's color triangle retains the particular emotional or psychological states as per his description. This color arrangement is also similar to the RYB color model [30]. Its warm colors are red, orange, and yellow, and its cool colors are green, blue, and purple. The three primary colors in the RYB color model are red, yellow, and blue. Mixing the primary colors causes the mixtures to absorb light wavelengths to create other colors. The three secondary colors are orange, green, and purple. Mixing the three primary colors together creates an almost black color. We have considered the color gamut for our proposed system based on these color models as shown in Figure 1d,e for simplified and extended versions of watch patterns, respectively. This primary and secondary color information is extended with the integration of color tones of light, saturated, and dark from Munsell's color system.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 1. (**a**) Lighted (L), saturated (S), and Dark (D) color tones for red and achromatic colors. (**b**) 3D representation of Munsell color system renotations with a slice cut away for visualization [31]; (**c**) Goethe's Colour Triangle with primary and secondary colors; (**d**) color wheel for simplified watch pattern; (**e**) color wheel for extended watch pattern.

1.4. Proposed System

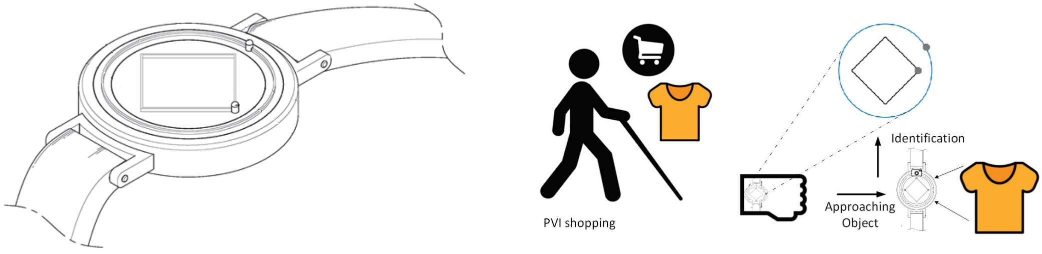
We have investigated tactile actuation for co-centric protruding points at different angular positions in our preliminary study [32]. The swell-paper relief pattern utilized dots on the boundary of a circle and a square, with uniform angular spacing between tactile dots at 45°and 90°, respectively. The focus group experiments revealed the discernment ability for angular tactile patterns, and the learnability of colors assigned to those patterns. Based on the outcomes of the preliminary study, we propose an integrated color pattern scheme for angular tactile color translation. The color gamut codes primary colors (red, blue, and yellow) and secondary colors (orange, green, and purple) from Goethe's color triangle. These six basic color hues are further coded into three levels of chroma and value (light, saturated, dark) as the color tone for each of the color hues. In addition, six levels of gray-scale colors ranging from white to black are included as achromatic colors from Munsell's color system. These integrated colors are coded into an angular tactile color pictogram, wherein the orientation of the tactile dot determines the color information. Our proposed system also integrates the TCP with an analog wearable watch for PVI.

The ColorWatch interface consists of two disks in distinct shapes of round and square, each with a tactile dot at their boundary, the angular position of which indicates the color hue and color tones, respectively. In contrast to related works on TCPs, the spatial tactile system automatically interprets reference color. The reference color acquired through a color sensor from artwork or real-life objects actively alters the angular position of rotatable disks, developing the tactile pattern corresponding to reference color for the appreciation of PVIs. The cross-modular association of tactile color perception considers tactile actuation, design aspects of color significance, and color placement. The proposed analog watch design with TCP integration is intuitively understandable, which makes it easily learnable for PVI. A prototype for the proposed interface has been developed and color identification tests have been performed. The results for identification and usability tests, work-load assessment, and qualitative feedback suggest that the developed scheme can be helpful for PVIs in tactile color perception.

2. Methodology

2.1. Concept Design

We propose the design of ColorWatch, which is a wearable wrist-worn analog tactile device for people with visual impairments (PVI). The concept of ColorWatch design includes two disks that can be rotated independently: one marked square disk which is on top of and smaller in size than a round disk, which functions as an hour hand in analog watches while the marked round disk functions as a minutes hand. Both disks are marked at one position each, with markers pointing to the reference positions for time. These disks and marked pointers can be touched by PVIs to convey tactile information. In addition to conventional time checking, the function of ColorWatch extends to represent object colors by the tactile interface. An integrated color sensor or smartphone camera can be used to detect color information of a reference object, which can then be wirelessly communicated to the ColorWatch. Hybrid smartwatches available in the market already exhibit wireless communication and control capabilities along with traditional watch hands. The ColorWatch design proposed a square and a round disk instead of conventional arms to enable PVIs to interact with them without external protective glass and minimizing the risk of damage or heavy touch to conventional arms. The control circuitry in ColorWatch can then direct the watch disks to point to the corresponding position associated with the detected color and color tone. A maintained contact push toggle switch can be used to toggle between timepiece and ColorWatch mode. The ColorWatch tactile design is effective for drawing less attraction to PVI, and for not being disruptive to others in contrast to talking watches. The analog disk design of ColorWatch also makes it easy to read, less fragile, and less susceptible to damage in contrast to a traditional analog watch hands. Figure 2a displays the schematic configuration and perspective view of the ColorWatch design concept, with Figure 2b outlining an application scenario for the ColorWatch. More features can theoretically be added to the ColorWatch; such as smartphone call notifications, incoming call pick and drop, and alarms through haptic feedback. The push button and mechanical dial of the ColorWatch can also hypothetically function as a tactile input method for smartphone scrolling. In this setting, PVI can wirelessly navigate smartphone features through ColorWatch dials discrete rotation, select desired options through one or two buttons or ColorWatch, and get audio feedback through smartphone speakers. However, this study focuses on the automatic spatial color translation framework for the PVI.



(**a**)

(**b**)

Figure 2. Perspective view of ColorWatch design concept; (**a**) PVI use scenario of ColorWatch during shopping (**b**).

2.2. Proposed Color Selection and Tactile Representation Scheme

The tactile association of color proposed in this study considers six distinct hues namely red (R), orange (O), yellow (Y), green (G), blue (B), and purple (P) from Goethe's color triangle. These hues are represented by the round disk of the prototype which dynamically points to the reference color. It can be seen from Figure 3 that complementary colors are arranged on opposite ends to each other, and mixing any two primary colors results in the secondary color between them. An emerged tactile dot of half-sphere on edge of the round disk resembling braille embossment is used as a pointer for tactile color marking. These tactile-color associations are further expanded by Munsell color systems's three dimensions of light (L), saturated (S), and dark (D) as a color tone of each monochromatic basic color or hue from Goethe's color triangle. The color tones are represented by the square disk of the ColorWatch design, where the emerged tactile dot at one corner points to the reference color tones, marked 90 degrees apart at 0, 90, and 180 degrees for L, S, and D, respectively. The remaining position at 270 degrees of square disk pointer is used to represent achromatic colors. When a reference to achromatic color is detected on the square disk, the round disk represents each of the achromatic colors given by White, Gainsboro, Light Gray, Dark Gray, Dim Gray, and Black (Table 1). The arrangement of these achromatic colors, and the six color-hues for monochromatic colors on round disk, depending on the color mode is provided in Table 2.

The analog clock dial design holds twelve marks separated by 30 degrees universally. These six markers for round disk and four markers for square disk can be incorporated on the analog clock design as the common factor of twelve markers. This association of perceptual colors and their tactile placement can be very convenient for learnability and ease of use of users and the user-centric design.

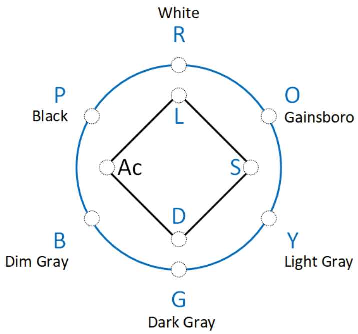


Figure 3. ColorWatch pattern for color hues and color tones. Blue and black text represents monochromatic and achromatic color modes, respectively.

Table 1. Colors tones or color mode according to the angle of the square disk.

Angle (Square)	Color Tone	Color Mode
0	Light	Monochromatic
90	Saturated	Monochromatic
180	Dark	Monochromatic
270	Achromatic	Achromatic

Table 2. Round disk angle chroma-lightness levels for monochromatic and achromatic colors (0–180° and 270° on the square disk, respectively).

Angle (Round)	Monochromatic Mode Color	Achromatic Mode Color
0	Red	White
60	Orange	Gainsboro
120	Yellow	Light Gray
180	Green	Dark Gray
240	Blue	Dim Gray
300	Purple	Black

2.3. Materials

We have developed a large-scale hardware prototype based on the ColorWatch concept which exhibits the represented color while exploring different objects. Figure 4 displays the working principle of a large-scale prototype. The two rotating disks in the round and square shape are used to represent both levels of primary colors and color tones as shown in Figure 5a. Figure 5b provides a demonstration for the prototype in use. The hardware prototype is a 110 × 170 × 70 cm encased acrylic box that encapsulates the control electronic circuit board, two stepper motors, and their motor drivers (Figure 5c). A set of wires is extended from the box which connects the color sensor module (Figure 5d). The ISL29125 RGB color sensor is used in the color sensor module for color data acquisition. The color sensor is designed to operate in diverse luminance environments ranging from darkrooms to sunlight by rejecting IR in light sources. The color sensor has low-power and needs 56 and 0.5 μA current for operation and power-down mode, respectively. The integrated ADC of the color sensor also rejects flicker caused by artificial light sources. The 2.25–3.6 V logic levels of the color sensor need to be converted if used with a 5 V Arduino board, and a bi-directional logic level converter is used for this purpose. The color sensor is housed in an acrylic casing that limits the incoming light from the target angles only. Four LED lights have been installed into the color sensor module for providing balanced luminance for target objects. An Arduino Uno microcontroller (Arduino, Somerville, MA, USA) is used as a control unit, which takes its inputs from the color sensor module. The 16 bit RGB color values are then converted to Hue, Saturation, and Value-based color model. The color gamut of ColorWatch is calibrated on color samples and these calibration values are compared by a nearest-neighbor based algorithm for classification of any target object color and color tones. The target object color and color tones then index the corresponding angular position of motors from a lookup array. Based on the difference of current and target angular position, the required angular rotations are then conveyed to stepper motors via motor drivers.

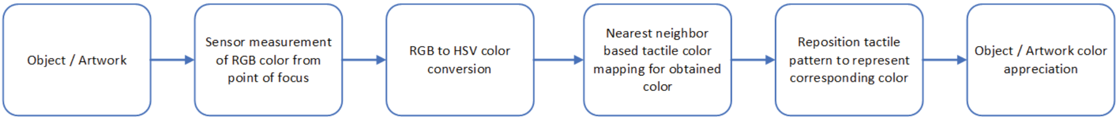
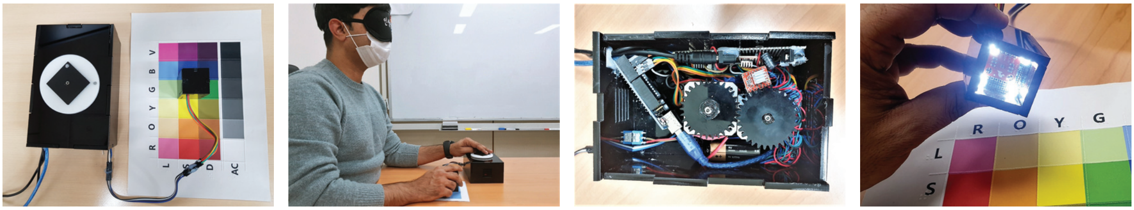


Figure 4. Block diagram of the prototype's working principle.



(**a**) (**b**) (**c**) (**d**)

Figure 5. The tactile perception and color sensor modules of prototype with a color identification training relief paper (**a**); a demonstration of the prototype in use (**b**); the internal hardware for tactile perception and color sensor modules (**c**,**d**).

3. Experimental Investigation

We have adopted a triangulation based method [33] by covering multiple approaches to address a research question, which offers an appropriate and reliable assessment. Based on our preliminary study for tactile-color cross-modular association, an improved solution has been proposed in the preceding section. The tactile perception large-scale prototype was developed which can produce dynamic results based on reference color. User experiments for design, methodology, and cross-modular association have been performed to evaluate usability and workload assessment, while qualitative feedback is also recorded for the proposed method and exploratory procedure.

3.1. Experiment Design

The experimental evaluation was performed in three stages; learning phase, tests, and feedback. Fifteen volunteers of ages between 19 and 29 years took part in the experimental evaluation of the ColorWatch design and prototype. The volunteers were recruited through school notice board announcement, and twelve of them were males while three were females. All of the subjects had normal vision, did not declare any associated cognitive or psychiatric disorders, and they performed the experiments after putting blindfolds on. Firstly, the test subjects were told about the ColorWatch concept, prototype function, and spatial tactile-color associations. The participants were then allowed free time to get familiarized with the prototype and its function with color identifications. All the colors and their color tones were provided as color sample cards of the size of a quarter of standard A4 size paper each. The subjects were encouraged to use a ColorWatch prototype on real-life objects during the learning phase. The color identification data during the learning phase were not recorded. However, subjects mainly tried to identify colors for table, paper, PC monitor, phone back cover, and their clothes through the tactile interface. After comparing the identifying color with their memory, some of them talked about the challenges PVI might face due to a lack of vision. It was interesting to note that many subjects tried color identifications with blindfolds put on without requirement, once they got familiarized with the prototype. The same color samples were used in the user tests for standardization of the color set for all users. After completing the learning phase, the subjects were then asked to put blindfolds on for user tests. At this point, color cards for all the twenty-four color samples in random order were provided one at a time. The subjects were asked to identify the given color by identifying the tactile pattern from the ColorWatch prototype. The primary color and color tone for the given card and the corresponding identification by each subject were carefully registered, and this procedure was repeated for all the color samples. The subjects were not provided with feedback during tactile identification tests. Finally, subjects evaluated the usability of ColorWatch, provided any remarks they had, and responded to the qualitative reasoning queries regarding their tactile following scheme and usability evaluations.

3.2. Identification Tests

As the subjects were given a random color sample, they were asked to identify the color and its tone corresponding to the cross-modular tactile pattern of ColorWatch. Time slots of fifteen minutes each for both the learning phase and color identification tests were allocated per subject, based on preliminary evaluation and improvisations from our group. To avoid prejudices from previous tests, and to make participants feel comfortable in their subjective identifications; the subjects were informed of flexible time limits. However, all the subjects completed the learning phase and identification tests within fifteen minutes for each stage. The square and the round disk could be easily distinguished from each other, in addition to the tactile embodiment on them since they are placed at distinct positions intersecting traditional clock hands. The six color-hues are located such as the complementary colors are on opposite ends from each other. Hence, it all came down for subjects to remember only three colors with their respective positioning on ColorWatch, and understanding the key idea for monochromatic and achromatic colors, and their tones. The subjects were quick to learn the pattern and reported no trouble in getting familiarized with the ColorWatch idea and its large scale prototype. During color identification tests on fifteen subjects, no feedback for their identifications was provided to them during tests, and no color sample or their identification was repeated for any case. All the identifications made were correct as shown in Table 3, except a total of three wrong identifications. The subject 'S2' misidentified light versus dark yellow and subject 'S13' misidentified a color tone for achromatic color, yielding the total correct identification rate at 99.17%.

Table 3. The correct identifications out of total color identifications made by fifteen subjects for ColorWatch color gamut.

Monochromatic – Achromatic Colors	**Light**	**Saturated**	**Dark**	**Achromatic**
Red–White	15/15	15/15	15/15	15/15
Orange–Gainsboro	15/15	15/15	15/15	15/15
Yellow–Light Gray	14/15	15/15	14/15	15/15
Green–Dark Gray	15/15	15/15	15/15	15/15
Blue–Dim Gray	15/15	15/15	15/15	14/15
Purple–Black	15/15	15/15	15/15	15/15
Correct Identifications (%)	98.89	100	98.89	98.89
Average Correct Identifications	357/360 (99.17%)			

3.3. Workload Assessment

The NASA Task Load Index (NASA-TLX) is considered the gold standard for subjective workload measurement. It was developed by the Human Performance Group at NASA's Ames Research Center and can evaluate task demands and workload of an individual performing it based on task, behavior, and subject-related scales [34]. The task-related scales are used to measure the objective demands of the task, the behavior-related scale reflects upon an individual's subjective evaluation of the task, while the subject related scale accounts for the psychological impact on the individual. The NASA-TLX test can be applied to evaluate quantitative subjective mental workload assessment of a service, system, or task based on six indicators. The indicators and evaluated scores for subjects are shown in Table 4. The scale from 1 to 10 points is chosen for the ease and familiarity of participants, with 1 ranging from very low to 10 being very high. Typically, the measured scores for different indicators are assigned weights corresponding to their relative importance, but some studies such as [35] have used raw TLX tests. Considering the subjective nature of the task in this study, we have evaluated the raw TLX test using uniform weights for all indicators.

Table 4. NASA-TLX test questions with subject wise score selection.

Indicator	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
1. Mental Demand	7	2	4	2	3	8	2	3	4	1	4	6	8	4	3
How much mental and perceptual activity was required? Was the task easy or demanding, simple, or complex? (memory ability of pattern, color, etc.)															
2. Physical Demand	6	1	2	4	3	1	2	5	1	1	2	1	3	2	3
How much physical activity was required? Was the task easy or demanding, slack or strenuous? (tactile cognition)															
3. Temporal Demand	7	3	1	2	3	2	1	3	2	1	3	8	7	3	2
How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?															
4. Overall Performance	8	9	10	9	10	10	10	9	10	10	9	9	3	10	7
How successful were you in performing the task? How satisfied were you with your performance?															
5. Effort	9	6	1	6	7	2	1	5	3	1	4	3	7	4	4
How hard did you have to work (mentally and physically) to accomplish your level of performance?															
6. Frustration Level	6	3	1	1	1	5	1	2	1	1	3	1	1	1	3
How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?															

Figure 6 summarizes workload assessment scores for subjects under the NASA-TLX test. The indicators for mental, physical, and temporal demands, effort, and frustration all score less than the median value for the scoring range of 1–10. Most of the participants were very quick to learn the tactile interface for color translation and they requested to start the tests without consuming most of the learning phase time of fifteen minutes given to them. Given that the physical and temporal demands are very low for the proposed system, the mental demand and efforts put on by the users may vary from person to person, as it has been reflected by their respective standard deviations. The mean scores for both mental demand and effort lie in the medium-low of scoring range for the test, and less than the median for scoring range. However, few subjects suggested that it might be easier with more practice. The proposed system is promising in terms of user satisfaction as the performance indicator for user satisfaction scored highest while the frustration indicator scored lowest among all the indicators for this test.

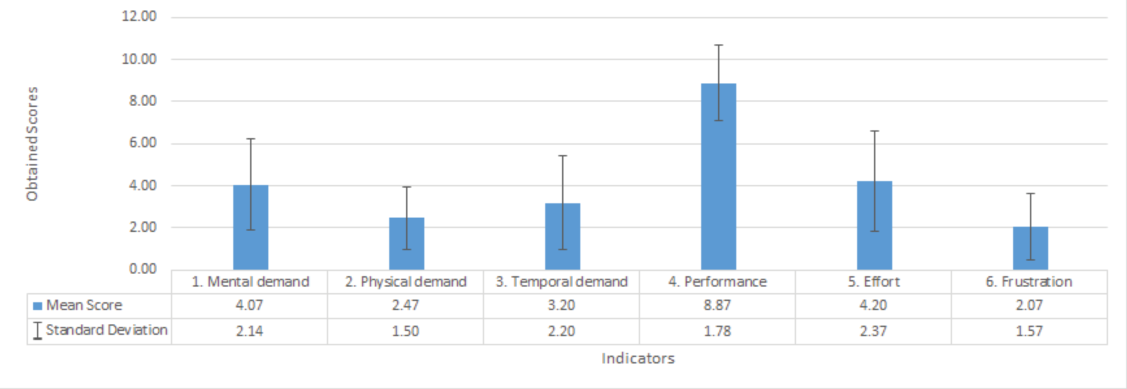


Figure 6. Workload assessment scores for NASA-TLX test question indicators.

3.4. Usability Test

System Usability Scale tests the usability of a product, service, or system using a five-question Likert scale based on responses to ten standardized questions [36]. The test questions are non-complex with responses that include strongly disagree, disagree, neutral, agree, and strongly agree, with index scores ranging from 1 to 5, respectively. The questions are given below for reference. The even and odd-numbered questions in the SUS test reflect a negative and positive attitude. During SUS score calculation, 1 is subtracted from user response for odd-numbered questions, while user response is subtracted from 5 for even-numbered questions. This scales them from 0 to 4 with 4 being the most positive response. Responses for all ten questions for each user is then added, making the sum in the range of 0 to 40. This sum is then multiplied by 2.5 to normalize the response for each user in the 0–100 range. These obtained raw SUS scores can be converted to individual percentile ranking to make relative judgement of usability by normalization or grading on the curve [37]:

I1 I think that I would like to use this service frequently.
I2 I found the service unnecessarily complex.
I3 I think the service was easy to use.
I4 I think that I would need the support of a technical person to be able to use this service.
I5 I found that the various functions in this service were well integrated.
I6 I thought there was too much inconsistency in this service.
I7 I would imagine that most people would learn to use this service very quickly.
I8 I found the service very cumbersome to use.
I9 I felt very confident using the service.
I10 I needed to learn a lot of things before I could get going with this service.

The usability test scores breakdown for all participants out of 1–5 (strongly disagree through strongly agree) are given in Table 5. Although the subjects were not accustomed to using their sense of touch instead of sense of vision for perceiving objects. However, positive indices outperform negative indices by obtaining negative scores about half the order of positive overall scores. Here, the positive and negative indices refer to odd and even-numbered questions of the SUS test which reflect the positive and negative attitude, respectively. The overall SUS score of 72.73 converts into the percentile range of 65–69% that ranks at ‘Good’ objective rating as a measure of user’s perception of the usability of system [37,38]. The individually scored SUS results are provided in Figure 7.

Table 5. Usability test subject-wise selection for SUS indices.

Measured Scores		**SUS Indices**									
		I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**
Subjects	**S1**	4	3	2	4	5	1	2	4	3	4
	S2	4	2	4	3	3	2	4	3	4	2
	S3	5	2	5	2	4	1	5	1	5	2
	S4	4	2	5	2	2	2	5	3	4	1
	S5	4	1	4	5	5	1	4	2	5	2
	S6	2	2	4	3	3	2	5	4	3	2
	S7	5	2	5	2	3	1	4	4	4	2
	S8	4	2	4	3	4	2	4	3	4	2
	S9	3	2	5	1	4	1	3	2	5	1
	S10	5	1	5	5	4	1	5	2	5	1
	S11	4	2	4	3	3	2	4	4	4	1
	S12	5	1	5	3	5	1	5	1	5	3
	S13	4	2	3	1	5	2	5	1	4	4
	S14	5	2	4	3	4	1	4	1	5	1
	S15	5	1	3	3	3	1	5	1	3	3

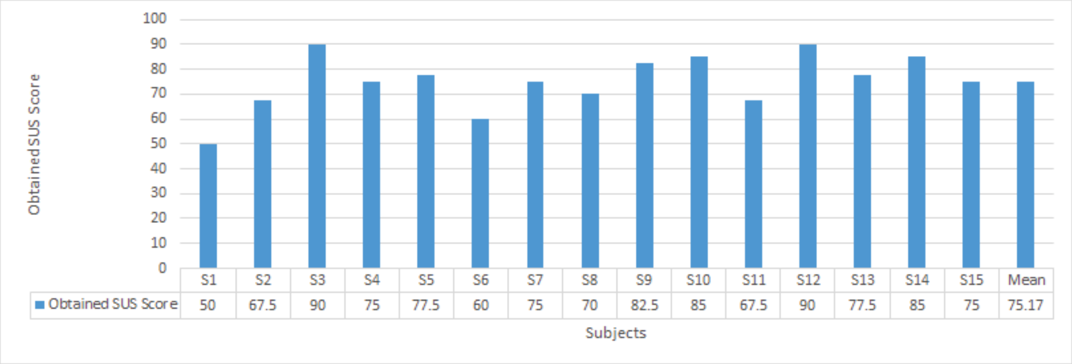


Figure 7. Usability test individual and mean obtained scores.

3.5. Subjective Reasoning

We observed the exploratory scheme for subjects during the tactile identification tests. Geometric and semantic identifications were noted whenever the subjects were able to correctly distinguish round and square disks and were able to locate and describe the tactile marker on disks, respectively. Subjects also provided feedback on the improvement of ColorWatch and whether it might be helpful for PVIs or not. They were then asked about challenges in geometric and semantic identifications. A few of them also provided ideas on helping PVI in color identifications after completion of tests.

Generally, the subjects reported the concept and its application through the rotating disks to be proper and easy. The subjective feedback from test participants was as follows;

- *The square was easier than the round disk.*
- *Because it was a large prototype and could not be recognized all at once, you have to touch it several times.*
- *It wasn't difficult, but I felt confused a couple of times. However, I think there will be no problem if you use it frequently.*
- *It was easy assuming that the disks were stationary.*

The participants' subjective feedback about conveying the colors through tactile associations, and the tactile perception is given below. It is stated that many of the participants deemed the color identifications through ColorWatch to be very easy or easy, hence only the critical or imperative responses are provided here:

- *It is divided into a circle and a square shape, so there was no big inconvenience.*
- *It was easier to distinguish by touching than it was to memorize.*
- *The Red and Green portions overlapped with the corners of the square and were not easily recognized compared to other positions.*
- *The distinction between square and circle is good.*
- *It seems that it was a little difficult to recognize the raised dot.*

Subjects felt confident about using the tactile interpretation of color when asked about suggestions for improvements regarding tactile or any other modality as a color perception solution for PVI. The critical comments are as follows:

- *It would be helpful if the spectrum of recognizable colors widens.*
- *I think it will be easier to distinguish if the outer disk is also not round but has edges or angles like the square disk.*
- *Product miniaturization.*
- *If voice is added, it will be helpful for recognition.*
- *Vibration and temperature might also be used.*

Finally, subjects' opinion about being the ColorWatch helpful for PVIs is as follows:

- *I'm not sure if there's a product like this, but, if there isn't, it might help.*

- *It seems to be helpful for the visually impaired in that it can select clothes suitable for TPO (Time Place Occasion) and reflect preferences for fashion.*
- *It can be needed for scenarios like if someone complimented your shirt, and you associate this emotional response with the shirt color.*
- *It seems to be helpful because even the visually impaired may have to recognize colors.*
- *The advantage that people who do not know Braille can easily recognize the color.*
- *I think it would be good to make it easier to recognize if it is less affected by ambient light.*
- *I think it will help. Because there is a lot you can find out from color.*
- *I think it will be helpful enough. You can quickly recognize colors.*
- *It seems to be helpful in color recognition. It was placed in a circle like a color scheme to make it relevant.*

4. Discussion

According to the National Federation of the Blind, braille literacy has been on the decline. Out of 1.3 million legally blind people in the USA, only 10% of them can read braille. Moreover, only 10% or blind children are learning braille, calling it "The Braille literacy crisis in America", which limits the prospects of braille based designs [39]. The braille watches [22–24] can theoretically be used as a digital braille interface if combined with color sensing and color recognition systems; however, no such scheme has been presented by the developers. A summary and comparison of these have been provided in Table 6. Analog time watches with conventional rotating hands exhibit gradually altering angular orientation because the hands move continuously with passing time, whereas digital time watches numerically convey time. The continuously rotating analog watch hands provide a means to the human brain for instantaneous spatial recognition, such as the relative difference between two time-stamps and the apprehension of remaining time until an event. Situations like these for an analog watch do not strictly engage mental numeric calculations, which is conversely true for a digital watch where brief mental calculations are required for interpreting these events. In terms of human perception, it is a matter of spatial recognition versus mental numeric calculations. Likewise, a mere description of color in text or braille manner in a theoretical scenario of inclusion of color information through braille watch might not be sufficient for PVI perception as it lacks the essence of color nature, their relative relationships, and interpretation.

Related works are focused on the tactile translation of colors in artworks, which limits the usability of those schemes in daily life. They also have high learnability requirements, and cannot be used for the color perception of arbitrary objects. We have integrated perceptual color patterns with analog watch design in an intuitive arrangement. The subjects' feedback validated the effectiveness of ColorWatch TCP as there were only a few incorrect color identifications, and almost all of the identifications were correct. The subjects were also able to understand the tactile-color association and the way they are organized on the watch design. It enabled them to quickly recall the correct related color for the detected angular position on the disk. The subjects reported the ease of tactile dot detection on the square disk in comparison with the round disk. This might be caused by a fewer number of probable positions for the tactile dot on the square disk rather than the round disk. Another possible reason for that may be the shape of the square disk which conserves its apparent tactile shape perception even after rotations at 90°, and the angular position of the tactile dot at one corner of it can be identified relative to its distinct and conserved geometry. A similar effect can be achieved for one tactile dot detection of any regular polygon shape, if the number of distinct positions or the number of fixed angle rotations it takes for one full revolution is a factor of the number of polygon vertices. The reason for the round shape of the larger disk for color hue in color perception, and the minute pointer in timepiece mode is chosen instead of a polygon shape is to help identify continuous rotations for the minute pointer in timepiece mode. Moreover, the continuous movement of round disk or minute hand can provide more precise time, instead of fixed rotations for every five minutes, for example. The color gamut of ColorWatch might be further extended to represent eight color hues, each for three levels of color tones along with eight levels of achromatic colors as shown in Figure 1e. In this way, the colors are placed at 60°angles on the color wheel, capable of presenting 32 colors. The dynamic nature of spatial color recognition for reference color and easy learnability makes it particular among TCPs from relevant research. A brief comparison for relevant works is provided in Table 7. This might help PVIs as an assistive device as the overall rating of 'Good' is reflected by test subjects' responses. The workload assessment test also validates it to be suitable for a broad population of potential PVI users with low learnability requirements, as indicated by below-average scores of mental demand coupled with low frustration, and higher satisfaction about performing the task correctly.

Table 6. Features and comparison of tactile watches.

Product	Output(s)	Mode	Features	Workload	Learnability
Tyler K. [22]	Timepiece	Numeric 4-dot condensed braille code	Time and date	Mental numeric calculation	The Braille literacy crisis
Dot Inc. [23]	Timepiece	Numeric 6-dot braille code interface	Smartphone wireless connectivity	Mental numeric calculation	The Braille literacy crisis
Anderson N. L. et al. [24]	Timepiece	Numeric 4-dot condensed braille code	Caller ID and pick/drop for smartphone calls	Mental numeric calculation	The Braille literacy crisis
This work	Timepiece & Colorperception	Analoginterface	Traditional analog interface, Durable design, Color perception	Intuitive spatial recognition	Intuitive human perceptual design for time and color representation

Table 7. The overview and comparison with relevant works on tactile color pictograms.

TCP	Basic Patterns (Concepts)	Number of Colors Presented	Medium
Taras et al. [17]	Dots (braille)	23 (6 hues + 2 levels of lightness for each hues + 5 levels of achromatic)	Braille embossed surface pattern
Ramsamy-Iranah et al. [12]	Polygons (children's knowledge)	14 (6 hues+ 5 other colors + 3 levels of achromatic)	Embossed surface pattern
Stonehouse [40]	Geometric pattern and texture (traditional conventions)	11 color hues	Embossed surface pattern
Shin et al. [4]	Lines, orientation, grating. The first eight colors are divided into 20° angle (rainbow shape)	90 (8 hues + 4 levels of lightness and 5 levels of saturation for each hues + 9 levels of brown and achromatic)	Embossed surface pattern
Cho et al. [16]	Dots, lines, and curves (pictograms)	Simplified: 29 (6 hues + 2 levels of lightness and 2 levels of saturation for each hue + 5 levels of achromatic)	Embossed surface pattern
		Extended: 53 (12 hues + 2 levels of lightness and 2 levels of saturation for each hue + 5 levels of achromatic)	
This work	Simplified: RYB color wheel model: six colors are divided into 60° angles in the 6 RYB color wheel (watch type)	24 (6 hues + 3 levels of lightness for each hue + 6 levels of achromatic)	Automatic spatial representation as assistive wearable device, Embossed surface pattern
	Extended: eight colors are divided into 45° angles in the 8 RYB color wheel (watch type).	32 (8 hues + 3 levels of color tones for each hue + 8 levels of achromatic)	

The contributions of this study are listed here:

1. The design for ColorWatch is presented which can aide PVIs in color perception, in addition to time recognition.
2. The angular tactile pattern for color is proposed with associated angular positions of tactile dot to a range of achromatic colors and basic colors with hue, value, and chroma indication.
3. The combination of colors from Goethe's color triangle and Munsell color system has been integrated with the analog watch design interface. This interface is capable of instantly presenting automatic spatial tactile patterns for any detected reference color from the artworks as well as from real-life objects. This eliminates the need for static color translations of artwork onto tactile relief. The spatial color identification for arbitrary objects outranks the static interpretations for the artworks of existing TCPs.
4. The prototype has been developed, and the tests have been performed to investigate effectiveness in terms of accurate identification of color, workload requirements for pattern learning, and usability for color detection.
5. The color wheel depicts essentially the visible spectrum of colors enclosed by a circle, and is a useful tool for describing what happens when you mix colors, complementary color relationships, and adjacent colors. Traditional TCP requires embossed surface patterns to represent colors for each artwork. Our proposed system eliminates such embossed surface patterns, and can be used as a reconfigurable platform, providing better mobility, dissemination, and flexibility.

5. Conclusions

The development of assistive technologies for art appreciation for visually impaired people can enhance their cultural and perceptual appreciation. These opportunities result in better comprehension and accessibility at museums, exhibitions, and everyday life. These multisensory interactions may also offer enhanced usability, understanding and promote educational tools aiding synesthetic capabilities to promote creative thinking. Color associations with aspects such as symbolism, culture, and preferences play an influential role, demanding the promotion of PVIs' color comprehension in daily life as well. Tactile color pictograms using tactile sensing attain sensational conception along with other physical properties of artwork such as contour, size, texture, geometry, and orientation. Although several TCPs have been developed, they are limited to fixed tactile color interpretation, which requires outgoing resources. We have proposed the design for ColorWatch integrating colors from Goethe's color triangle and Munsell color system with analog wristwatch, allowing spatial color-to-tactile interpretation. We have associated achromatic and monochromatic colors with chroma and value levels to the cross-modular tactile interface. The tactile interface manifests angular positions of tactile patterns. These patterns can be transformed automatically corresponding to the reference color. The arrangement of the tactile pattern is based on intuitive learning, which is translated through analog wristwatch tactile interface. This integrated approach offers ease of learnability to provide the essence of particular emotional or psychological states. We developed a prototype and performed an identification test for proof of concept. The test results for color identification present good accuracy and validate our hypothesis. Usability tests based on system usability scale and workload assessment by NASA-TLX tests suggest that the proposed ColorWatch system can help people with visual impairments in color identification and reduce a factor that hinders their museums' accessibility and real-life color perception. The function of ColorWatch may be expanded to represent color gamut of forty-two colors with twelve color hues based on the RYK color wheel, originally described by Issac Newton. The six additional color hues can be represented at uniform 30°angular distances, alternating between existing chromatic color hues. We shall expand experiments with subjects for balanced gender and diverse PVI vision statuses, and explore their simultaneous cognition abilities for a multisensory appreciation of artworks as a future study.

Author Contributions: Conceptualization, M.S.J. and J.D.C.; methodology, software, M.S.J.; writing—original draft preparation, M.S.J.; data curation and validation, C.-H.L.; project administration, supervision, writing—review and editing, J.D.C.; All authors have read and agreed to the submitted version of the manuscript.

Funding: This research was funded by the Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea, Grant No. 2018M3C1B6061353.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Sungkyunkwan University (protocol code: 2020-11-005-001, 22 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available within the article.

Acknowledgments: We would like to thank all volunteers for their participation and the reviewers for their insights and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Striem-Amit, E.; Wang, X.; Bi, Y.; Caramazza, A. Neural representation of visual concepts in people born blind. *Nature communications* **2018**, *9*, 1–12. [CrossRef] [PubMed]
2. Bedny, M.; Caramazza, A. Perception, action, and word meanings in the human brain: The case from action verbs. *Ann. N. Y. Acad. Sci.* **2011**, *1224*, 81–95. [CrossRef] [PubMed]
3. Cytowic, R.E. *Synesthesia: A Union of the Senses*; MIT Press: Cambridge, MA, USA, 2002.
4. Shin, J.; Cho, J.; Lee, S. Please Touch Color: Tactile-Color Texture Design for The Visually Impaired. In Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 25–30 April 2020; pp. 1–7.
5. Cho, J.D. A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment. *Electronics* **2021**, *10*, 470. [CrossRef]
6. Cho, J.D.; Jeong, J.; Kim, J.H.; Lee, H. Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments. *Electronics* **2020**, *9*, 1981. [CrossRef]
7. Gilbert, A.N.; Martin, R.; Kemp, S.E. Cross-modal correspondence between vision and olfaction: The color of smells. *Am. J. Psychol.* **1996**, *109*, 335–351. [CrossRef] [PubMed]
8. Iranzo Bartolomé, J.; Cho, J.D.; Cavazos Quero, L.; Jo, S.; Cho, G. Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People. *Electronics* **2020**, *9*, 1939. [CrossRef]
9. Cavadini, R.; Cerina, L.; Bracco, M.I.; Santambrogio, M.D. A wearable device for blind people to restore color perception. In Proceedings of the 2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry, Modena, Italy, 11–13 September 2017; pp. 1–4.
10. Sathian, K.; Stilla, R. Cross-modal plasticity of tactile perception in blindness. *Restor. Neurol. Neurosci.* **2010**, *28*, 271–281. [CrossRef] [PubMed]
11. Lawrence, M.A.; Kitada, R.; Klatzky, R.L.; Lederman, S.J. Haptic roughness perception of linear gratings via bare finger or rigid probe. *Perception* **2007**, *36*, 547–557. [CrossRef] [PubMed]
12. Ramsamy-Iranah, S.; Rosunee, S.; Kistamah, N. Application of Assistive Tactile Symbols in a 'Tactile' Book on colour and Shapes for Children with Visual Impairments. *Int. J. Arts Sci.* **2017**, *10*, 575–590.
13. Bindi (Decoration). Available online: https://en.wikipedia.org/wiki/Bindi_(decoration) (accessed on 16 February 2021).
14. Grant, A.C.; Thiagarajah, M.C.; Sathian, K. Tactile perception in blind Braille readers: A psychophysical study of acuity and hyperacuity using gratings and dot patterns. *Percept. Psychophys.* **2000**, *62*, 301–312. [CrossRef] [PubMed]
15. Kitada, R.; Kito, T.; Saito, D.N.; Kochiyama, T.; Matsumura, M.; Sadato, N.; Lederman, S.J. Multisensory activation of the intraparietal area when classifying grating orientation: A functional magnetic resonance imaging study. *J. Neurosci.* **2006**, *26*, 7491–7501. [CrossRef] [PubMed]
16. Cho, J.D.; Quero, L.C.; Bartolomé, J.I.; Lee, D.W.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2021**, *46*, 103–116. [CrossRef]
17. Taras, C.; Ertl, T. Interaction with Colored Graphical Representations on Braille Devices. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, San Diego, CA, USA, 19–24 July 2009; pp. 164–173.
18. Axel, E.; Levent, N. Art beyond sight: Guidelines for making tactile diagrams & standard tactile patterns, lines, and icons. *Art Education for the Blind*; AFB Press: Arlington, VA, USA, 2003.
19. Hetherington, K. Museums and the visually impaired: the spatial politics of access. *Sociol. Rev.* **2000**, *48*, 444–463. [CrossRef]
20. Vaz, R.; Freitas, D.; Coelho, A. Blind and Visually Impaired Visitors' Experiences in Museums: Increasing Accessibility through Assistive Technologies. *Int. J. Incl. Mus.* **2020**, *13*, 57–80. [CrossRef]

21. Adnan, M.E.; Dastagir, N.M.; Jabin, J.; Chowdhury, A.M.; Islam, M.R. A cost effective electronic braille for visually impaired individuals. In Proceedings of the 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), New York, NY, USA, 21–23 December 2017; pp. 175–178.
22. Tyler, K. Active Braille Timepiece & Related Methods. U.S. Patent App. 13/180,351, 17 January 2013.
23. DOT Watch, a Braille Wearable Smartwatch. Available online: https://www.dotincorp.com/ (accessed on 20 January 2021).
24. Anderson, N.L.; Puccinelli, J.P.; Singh, C.; Jamar, K.S.J.; Milne, T.A. Braille Watch. U.S. Patent 8,483,018, 9 July 2013.
25. Twyman, M.; Mullenbach, J.; Shultz, C.; Colgate, J.E.; Piper, A.M. Designing wearable haptic information displays for people with vision impairments. In Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, Stanford, CA, USA, 15–19 January 2015; pp. 341–344.
26. Velázquez, R. Wearable assistive devices for the blind. In *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 331–349.
27. Tyler, J.E.; Hardy, A.C. An analysis of the original Munsell color system. *JOSA* **1940**, *30*, 587–590. [CrossRef]
28. Von Goethe, J.W. *Goethe's Theory of Colours*; Murray, J., Ed.; William Clowes and Sons: London, UK, 1840.
29. Albers, J. *Interaction of Color*; Yale University Press: New Haven, CT, USA, 2013.
30. RYB Color Model. Available online: https://en.wikipedia.org/wiki/RYB_color_model (accessed on 20 January 2021).
31. Munsell Color System: Explanation. Available online: https://en.wikipedia.org/wiki/Munsell_color_system (accessed on 20 January 2021).
32. Lee, C.H.; Jabbar, M.S.; Cho, J.D. ColorWatch: Color-Tactile Interpretation to Improve Color Perception and Accessibility for People with Visual Impairments. In Proceedings of the IEEE 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 21–23 October 2020; pp. 1632–1637.
33. Mackay, W.E.; Fayard, A.L. HCI, natural science and design: A framework for triangulation across disciplines. In Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, Amsterdam, The Netherlands, 18–20 August 1997; pp. 223–234.
34. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*; Elsevier: North Holland, The Netherlands, 1988; Volume 52, pp. 139–183.
35. Hart, S.G. NASA-Task Load Index (NASA-TLX); 20 Years Later. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Los Angeles, CA, USA, 16–20 October 2006; Volume 50, pp. 904–908.
36. Brooke, J. SUS: A "quick and dirty'usability. In *Usability Evaluation in Industry*; Taylor & Francis: London, UK, 1996; p. 189.
37. 5 Ways to Interpret a SUS Score. Available online: https://measuringu.com/interpret-sus-score/ (accessed on 20 January 2021).
38. Bangor, A.; Kortum, P.; Miller, J. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Stud.* **2009**, *4*, 114–123.
39. The Braille Literacy Crisis in America: Facing the Truth, Reversing the Trend, Empowering the Blind. In *Braille Monitor 52.5*; The National Federation of the Blind: Baltimore, MD, USA, 2009; Volume 52.
40. Tactile Colours: Scheme for the Tactile Representation of Colours. Available online: http://cpwstonehouse.com/tactile-colours/ (accessed on 20 January 2021).

electronics

MDPI

Article

Auditory *Uta-Karuta*: Development and Evaluation of an Accessible Card Game System Using Audible Cards for the Visually Impaired

Haruna Miyakawa [1], Noko Kuratomo [2], Hisham E. Bilal Salih [3] and Keiichi Zempo [3,*]

[1] Graduate School of Science and Technology, University of Tsukuba, Tsukuba 305-8573, Ibaraki, Japan; miyakawa@aclab.esys.tsukuba.ac.jp
[2] Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba 305-8573, Ibaraki, Japan; kuratomo@aclab.esys.tsukuba.ac.jp
[3] Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Ibaraki, Japan; hisham@aclab.esys.tsukuba.ac.jp
* Correspondence: zempo@iit.tsukuba.ac.jp

Abstract: Playing board games is important for people with a visually impairment, as it promotes interactive socialization and communication skills. However, some board games are not accessible to them at present. In this study, we proposed an auditory card game system that presents a card's contents with auditory stimuli to all players, towards playing equally with others, regardless of whether they have a visual impairment or not as one of the solutions to make board games accessible. This proposal contributes significantly to expand the range of inclusive board games for the visually impaired. The purpose of this paper is to determine whether the game allows for fair competition for people with visual impairments and to clarify the effects of the valuable parameters of the system on the players. The effectiveness of the proposed system was verified by having experimental participants play "Auditory *Uta-Karuta*". The results suggested that the proposed system has the potential for an accessible board game design regardless of visual impairment. In the following experiment, we investigated the impact of each valuable parameter of the system on the player's perception of the board games to clarify the appropriate audio cue design method. The results of this experiment will greatly assist in designing an appropriate board game using the proposed system.

Keywords: accessibility; assistive technology; auralization

Citation: Miyakawa, H.; Kuratomo, N.; Salih, H.E.B.; Zempo, K. Auditory *Uta-Karuta*: Development and Evaluation of an Accessible Card Game System Using Audible Cards for the Visually Impaired. *Electronics* **2021**, *10*, 750. https://doi.org/10.3390/electronics10060750

Academic Editors: Jun Dong Cho and Juan M. Corchado

Received: 26 January 2021
Accepted: 15 March 2021
Published: 22 March 2021

1. Introduction

Games as a means for social interaction and enjoyment were not initially included in digital technology. Multiple players had to exist in the same physical environment, and played game by manipulating tangible components made of wood or stone [1]. However, the development of digital technology has led to digital games, making it possible to play games without being in the same physical environment [1]. In spite of the shift from board games (which we define as all non-digital games) to digital games, board games have continued to be popular.

Playing board games is important for promoting interactive socialization of participants and improving problem-solving and communication skills [2]. Moreover, sharing a physical environment and game components facilitates concentration and verbal communication, which can help enhance relationships [3,4].

However, most board games are difficult to play with people with visual impairments because they are designed on the assumption that visual contact can be seen [5,6]. Further, board games can cause more difficulty due the need for interaction and special manipulation with the game components.

Although board games have these accessibility issues, the positive impact of board games on players is considered to be particularly significant for people with visual impairments. Most people with visual impairments experience more social isolation and difficulties in interpersonal relationships due to their disability [7]. The positive effects of board games, such as improving problem-solving and communication skills, can solve these problems. In addition, to manipulate game components one's own self and participate in the game without the help of anyone, it could be able to contribute to further strengthening autonomy and increasing satisfaction of people with visual impairments. These contributions will able to promote the ability to live powerfully with disabilities: resilience [8,9]. Thus, it is important to give people with visual impairments the chance to have access to games for effective inclusion and participation in society. The efforts to improve board game accessibility have been made by devising rules or structures for the cards/pieces used in board game. "The Arabian Pots" [10] and "Nyctophobia" [11], which are typical examples of board games whose rules are designed for people with visual impairments, require only the ability to distinguish sounds. On the other hand, the newer versions of "Uno" [12] and "Hanabi" [13] are games that ensure accessibility by devising structures for the cards/pieces that they use. They use Braille or larger-than-usual cards to represent different colors and numbers for people with visual impairments. The solutions to making the cards/pieces used during board games accessible for people with visual impairments are mainly limited to enlarging the cards/pieces themselves or using tactile presentation. The former is the simplest solution, but it has the disadvantage of not allowing inclusion for completely blind players. The latter has the advantage of the inclusion of all visually impaired people in existing board games, but this solution alone is not enough to apply to all board games. It is necessary to memorize Braille script, as well as textures and their corresponding information, in advance, which is not intuitive. In other words, a solution that only presents tactile sensations has a risk of diminishing the enjoyment of the board game when integrated into existing games.

In this paper, we propose an auditory card game system that presents the card's contents by sound [14], with the aim to develop a board game that can be played in the same way by all players, regardless of whether they have a visual impairment or not. This is one of the solutions for making cards/pieces used during board games accessible for people with visual impairments. The purpose of this paper is to determine whether the game allows for fair competition for people with visual impairments and to clarify the effects of the valuable parameters of the system on the players. The remainder of this paper is organized as follows: in Section 2, we describe existing research and products from the community on games that are accessible to people with visual impairments. In Section 3, we introduce the configuration of the proposed game system. In Section 4, we describe the overview and results of a subject experiment on a new game that applies the proposed system to an existing card game. In Section 5, we investigate each valuable parameter of the system and explain the appropriate game design based on the results of various existing studies. Finally, the limitations of this study are described, followed by the conclusion.

2. Related Works

To solve accessibility issues for people with visual impairments in digital/non-digital games, various efforts have been deployed. In this section, we first describe the efforts by researchers and the community on the accessibility of digital games for the visually impaired. Next, we describe these efforts with respect to a non-digital game, breaking it down by the role of the components used in the game. Finally, we explain where our proposed system stands and contribution in relation to previous works.

In the digital game domain, audiogames that can be played with audio only are mainstream. Although audiogames is very small market compared to that of computer games, a lot of people with visually impaired enjoy audiogames, and the players and developers community is enhanced on internet site by introducing new audiogames and sharing player reviews [15]. However, audiogames do not appeal as much to sighted people because they don't have visual information. In fact, most audiogames are not interesting for most people, and the majority of players are people with visual impairments.

To create accessible games that can interest both camps, researchers have been contributing to the development of games that are not only audio but also visual and/or tactile. For example, "Terraformers" [16] uses sound propagation to convey the proximity of obstacles. "Terraformers" [16] and "AudioBattleship" [17] give feedback for every action using a voice over. "Sonic-Badminton" [18] enables play regardless of visual impairment by presenting the position of the wings of the badminton with stereophonic sound. Examples of the use of tactile sensation include "TiM games" [19] and "Digital Clock Carpet" [20] which can navigate in 2D environments by presenting different tactile stimulation with various materials. Further, "Kinaptic" [21] has enabled people with visual impairments to play a chasing game in a virtual environment by using sound and tactile feedback. These games aim to have same level of win rates between people with visual impairments and sighted people.

On the other hand, board games are far behind digital games when it comes to accessibility. In addition, most board game accessibility efforts have come from the community such as "Board Game Geek" [22] and "Meeple Like us" [23]. These efforts have been established by devising the rule or structure of the cards/pieces used in the board game. "The Arabian Pots" [10], which is a typical example of a board game, whose rules are designed for people with visual impairments, requires only the ability to distinguish sounds. It was designed after repeated tests so that only blind people can actually play. "Nyctophobia" [11] is a tactile maze game that does not assume the use of visual information. Players cannot see the board and have to rely on touch to play the game. On the other hand, "Splendor" [24] is a typical example of a game that ensures accessibility by devising the structure of the cards/pieces used in the board game. The first version of "Splendor" [24] used color as the only factor required to acquire a card, so players who could not distinguish between colors could not play the game, but the addition of iconography to the newer version makes this the game accessible to this audience. Similarly, the newer version of "Uno" [12] uses Braille to represent different colors and number. "Hanabi" [13] designed for people with visual impairments, has been modified to use larger-than-usual cards and card holders, but cannot be played by blind people because they still do not use non-visual modalities.

These board game's pieces and cards are generally called components, and they play the role of input/output interfaces between the game world and the players. The functions and entertaining properties of the components are realized because the components are real objects. The functions of components can be broadly classified into the presentation of information to all players and the presentation of information to specific players. In this paper, we define "public information" as the information presented to all players. They corresponds to the function of pieces used in Othello, Chess and Shogi. In contrast, we define "personal information" as the information presented to specific players. They corresponds to the function of deal cards used in Uno and Poker. Although players progress through the board games by manipulating the components, the physical constraints of the components may cause accessibility problems.

Table 1 shows a comparison of the aforementioned accessible board games classified according to the function of the components and the input/output between the players and the interfaces.

In order for people with visual impairments to correctly recognize the components of board games, it is necessary to use hearing and tactile sensation instead of sight. The work of making components accessible can also make existing board games accessible, which can also be an opportunity for people with visual impairments to enter the board game community of the sighted. As shown in Table 1, tactile presentation has been the main solution when components present public information in previous studies. The only solution that uses auditory stimuli is online competition, which does not guarantee face-to-face communication. Similarly, when the component presents private information, tactile presentation is the main solution. However, unlike public information, private information cannot be presented to other players, so it is not appropriate to use auditory stimuli that can be transmitted to an indefinite number of players. Even if a headphone/earphone is used, it may interfere with face-to-face communication.

Table 1. Comparison between related works and proposal interfaces.

	UA-Chess [25]	**Splendor [24]**	**Hanabi [13]**	**Uno [12]**	**Auditory Uta-Karuta [14]**
The Role of Component	Public	Public/Private	Private	Private	Public
Equipment	PC	Card with Texture	Bigger Card	Card with Texture	Small Tablet-type Device
Input	Mouse, Keyboard, Voice	Touch Components	Look at Components	Touch Components	Game Master Operations
Output	Audio Cue, Screen reader	Haptic Sensation	Visual Sensation	Haptic Sensation	Auditory Sensation
Modality	Audio	Haptic	Visual	Haptic	Audio

In this study, we proposed an auditory card game system that presents a card's contents with auditory stimuli to all players, towards playing equally with others, regardless of whether they have a visual impairment or not. This is one of the solutions that presents public information of board games. Making board games components accessible will make it easier for visually impaired people to play existing board games with sighted people. Tactile presentation is the mainstream method for people with visual impairments to access board game components on their own, but it has the disadvantage that it requires touching every time the information is needed. This disadvantage cannot be applied when the game situation can change by touching board game components or when the game situation changes quickly. Furthermore, it is necessary to memorize Braille and textures and their corresponding information in advance, which is not intuitive. On the other hand, by presenting components with sound, players can obtain information about a component's contents and locations without touching them. Filho et al. [1] developed a game with textured card content to ensure accessibility for the visually impaired and designed guidelines, but some of their experimental participants requested audio feedback from the components. In addition to expanding the range of accessible board games, the use of auditory stimuli has the potential to increase immersion and make the game more intuitive and enjoyable. Furthermore, by digitizing the components of a board game, some of the physical constraints of the components can be lifted. This allows us to assign multiple components to a single device, or to increase the range of strategies by intentionally rewriting the information of the components. The final game design policy is to make the game fun for everyone, regardless of the degree of visual impairment.

3. Auditory Card Game System

3.1. Karuta

Karuta is a traditional Japanese playing card [26]. During the Heian period (794–1185), aristocrats had spent time writing elaborate poetry and playing *Kai-awase*. In *Kai-awase*, players depicted the poem or scenes related to the poem on the shells, face down, and competed to find the most pairs of shells. In the mid-1500s, Portuguese sailors introduced European playing cards that they called carta into Japan. During the ensuing Edo period (1600–1868), *Karuta* evolved from a Portuguese import into a Japanese traditional game, combining *Kai-awase* with carta.

One of the main types of *Karuta* is *Uta-Karuta*. Figure 1 shows the progression of *Uta-Karuta*. *Uta-Karuta* consists of multiple cards divided into two sets. One set is for reading and the other is for grabbing. Cards for reading are called *Yomifuda*, cards for grabbing are called *Torifuda*. *Yomifuda* and *Torifuda* exactly correspond to each other. The basic rule of *Uta-Karuta* is described below.

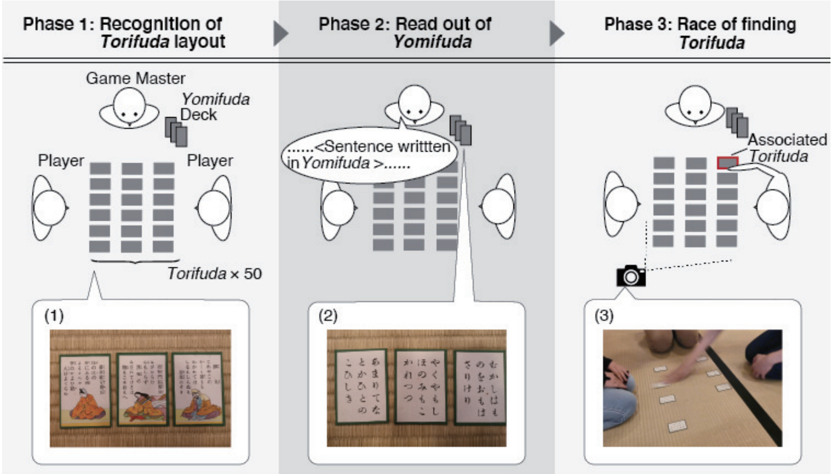


Figure 1. *Uta-Karuta* is a typical example of *Karuta*. *Uta-Karuta* consists of multiple cards divided into two sets: *Torifuda* and *Yomifuda*.*Yomifuda* and *Torifuda* exactly correspond to each other. (**1**) *Torifuda*: "grabbing cards" with pictures and/or written language, (**2**) *Yomifuda*: "reading cards" with written information, (**3**) The players position and cards placement during competition.

Phase 1: Recognition *Torifuda* layout
Arrange the *Torifuda* face up on a flat surface in front of the players. Players memorize the contents and position of the *Torifuda* as much as possible.

Phase 2: Read out of *Yomifuda*
The game master randomly draws a *Yomifuda* from the deck and starts reading it out.

Phase 3: Race of finding *Torifuda*
Players race to determine which *Torifuda* corresponds to the *Yomifuda* and touch/grab/claim what they think is the correct *Torifuda*.

Players repeat Phase 2 and Phase 3 until all the *Torifuda* are gone. The winner is the player who has the most *Torifuda*.

3.2. Auralization

While many researchers and communities have improved accessibility for the visually impaired by using sound in digital games, in the field of board games, they have designed accessible games mainly by using tactile feedback. Although some sound-based board games allow players to play against each other without relying on visual information by manipulating sound components, the idea of electronically generating sound from simple components such as cards and pieces made of paper or plastic is novel. The auditory card game system proposed in this study provides a new solution to board game accessibility by electronically generating sound from simple components. Figure 2 shows an overview of the auditory card game system. This system assumes a game with a game master. It comprises several "audible cards" and master device that is wireless connected to them. Small tablet devices, such as the iPod touch, can emit sound from individually mounted speakers. In addition, they are the same size as cards used in board games, making it easy for players to intuitively imagine them as cards. With this consideration, iPod touch devices were used for the audible cards. The audible cards correspond to the components used in the board game. The master device is a terminal operated by the game master to advance the game and make each audible card emit its sound. The audible cards hold the sound data used in the game and in our experiment. The master device operated by the game master and the multiple audible cards have a "one-to-many" type Bluetooth communication. When a device ID and sound data are specified by the master device, the specified audible card will start playing a specified sound from its own sound set. The application was developed in Swift with "Multipeer Connectivity" frameworks. Each audible card emits its own unique sound stimulus, much like how playing cards correspond to pictures and symbols. The role of an audible card as a component of the board game is the presentation of public information. The aim of this system is to place these audible cards on the board as the game's components, as well as to make the players recognize what each card represents and the position of the card through auditory stimuli.

Figure 2. System overview: (**a**) a small tablet device regarded as a card: an audible card. (**b**) The game master plays an audio signal from the audible card by wireless connected device. (**c**) Players recognize the information of the card through acoustic sensation.

Auralization has been used in many solutions for improving accessibility problems for the visually impaired [27–29]. The auditory card game system works by presenting the contents of the cards through sound rather than visual information. This is a process of making audible the components of a board game that originally relied on visual and tactile modalities.

3.3. Auditory Uta-Karuta

In normal card-style *Karuta*, both players need to recognize public information on the board without touching the cards. In order to do so, it is necessary to recognize the contents of the cards and the arrangement of the cards by looking at them, which causes a problem of accessibility for people with visual impairments. The aim of the auditory card game system is to represent information about the contents and arrangement of multiple cards using auditory stimuli rather than visual stimuli. This will improve accessibility for people with visual impairments because they can recognize the contents and arrangement of multiple cards simultaneously as a result of the presentation of sounds. We developed "Auditory *Uta-Karuta*" using an auditory card game system into *Uta-Karuta*, to create a new game that can give people with visual impairments access to certain games for effective inclusion. Figure 3 shows the system configuration and procedure of "Auditory *Uta-Karuta*". "Auditory *Uta-Karuta*" consists of multiple audible cards and a parent device operated by the game master. Multiple audible cards are called *Torifuda* devices, and correspond to *Torifuda* in *Uta-Karuta*. The parent device is called a *Yomifuda* device, and corresponds to *Yomifuda* in *Uta-Karuta*. The *Yomifuda* device can not only read out *Yomifuda* but also make *Torifuda* devices emit sound by wireless connection to each *Torifuda* device. Each *Torifuda* device emits a unique sound stimulus, much like playing cards correspond to pictures and symbols, according to the *Yomifuda* device's instruction. *Yomifuda* presented by *Yomifuda* device, and *Torifuda* devices exactly correspond to each other. In this experiment, no information was shown on the screen of the audible cards. The rules of "Auditory *Uta-Karuta*" are described below.

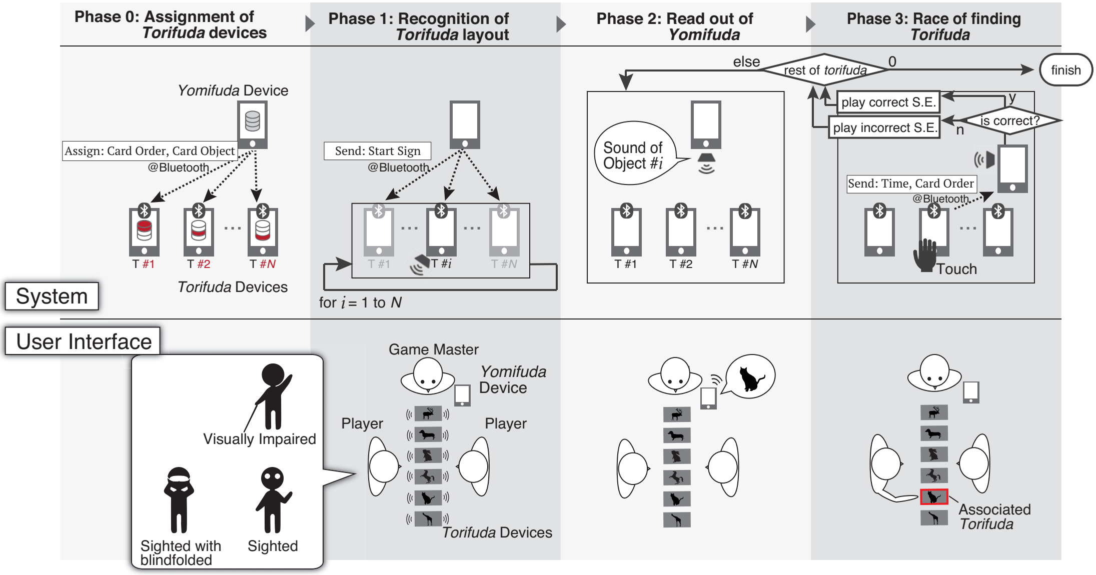


Figure 3. The behavior of the auditory card game system when playing "Auditory *Uta-Karuta*" and the procedure of the game. All devices are communicated by signal. A game master operates the *Yomifuda* device to proceed the game.The players obtain the card by touching/grabbing the *Torifuda* device corresponding to the sound presented by the *Yomifuda* device.

Phase 0: Assignment of *Torifuda* devices

The first step is to establish wireless communication between the *Torifuda* devices and the *Yomifuda* device. The wireless communication was implemented using Apple's "MultiPeer Connectivity" framework, and all communication was done in signal to avoid delays as much as possible. After communication is established, the *Yomifuda* device randomly assigns content from its own sound data set for each of the *Torifuda* devices. At the same time, *Yomifuda* device gives the numbers to the *Torifuda* devices. This number indicates the order in which the *Torifuda* devices are presented with the sounds in Phase 1.

The *Torifuda* devices hold the sound data used in the experiment. When given signal indicating a particular sound from the *Yomifuda* device, the *Torifuda* devices find the appropriate sound in its own sound data and prepares to play it. Similarly, when given signal indicating the order of read out, the *Torifuda* devices set the waiting time until playback according to the value of the signal.

Phase 1: Recognition of *Torifuda* Layout

Arrange the *Torifuda* devices face up on a flat surface in front of the players. When the game master presses the "Start" button displayed on the *Yomifuda* device, a signal the start of read out is given to all connected *Torifuda* devices. When the *Torifuda* devices receive a signal the start of read out from the *Yomifuda* device, they start reading based on the pre-set wait time and sound data. Players memorize the contents and position of the *Torifuda* devices as much as possible.

Phase 2: Read out of *Yomifuda*

The screen of the *Yomifuda* device shows the same number of buttons as the number of sounds used in the experiment. Each button corresponds to a sound, and when a button is pressed, the corresponding sound is played from the *Yomifuda* device. The game master operates the *Yomifuda* device, selects the *Yomifuda*, and presents the corresponding audio signal through the *Yomifuda* device.

Phase 3: Race of Finding *Torifuda*

> Players race to determine which *Torifuda* devices corresponds to the *Yomifuda*, and touch/grab what they think are the correct *Torifuda* devices. In this phase, only the touched *Torifuda* device gives its own card order to the *Yomifuda* device. The *Yomifuda* device determines whether or not the correct *Torifuda* device was touched based on the card order given, and plays the correct/incorrect sound effect.

Subsequently, players repeat Phase 2 and Phase 3 until all the *Torifuda* devices are gone. If a player selects incorrect *Torifuda* device, the incorrect *Torifuda* device is re-arranged among the players, and his/her opponent obtain the correct *Torifuda* device. The winner is the player who has the most *Torifuda* devices. "Auditory *Uta-Karuta*" was prepared in two types: Type α and Type β. The details of these types are described below.

Type α: In Phase 2, all *Torifuda* devices do not present the sound.

Type β: In Phase 2, all *Torifuda* devices present the sound.

People with visual impairments are thought to have a higher capacity for processing auditory information than sighted people because they spend the majority of their daily lives using their hearing ability. In other words, auditory memory can also vary due to players' characteristics. Based on this possibility, Type α and Type β were prepared. Type α does not present the sounds from the audible cards in Phase 2, so players need to remember the arrangement of the six audible cards and what the sounds indicate. On the other hand, Type β is designed to ignore the differences in the players' auditory memory abilities because the audible cards present the cards' contents with sounds in Phase 2. As there is a difference between these conditions, we can conclude that the positions of the six audible cards and the recognition of their contents can vary according to the characteristics of the players.

4. System Evaluation

4.1. Overview

In this paper, we propose an auditory card game system that presents the card's contents by sound so as to make a game accessible for people with visual impairments. We developed a game called "Auditory *Uta-Karuta*" by using the proposed system, and conducted an evaluation experiment in which participants played against each other. The purpose of this experiment is to test the applicability of auditory card game system to board games. For this purpose, the game performance of three different types of players was revealed thorough the experiment. Since the proposed system assumes that people with visual impairments can participate in a board game together with sighted people, we prepared sighted people and people with visual impairments as experimental participants. In addition to those, we also prepared sighted players with blindfolds and conducted a game between sighted players and sighted players with blindfolds to clarify how much their game performance is affected due to the presence of vision. Furthermore, by examining a game played between a visually impaired person and a sighted person wearing a blindfold, it was possible to examine the difference in auditory abilities between people with visual impairments and sighted people. In this experiment, since the participants were required to touch/grab audible cards, the sighted participants who were able to recognize the positions of the audible cards through their vision were considered to have an advantage in the card-grabbing situation. On the other hand, visually impaired participants spend their daily lives using their hearing ability and, thus, have a higher capacity for recognizing auditory information than sighted participants. "Auditory *Uta-Karuta*" does not present visual contents to all of the participants, and it is necessary to recognize the audible cards mainly with one's hearing ability. As mentioned above, while sighted participants are expected to be better at touching/grabbing the audible cards, visually impaired participants are expected to be better at detecting the correct audible cards due to their high auditory abilities. Therefore, an evaluation experiment was conducted to find the extent to which the respective advantages and disadvantages of people with visual impairments and sighted people affect the game's results. As a comparison, we prepared Braille *Karuta*, which is representative of existing tactile presentation methods.

4.2. Stimuli and Equipment

For each audible card, an iPod touch (6th generation) was used. Six audible cards were placed horizontally in a row between two participants, with a distance of 70 mm between each card. No information was shown on the screen of the audible cards, and the players had to obtain information only by hearing. The two participants sat face to face with a distance of about 1 m between them. The sounds used as stimuli were consisted of six different animal crying sounds obtained from [30]. All sounds were 5 s in length at Phase 1 to ensure clarity of the card's contents and positions. Sounds were sampled at 44.1 kHz with 16-bit resolution, normalized for amplitude across conditions, and presented in mono. The experiment was programmed in Swift and presented using the speaker of the iPod touch (6th generation).

We used bridge cards (56 mm × 88.9 mm), one of the common standards for playing cards, as the Braille cards for comparison. These cards had animal names written in Braille and Japanese characters on both ends of the card so that participants could understand them, whether they have a visual impairment or not. They were placed horizontally in a row between two participants, in the same position as the audible cards.

There were 18 participants (14 males and four females) aged from 19 to 42 years old. Among the participants, there were six people with clear vision (four males and two females), six people with visual impairment (five males and one female), and six people with a blindfold (five males and one female). All visually impaired participants could read Braille. In contrast, all sighted participants could not read Braille. All participants had normal hearing and no physical or language disabilities. They had practiced the game before the experiment and understood the rules well.

4.3. Procedure and Evaluation

At the beginning, the participants listened to the six sounds used in the experiment and grasped the meaning of the cards they indicated. For example, a cat crying sound corresponded to the card's content "Cat". The participants sat face-to-face and checked the positions of the six audible/Braille cards in front of them, either visually or by touching them with their hands. In addition, a test phase was set up—the sound used in the experiment was randomly generated by audible cards laid out in front of the participants, and the participants were asked to answer with the meanings of the sounds. The purpose of this phase was to make sure that the participants heard each audible card correctly. After confirming the position of the audible cards and the sounds and practicing touching and grabbing the audible cards several times, they moved on to the main experiment. In the main experiment, the participants played "Auditory *Uta-Karuta*" using the procedure described in Section 4.2. In addition, they played both Type α and Type β. The six sounds used in each experiment were not changed, but the order of the six types of cards and the cards to be read were randomly changed each time. In the case of using Braille cards, which is the comparison object, the sighted participants grasped the contents of the cards visually, and the visually impaired and blindfolded participants grasped the contents of the cards by touching the Braille. In this case, the time used in Phase 1 was 30 s, as in the case of audible cards. The game was played in the following three states.

Case 1: Person with a visual impairment vs. sighted person

Case 2: Person with a visual impairment vs. sighted person with a blindfold

Case 3: Sighted person vs. sighted person with a blindfold

Case 1 was a competition between a person with a visual impairment and a sighted person, the aim of which was to verify whether the "Auditory *Uta-Karuta*" allowed people with visual impairments to be equally competitive. If there was a difference in the results in Case 1, we could conclude that the accessibility of "Auditory *Uta-Karuta*" for people with visual impairments was inadequate. Case 2 was a competition between a person with a visual impairment and a sighted person with a blindfold, the aim of which was to verify whether an equal match was possible when participants were completely deprived of visual information. Furthermore, if there was a difference in the results in Case 2, the difference in hearing ability between people with visual impairments and sighted people would be confirmed. Case 3 was a competition between a sighted person and a sighted person with a blindfold, the aim of which was to clarify the extent to which the sighted person was affected by their vision. If there was a difference in the results in Case 3, we could conclude that game performance in "Auditory *Uta-Karuta*" is affected due to the presence of vision for sighted players. The participants were informed about the degree of visual impairment of their opponents before the experiment.

The evaluation method is described below. One of the theories that examines the situations in which people find games interesting is flow theory [31]. According to this theory, people feel enjoyment when the difficulty of the challenge is in competition with their own abilities, and when it is close to the limit of their abilities. These "flow states" have been adopted in theories of game enjoyment [32–34]. Abuhamdeh [35] found that games with better opponents and closer games gave more enjoyment to people. In the game of *Karuta*, the winning rate and the strength of the opponent should be close to one's own, which is considered to be a condition that enables the game to be played for an interesting and close game. In addition, since *Karuta* is a game that includes an element of luck—it is possible to guess a card correctly before the read-out phase—it is hard to consider that the winning rate and the feeling of competitiveness are in perfect correspondence. Therefore, for the system evaluation experiment, we used the winning rate as a quantitative evaluation and the subjective evaluation of the opponent as a qualitative evaluation.

(a) Winning Rate:

Winning rate of the people with visual impairments in Cases 1 and 2 and of those with a blindfold in Case 3.

(b) Subjective Evaluation of the Opponent:

The opponent's strength [0, 100] when the player strength is 50.

The winning rate is determined based on players who do not have visual information (people with visual impairments and people with a blindfold) in *Karuta*. The more equal the competition is, the closer the winning rate will be to 50%. Subjective evaluations of the opponents were used for the purpose of subjectively evaluating the strength of an opponent's game performance based on the competition, regardless of the winning rate. The participants' own strength was defined as 50, and they chose an integer value from 0 to 100 to rate the strength of their opponent. The value of 0 is the weakest and 100 is the strongest. Thus, if the value is close to 50, which is the strength of the players themselves, the competition can be considered psychologically equal.

4.4. Results and Discussion

In this experiment, we evaluated the results by having participants play "Auditory *Uta-Karuta*" with three types of experiments: sighted person, visually impaired person, sighted person with a blindfold. Figure 4 shows the result of the evaluation experiment. For the winning rate, a Welch's *t*-test was conducted to examine the significant differences ($p < 0.05$) from the winning percentage of 50%. For the subjective evaluation of the opponent, Fisher's Least Significant Difference (LSD) test was conducted to examine the significant differences ($p < 0.05$) from the existing method using Braille cards.

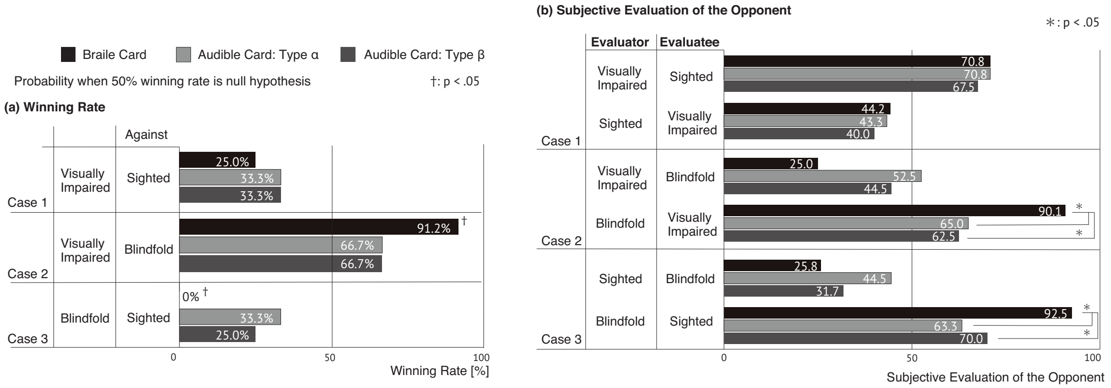


Figure 4. (**a**) Winning rate of the people with visual impairments in Cases 1 and 2 and those with a blindfold in Case 3, (**b**) subjective evaluation of the opponent: The opponent's strength [0, 100] when the player strength is 50.

Figure 4a shows the winning rate of the evaluation experiment. Although there were significant differences in the method using Braille cards, there was no significant difference in Cases 2 and 3. In other words, the null hypothesis that the winning rate was 50%, was rejected when the tactile presentation was used, whereas the null hypothesis was not rejected when the proposed method was used. Therefore, it was possible to enable fair competition, in which the players can be given the same degree of victory opportunity, even without visual information.

Figure 4b shows the subjective evaluation of the opponent of the evaluation experiment. There were significant differences between using Braille cards and using audible cards in Cases 2 and 3. In other words, the null hypothesis that there is no difference in the evaluation value of the opponents between the previous and proposed methods, was rejected. In addition, the results in the proposed method were close to 50. Therefore, the proposed method makes it possible to boost the sense of rivalry, even without visual information compared to using Braille *Karuta*.

Furthermore, Figure 5 shows the questionnaire results of the enjoyment of each game. The participants rated their enjoyment of playing Braille *Karuta* and "Auditory *Uta-Karuta*" on a scale of 1 to 5. Welch's t-test was conducted to examine the significant differences ($p < 0.05$) from the existing method using Braille cards. For all the participants in the experiment, "Auditory *Uta-Karuta*" was more enjoyable. While the sighted people can recognize the Braille cards by reading the characters on them, the people with visual impairments and the blindfolded people can only recognize Braille cards by touching them. Therefore, the sighted people could easily win the game when they played Braille *Karuta*, and could not feel enough enjoyment. Although the people with visual impairments can understand Braille cards, they were unable to win in a match against a sighted player. Similarly, when playing against a blindfolded player, it was impossible to realize close competition because blindfolded players could not understand Braille cards. For these reasons, it is assumed that the people with visual impairments were not able to feel enjoyment playing Braille *Karuta*. The blindfolded participants were not familiar with Braille and could not fully understand the Braille cards, so they did not enjoy Braille *Karuta*. On the other hand, in the game of "Auditory *Uta-Karuta*" all the participants recognized the cards by sound, and the game became closer to an equal match, which resulted in a sense of enjoyment. Furthermore, compared to Braille *Karuta*, the difference in the evaluation of enjoyment between the attributes of the experimental participants tended to be smaller for "Auditory *Uta-Karuta*". It is assumed that this is because the audible cards can make the performance of recognizing them closer to the same level regardless of the attributes of the experimental participants compared to the Braille cards.

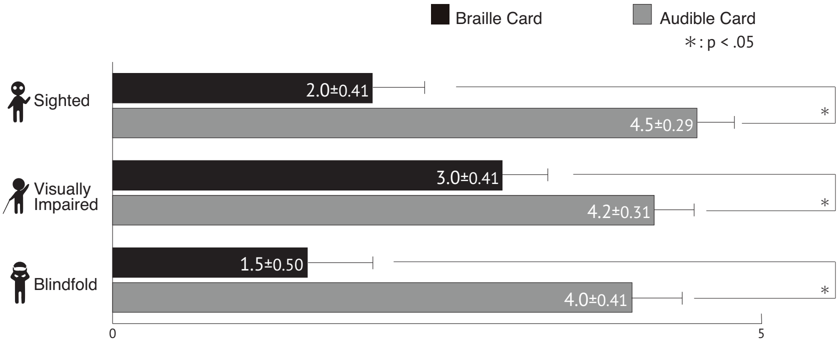


Figure 5. Results of the questionnaire. The participants of the experiment rated their enjoyment of playing Braille *Karuta* and "Auditory *Uta-Karuta*" on a scale of 1 to 5.

Thus, the results showed that "Auditory *Uta-Karuta*" has the potential to be played equally regardless of whether they have a visual impairment or not compared to using Braille cards. Furthermore, it proved to be possible to realize not only equal competition, but also more enjoyable games for the players. This result suggests that the proposed system may be used as one of the means to convey public information in board games that are accessible to the visually impaired. Although they showed that the game was significantly effective in a match between a visually impaired person and a sighted person with a blindfold, there was no significant difference in a match between a visually impaired person and a sighted person. In other words, the experimental game could not completely enable fair competition, regardless of visual impairment. However, it is interesting that the visually impaired persons had a higher winning rate compared with sighted persons with a blindfold, even though neither of them could obtain visual information. Similarly, the visually impaired persons were highly evaluated by sighted persons with a blindfold. These results suggest that the people with visual impairments may have better auditory memorization and directional resolution abilities than the sighted people. This suggestion is assumed would be proper in light of the fact that the visually impaired live mainly using their sense of hearing and touch instead of their sense of sight. This is verified by previous studies [36,37]. In spite of the above facts, the reason for the lack of significant difference between the visually impaired and the sighted is due to the way they touch/grab the audible cards. In fact, in the experiment, it was observed that many visually impaired participants failed to touch/grab the cards. In other words, in this game of "Auditory *Uta-Karuta*", the advantage of sighted people's ability to accurately recognize the positions of the audible cards had a greater impact on the outcome of the game than the high auditory abilities of the visually impaired people. However, it is easy to assume that it would be difficult for visually impaired people to perform as well as or better than sighted people in the card-grabbing situation, unless the sighted people were given a handicap or the visually impaired people were given an advantage. As mentioned above, the degree of usefulness of an auditory card game system can change depending on the characteristics of the players and the rules of the game. Therefore, in the future, it is possible to consider a flexible game design that presents images in order to control the game performance of sighted players. In order to design such a flexible game in the future, it is first necessary to verify the effects of an auditory card game system that uses only auditory cues for the players by changing each valuable parameter. This process reveals the values of the parameters that allow each player to maximize the recognition of audible cards, as well as the parameters that affect game performance depending on the players' characteristics. In the following sections, we investigate each valuable parameter of the system and explain the limitations of the proposed system and appropriate game design based on the results of various existing studies.

5. Suggestion for Suitable Auditory Cues in Auditory Card Game System

5.1. Overview

In the previous sections, we developed "Auditory *Uta-Karuta*" and verified the usefulness of this system by actually having the game played by visually impaired people, sighted people, and sighted people with a blindfold. As a result, we could not conclude that "Auditory *Uta-Karuta*" had a completely equal game design, as it favored the sighted participants. However, compared to Braille *Karuta*, "Auditory *Uta-Karuta*" approached a more favorable result. This result suggests that, compared to games that use tactile means of presentation, "Auditory *Uta-Karuta*" has a greater potential to make games accessible, regardless of visual impairment. Although we used animal calls and arranged the audible cards horizontally in "Auditory *Uta-Karuta*", we believe that we can find specific valuable parameters that will enable even more equal games by considering a comprehensive range of possible game designs. In other words, it is necessary to search for the elements required for an equal game design that takes the characteristics of sighted people—who have vision—and visually impaired people—who have high hearing abilities—into account.

In this section, we investigate each valuable parameter of the system and explain the limitations of the proposed system and appropriate game design based on the results of various existing studies.

5.2. Valuable Parameters

The main components of our system are the degree of visual impairment, the selection of the audio type, and the style of sound playback. Details are described below.

(1) **Degree of visual impairment:** Since the proposed system assumes that people with visual impairments can participate in the same board game together with sighted people, we prepared sighted people and people with visual impairments as parameters indicating the degree of visual impairment. Furthermore, in order to clarify how much the visual information contributes to grasping the contents and position of the components for the sighted, the sighted people with a blindfold were prepared.

(2) **Audio types:** As shown in Table 2, the presentation of information using auditory stimuli can be divided into verbal sounds and nonverbal sounds, the latter of which can be further divided into representational sounds and abstract sounds. A typical of representational sounds is "auditory icons" [38], and a typical of abstract sounds is "earcons" [39]. Representational sounds are real world sounds that have a direct association with an object, and abstract sounds are synthetic sounds that have no direct association with an object. Each of these types of sound have advantages and disadvantages. The advantage of verbal sounds is the meaning of the message is relatively unambiguous [38]. However, users have to listen to the whole message to understand the meaning. Moreover, verbal sounds have language limitations. The advantage of representational sounds is that they can convey complex messages in a single sound. In addition, representational sounds are easy to identify by analogy. However, they cannot be assigned real world sounds for all events. The advantage of abstract sounds is flexibility. They can present information in a systematic way [38]. Further, they can represent hierarchies by controlling their parameters such as timbre and pitch. However, they have difficulties associated with learning and remembering. In this experiment, we prepared three types of audio to be used: A: verbal sounds, B: representational sounds, and C: abstract sounds.

(3) **The style of sound playback:** We prepared two types of sound playback. Type α, presents the auditory stimuli one by one in order, and Type β presents the sound for all audible cards at the same time. While Type α has the advantage that players can clearly recognize the information of the audible card one by one, Type β has the advantage that players can obtain the information of multiple audible cards at once.

Hence, as shown in Table 3, we prepared parameters for three types of degree of visual impairment, three audio types, and two types of sound playback.

Table 2. The list of audio types using experiment.

	A: Verbal Sounds	Non-Verbal Sounds	
		B: Representational Sounds	**C: Abstract Sounds**
Summary	Sound presentation using real world sounds	Real world sounds that have a direct association with an object	Synthetic sounds with artificial correspondence without using real world sounds
Advantage	Provide clear information	Intuitive understanding by analogy	Systematic presentation
Disadvantage	Language limitations	Cannot represent all the events	Lack of meaningful relationship with their referent

Table 3. The composition of valuable parameters.

Degree of Visual Impairment	Audio Types	The Style of Sound Playback
1: Sighted	A: Verbal	Type α: Play sounds one by one in order
2: Visually impaired	B: Representational	Type β: Play sounds simultaneously
3: Sighted with blindfold	C: Abstract	

5.3. Stimuli and Equipment

For the audible cards, iPod touch devices (6th generation) were used. A total of 12 audible cards were placed in front of the participants, and each card was arranged as shown in Figure 6. No information was shown on the screen of the audible cards, and the participants had to obtain information only by hearing. Sounds were sampled at 44.1 kHz with 16-bit resolution, normalized for amplitude across conditions, and presented in mono. The experiment was programmed in Swift and presented using the speaker of the iPod touch (6th generation). The representational and abstract sounds were obtained from [30]. The verbal sounds were an adult female voice (Kyoko) in Japanese, created using Apple's speech reading software. Representational sounds were environmental sounds which were selected based on a strong relationship between the sound and the event with which it was associated. For example, a cat crying sound was used for the card's content "Cat". The detail of correspondence between the card's contents and the verbal/representational sound is shown in Table 4. Abstract sounds were designed to enable discrimination between multiple abstract sounds and to ensure localization performance, referring to guidelines and experimental results in previous studies [40–42]. Brewster [40] suggested using musical timbres with multiple overtones and including a range from 0.125 to 5 kHz in his guidelines for abstract sound design. In addition, Brewster recommended to add delay of about 0.1 s between each abstract sound when playing them one after another. Patterson [41] found that similar rhythms can confuse abstract sounds even when there are large differences in spectra. Morikawa [42] concluded that auditory stimuli consisting of only frequencies below 2 kHz or above 12 kHz cannot provide sufficient horizontal sound localization.

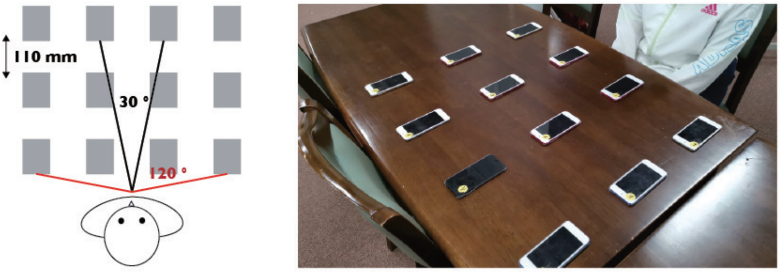


Figure 6. The state of experiment: The subject sat on the chair and described the content and position of the audible cards laid out in front of them.

Table 4. Verbal Sound and representational sound sets used in the experiment. Verbal Sound sets are pronounced in Japanese, and written in international phonetic symbols.

Card Contents	Verbal Sound	Representational Sound
Cat	ne\ko	Cat's meow
Glass	gɯᵝ \ɾasɯᵝ	Pouring water into a glass
Cow	ɯᵝ \ɕi⁻	Cow's bark
Toad	ka\eɾɯᵝ⁻	Frog's croak
Knife	na\iɸɯᵝ	Cutting food with a knife
Keyboard	kʲi/ːbo\ːdo	Tapping a keyboard
Crow	kaɾasɯᵝ	Crow's caw
Dog	i/nɯᵝ \	Dog's bark
Scissors	hasamʲi	Scissors cutting paper
Dentifrice	ha/bɯᵝ \ɾaɕi	Brushing one's teeth.
Hair Dryer	do\ɾaijaː⁻	Drying one's hair with a hair dryer
Cicada	semʲi	Cicada noise

Based on these guidelines and results, the abstract sounds were designed as follows. They consisted mainly of musical timbres with a wide bandwidth from 0.5 to 16 kHz and overtones as much as possible to ensure both the guideline and the localization performance. Furthermore, each abstract sound was designed with a different rhythm to reduce confusion. In the Type α experiment, a 0.1 s gap was inserted between each sound to enable the participant to determine when one sound ends and the next begins. Abstract sounds were randomly assigned to the contents of the card.

Participants were tested on six stimulus combinations of three audio types and two methods of presenting multiple sounds. Blocks of demonstrations and experiments were presented to participants with a new random order of stimuli in each block. In the blocks of demonstrations, the participants listened to the 12 sounds used in the experiment and grasped the meaning of the cards they indicated. The maximum memorization time was 10 min for each audio type. The participants sat and checked the positions of the 12 audible cards in front of them, either visually or by touching them with their hands. In addition, the test sounds were played from each of the 12 cards in turn, and the participants had to confirm correspondence between the sound source and the card's position. In the experiment block, each of the 12 audible cards emitted a different sound. In the Type α experiment, all sounds were presented randomly for 5 s each. The participants described orally the meaning and the position of the sound presented by each audible card. There was no time limit for this, and the participants were asked to describe as much as possible. In this experiment, the identification rate and recognition rate were evaluated. The identification rate indicates how many of the 12 audible cards the participants were able to distinguish. The recognition rate indicates how many of the 12 audible cards of which the participants were able to correctly grasp the location and contents.

There were 18 participants (10 males and eight females) aged from 19 to 42 years old. Among the participants, there were five people with clear vision (four males and one females), seven people with visual impairments (three males and four females) and six people with a blindfold (three males and three females). All participants had normal hearing and no physical or language disabilities.

5.4. Results

Figure 7 shows the differences in experimental performance according to the degree of visual impairment. In this analysis, a Tukey Honestly Significant Difference (HSD) test was conducted to examine the significance of the differences ($p < 0.05$). In the case of Type α, both the identification rate and recognition rate were significantly different for the visually impaired compared to the sighted or blindfolded sighted. In other words, the null hypothesis that there is no difference in the ability to recognize sounds between sighted and visually impaired was rejected. This shows that the people with visual impairments had a better ability to discriminate and localize sounds in verbal and representational sounds than sighted people and blindfolded people. In the case of Type β, there was no significant difference in either the identification rate or the recognition rate.

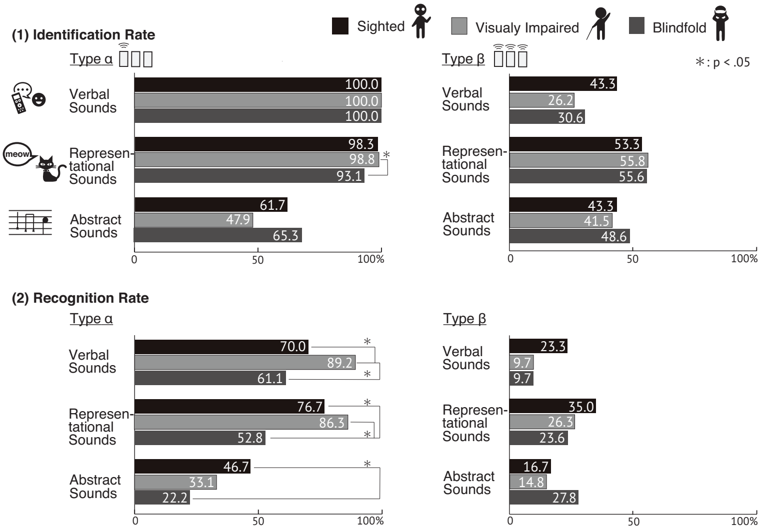


Figure 7. Differences in performance by degree of visual impairment.

Figure 8 shows differences in experimental performance according to the audio type. In this analysis, a Tukey HSD test was conducted to examine the significance of the differences ($p < 0.05$). In the case of Type α, the abstract sounds were significantly different from the other two audio types for all participants in both the identification rate and the recognition rate. In other words, the null hypothesis that the effect of the abstract sound on the identification/recognition rate of the player is not different from that of the other two types of sound was rejected. This suggests that verbal and representational sounds were effective in this system in presenting the content and location of the audible cards for all experimental participants. In the case of Type β, there was no significant difference in either identification rate or recognition rate. As a general tendency, representational sounds tended to obtain higher identification and recognition rates. In addition, we obtained the opinion from the participants that "Verbal sounds were mixed up and became unintelligible when they were emitted at the same time". These results suggest that it may be inappropriate to use verbal sounds in Type β.

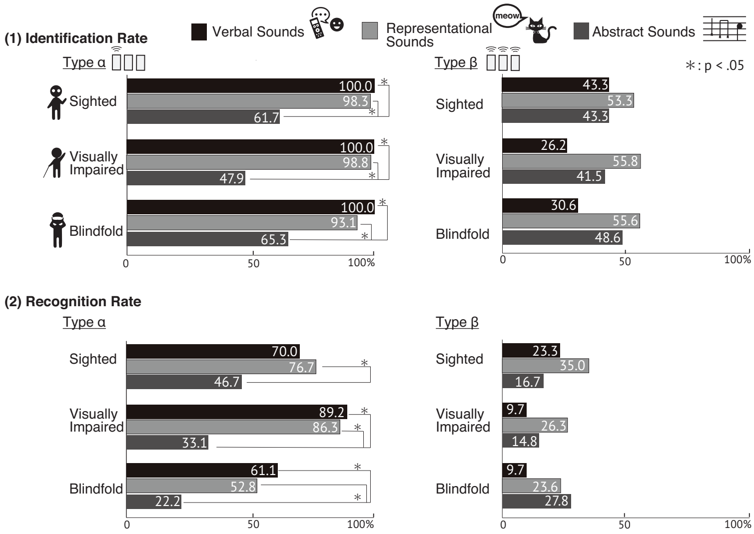


Figure 8. Differences in performance by audio types.

Figure 9 shows the differences in experimental performance according to the style of sound playback. In this analysis, Welch's *t*-test was conducted to examine the significance of the differences ($p < 0.05$). Type β has the advantage that players can obtain the information of multiple audible card at once compared with Type α. There was a significant difference for all participants in both the identification rate and the recognition rate compared to Type α. In other words, the null hypothesis that there is no difference in the effect of the style of sound playback on the player's identification/recognition rate was rejected. This suggests that Type β is not effective for accurately recognizing multiple audible cards.

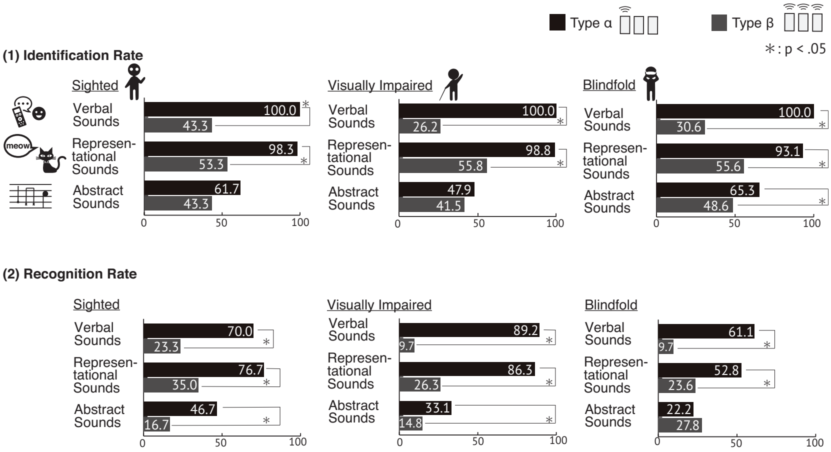


Figure 9. Differences in performance by the style of sound playback.

5.5. General Discussion

This experiment was conducted to clarify the effect of each valuable parameter of the proposed system on the player's game performance. We prepared parameters for three types of degree of visual impairment, three audio types, and two types of the style of sound playback.

Comparing the experimental results according to the difference in the degree of visual impairment showed that the people with visual impairments had a better sound discrimination and localization ability than the sighted people. In addition, a better sound discrimination and localization ability of people with visual impairments has been mentioned in previous studies [43,44]. Therefore, people with visual impairments were able to recognize the audible cards as well as or better than the sighted people. In other words, we found that sound presentation was effective as a solution for presenting public information indicated by components in board games.

Comparing the experimental results according to audio types, the results showed that representational and verbal sounds enabled players to accurately recognize the information indicated by the components. Although the verbal sounds were able to convey the information correctly to all the participants in the experiment, one participant commented that the verbal sounds prompt confusion when they are produced simultaneously. In addition, the fact that verbal sounds are easily masked by background noise has been mentioned in previous studies [45]. Thus, it is undesirable to use multiple verbal auditory stimuli in this system. Further, verbal sounds should not be used in board games where verbal communication is active, because players can miss the information indicated by the components. In other words, verbal sounds should only be used as audio feedback at times when verbal communication is not occurring, such as before and after a competition. As with the verbal sounds, the representational sounds were able to convey information correctly to all participants. This shows that representational sounds have a great potential to be used effectively in an auditory card game system. In addition, they are superior in terms of intuitiveness, learnability, memorability, and user preference according to previous work [45–47]. These advantages suggest that representational sound has the potential to make board games even more fun. Compared to verbal sounds, they are less likely to be masked by background speech [45], so they are more effective when presenting multiple sounds stimuli simultaneously. When representational sounds cannot fully express information, abstract sounds should be used. Since abstract sounds are not easily masked by background sounds as well as representational sounds [45], they can also be effective when multiple auditory stimuli are required. However, unlike representational sounds, abstract sounds are obviously limited in the size they can be learned. The experimental results obtained in this study also confirmed that abstract sounds can cause recognition errors. According to the experimental results, when abstract sounds were used as stimuli, the maximum number of audible cards correctly recognized by the participants was from about 5 to 8, regardless of the degree of visual impairment. Similarly, previous studies [41,48] have recommended limiting the set of abstract sounds to a maximum size of 5 to 8. For these reason, it is appropriate to keep the number of abstract sounds within the range of 5 to 8. We recommend the use of abstract sounds as feedback for player actions and changes in the game, rather than indicating component contents.

Comparing the experimental results according to the style of sound playback showed that using many cards at the same time has a negative impact on recognizing components. The number of audible cards that the participants could recognize correctly was about 3, even if the auditory stimuli were given at the same time. Therefore, it was suggested that the players could recognize the audible cards correctly if the number of these was about 3. However, verbal sounds are easily masked by background speeches, so representational/abstract sounds should be used when multiple audible cards are presented simultaneously.

Finally, in order to clarify what caused the errors of the participants, we investigated the error composition in the experiment. Errors consist of the following three types.

1. Sound identification error: experimental participants misinterpreted the meaning of the audible card presented them.

2. Vertical localization error: experimental participants misplaced the audible card presented them in the vertical direction.
3. Horizontal localization error: experimental participants misplaced the audible card presented them in the horizontal direction.

Figure 10 shows the average number of errors per trial for the experimental participants. In this analysis, a Tukey HSD test was used to examine the significance differences ($p < 0.05$) in the number of errors between players. The null hypothesis is that there is no difference in the average number of errors between players.

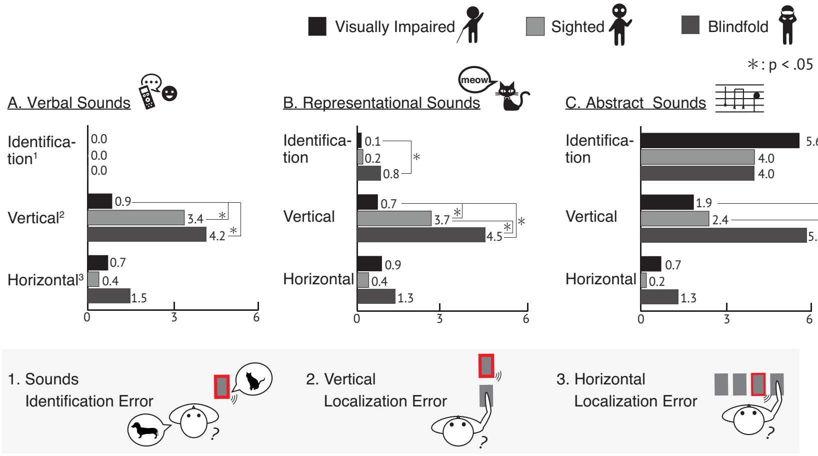


Figure 10. The average number of errors per trial for the experimental participants. There were three types of errors: sound identification errors, vertical localization errors, and horizontal localization errors.

Sound identification errors were generally not significantly different regardless of the audio type. However, there was a significant difference between the visually impaired and the blindfolded in the representational sound. This suggests that the presence/absence of visual information can have an effect on the sound discrimination itself.

The number of vertical localization errors for the visually impaired was significantly lower in verbal and representational sounds than for the sighted. The reason for this result is considered to be that the speaker of the iPod touch is located at the bottom of the device. It was predicted that the sighted people were influenced by the visual information and mistook each device for the lower device. However, the number of errors for the visually impaired was significantly lower than that for the blindfolded participants, who were excluded from visual information. In other words, people with visual impairments showed a higher vertical directional resolution than the sighted people. Ohuchi [43] and Voss [44] concluded that people with visual impairments tend to have higher horizontal directional resolution and distance resolution than sighted people. Therefore, the results of the high vertical resolution of the visually impaired in this experiment can be attributed to their high distance resolution. However, in the vertical localization of abstract sounds, people with visual impairments did not show a significant difference in the number of errors compared to the sighted people. Therefore, the results suggest that people with visual impairments can have difficulty with the source vertical localization of abstract sounds compared to verbal and representational sounds. In other words, the use of abstract sounds in the proposed system can interfere with the high vertical resolution of people with visual impairments. As a result, it is not desirable to use abstract sounds in situations where we want to ensure the vertical directional resolution of visually impaired people, such as when we want to accurately convey the position of a component. However, abstract sounds have the advantage of being freely designable, so reflecting the component's position in the rhythm, pitch, and timbre of the sound could make it easier to locate the component.

Horizontal localization errors did not differ significantly among participants, regardless of the audio type. Horizontal localization seems to be easier for all subjects because it reduces misrecognition of sighted participants due to the position of the speaker and increases auditory cues such as interaural level difference and interaural time difference. Therefore, if the component's position is to be shared by all players, it is recommended that the components be aligned as horizontally as possible. Furthermore, it is possible to control the placement of the cards to give an advantage to people with visual impairments.

To allow for more flexible board game design, it is also possible to include visual information on the audible cards. At this time, it is easy to assume that visual stimuli affect the recognition of components to sighted people. In fact, in previous studies [49], it has been concluded that the type of sound and its congruency with visual information can affect reaction time. Therefore, it is possible to induce confusion by providing visual content that differs from auditory content to sighted people.

5.6. Expandability

The components used in board games are responsible for presenting public/private information to the players. The auditory card game system proposed in this paper is a means of presenting public information to players. The previous solution was tactile presentation methods, such as adding Braille and textures to game components. The situation where the proposed method is superior to the tactile presentation method is considered to be when the components of the board game present public information to the player. The public information is presented to all players continuously, and its content changes as the game progresses. Therefore, as the game progresses, the players need to touch each and every game component to obtain public information. This behavior of the players can cause the game to lose its fun and gameplay. On the other hand, the proposed method enables the player to progress the game without touching the components of the board games. Therefore, this system can provide positive impact to a style of board game similar to that of *Uta-Karuta*, in which a small amount of public information is presented by appropriately controlling parameters, such as the placement of components and the type of sound to be presented. For example, "Concentration" [50] and "Spot it!" [51] are considered to be suitable games for auditory card game systems because they are simple games consisting of only a small amount of public information. Furthermore, the proposed method has the advantage of promoting intuitive understanding by using representational sounds. The typical cards used in the *Uta-Karuta* contain poems on themes such as love and nature. In other words, *Uta-Karuta* is not only a memorization game, but also an emotional fun game. In order to keep the original fun of the *Uta-Karuta* as much as possible, it will be important for players to be able to intuitively understand the components when they get information about them. Thus, games such as "Cobra Paw" [52], which is a speed game, can be considered in order to improve accessibility while maintaining the immersiveness of the game.

As described above, the auditory card game system is preferably used for presenting small amounts of public information, but it can also be used when private information is included. The solution is to wear devices such as earphones or headphones. However, a board game that includes private information requires more strategy than a game that consists only of public information. In this case, communication, such as conversations between players, is necessary, and using earphones or headphones may interfere with such communication. Therefore, it is possible to apply auditory card game systems in board games that include private information, but it may impair the enjoyment and elements of the game.

When an auditory card game system is applied to a board game with a large amount of information, the disadvantage is that it is necessary to play the sounds used in the game in order, as shown in our results, and it takes time to present all the information to the players. Although it is not impossible to apply an auditory card game system when there is a large amount of information, such as by orally explaining the movements over squares when playing a chess game without a board or pieces, it is feared that intuitiveness may be impaired.

Therefore, in order for auditory card game systems to play a role in improving accessibility while effectively guaranteeing the enjoyment of a game, we believe that games should use small amounts of public information and be played in fast-paced situations that require intuition.

6. Conclusions

We proposed an auditory card game system that presents the card's contents by sound, towards playing equally with others, regardless of whether they have a visual impairment or not. This is one of the solutions that presents public information of board games. Although this proposal cannot improve the accessibility of all board games, it contributes to expand the range of inclusive board games for the visually impaired. Furthermore, clarifying the impact of element of the system to the players will greatly assist in designing an appropriate board game using the proposed system. The purpose of this paper is to determine whether the game allows for fair competition for people with visual impairments and to clarify the effects of the valuable parameters of the system on the players. In the system evaluation, we developed "Auditory *Uta-Karuta*" and verified the usefulness of this system by having the game played by visually impaired people, sighted people, and sighted people with a blindfold. The purpose of this experiment is to test the applicability of auditory card game system to board games. We found that "Auditory *Uta-Karuta*" has the potential to be played by all players, regardless of whether they have a visual impairment or not. The following experiment was conducted to measure the effect of the design of the audio cue on the player in the proposed system, and several major considerations were obtained.

1. Abstract sounds can adversely affect sound localization abilities and should not be used where component location sharing is required.
2. In order for all players to correctly recognize the position of the components, audible cards should be aligned horizontally.
3. It is preferable that the sounds stimuli used in the audible card be composed of representational sounds as much as possible.
4. When using abstract sounds, it is recommended to limit the number of auditory stimuli from about 5 to 8 in order to reduce recognition error.
5. It is effective to sound the audible cards one by one so that players can accurately recognize the location and information of multiple audible cards.
6. When presenting multiple audible cards at the same time, it should be limited to about three cards, and no verbal sound should be used in this case.

These results show that it is necessary to design audio cues appropriately due to the characteristic of public information indicated by the components in order to make the auditory card game system effective. In particular, for a board game that requires accurate information presentation for players, such as *Uta-Karuta*, it is considered that playing representational sounds one by one is effective. However, in Section 4, despite the game design described above, the sighted players had an advantage, so it would be effective to adopt a vertical layout that gives an advantage to the visually impaired.

As described above, the results of the experiments in Section 5 can be used to create an appropriate board game design that allows players with different visual states to achieve the same level of game performance. However, the enjoyment of a game depends not only on the winning rate, but also on factors such as strategy and communication. Therefore, further experiments will be needed to construct a board game that is fun for players regardless of whether they have visually impaired or not, with a view to the qualitative factors that may influence the enjoyment of the game. In addition, as mentioned in Section 5.6, the results of this study can even be used for games in which private information is used, and we will consider applying them to the case of games that include private information in future research. Furthermore, in order to reduce the cost and broaden the range of board games that can be supported, it is necessary to consider not only the iPod touch, but also many other component shapes and types.

Author Contributions: Methodology, H.M., N.K., and K.Z.; software, H.M.; investigation, H.M., N.K., and H.E.B.S.; writing—original draft preparation, H.M.; writing—review and editing, N.K., H.E.B.S., and K.Z.; supervision, K.Z.; project administration, K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by Ethics Review Committee of the University of Faculty of Engineering, Information and Systems, University of Tsukuba (2019R314, 24 May 2019).

Informed Consent Statement: Informed consent was obtained from all participants in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. da Rocha Tomé Filho, F.; Mirza-Babaei, P.; Kapralos, B.; Moreira Mendonça Junior, G. Let's Play Together: Adaptation Guidelines of Board Games for Players with Visual Impairment. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; p. 631.
2. Zan, B. Interpersonal Understanding Among Friends: A Case-Study of Two Young Boys Playing Checkers. *J. Res. Child. Educ.* **1996**, *10*, 114–122. [CrossRef]
3. Konieczny, P. Book Review: Stewart Woods, Eurogames: The Design, Culture and Play of Modern European Board Games. *Int. Sociol.* **2017**, *32*, 671–673. [CrossRef]
4. Potter, R.J. Encounters: Two Studies in the Sociology of Interaction. Erving Goffman. *Am. J. Sociol.* **1962**, *68*, 125–126. [CrossRef]
5. Yuan, B. Towards Generalized Accessibility of Video Games for the Visually Impaired. Ph.D. Thesis, University of Nevada, Reno, NV, USA, 2009.
6. Andrade, R.; Rogerson, M.J.; Waycott, J.; Baker, S.; Vetere, F. Playing Blind: Revealing the World of Gamers with Visual Impairment. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; p. 116.
7. Maia, J.; Del Prette, Z.; Freitas, L. Social skills of visually disabled people. *Rev. Bras. Ter. Cogn.* **2008**, *4*. [CrossRef]
8. Werner, E.E.; Smith, R.S. *Journeys from Childhood to Midlife: Risk, Resilience, and Recovery*; Cornell University Press: Ithaca, NY, USA, 2001.
9. Jessup, G.M.; Cornell, E.; Bundy, A.C. The treasure in leisure activities: Fostering resilience in young people who are blind. *J. Vis. Impair. Blind.* **2010**, *104*, 419–430. [CrossRef]
10. Hamada, T. The Arabian Pots. Available online: https://boardgamegeek.com/boardgame/179956/arabian-pots (accessed on 17 March 2021).
11. Stippell, C. Nyctophobia. Available online: https://boardgamegeek.com/boardgame/249505/nyctophobia (accessed on 17 March 2021).
12. Robbins, M. Uno. Available online: https://boardgamegeek.com/boardgame/2223/uno (accessed on 17 March 2021).
13. Bauza, A. Hanabi. Available online: https://boardgamegeek.com/boardgame/98778/hanabi (accessed on 17 March 2021).
14. Miyakawa, H.; Kuratomo, N.; Salih, H.E.B.; Zempo, K. Auditory Uta-KARUTA: Sonificated Card Game Towards Inclusive Design. In *The Adjunct Publication of the 32Nd Annual ACM Symposium on User Interface Software and Technology*; ACM: New York, NY, USA, 2019; pp. 90–92. [CrossRef]
15. AudioGames.net. Available online: https://audiogames.net/ (accessed on 17 March 2021).
16. Westin, T. Game accessibility case study: Terraformers—A real-time 3D graphic game. In Proceedings of the 5th International Conference on Disability, Virtual Reality and Associated Technologies, Oxford, UK, 20–22 September 2004.
17. Sánchez, J.; Baloian, N.; Hassler, T.; Hoppe, H. AudioBattleship: Blind learners collaboration through sound. In Proceedings of the CHI'03 Extended Abstracts on Human Factors in Computing Systems, Austin, TX, USA, 30 April–4 June 2003; pp. 798–799. [CrossRef]
18. Kim, S.; Lee, K.P.; Nam, T.J. Sonic-badminton: Audio-augmented badminton game for blind people. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 1922–1929.
19. Archambault, D. The TiM Project: Overview of Results. In *Computers Helping People with Special Needs*; Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 248–256.
20. Sánchez, J.; Sáenz, M.; Garrido-Miranda, J. Usability of a Multimodal Video Game to Improve Navigation Skills for Blind Children. *Taccess* **2010**, *3*, 7. [CrossRef]

21. Grabski, A.; Toni, T.; Zigrand, T.; Weller, R.; Zachmann, G. Kinaptic-Techniques and insights for creating competitive accessible 3D games for sighted and visually impaired users. In Proceedings of the 2016 IEEE Haptics Symposium (HAPTICS), Philadelphia, PA, USA, 8–11 April 2016; pp. 325–331.
22. Board Game Geek. Available online: https://boardgamegeek.com/ (accessed on 17 March 2021).
23. Meeple Like Us. Available online: https://www.meeplelikeus.co.uk/ (accessed on 17 March 2021).
24. André, M. Splendor. Available online: https://boardgamegeek.com/boardgame/148228/splendor (accessed on 17 March 2021).
25. Grammenos, D.; Savidis, A.; Stephanidis, C. UA-Chess: A Universally Accessible Board Game. In Proceedings of the 3rd International Conference on Universal Access in Human-Computer Interaction, Las Vegas, NV, USA, 22–27 July 2005.
26. Bull, D. Karuta: Sport or culture? *Jpn. Q.* **1996**, *43*, 67.
27. Rector, K.; Salmon, K.; Thornton, D.; Joshi, N.; Morris, M.R. Eyes-Free Art: Exploring Proxemic Audio Interfaces For Blind and Low Vision Art Engagement. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technology (IMWUT 2017), New York, NY, USA, 3 September 2017.
28. Osmanovic, N.; Hrustemovic, N.; Myler, H.R. A testbed for auralization of graphic art. In Proceedings of the Annual Technical Conference IEEE Region 5, Djerba, Tunisia, 11 April 2003; pp. 45–49. [CrossRef]
29. Ferati, M.; Vogel, B.; Kurti, A.; Raufi, B.; Astals, D.S. Web accessibility for visually impaired people: Requirements and design issues. In *Usability-and Accessibility-Focused Requirements Engineering*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 79–96.
30. SoundEffectLab. Available online: https://soundeffect-lab.info (accessed on 17 March 2021).
31. Csikszentmihalyi, M. *Beyond Boredom and Anxiety*; Jossey-Bass: San Francisco, CA, USA, 2000.
32. Sweetser, P.; Wyeth, P. GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.* **2005**, *3*, 3. [CrossRef]
33. Yannakakis, G.N.; Hallam, J. Towards optimizing entertainment in computer games. *Appl. Artif. Intell.* **2007**, *21*, 933–971. [CrossRef]
34. Malone, T.W. Toward a theory of intrinsically motivating instruction. *Cogn. Sci.* **1981**, *5*, 333–369. [CrossRef]
35. Abuhamdeh, S.; Csikszentmihalyi, M. The Importance of Challenge for the Enjoyment of Intrinsically Motivated, Goal-Directed Activities. *Personal. Soc. Psychol. Bull.* **2012**, *38*, 317–330. [CrossRef]
36. Lessard, N.; Paré, M.; Lepore, F.; Lassonde, M. Early-blind human subjects localize sound sources better than sighted subjects. *Nature* **1998**, *395*, 278–280. [CrossRef] [PubMed]
37. Doucet, M.E.; Guillemot, J.P.; Lassonde, M.; Gagné, J.P.; Leclerc, C.; Lepore, F. Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Exp. Brain Res.* **2005**, *160*, 194–202. [CrossRef]
38. Gaver, W.W. The SonicFinder: An interface that uses auditory icons. *Hum. Comput. Interact.* **1989**, *4*, 67–94. [CrossRef]
39. Blattner, M.M.; Sumikawa, D.A.; Greenberg, R.M. Earcons and Icons: Their Structure and Common Design Principles. *Hum. Comput. Interact.* **1989**, *4*, 11–44. [CrossRef]
40. Brewster, S.A.; Wright, P.C.; Edwards, A.D. A detailed investigation into the effectiveness of earcons. In *Santa Fe Institute Studies in The Sciences Of Complexity-Proceedings Volume*; Addison-Wesley Publishing Co: Boston, MA, USA, 1994; Volume 18, p. 471.
41. Patterson, R. *Guidelines for Auditory Warning Systems on Civil Aircraft*; Civil Aviation Authority: London, UK, 1982.
42. Morikawa, D.; Hirahara, T. Signal frequency range necessary for horizontal sound localization. *Acoust. Sci. Technol.* **2010**, *31*, 417–419. [CrossRef]
43. Ohuchi, M.; Iwaya, Y.; Suzuki, Y.; Munekata, T. A comparative study of sound localization acuity of congenital blind and sighted people. *Acoust. Sci. Technol.* **2006**, *27*, 290–293. [CrossRef]
44. Voss, P.; Lassonde, M.; Gougoux, F.; Fortin, M.; Guillemot, J.P.; Lepore, F. Early-and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr. Biol.* **2004**, *14*, 1734–1738. [CrossRef]
45. Bonebright, T.; Nees, M. *Memory for Auditory Icons and Earcons with Localization Cues*; Georgia Institute of Technology: Atlanta, GA, USA, 2007.
46. Garzonis, S.; Jones, S.; Jay, T.; O'Neill, E. *Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference*; CHI '09; Association for Computing Machinery: New York, NY, USA, 2009; pp. 1513–1522. [CrossRef]
47. Leung, Y.; Smith, S.E.; Parker, S.; Martin, R. *Learning and Retention of Auditory Warnings*; Georgia Institute of Technology: Atlanta, GA, USA, 1997.
48. McCormick, E.J.; Sanders, M.S. *Human Factors in Engineering and Design*; McGraw-Hill Companies: New York, NY, USA, 1982.
49. Bussemakers, M.P.; De Haan, A. When it sounds like a duck and it looks like a dog... auditory icons vs. earcons in multimedia environments. In Proceedings of the International Conference on Auditory Display, Atlanta, GA, USA, 2–5 April 2000; pp. 184–189.
50. O'Connor, J. Concentration. Available online: https://boardgamegeek.com/boardgame/6798/concentration (accessed on 17 March 2021).
51. Denis, B.; Jacques Cottereau, G.G.N. Spot it! Available online: https://boardgamegeek.com/boardgame/63268/spot-it (accessed on 17 March 2021).
52. Weston, D. Cobra Paw. Available online: https://boardgamegeek.com/boardgame/220700/cobra-paw (accessed on 17 March 2021).

electronics

Article

3D Sound Coding Color for the Visually Impaired

Yong Lee [1], Chung-Heon Lee [2] and Jun Dong Cho [3,*]

[1] Department of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon 16419, Gyeonggi-do, Korea; leee@skku.edu
[2] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Gyeonggi-do, Korea; ch.lee@skku.edu
[3] Department of Human Information and Cognition Technology Convergence, Sungkyunkwan University, Suwon 16419, Gyeonggi-do, Korea
* Correspondence: jdcho@skku.edu; Tel.: +82-10-4332-7127

Abstract: Contemporary art is evolving beyond simply looking at works, and the development of various sensory technologies has had a great influence on culture and art. Accordingly, opportunities for the visually impaired to appreciate visual artworks through various senses such as auditory and tactile senses are expanding. However, insufficient sound expression and lack of portability make it less understandable and accessible. This paper attempts to convey a color and depth coding scheme to the visually impaired, based on alternative sensory modalities, such as hearing (by encoding the color and depth information with 3D sounds of audio description) and touch (to be used for interface-triggering information such as color and depth). The proposed color-coding scheme represents light, saturated, and dark colors for red, orange, yellow, yellow-green, green, blue-green, blue, and purple. The paper's proposed system can be used for both mobile platforms and 2.5D (relief) models.

Keywords: visual impairment; accessibility; aesthetics; color; multi-sensory; museum exhibits

Citation: Lee, Y.; Lee, C.-H.; Cho, J.D. 3D Sound Coding Color for the Visually Impaired. *Electronics* **2021**, *10*, 1037. https://doi.org/10.3390/electronics10091037

Academic Editor: Giovanni (Gianni) Ramponi

Received: 19 March 2021
Accepted: 22 April 2021
Published: 27 April 2021

1. Introduction

According to the 2020 data on visual impairment from the WHO, globally, the number of people of all ages visually impaired is estimated to be 285 million, of whom 39 million are blind [1]. People with visual impairments are interested in visiting museums and enjoying visual art [2]. Although many museums have improved the accessibility of their exhibitions and artworks through specialized tours and the access to tactile representations of artworks [3–5], it is still not enough to meet the needs of the visually impaired [6].

Multisensory (or multimodal) integration is an essential part of information processing by which various forms of sensory information, such as sight, hearing, touch, and proprioception (also called kinesthesia, the sense of self-movement and body position), is combined into a single experience [7]. The cross sensation between sight and other senses here refers to the representation of sight and other senses at the same time, but the aim of this paper is to use more than two other senses, such as touch and audio, besides the visual one to perform at the same time. Making art accessible to the visually impaired requires the ability to convey explicit and implicit visual images through non-visual forms. It argues that a multi-sensory system is needed to successfully convey artistic images. What art teachers wanted to do most with their blind students is to have them imagine colors using a variety of senses—audio, touch, scent, music, poetry, or literature.

In viewing artworks by the visually impaired, museums generally provide visually impaired people with audio explanatory guides that focus on the visual representation of the objects in paintings [8]. Brule et al. [9] created a raised-line overlaying multisensory interactive map on a capacitive projected touch screen for visually impaired children after a five-week field study in a specialized institute. Their map consisted of several multisensory tangibles that can be explored in a tactile way but can also be smelled or tasted, allowing users to interact with them using touch, taste, and smell together. A sliding gesture in the dedicated menu in Mapsense filters geographical information (e.g., cities, seas, etc.). Additionally, the Mapsense design used conductive tangibles that can be detected. Some tangibles can be filled with “scents”, such as olive puree, mashed raisins, and honey, which means that they use different methods (scent and taste) to promote reflexive learning and use objects to support storytelling. The Metropolitan Museum of Art in New York has displayed replicas of the artworks exhibited in the museum [10]. The Art Talking Tactile Exhibit Panel in the San Diego Museum allows visitors to touch Juan Sánchez Cotán’s master still-life, “Quince, Cabbage, Melon, and Cucumber”, painted in Toledo, Spain, in 1602 [11]. If the users touch one of these panels with bare hands or wearing light gloves, they can hear information about the touched part. This is like tapping on an iPad to make something happen; however, instead of a smooth, flat touch screen, these exhibit panels can include textures, bas-relief, raised lines, and other tactile surface treatments. Dobbelstein et al. [12] introduced inScent, a wearable olfactory display that allows users to receive notifications through scent in a mobile environment. Anagnostakis et al. [13] used proximity and touch sensors to provide voice guidance on museum exhibits through mobile devices. Reichinger et al. [14] introduced the concept of a gesture-controlled interactive audio guide for visual artworks that uses depth-sensing cameras to sense the location and gestures of the user’s hands during tactile exploration of a bas-relief artwork model. The guide provides location-dependent audio descriptions based on user hand positions and gestures. Recently, Cavazos et al. [15] provided an audio description as well as related sound effects when the user touched a 2.5D-printed model with their finger. Thus, the visually impaired could enjoy it freely, independently, and comfortably through touch to feel the artwork shapes and textures and to listen and explore the explanation of objects of their interest without the need for a professional curator.

The use of binaural techniques that have been used to express the direction of sound is rarely used to express colors in works of art for the visually impaired. However, the connection between color and spatial audio using binaural recordings [16] of audio when appreciating colors in artworks using binaural sound has not been addressed. When using spatial audio to artificially represent the color wheel, it is necessary to investigate whether it is confusing or has a positive effect on color perception. Binaural technology allows the augmentation of spatial positioning of sound with the usage of a simple pair of headphones. Binaural recording and rendering refer specifically to recording and reproducing sounds in two ears [16]. It is designed to resemble the human two-ear auditory system and normally works with headphones [17]. Lessard et al. [18] investigated how the three-dimensional spatial mapping is carried out by early blind individuals with or without residual vision. Subjects were tested under monaural and binaural listening conditions. They found that early blind subjects could map their auditory environment with equal or better accuracy than sighted subjects. In [19], 3D-Sound was useful for visually impaired people; they felt significantly higher confidence in 3D-Sound.

This paper proposes a tool to intuitively recognize and understand the three elements of color: hue, value, and saturation using spatial audio. In addition, when touching objects in artwork with a finger, the description of the work is provided by voice, and the color, brightness, and depth of the object are expressed through the modulation of the voice.

2. Background and Related Works

2.1. Review of Tactile and Sound Coding Color

In order to convey color to visually impaired people, a method of coding color with tactile patterns or sounds has been proposed [20–23]. Taras et al. [20] presented a color code created for viewing on braille devices. The primary colors, red, blue, and yellow, are each coded by two dots. Mixed colors, for example, violet, green, orange, and brown, are coded as combinations of dots representing the primary colors. Additionally, the light and dark shades are added by using the second and third dots in the left column of the Braille cell.

Ramsamy-Iranah et al. [21] designed color symbols for children. The design process for the symbols was influenced by the children's prior knowledge of shapes and linked to their surroundings. For example, a small square box was associated with dark blue, reflecting the blue square soap, a circle represented red because it was associated with the red "dot" called "bindi" on the forehead of a Hindu woman. Yellow was represented by small dots reflecting the pollen of flowers. Orange is a mixture of yellow and red; therefore, circles of smaller dimensions were used to represent orange. Horizontal lines represented purple, and curved lines were associated with the green representative of bendable grass stems.

Shin et al. [22] coded nine colors (pink, red, orange, yellow, green, blue, navy, purple, brown, and achromatic) using a grating orientation (a regularly spaced collection of identical, parallel, elongated elements). The texture stimuli for color were structured by matching variations of orientation to hue, the width of the line to chroma, and the interval between the lines to value. The eight chromatic colors were divided into 20° angles and were achromatic at 90°. Each color had nine levels of value and of chroma.

Cho et al. [23] developed a tactile color pictogram that used the shape of the sky, earth, and people derived from thoughts of heaven, earth, and people as metaphors. Colors could thus be recognized easily and intuitively by touching the different patterns. An experiment comparing the cognitive capacity for color codes found that users could intuitively recognize 24 chromatic and 5 achromatic colors with tactile codes [23].

Besides tactile patterns, sound patterns [24–27] use classical music sounds played on different instruments. Cho et al. [27] considered the tone, intensity, and pitch of melody sound extracted from classic music to express the brightness and saturation of colors. The sound code system represented 18 chromatic and 5 achromatic colors using classical music sounds played on different instruments. While using sound to depict color, tapping a relief-shaped embossed outline area transformed the color of that area into the sound of an orchestra instrument. Furthermore, the overall color composition of Van Gogh's "The Starry Night" was expressed as a single piece of music that accounted for color using the tone, key, tempo, and pitch of the instruments. The shape could be distinguished by touching it with a hand, but the overall color composition could be conveyed as a single piece of music, thereby reducing the effort required to recognize color from needing to touch each pattern one by one [27].

Jabber et al. [28] developed an interface that automatically translated reference colors into spatial tactile patterns. A range of achromatic colors and six prominent basic colors were represented with three levels of chroma and values through a color watch design. The color was represented through combination discs that represented the color hue, and square discs that represented lightness, and were perceived by touch.

This paper introduces two sound color codes, a six-color wheel and an eight-color wheel, created with 3D sound, based on the aforementioned observations. Table 1 shows a comparison between the previous color codes and the two sound color codes proposed in this paper.

2.2. Review of HRTF Systems

The Head-Related Transfer Function (HRTF) is a filter defined on a spherical area that describes how the shape of the listener's head, torso, and ears affects incoming sound from all directions [29]. When sound hits the listener, the size and shape of the head, ears and ear canal, the density of the head, and the size and shape of the nasal and oral cavity all alter the sound and affect the way the sound is perceived, raising some frequencies and attenuating others. Therefore, the time difference between the two ears, the level difference between the two ears, and the interaction between sound and personal body anatomy are important for HRTF calculation. In this way, the ordinary audio is converted to 3D sound. Although binaural synthesis with HRTFs has been implemented in real-time applications, only a few commercialized applications utilize it. Limited research exists on the differences between audio systems that use HRTF, compared to systems that do not [30]. Systems that do not use HRTF in their binaural synthesis instead often use a simplified interaural intensity difference (IID) [30]. This simplified IID alters the amplitude equally for all frequencies, relative to orientation and distance from the audio source to both ears of the listener. These systems do not utilize any audio cues for vertical placement and will therefore be referred to as "panning systems", while systems that use HRTF do have cues for vertical placement, and will therefore be referred to as "3D audio systems". Three-dimensional audio systems will show a difference in human localization performance compared to a panning system, because these systems utilize more precise spatial audio cues than panning systems. These results suggest that 3D audio systems are better than panning systems in terms of precision, speed, and navigation, in an audio-exclusive virtual environment [31]. Additionally, the non-individualized HRTF filters currently in use may lack the published accuracy [32], but a better-personalized HRTF will increase the accuracy. Most of the virtual auditory displays employ generic or non-individualized HRTF filters that lead to a decreased sound localization accuracy [33].

Table 1. Existing color codes with instruments and the color codes in this paper.

Developer (Sense Used)	Basic Patterns (Concepts)	# of Colors Presented
Taras et al. 2009 [20] (Touch)	Dots (Braille)	23 (6 Hues + 2 levels of lightness for each hue + 5 levels of achromatic colors)
Ramsamy-Iranah et al. 2016 [21] (Touch)	Polygons (Children's Knowledge)	14 (6 Hues + 5 other colors + 3 levels of achromatic colors)
Shin et al. 2019 [22] (Touch)	Lines (Orientation, Grating) The first eight colors are expressed in eight different angles of directionality by dividing a rainbow-shaped semicircle at intervals of 20 degrees	90 (8 hues + 4 levels of lightness, and 5 levels of saturation for each hue + 10 levels of brown and achromatic colors)
Cho et al. 2020 [23] (Touch)	Dots, lines, and curves (pictograms)	Simplified: 29 (6 hues + 2 levels of lightness, and 2 levels of saturation for each hue + 5 levels of achromatic colors) Extended: 53 (12 hues +2 levels of lightness, and 2 levels of saturation for each hue + 5 levels of achromatic colors)
Cho et al. 2020 [27] (Hear)	Classical music melodies played on different instruments	23 (6 Hues + 2 levels of lightness for each hue + 5 levels of achromatic colors)
Jabber et al. [28] (Touch)	Embossed surface pattern by color wheel	Simplified: 24 (6 hues + 3 levels of lightness for each hue + 6 levels of achromatic colors) Extended: 32 (8 hues + 3 levels of lightness for each hue + 8 levels of achromatic colors)
This paper: 6-color wheel (Hear)	Spatial sound representation using binaural recoding in virtual environment	21 (6 hues + 3 levels of lightness for each hue + 3 levels of achromatic colors)
This paper: 8-color wheel (Hear)		27 (8 hues + 3 levels of lightness for each hue + 3 levels of achromatic colors)

Use cases of individualized HRTFs can be found for hearing aids [34], dereverberation [35], stereo recording enhancements [36], emotion recognition [37], 3D detection assisting blind people to avoid obstacles [38], etc.

In [18,19,38], spatial sound was proven useful for visually impaired people, and they felt significantly higher confidence with spatial sound. This paper reveals through experiments that spatial sound expressing colors through HRTF is an effective way to convey color information. The paper's spatial sound strategy is based on cognitive training and sensory adaptation to spatial sounds synthesized with a non-individualized HRTF. To the best of our knowledge, no HRTF has been applied to represent color wheels.

Drossos et al. [39] used binaural technology to provide accessible games for blind children. In the game of Tic-Tac-Toe, they used binaural processing of selected audio material performed by the utilization of a KEMAR HRTF library [40], and through three kinds of sound presentation methods to carry out the information transmission and feedback in the game. The first method was to use eight different azimuths in the 0° elevation plane to represent the Tic-Tac-Toe chessboard shown in Figure 1. The second method was to use a combination of three elevations and three azimuths to simulate a Tic-Tac-Toe chessboard standing upright in front of the user. The third method was the same as the second method, but used pitch instead of elevation.

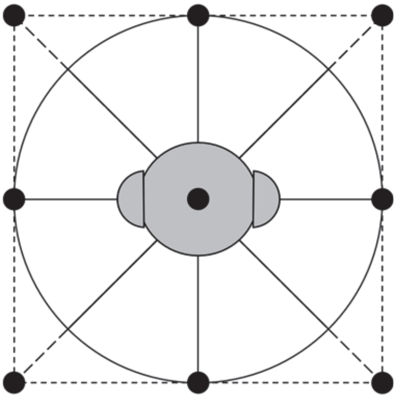


Figure 1. Illustration of a sound spatial positioning from Drossos et al. [39].

2.3. Review of the Sound Representations of Colors

Newton's *Opticks* [41] showed that the colors of the spectrum and the pitches of musical scales are similar (for example, "red" and "C"; "green" and "Ab"). Maryon [42] also explored the similarity between the ratio of each tone to the wavelength of each color to connect them. This method of associating the pitch frequency of the scale with color can be a way of substituting colors and notes for one another [43]. However, the various sensibilities that can be obtained through color are limited by simply substituting colors into the musical scale. Lavigna [44] suggested that the technique of a composer in organizing an orchestra seems very similar to the technique of a painter applying colors. In other words, a musician's palette is a list of orchestral instruments.

A comprehensive survey of associations between color and sound can be found in [45], including how different color properties such as value and hue are mapped onto acoustic properties such as pitch and loudness. Using an implicit associations test, those researchers [45] confirmed the following cross-modal correspondences between visual and acoustic features. Pitch was associated with color lightness, whereas loudness mapped onto greater visual saliency. The associations between vowels and colors are mediated by differences in the overall balance of low- and high-frequency energy in the spectrum rather than by vowel identity as such. The hue of colors with the same luminance and saturation was not associated with any of the tested acoustic features, except for a weak preference to match higher pitch with blue (vs. yellow). In other research, high loudness was associated with orange/yellow rather than blue, and the high pitch was associated with yellow rather than blue [46].

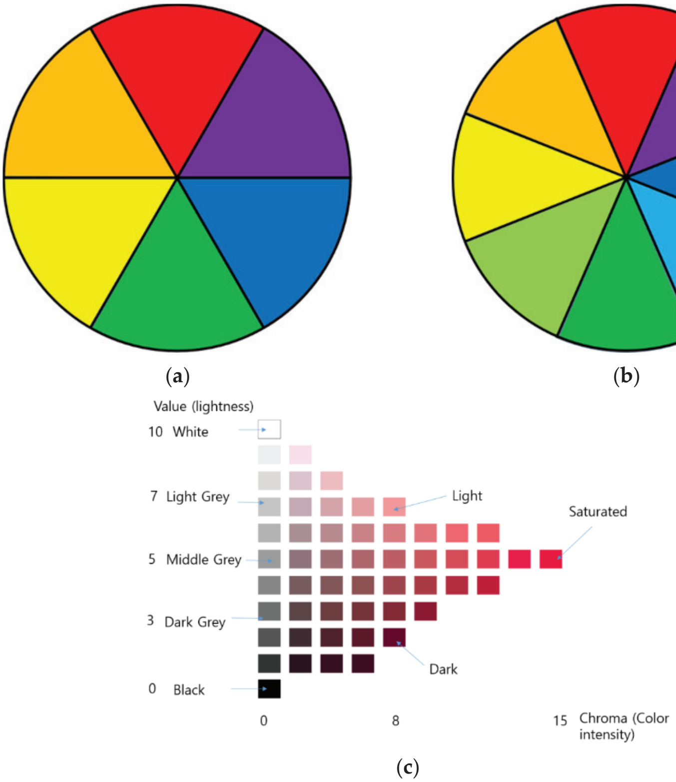
Chroma has a relationship with sound intensity [46,47]. When the intensity of a sound is strong and loud, its color is close, intense, and deep. However, when the sound intensity is weak, the color feels pale, faint, and far away. A higher value is associated with higher pitch [48,49]. Children of all ages and adults matched pitch to value and loudness to chroma. The value (i.e., lightness) is high and heavily dependent on the light and dark levels of the color. Using the same concept in music, sound is divided into light and heavy feelings according to the high and low octaves of a scale. Another way to match color and sound is to associate an instrument's tone with color, as in Kandinsky [24]. A low-pitched cello has a low-brightness dark blue color, a violin or trumpet-like instrument with a sharp tone feels red or yellow, and a high-pitched flute feels like a bright and saturated sky blue.

3. Binaural Audio Coding Colors with Spatial Color Wheel

3.1. Spatial Sound Representations of Colors

The purpose of this study is to convey the concept of the spatial dimension of the color wheel. In other words, a timepiece watch makes it easy to familiarize oneself with the concept of relative time, and helps the reader understand the adjacency and complementarity of time. Similarly, this paper uses this concept for color presentation. In particular, for secondary colors such as orange, green, and purple, the basic concept of how the primary and secondary colors are created can be expressed simultaneously through the color wheel.

Figure 2 illustrates the RYB color wheel that was created by Johannes Itten [50]. There are two simplified color wheels that we want to express using 3D sound. One is a 6-color wheel composed of three primary colors (red, yellow, blue) and three secondary colors (orange, green, purple) as shown in Figure 2a, and the other as shown in Figure 2b is an 8-color wheel consisting of 8 colors (red, orange, yellow, yellow-green, green, blue-green, blue, purple). In addition, for each color (hue), three color tones (light, saturated, dark) as shown in Figure 2c are expressed in 3D sound. In addition, three achromatic colors of white, black and gray are expressed in 3D sound.

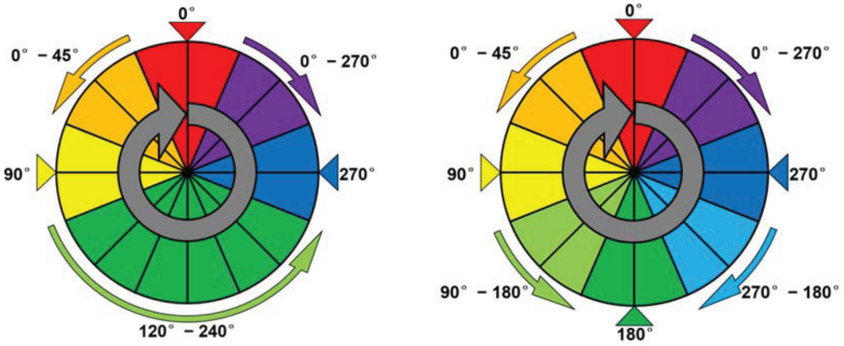


(**a**) (**b**)

(**c**)

Figure 2. (**a**) 6-color wheel; (**b**) 8-color wheel; (**c**) saturated (S), light (L), and dark (D) for red.

For easy identification of the color code, HRTF is used for the color representation with different fixed azimuth angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°) to represent each color. However, the difference in the effect of the same HRTF for each person makes it possible to confuse 45°, 135°, 225°, and 315° with the adjacent angles. This effect is not ideal. Therefore, the primary colors are represented by a fixed 3D sound, and the secondary colors are represented by a moving 3D sound to make it easier to recognize how the two primary colors are mixed. The color representation of the six-color wheel codes is shown in Figure 3a and Table 2, and the eight-color wheel codes are shown in Figure 3b and Table 3. The six-color wheel codes are not represented like the six-color wheel in Figure 2a, because the fixed azimuth angles of 120° and 240° are relatively vague and not as accurate as 90° and 270°. Thus, yellow and blue are represented by 90° and 270°, and the range of green is relatively expanded.



(**a**) (**b**)

Figure 3. (**a**) Sound representations of 6-color wheel codes; (**b**) sound representations of 8-color wheel codes.

Table 2. Sound representations of 6-color wheel codes.

Azimuth/Pitch	−3	0	3
0°	1. Dark red	2. Saturated red	3. Light red
0°–90°	4. Dark orange	5. Saturated orange	6. Light orange
90°	7. Dark yellow	8. Saturated yellow	9. Light yellow
120°–240°	10. Dark green	11. Saturated green	12. Light green
270°	13. Dark blue	14. Saturated blue	15. Light blue
360°–270°	16. Dark violet	17. Saturated violet	18. Light violet
360°–0°	19. Black	20. Gray	21. White

The sound files developed in this research are provided separately as a Supplementary Materials.

Table 3. Sound representations of 8-color wheel codes.

Azimuth/Pitch	−3	0	3
0°	1. Dark red	2. Saturated red	3. Light red
0°–90°	4. Dark orange	5. Saturated orange	6. Light orange
90°	7. Dark yellow	8. Saturated yellow	9. Light yellow
90°–180°	10. Dark yellow-green	11. Saturated yellow-green	12. Light yellow-green
180°	13. Dark green	14. Saturated green	15. Light green
270°–180°	16. Dark blue-green	17. Saturated blue-green	18. Light blue-green
270°	19. Dark blue	20. Saturated blue	21. Light blue
360°–270°	22. Dark violet	23. Saturated violet	24. Light violet
360°–0°	25. Black	26. Gray	27. White

The sound files developed in this research are provided separately as a Supplementary Materials.

Eight colors have three levels of brightness, expressed by changing the pitch of the sound. A normal audio sound represents saturated colors, an audio sound that raises three semitones represents a lighter color, and an audio sound that decreases three semitones represents a darker color. In this way, this paper proposes a color-coding system that can represent 24 chromatic colors and three achromatic colors. The strategy complies with the definition of light and dark colors in the Munsell color system, as shown in Figure 2c. The reason for raising or lowering the three semitones is that the three semitones have little effect on the pitch characteristics of the original sound. For achromatic colors, gray is represented by 3D sound from 360° to 0°. The black is the gray sound decreasing three chromatic scales, and white is the gray sound raising three chromatic scales.

There are many types of HRTF databases, such as the CIPIC HRTF-database [51], Listen HRTF-database [52], MIT HRTF-database [53], etc. This paper used the ITA HRTF-database [54,55] to change the audio direction by MATLAB. Additionally, Adobe Audition was used to change the sound of the pitch.

3.2. Sound Representations of Depth

In order to find the most suitable sound variables to express depth, the paper tested them experimentally and applied them to the sound code.

3.2.1. Matching Test

Sound Stimuli

According to the abstract sound variables [56], Table 4 indicates the changes in various related information such as language, direction, music, pitch, speed, size, depth, and special effects. This study used sound variables such as loudness, pitch, velocity, length, attack/decay for the matching with depth.

Table 4. The abstract sound variables [56].

Sound Variables	Introduction
Location	The location of a sound in a two- or three-dimensional space.
Loudness	The magnitude of a sound.
Pitch	The highness or lowness (frequency) of a sound.
Register	The relative location of a pitch in a given range of pitches.
Timbre	The general prevailing quality or characteristic of a sound.
Duration	The length of time a sound is (or is not) heard.
Rate of change	The relationship between the duration of sound and silence over time.
Order	The sequence of sounds over time.
Attack/Decay	The time it takes a sound to reach its maximum/minimum.

Semantic Stimuli

The purpose of this experiment is to find the most suitable sound variables to express depth. To obtain the association of sound variables and depth, this paper used the explicit association + implicit association test. That is, the explicit association is used first for match detection, and if no match can be made, the implicit association test is performed to match implicitly with other adjective pairs. Osgood [57] simplified the semantic space of the relative adjectives into three aspects, which are (1) evaluation (like–dislike), (2) potency (strong–weak), and (3) activity (fast–slow). The adjectives adopted in this research are pairs of adjectives with which people are familiar, such as emotion, shape, location, activity, texture, contrast, temperature, sound characteristics, etc. Thus, the simplified concept pairs of adjectives are chosen per aspect, shown in Table 5. Note that 11 pairs among them are related to sound attributes, as shown in Table 6.

This paper used sound variables such as loudness (Small~Loud), pitch (Low~High), velocity (Fast~Slow), length (Short~Long), and attack/decay (Decay~Attack) for this test. For each sound variable, participants received several audio segments with different levels of variability. Participants in the experiment used this audio file to recognize sound variables and evaluate how well those sound variables matched adjectives. In each of these 11 pairs of concepts, the score for the feeling conveyed by the sound attribute stimulus is 2 points when chosen as most positively consistent with the feeling of depth, −2 points when chosen as most negatively consistent with the feeling of depth, and 0 when chosen as least consistent with the feeling of depth. These score points were computed for each subject for each of the 11 sound-attribute stimuli.

Table 5. Modified concept pairs of adjectives originally from Osgood [57]

Evaluation	Potency	Activity
Bright~Dark	Strong~Weak	Fast/Agile~Slow/Dull
Clear~Cloudy	Hard~Soft	Noisy~Quiet
Joyful~Depressed	Rough~Smooth	Extroverted~Introverted
Calm~Tense	Pointed (Kiki)~Round (Bouba) Sharp~Dull	Centrifugal~Centripetal Dilated~Constricted
Comfortable~Anxious	Far~Near	Passionate~Depressed
Warm~Cool	High~Low (e.g., high-pitch~low-pitch)	Active~Inactive

Table 6. The adjective pairs used in the experiments.

Number	Sound Attributes
1	Fast/Agile~Slow/Dull
2	Strong~Weak
3	Warm~Cool
4	Tense~Calm
5	Active~Inactive
6	Noisy~Quiet
7	Clear~Cloudy
8	Pointed (Kiki)~Round (Bouba)Sharp~Dull
9	Dilated~Constricted(Centripetal~Centrifugal)
10	High~Low (e.g., high-pitch~low-pitch)
11	Near~Far

Experiment Participants and Results

Seven members of Sungkyunkwan University were recruited as experiment participants. The gender split of the participants was 4 men and 3 women, and the average age was 22.29 years old (minimum 21 years old, maximum 24 years old). When participating in the experiment, side effects such as headaches could occur due to repeated auditory stimulation, and if they felt physical or mental discomfort; the experiment was conducted only after notifying the participants in advance that they could request to stop the experiment at any time.

Test results are shown in Table 7 and Figure 4. For each of the 11 pairs of adjective concepts in Table 7, the scores for the sense of depth transmitted by the sound stimulus are between −1 and 1. In other words, the absolute value of 1 is given when the sound stimulus feels the most consistent with the sense of depth, and 0 points are given when the sound stimulus is the most inconsistent with the sense of depth. By matching the results of sound variables with adjective pairs and matching results of sound variables with depth, this paper can conclude that there is a strong correlation between sound intensity and depth. That is, when the sound is loud it is associated with proximity, while when the sound intensity is small it is associated with depth.

3.2.2. Sound Representations of Color and Depth

With the results of the previous experiments, this study used the sound size variation to represent the sense of depth. To deepen the sense of depth, the paper added a reverberation effect while changing the sound size to make the sound depth more obvious. To make it easier to recognize the depth information expressed in velocity, only 3 distance levels (far, mid, and near) were used. The near level was set to the normal sound speed. The mid-level was set to 80% dry, 50% reverberation, and 10% early. The far level was set to 30% dry, 15% reverberation, and 10% early. The reverb setting was 1415 ms decay time, 57 ms pre-decay time, 880 ms diffusion, 22 perception, 1375 m^3 room size, 1.56 dimensions, 13.6% left/right location, and 80 Hz high pass cutoff.

Table 7. Matching results of sound variables with adjective pairs.

	Loudness (Small Sound~Loud Sound)	**Pitch (Low Sound~High Sound)**	**Velocity (Fast Sound~Slow Sound)**	**Length (Short Sound~Long Sound)**	**Attack/Decay (Decay~Attack)**
Fast/Agile~Slow/Dull	−0.29	−0.71	1.43	0.29	0
Strong~Weak	−1.71	−0.14	0.43	0	−0.71
Warm-Cool	−0.14	−0.14	−0.71	−0.57	−0.14
Tense~Calm	−0.57	−0.57	1	1.29	0
Active~Inactive	−0.86	−1.14	1.14	0.14	−1
Noisy~Quiet	−1.14	−0.29	0.57	0.14	−0.57
Clear~ Cloudy	0	−0.57	0.29	0.14	−0.71
Pointed (Kiki)~Round (Bouba) Sharp~Dull	0	0.43	−0.43	−0.57	−0.29
Dilated~Constricted (Centripetal~Centrifugal)	0	0.14	0.71	0.86	−0.57
High~Low (e.g., high-pitch~low-pitch)	−0.57	−1.43	0.14	−0.14	−0.71
Near~Far	−1.71	0	0.43	0.57	−1.14

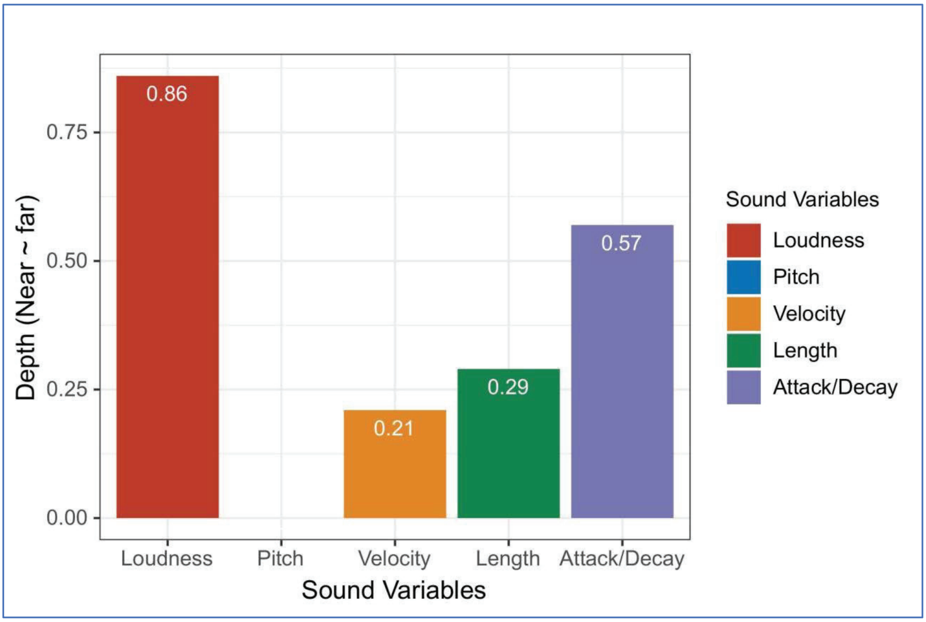


Figure 4. Matching results of sound variables with depth (Near~Far).

3.3. Prototyping Process

We have created an Android mobile application as a tool to deliver the proposed sound code to users. Figure 5 shows the prototyping process for creating a mobile application used as a tool for expressing color using the proposed sound code.

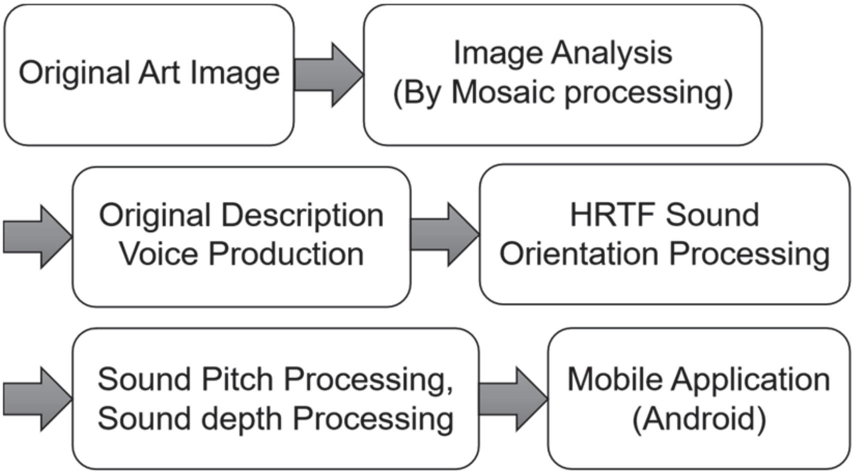
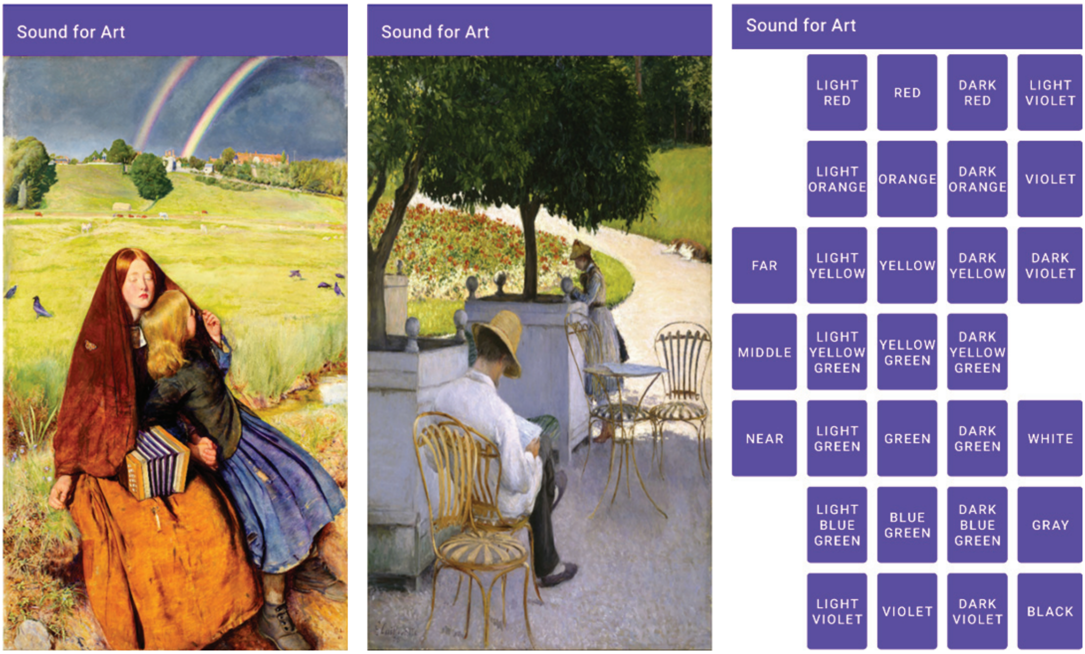


Figure 5. Prototyping process.

The first step was to analyze the images of the whole work. The specific method was to use software such as Photoshop to divide the artwork into specific grids (e.g., 15 × 19 or 12 × 15) and analyze the name, color, and depth of each object along with the artwork introduction. The analysis selected a specific name, color and brightness level, and depth level. The second step was to create an audio file corresponding to the spoken instructions with names for all analyzed objects. The third step was to apply HRTF to each part of the audio file corresponding to the object name to represent the object's color in 3D sound. The fourth step was to use Adobe Audition audio processing software to perform pitch scaling without time scaling processing and reverberation processing on each part's voice-described audio file through the lightness of the color and depth levels. The fifth step was to create a mobile application using Android Studio software. The basic making method was to split artworks as buttons in the way described above and add processed audio files to each part. The artworks used in this prototype as examples were John Everett Millais' "The Blind Girl" and Gustave Caillebotte's "The Orange Trees." The prototype application interface is shown in Figure 6. Figure 6a,b shows where the user could apply the 3D sound coding to the artwork for viewing. By clicking on any part of the artwork, the user could access the audio description of the clicked area. Additionally, each voice description used sound coding in this paper. It was possible to obtain information about color, brightness, and depth while receiving the voice description. Figure 6c shows the listening test of the application. The user could perform headphone tests and sound learning in this interface.



(**a**) (**b**) (**c**)

Figure 6. Application interface. (**a**) The Blind Girl interface; (**b**) The Orange Trees interface; (**c**) test interface.

4. User Test and Results

4.1. Participants

Ten students were recruited as participants of the experiment. The gender split of the participants was five males and five females, and the average age was 22.5 years (minimum 20 years old, maximum 25 years old). While participating in this experiment, repeated auditory stimulation may have caused side effects such as headaches, and if physical or mental discomfort was felt, participants were informed in advance that they could request to stop the experiment at any time. All participants used their own cell phones and earphones for the experiment. Five participants used the six-color wheel codes, and the other five participants used the eight-color wheel codes.

The experimental evaluation was performed in three stages: learning phase, tests, and feedback. During the learning phase, experiment participants learned and became familiarized with the sound codes through explanations, schematic, and sample audio. Test part divided into color, color + lightness, color + lightness + depth to be tested separately. In the test part, the participants were asked to perceive color, lightness, and depth through sound alone without looking at pictures such as the color wheel. After that, the participants evaluated the workload assessment and usability test.

4.2. Identification Tests

In experiment 1, experiment participants performed color identification on random sound samples that only transformed the color variable. As shown in Table 8, the color identification rate of Group A using six-color wheel codes was 100%. Additionally, for Group B with eight-color wheel codes used, the color identification rate was 86.67%.

Table 8. Six-color wheel and eight-color wheel code identification test in experiment 1 (color).

Colors Sound	Color Dimensions Left: 6-Color Wheel; Right: 8-Color Wheel)															
	Red		Orange		Yellow		Yellow-Green	Green		Blue-Green	Blue		Violet		Gray	
Red	5	4							1							
Orange			5	4						1						
Yellow					5	5										
Yellow-green						1	4									
Green		1						5	4							
Blue-green										5						
Blue											5	5				
Violet													5	5		
Gray							1								5	4
Average correct answers (%)	100	80	100	80	100	100	80	100	80	100	100	100	100	100	100	80
Total (%)					100							86.67				

In experiment 2, experiment participants performed color and lightness identification on random sound samples that transformed the color variable and the lightness variable. As shown in Table 9, the color discrimination rate and brightness discrimination rate of both groups A and B were 100%.

Table 9. total color codes identification test in experiment 2 (color + lightness).

Color + Lightness Colors Sound	Color Dimensions—Color			Color Dimensions—Lightness		
	Red	Yellow	Blue	Dark	Saturated	Light
Red—Dark	10			10		
Red—Saturated	10				10	
Red—Light	10					10
Yellow—Dark		10		10		
Yellow—Saturated		10			10	
Yellow—Light		10				10
Blue—Dark			10	10		
Blue—Saturated			10		10	
Blue—Light			10			10
Average correct answers (%)	100	100	100	100	100	100
Total (%)		100			100	

In experiment 3, participants performed color, brightness, and depth identification on random sound samples representing color, brightness, and depth variables. As shown in Table 10, there is confusion between red and blue in a multivariate situation. It is possible that the sound on the right side of the HRTF sample is a bit louder than the sound on the left side, which makes the right side similar to the front sound in the case of reverberation. Additionally, in the multivariate case, the depth variable may show a small recognition error.

When we analyzed the identification test results shown in Table 11, we found that the identification rate of S3 participants was significantly lower than that of other participants. This may be due to the headset brought by the individual participant. Excluding the S3 participants, the discrimination rate results were much better.

Table 10. Total color codes identification test in experiment 3 (color + lightness + depth).

Color + Lightness + Depth Colors Sound	**Color Dimensions—Color**			**Color Dimensions—Lightness**			**Color Dimensions—Depth**		
	Red	**Yellow**	**Blue**	**Dark**	**Saturated**	**Light**	**Near**	**Mid**	**Far**
Red—Dark—Near	9		1	10			9	1	
Red—Dark—Mid	8		2	10				9	1
Red—Dark—Far	8		2	10					10
Red—Saturated—Near	10				10		10		
Red—Saturated—Mid	10				10			10	
Red—Saturated—Far	9		1		10				10
Red—Light—Near	10			1		9	10		
Red—Light—Mid	9		1			10		10	
Red—Light—Far	9		1			10			10
Yellow—Dark—Near		10		10			10		
Yellow—Dark—Mid		10		10			1	8	1
Yellow—Dark—Far	1	9		10					10
Yellow—Saturated—Near		10			10		9	1	
Yellow—Saturated—Mid		10			10			10	
Yellow—Saturated—Far		10			10				10
Yellow—Light—Near		10				10	10		
Yellow—Light—Mid	1	9				10		10	
Yellow—Light—Far		10				10	1	1	8
Blue—Dark—Near			10	10			10		
Blue—Dark—Mid			10	10			1	8	1
Blue—Dark—Far			10	10					10
Blue—Saturated—Near			10		10		10		
Blue—Saturated—Mid			10		10			10	
Blue—Saturated—Far			10		10				10
Blue—Light—Near			10			10	9	1	
Blue—Light—Mid			10			10		10	
Blue—Light—Far	2		8	1		9		1	9
Average correct answers (%)	91.11	97.78	97.78	100	100	97.78	96.67	94.44	96.67
Total (%)		95.56			99.26			95.93	

Table 11. Identification test results for each participant.

Total Tests		**6-Color Wheel (43 Tests)**					**8-Color Wheel (45 Tests)**				
		S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
Color	Correct answer	40	43	43	37	42	45	45	45	41	42
	Rate (%)	93.02	100	100	86.05	97.67	100	100	100	95.35	97.67
Lightness	Correct answer	36	36	36	34	36	36	36	36	36	36
	Rate (%)	100	100	100	95.35	100	100	100	100	100	100
depth	Correct answer	23	27	27	20	27	27	27	27	27	27
	Rate (%)	90.70	100	100	83.72	100	100	100	100	100	100
Rate (%)		94.57	100	100	88.37	99.22	100	100	100	98.45	99.22
Total Rate (%)				96.43					99.53		
						97.98					

4.3. Workload Assessment

The Official NASA Task Load Index (TLX) is a subjective workload assessment tool that is used in various human–machine interface systems [58]. By incorporating a multi-dimensional rating procedure, NASA TLX derives an overall workload score based on a weighted average of ratings on six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The scale from 0 to 10 points is chosen for the ease and familiarity of participants, with 0 ranging from very low to 10 being very high. TheTLX test was performed using uniform weights for all metrics. The six-color wheel codes achieved 43.75 points, and the eight-color wheel codes achieved 48.75 points.

Figure 7 summarizes workload assessment scores for subjects under the NASA-TLX test. The scores for Mental Demand, Temporal Demand, Overall Performance, and Effort were in the middle or upper-middle. This was because adding three variables to speech sound made it relatively more difficult to use while increasing efficiency. More time needs to be invested in practice and training based on understanding the principles. This also makes the task more demanding for first-time participants and can feel relatively difficult with insufficient learning, which can increase frustration in use. Gradual learning over time makes it less difficult. The reduction in variables and improvements in sound production methods will further reduce the difficulty of use.

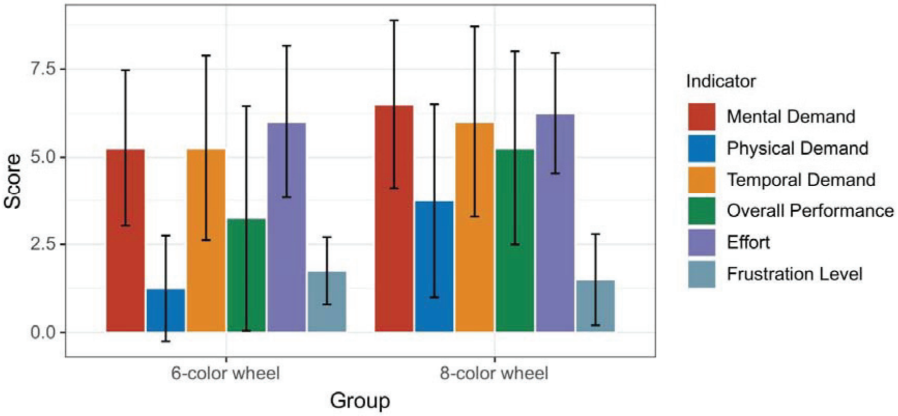


Figure 7. Workload assessment scores for NASA-TLX test question indicators.

4.4. User Experience Test

User experience (UX) testing of participants in the experiment was performed by modifying the System Usability Scale approach to match the purpose of the experiment. The System Usability Scale (SUS) is a questionnaire that is used to evaluate the usability of products and services. These survey questions are used as a quantitative method to evaluate and gain actionable insights on the usability of a wide variety of new systems which may be either software or hardware [59].

Participants were asked to rate the following seven items:

(1) I think that I would like to use this system frequently;
(2) I found the complexity in this system appropriate;
(3) I thought the system was easy to use;
(4) I found that the various functions in this system were well integrated;
(5) I thought that there was consistency in this system;
(6) I would imagine that most people would learn to use this system very quickly;
(7) I think this system was light to use.

The UX test scores were broken down for all participants out of 1–5 (strongly disagree through strongly agree). By converting on a hundred-point scale, the six-color wheel code score was 72.32 points, and the eight-color wheel code score was 71.43 points. The scored user experience results are provided in Figure 8. The average Q1 score was 2.5, Q2 was 3, Q3 was 2.63, Q4 was 3.25, Q5 was 3.5, Q6 was 2.75, and Q7 was 2.5. As with the questions discussed in the previous NASA-TLX section, the lack of time and unfamiliarity with the use of the program resulted in relatively low ratings for individual questions.

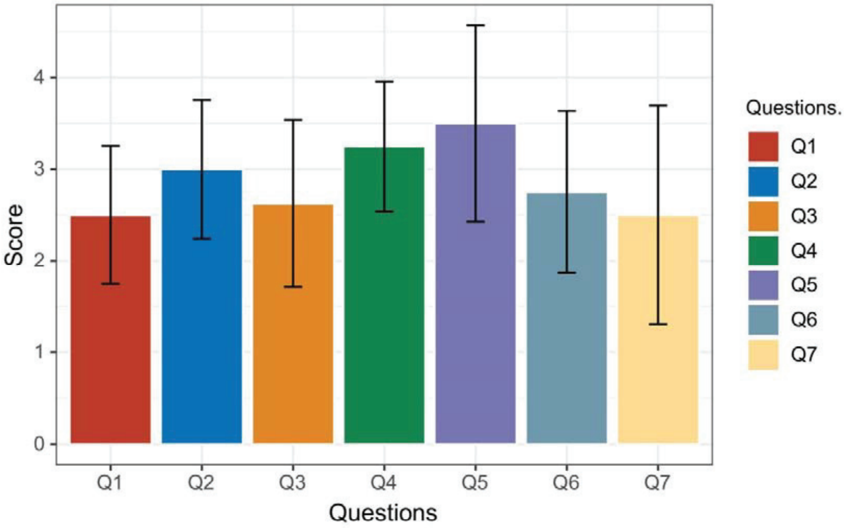


Figure 8. Scores obtained for UX test questions.

Table 12 lists the participants' positive and negative feedback.

Table 12. Positive and negative user feedback from the UX Test.

Positive User Feedback	Negative User Feedback
I do not think it's too complicated. Once you get used to it, it's easy to use.	It takes a while to get used to it at first and requires frequent viewing of the photos.
The distinction between color, brightness, and depth is very clear.	In some cases, sound confusion can occur.
It's very easy to use with just a good headset.	The sounds used in the experiment were too monotonous. The experience should be better with the prototype.
Expressing all three characteristics at the same time allows you to convey information efficiently.	For congenitally visually impaired people, there is a lack of experience with color. Therefore, for them, this method may not make much sense.
It's interesting to feel the depth with the sound.	There is no difficulty in distinguishing, but it was a little difficult to distinguish when hearing fatigue occurred.

5. Discussion

In experiment 1, the color identification rate of Group A with six-color wheel codes used was 100%. Additionally, for Group B with eight-color wheel codes used, the color identification rate was 86.67%. In experiment 2, the color identification and the lightness rate of Group A and Group B was 100%. Additionally, in experiment 3, the color, lightness, and depth identification of participants were 95.56%, 99.26%, 95.93%. The overall recognition rate was very high and still performed well with multiple sound variables. However, because the number of colors in the eight-color wheel was more than that of the six-color wheel, the recognition rate was 13.33% lower than that of the six-color wheel. If the distinction between confused colors is strengthened, the recognition rate will be better.

In the workload assessment test, the six-color wheel codes scored 43.75 points, and the eight-color wheel codes scored 48.75 points; the total score was 46.25 points. The lower the rating, the less the load on the user. With a full score of 100, the overall score tended to be in the middle, i.e., the user load was medium. In the user experience test, the six-color wheel codes scored 72.32 points and the eight-color wheel codes scored 71.43 points. The higher the score, the better the user's sense of use. With a score of 100 out of 100, the overall score was good. As described in the Results section, the experiment may feel relatively loaded and difficult to use due to insufficient learning and familiarity with the design. Additionally, due to the excessive number of variables used in the experiment, there may be a degree of fatigue for the participants. Therefore, a better HRTF matching seems necessary, which will make the effect more visible, and the participants could more clearly distinguish the colors. Additionally, audio optimization to make the audio more accurate and friendly is also a method, while simplifying the design will also optimize the user's perception. Table 13 shows conflicting user feedback and future work to resolve the conflict.

Table 13. Critical user feedback and future works.

Conflicted User Feedbacks	Conflict Resolution (Future Works)
It takes a while to get used to it at first and requires frequent viewing of the photos.	The unfamiliarity of first-time use may take some time for the user to adapt. Therefore, it is necessary to provide a concise learning tutorial along with the mobile app.
In some cases, sound confusion can occur.	It is possible that the sound on the right side of the HRTF sample is a bit louder than the sound on the left side, which makes the right side similar to the front sound in the case of reverberation. Early users cannot rule out the possibility that the color is difficult to recognize when adding a depth variable to the voice modulation. For this reason, firstly, the ratio and setting of the volume and reverberation variables in the depth variables will be adjusted so that the effect of the addition of the depth variable on the other variables is reduced. Secondly, individual sounds that are particularly similar will be adjusted accordingly.
The sounds used in the experiment were too monotonous. The experience should be better with the prototype.	It is correct to carry out the development of mobile applications. The final version will be complete and tested with the mobile app after the audio is improved later. Additionally, the study will add more artworks for practical application.
For congenitally visually impaired people, there is a lack of experience with color. Therefore, for them, this method may not make much sense.	Congenitally blind people understand colors through physical and abstract associations. Color audition means the reaction of feeling color in one sound [27]. In the future, the study will not only focus on functionality but will also add emotional things into it. Adding sensual sounds such as music to connect colors with emotions will make the color expression more vivid.
There is no difficulty in distinguishing, but it was a little difficult to distinguish when hearing fatigue occurred.	Switching between the simultaneous performance of multiple variables and performance of a single variable will be added, reducing user auditory fatigue.

This study has several advantages over other ones:

(1) This study presented color, lightness, and depth information at the same time with 3D sound and voice modulation;
(2) The virtual color wheel with 3D sound will help the user to understand the color composition;
(3) Our method can be combined with tactile tools for multiple art enjoyment facets.

However, this study has some limitations:

(1) The relative use of many variations of sound, which also makes it relatively more complex than other single variable methods, and also has basic requirements for the level of hearing. Additionally, the quality of the headphones will also directly affect the use of the effect;

(2) The existing and publicly available HRTF methods still have some drawbacks, i.e., they may have some effects when the gap with the selected HRTF specimen is too large. This study simplified the design of this, but there are still some limitations;

(3) The focus on function and lack of emotion may be useful for people with acquired visual impairment, while people with congenital visual impairment may lack empathy for color perception.

Through quiz tests and user evaluations, the sound code in future work could be improved in the following ways:

(1) The audibility and accuracy of the sound can be improved. Finding a more popular HRTF conversion method, or exploring the private custom HRTF, will lead to improvements in sound accuracy. Additionally, a better way to create sound accurately will greatly improve the user experience;

(2) While implementing complex functions, a simplified solution is needed to alleviate the user's difficulty in using them. The solution is to reduce the content of the expression to reduce the sound variables. Another is to use single-variable audio in the form of different forms of touch by the mobile app to play the corresponding variable audio;

(3) In this work, there were no large-scale tests using mobile applications. However, from the feedback of previous mobile applications, it is clear that the mobile application format will greatly increase the usability of the sound code we developed in this paper.

6. Conclusions

In this paper, we presented a methodology of 3D sound color coding using HRTF. The color hue is represented by the sound simulation of the position of the color wheel, and the lightness of color is reflected by the use of the sound pitch. The correlation between sound loudness and depth was found through experiments on the correlation between sound variables and depth, and the correlation was used to represent depth by changing the sound loudness and increasing the reverberation in addition to the original sound codes. Additionally, an identification test and system usability test were conducted in this study. A total of 97.88% of the identification test results showed that the system has excellent recognition. The results of the NASA TLX test and user experience test also showed the good usability of the system. Experiments with visually impaired subjects will be implemented in future studies.

This is a new attempt to express color. Although there are many ways to use sound to express color, there are few ways to use changes in a sound position to express color accurately. The variable of sound position is very common and familiar to the visually impaired. The use of this method also opens up a new direction in the way that art can be experienced by the visually impaired. However, there is still room for improvement in this method. Further refinements will increase the accuracy and usability. Future improvements in sound processing will also make recognition easier.

Neither sighted people nor people with visual impairment had experienced the proposed 3D sound coding colors before; therefore, it was judged that there were no significant differences in the perception ratings between sighted and visually impaired test people. However, future extended testing will be necessary to analyze the differences in the speed of perception between those two groups. Regarding the size of test participants, ten users who participated in this study's experiments may not be enough even though the magic number 5 rule (Nielsen & Landauer [60]) is vastly known and used for usability testing. The sample size is a long-running debate. Lamontagne et al. [61] investigated how many users are needed in usability testing to identify negative phenomena caused by a combination of the user interface and the usage context. They focused on identifying psychophysiological pain points (i.e., emotionally irritant experienced by the users) during human–computer interaction. Fifteen subjects were tested in a new user training context and results show that out of the total psychophysiological pain points experienced by 15 participants, 82% of them were experienced with nine participants. In the implicit association test done by Greenwald et al. [62], thirty-two (13 male and 19 female) students from introductory psychology courses. Therefore, as future work, we will also further perform scaled experiments on sighted participants and people with visual impairment.

The visual perception of artwork is not just bound to distance and color, but to a collection of different tools that artists use to generate visual stimuli. These, for example, are color hue, color value, texture, placement, size, contrast changes, cool vs. warm colors, etc. A better understanding of how these tools affect the visual perception of artwork may in the future enable the implementation of experiments that employ new visual features which may help to achieve enhanced "visual understanding" through sound. Schifferstein [63] observed that vivid images occur in all sensory modalities. The quality of some types of sensory images tends to be better (e.g., vision, auditory) than of others (e.g., smell and taste) for sighted people. The quality of visual and auditory images did not differ significantly. Therefore, training these multi-dimensional auditory experiences and incorporating color hue, near/far (associated with warm/cool), and light/dark introduced in this paper may lead to more vivid visual imageries, incorporating color or seeing them with the mind's eye. This study leaves other visual stimuli such as texture, placement, size, and contrast changes for the future work.

Synesthesia is a transition between senses in which one sense triggers another. When one sensation is lost, the other sensations not only compensate for the loss, but the two sensations are synergistic by adding another sensation to one [64]. Taggart et al. [65] found that artists, novelists, poets, and creative people have seven times more synesthesia than other fields. Artists often connect unconnected realms and blend the power of metaphors with reality. Synesthesia appears in all forms of art and provides a multisensory form of knowledge and communication. It is not subordinated but can expand the aesthetic through science and technology. Science and technology could thus function as a true multidisciplinary fusion project that expands the practical possibilities of theory through art. Synesthesia is divided into strong synesthesia and weak synesthesia. Strong synesthesia is characterized by a vivid image in one sensory modality in response to the stimulation of another sense. Weak synesthesia, on the other hand, is characterized by cross-sensory correspondences expressed through language or by perceptual similarities or interactions. Weak synesthesia is common, easily identified, remembered, and can be manifested by learning. Therefore, weak synesthesia could be a new educational method using multisensory techniques. Synesthetic experience is the result of a unified sense of mind; therefore, all experiences are synesthetic to some extent. The most prevalent form of synesthesia is the conversion of sound into color. In art, synesthesia and metaphor are combined. Through art, the co-sensory experience became communicative. The origin of the co-sensory experience can be found in painting, poetry, and music (visual, literary, musical). To some extent, all forms of art are co-sensory [66]. The core of an artwork is its spirit, but grasping that spirit requires a medium which can be perceived not only by the one sense intended, but also through various senses. In other words, the human brain creates an image by integrating multiple nonvisual senses and using a matching process with previously stored images to find and store new things through association. So-called intuition thus appears mostly in synesthesia. To understand reality as much as possible, it is necessary to experience reality in as many forms as possible; thus, synesthesia offers a richer reality experience than the separate senses, and that can generate unusually strong memories. Kandinsky said that when observing colors, all the senses (taste, sound, touch, and smell) are experienced together. An intensive review on Multi-sensory Experience and Color Recognition in Visual Arts Appreciation of Person with Visually Impairment can be found in Cho [67]. Therefore, a method for expressing colors through 3D audio could be developed, as has been presented in this paper. These weak synesthetic experiences of interpreting visual color information through 3D sound information will positively affect color perception for people with visual impairments.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/electronics10091037/s1.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L. and J.D.C.; software, Y.L.; validation, Y.L., J.D.C. and C.-H.L.; formal analysis, Y.L. and C.-H.L.; investigation, C.-H.L. and Y.L.; resources, Y.L.; data curation, Y.L.; writing—original draft preparation, Y.L. and J.D.C.; writing—review and editing, Y.L. and J.D.C.; visualization, Y.L.; supervision, J.D.C.; project administration, J.D.C.; funding acquisition, J.D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea, grant number 2018M3C1B6061353.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Sungkyunkwan University (protocol code: 2020-11-005-001, 22 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The sound files developed in this research are provided separately as Supplementary Materials.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Global Data on Visual Impairment 2010. Available online: https://www.who.int/blindness/publications/globaldata/en/ (accessed on 6 April 2021).
2. Coates, C. Best Practice in Making Museums More Accessible to Visually Impaired Visitors. 2019. Available online: https://www.museumnext.com/article/making-museums-accessible-to-visually-impaired-visitors/ (accessed on 30 November 2020).
3. Accessible Guides at The British Museum. Available online: https://www.britishmuseum.org/visit/audio-guide (accessed on 26 January 2021).
4. Audio Guides by The Metropolitan Museum of Art. Available online: https://www.metmuseum.org/visit/audio-guide (accessed on 26 January 2021).
5. Art inSight and MoMA Audio by Museum of Modern Art. Available online: https://www.moma.org/visit/accessibility/index (accessed on 26 January 2021).
6. Samantha Silverberg. A New Way to See: Looking at Museums through the Eyes of The Blind. 2019. Available online: https://www.pressreleasepoint.com/new-way-see-looking-museums-through-eyes-blind (accessed on 30 November 2020).
7. Vaz, R.; Freitas, D.; Coelho, A. Blind and Visually Impaired Visitors' Experiences in Museums: Increasing Accessibility through Assistive Technologies. *Int. J. Incl. Mus.* **2020**, *13*, 57–80. [CrossRef]
8. Jadyn, L. Multisensory Met: The Development of Multisensory Art Exhibits. Available online: http://www.fondazionemarch.org/multisensory-met-the-development-of-multisensory-art-exhibits.php (accessed on 30 November 2020).
9. Brule, E.; Bailly, G.; Brock, A.; Valentin, F.; Denis, G.; Jouffrais, C. MapSense: Multi-Sensory Interactive Maps for Children Living with Visual Impairments. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 445–457.
10. Hayhoe, S. *Blind Visitor Experiences st Art Museums*; Rowman & Littlefield: Lanham, MD, USA, 2017.
11. San Diego Museum of Art Talking Tactile Exhibit Panels. Available online: http://touchgraphics.com/portfolio/sdma-exhibit-panel/ (accessed on 23 April 2021).
12. Dobbelstein, D.; Herrdum, S.; Rukzio, E. Inscent: A Wearable Olfactory Display as An Amplification for Mobile Notifica-Tions. In Proceedings of the 2017 ACM international Symposium on Wearable Computers, Maui, HI, USA, 11–15 September 2017.
13. Anagnostakis, G.; Antoniou, M.; Kardamitsi, E.; Sachinidis, T.; Koutsabasis, P.; Stavrakis, M.; Vosinakis, S.; Zissis, D. Ac-Cessible Museum Collections for the Visually Impaired: Combining Tactile Exploration, Audio Descriptions and Mobile Gestures. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Florence, Italy, 6–9 September 2016; pp. 1021–1025.
14. Reichinger, A.; Fuhrmann, A.; Maierhofer, S.; Purgathofer, W. *A Concept for Reuseable Interactive Tactile Reliefs. Computers Helping People with Special Needs*; Miesenberger, K., Bühler, C., Penaz, P., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 108–115.
15. Cavazos Quero, L.; Bartolomé, L.C.; Cho, J.D. Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics. *Electronics* **2021**, *10*, 297. [CrossRef]
16. Hammershoi, D.; Møller, H. Methods for binaural recording and reproduction. *Acta Acust. United Acust.* **2002**, *88*, 303–311.
17. Ranjan, R.; Gan, W.S. Natural listening over headphones in augmented reality using adaptive filtering techniques. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1988–2002. [CrossRef]
18. Lessard, N.; Paré, M.; Lepore, F.; Lassonde, M. Early-blind human subjects localize sound sources better than sighted subjects. *Nature* **1998**, *395*, 278–280. [CrossRef]

19. Dong, M.; Wang, H.; Guo, R. Towards understanding the differences of using 3d auditory feedback in virtual environments between people with and without visual impairments. In Proceedings of the 2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE), Los Angeles, CA, USA, 19 March 2017.
20. Taras, C.; Ertl, T. Interaction with Colored Graphical Representations on Braille Devices. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, San Diego, CA, USA, 19–24 July 2009; pp. 164–173.
21. Ramsamy-Iranah, S.; Rosunee, S.; Kistamah, N. Application of assistive tactile symbols in a ’Tactile book’ on color and shapes for children with visual impairments. *Int. J. Arts Sci.* **2017**, *10*, 575–590.
22. Shin, J.; Cho, J.; Lee, S. Please Touch Color: Tactile-Color Texture Design for the Visually Impaired. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020.
23. Cho, J.D.; Quero, L.C.; Bartolomé, J.I.; Lee, D.W.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2021**, *46*, 103–116. [CrossRef]
24. Kandinsky, V. *Concerning the Spiritual in Art*; Dover Publications: Mineola, NY, USA, 1977.
25. Donnell-Kotrozo, C. Intersensory perception of music: Color me trombone. *Music Educ. J.* **1978**, *65*, 32–37. [CrossRef]
26. Deville, B.; Deville, B.; Bologna, G.; Bologna, G.; Vinckenbosch, M.; Vinckenbosch, M.; Pun, T.; Pun, T. See color: Seeing colours with an orchestra. In *Human Machine Interaction*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 251–279.
27. Cho, J.D.; Jeong, J.; Kim, J.H.; Lee, H. Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments. *Electronics* **2020**, *9*, 1981. [CrossRef]
28. Jabbar, M.S.; Lee, C.H.; Cho, J.D. ColorWatch: Color Perceptual Spatial Tactile Interface for People with Visual Impairments. *Electronics* **2021**, *10*, 596. [CrossRef]
29. Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization*; MIT Press: Cambridge, MA, USA, 1997.
30. Murphy, D.; Neff, F. Spatial sound for computer games and virtual reality. In *Game Sound Technology and Player Interaction: Concepts and Developments*; IGI Global: Hershey, PA, USA, 2011; pp. 287–312.
31. Jenny, C.; Reuter, C. Usability of individualized head-related transfer functions in virtual reality: Empirical study with perceptual attributes in sagittal plane sound localization. *JMIR Serious Games* **2020**, *8*, e17576. [CrossRef] [PubMed]
32. Larsen, C.H.; Lauritsen, D.S.; Larsen, J.J.; Pilgaard, M.; Madsen, J.B. Differences in human audio localization performance between a HRTF-and a non-HRTF audio system. In Proceedings of the 8th Audio Mostly Conference; 2013; Association for Computing Machinery, New York, NY, USA, 8 September 2013; Article 5. pp. 1–8. [CrossRef]
33. Mendonça, C.; Campos, G.; Dias, P.; Vieira, J.; Ferreira, J.P.; Santos, J.A. On the improvement of localization accuracy with non-individualized HRTF-based sounds. *J. Audio Eng. Soc.* **2012**, *60*, 821–830.
34. Desloge, J.; Rabinowitz, W.; Zurek, P. Microphone-array hearing aids with binaural output. I. fixed-processing systems. *IEEE Trans. Speech Audio Process.* **1997**, *5*, 529–542. [CrossRef]
35. Jeub, M.; Vary, P. Binaural dereverberation based on a dual-channel wiener filter with optimized noise field coherence. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14–19 March 2010.
36. Drossos, K.; Mimilakis, S.; Floros, A.; Kanellopoulos, N. Stereo goes mobile: Spatial enhancement for short-distance loudspeaker setups. In Proceedings of the 2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Piraeus-Athens, Greece, 18–20 July 2012; pp. 432–435.
37. Drossos, K.; Floros, A.; Giannakoulopoulos, A.; Kanellopoulos, N. Investigating the impact of sound angular position on the listener affective state. *IEEE Trans. Affect. Comput.* **2015**, *6*, 27–42. [CrossRef]
38. Li, B.; Zhang, X.; Muñoz, J.P.; Xiao, J.; Rong, X.; Tian, Y. Assisting blind people to avoid obstacles: An wearable obstacle stereo feedback system based on 3D detection. In Proceedings of the 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 6–9 December 2015; pp. 2307–2311.
39. Drossos, K.; Zormpas, N.; Giannakopoulos, G.; Floros, A. Accessible games for blind children, empowered by binaural sound. In Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, Corfu Greece, 1–3 July 2015; pp. 1–8.
40. Gardner, B.; Martin, K. Hrtf Measurements of a Kemar Dummy-Head Microphone; Vol. 280. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology. 1994. Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.9751&rep=rep1&type=pdf (accessed on 10 February 2021).
41. Newton, I. *Opticks, or, a Treatise of the Reflections, Refractions, Inflections & Colours of Light*; Courier Corporation: Chelmsford, MA, USA, 1952.
42. Maryon, E. *MARCOTONE the Science of Tone-Color*; Birchard Hayes & Company: Boston, MA, USA, 1924.
43. Peacock, K. Synesthetic perception: Alexander Scriabin’s color hearing. *Music Percept.* **1985**, *2*, 483–505. [CrossRef]
44. Lavignac, A. *Music and Musicians*; Henry Holt and Company: New York, NY, USA, 1903.
45. Anikin, A.; Johansson, N. Implicit associations between individual properties of color and sound. *Atten. Percept. Psychophys.* **2019**, *81*, 764–777. [CrossRef]
46. Hamilton-Fletcher, G.; Witzel, C.; Reby, D.; Ward, J. Sound properties associated with equiluminant colors. *Multisens. Res.* **2017**, *30*, 337–362. [CrossRef] [PubMed]
47. Giannakis, K. Sound Mosaics: A Graphical User Interface for Sound Synthesis Based on Audio-Visual Associations. Ph.D. Thesis, Middlesex University, London, UK, 2001.

48. Jonas, C.; Spiller, M.J.; Hibbard, P. Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychon. Bull. Rev.* **2017**, *24*, 1104–1112. [CrossRef] [PubMed]
49. Cogan, R.D. *Sonic Design: The Nature of Sound and Music*; Prentice Hall: Upper Saddle River, NJ, USA, 1976.
50. Itten., J. *The Art of Color: The Subjective Experience and Objective Rationale of Color*; Wiley: New York, NY, USA, 1974.
51. Algazi, V.R.; Duda, R.O.; Thompson, D.M.; Avendano, C. The cipic hrtf database. In Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575), New Platz, NY, USA, 24–24 October 2001.
52. Listen HRTF Database. Available online: http://recherche.ircam.fr/equipes/salles/listen/index.html (accessed on 10 February 2021).
53. HRTF Measurements of a KEMAR Dummy-Head Microphone. Available online: https://sound.media.mit.edu/resources/KEMAR.html (accessed on 10 February 2021).
54. Bomhardt, R.; Klein, M.D.L.F.; Fels, J. A high-resolution head-related transfer function and three-dimensional ear model database. *Proc. Meet. Acoust.* **2016**, *29*, 050002.
55. Berzborn, M.; Bomhardt, R.; Klein, J.; Richter, J.G.; Vorländer, M. The ITA-Toolbox: An open source MATLAB toolbox for acoustic measurements and signal processing. In Proceedings of the 43rd Annual German Congress on Acoustics, Kiel, Germany, 6 May 2017; Volume 6, pp. 222–225.
56. Krygier, J.B. Sound and geographic visualization. In *Visualization in modern cartography*; MacEachren, A.M., Taylor, D.R.F., Eds.; Pergamon: Oxford, UK, 1994; pp. 149–166.
57. Osgood, C.E.; Suci, G.J.; Tannenbaum, P.H. *The Measurement of Meaning*; University of Illinois: Urbana, IL, USA, 1957; Volume 12.
58. NASA Task Load Index. Available online: https://humansystems.arc.nasa.gov/groups/tlx/index.php (accessed on 15 March 2021).
59. System Usability Scale. Available online: https://www.questionpro.com/blog/system-usability-scale (accessed on 15 March 2021).
60. Nielsen, J.; Landauer, T.K. A mathematical model of the finding of usability problems. In Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA; 1993; pp. 206–213. [CrossRef]
61. Lamontagne, C.; Sénécal, S.; Fredette, M.; Chen, S.L.; Pourchon, R.; Gaumont, Y.; Léger, P.M. User Test: How Many Users Are Needed to Find the Psychophysiological Pain Points in a Journey Map? In *Human Interaction and Emerging Technologies*; Springer: Cham, Switzerland, 2019; pp. 136–142.
62. Greenwald, A.G.; Nosek, B.A.; Banaji, M.R. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Personal. Soc. Psychol.* **2003**, *85*, 197. [CrossRef]
63. Schifferstein, H.N. Comparing mental imagery across the sensory modalities. *Imagin. Cogn. Personal.* **2009**, *28*, 371–388. [CrossRef]
64. Brang, D.; Ramachandran, V.S. How do crossmodal correspondences and multisensory processes relate to synesthesia? In *Multisensory Perception*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 259–281.
65. Taggart, E. Synesthesia Artists Who Paint Their Multi-Sensory Experience. 2019. Available online: https://mymodernmet.com/synesthesia-art/ (accessed on 23 April 2021).
66. Lawrence, E. *Marks the Unity of the Senses/Interrelationships Among the Modalities, Series in Cognition and Perception*; Academic Press: New York, NY, USA, 1978.
67. Cho, J.D. A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment. *Electronics* **2021**, *10*, 470. [CrossRef]

electronics

MDPI

Article

ColorPoetry: Multi-Sensory Experience of Color with Poetry in Visual Arts Appreciation of Persons with Visual Impairment

Jun-Dong Cho * and Yong Lee

Department of Electronic and Electrical Engineering, Sungkyunkwan University,
Suwon, Gyeonggi-do 16419, Korea; lee@skku.edu
* Correspondence: jdcho@skku.edu; Tel.: +82-10-4332-7127

Abstract: Visually impaired visitors experience many limitations when visiting museum exhibits, such as a lack of cognitive and sensory access to exhibits or replicas. Contemporary art is evolving in the direction of appreciation beyond simply looking at works, and the development of various sensory technologies has had a great influence on culture and art. Thus, opportunities for people with visual impairments to appreciate visual artworks through various senses such as hearing, touch, and smell are expanding. However, it is uncommon to provide a multi-sensory interactive interface for color recognition, such as integrating patterns, sounds, temperature, and scents. This paper attempts to convey a color cognition to the visually impaired, taking advantage of multisensory coding color. In our previous works, musical melodies with different combinations of pitch, timbre, velocity, and tempo were used to distinguish vivid (i.e., saturated), light, and dark colors. However, it was rather difficult to distinguish among warm/cool/light/dark colors with using sound cues only. Therefore, in this paper, we aim to build a multisensory color-coding system with combining sound and poem such that poem leads to represent more color dimensions, such as including warm and cool colors for red, orange, yellow, green, blue, and purple. To do this, we first performed an implicit association test to identify the most suitable poem among the candidate poems to represent colors in artwork by finding the common semantic directivity between the given candidate poem with voice modulation and the artwork in terms of light/dark/warm/color dimensions. Finally, we conducted a system usability test on the proposed color-coding system, confirming that poem will be an effective supplement for distinguishing between vivid, light, and dark colors with different color appearance dimensions, such as warm and cold colors. The user experience score of 15 college students was 75.1%, that was comparable with the color-music coding system that received a user experience rating of 74.1%. with proven usability.

Keywords: visual impairment; accessibility; aesthetics; color; multi-sensory; museum exhibits

Citation: Cho, J.-D.; Lee, Y. ColorPoetry: Multi-Sensory Experience of Color with Poetry in Visual Arts Appreciation of Persons with Visual Impairment. *Electronics* **2021**, *10*, 1064. https://doi.org/10.3390/electronics10091064

Academic Editors: Juan M. Corchado and Rui Pedro Lopes

Received: 2 March 2021
Accepted: 25 April 2021
Published: 30 April 2021

1. Introduction

Multimodal (or multisensory) integration refers to the neural integration or combination of information from different sensory modalities (the classic five senses of vision, hearing, touch, taste, and smell, and, perhaps less obviously, proprioception, kinesthesis, pain, and the vestibular senses), which gives rise to changes in behavior associated with the perception of and reaction to those stimuli. Information is typically integrated across sensory modalities when the sensory inputs share certain common features [1].

Cross-modality refers to the interaction between two different sensory channels. Cross-modal correspondence is defined as the surprising associations that people experience between seemingly unrelated features, attributes, or dimensions of experience in different sensory modalities.

The intuitive strategies based on cross-modal analogy, association, and symbolism are suitable for creating a design that provides connections between the senses, which directly appear appropriate and easy to interpret [2].

Although many studies have been conducted on the cross-sensation between the sight and other senses, there are not many studies on the cross-modality between the non-visual senses. Through the investigation of weak synesthesia that is perceived at the same time by intersecting various senses such as audio, touch, smell, etc., other sensory information can be connected to a specific form of color information. In this task, so far, efforts have been made to create different single-mode sensory perceptions for color, for example, color–sound, color–tactile pattern, color–odor, etc.

In our previous studies, colors were expressed as embossed tactile patterns for recognition by finger touches, temperature, sound, and smell to provide a rich art experience to person with visual impairments. The Black Book of Colors by Cottin [3] described the experience of a fictional blind child named Thomas, who describes color through association with certain elements in his environment. This book highlights the fact that blind people can gain experience through multisensory interactions: "Thomas loves all colors because he can hear, touch, and taste them."

An accompanying audio explanation provides a complementary way to explore the overall color composition of an artwork. When tactile patterns are used for color transmission, the image can be comprehensively grasped by delivering graphic patterns, painted image patterns, and color patterns simultaneously [4].

This suggests to us the possibility and the justification for developing a new way of appreciating works in which the colors used in works are subjectively explored through the non-visual senses. In other words, it can be inferred that certain senses will be perceived as being correlated with certain colors and concepts through unconscious associations constructed with the concepts. However, while this helps visually impaired users perceive and understand colors in completely different ways, it becomes necessary to integrate all perceptual sensations into a single multi-sensory system, where all the senses are perfectly connected and interchanged. This concept is in line with the theory of emotional intelligence, which defends the organic functioning of people, as a whole, in interaction between the emotional, the sensory, and the cognitive [5,6]. According to Damásio [5], feelings are basically the mind's interpretation of the state of the body, and in order for reason to be truly rational, it must be based on emotional cues from the body. Emotional memories are not easily erased from memory and remain the longest. Tacit memories that did not easily emerge above consciousness can suddenly come to mind through the opportunity of contact with any part of one's body [5].

Audio description does provide a useful service for those with low or no visual acuity [7,8]. Schifferstein [9] observed that vivid images occur in all sensory modalities. The quality of some types of sensory images tends to be better (e.g., vision, audition) than of others (e.g., smell and taste) for sighted people. The quality of visual and auditory images did not differ significantly. Therefore, training these multi-sensory experiences introduced in this paper may lead to more vivid visual imageries or seeing with the mind's eye. However, such descriptions are often monolithic accounts, missing the opportunity for alternative messages to be delivered and received. Making art accessible to the visually impaired requires the ability to convey explicit and implicit visual images through non-visual forms. It argues that a multi-sensory system is needed to successfully convey artistic images. It also designed a poetic text audio guideline that blends sounds and effects to translate the work into an ambiguous artistic sound text [9].

Audio or verbal description on an artwork generally attempts to explain the painting without expressing the individual subjectivity. When making visual art accessible to the visually impaired, it is not enough to describe the colors and situations portrayed because that objective information does not attach to anything in their experience. Using expressions that contain the sensibility of poems, color, and situation can be matched with the parts of each painting.

Poem is a piece of writing that has features of both speech and song, whereas poetry is the art of creating these poems. Throughout the poem, seeing and hearing are used to understand color. Imagery is a literary device that refers to the use of figurative language to evoke a sensory experience or create a picture with words for a reader. By utilizing effective descriptive language and figures of speech, writers appeal to a reader's senses of sight, taste, smell, touch, and sound, as well as internal emotion and feelings. Poetic imagery provides sensory details to create clear and vivid descriptions. This appeals to a reader's imagination and emotions as well as their senses [10].

The connection between color and poem has not been studied through existing research. In this paper, we explore how color can be explained by poetry so that visually impaired people feel or perceive color through poetry. However, the connection between color and poetry should be clearly matched. When appreciating color in artworks using poetry, we need to explore whether color concept and poem interfere with each other, causing confusion in color perception, or synergizing with each other to positively affect color perception. A system usability test is performed to evaluate whether it helps to easily recognize the color.

2. Background

2.1. Review on Ekphrasis

In poetry, synesthesia refers specifically to figurative language that includes a mixing of senses. For example, saying "He wore a loud yellow shirt" is an example of synesthesia, as it mixes visual imagery (yellow) with auditory imagery (loud) [11]. Here is another example "The Loudness of Color" by Jennifer Betts [12]:

The music of white dances softly around
The soft silence and blue are bound
Purple is calm, the sound soft and sweet
The color lightness of a rainbow is a hypnotic beat
In yellow, the silence is loud
While red is a yell, robust and proud

Red does not actually yell, but many would describe it as loud. Using sound to describe color makes it fun and interesting for the reader. The art of writing poetry about paintings is known as ekphrasis [13–15]. Ekphrasis' technology was used to convey description, imagination, and emotion (vibrance) to direct visual artworks and played a role in enhancing the value of art. Edward Hirsch (1994) said, "works of art initiate and provoke other works of art" [16].

Color theory is associated with the Bauhaus artists Johannes Itten, Josef Albers, and Paul Klee. Itten wrote "The Art of Color", The Elements of Color and Albers' Interaction of Color, and Paul Klee notably produced picture poems, concerned with colors and words.

The work of many contemporary poets has also been influenced by the visual arts: Rainer Maria Rilke, W.B. Yeats, W.H. Auden, and William Carlos Williams [17], Jane Flanders, Anne Sexton, X. J. Kennedy, and Gershon Hepner (Table 1), are a few. The examples of ekphrases are shown in Table 1.

Koch [18] and Routman [19] made common observations on the practice of teaching poetry, specifically in regard to the importance of utilizing a range of quality poetry written by children and adults. Koch prefers to read poetry examples and then allow students to be influenced and imitate other poets. It makes them feel more open to understanding what others have written. Previous works mainly focus on generating poetry given keywords or other text information, while visual inspirations for poetry have been rarely explored.

Table 1. Examples of ekphrasis (**a**): artworks, (**b**): poems.

(a) **Artworks**			
Vincent van Gogh Arles, 1888 Van Gogh Museum 	Vincent van Gogh The Starry Night, 1889 Modern Museum of Art 	Marcel Duchamp Nude Descending a Staircase No.2, 1912 Philadelphia Museum of Art 	Mark Rothko No. 13 (White, Red on Yellow), 1958 Museum of Modern Art 
(b) **Poems**			
Van Gogh's Bed Jane Flanders, 2015 [20]	The Starry Night Anne Sexton, 1961 [21]	Nude Descending a Staircase X. J. Kennedy, 1961 [22]	White, Red on Yellow Gershon Hepner, 2008 [23]
is orange, like Cinderella's coach, like the sun when he looked it straight in the eye. is narrow, he sleeps alone, tossing between two pillows, while it carried him bumpily to the ball. is clumsy, but friendly. A peasant built the frame; and old wife beat the mattress till it rose like meringue. is empty, morning light pours in like wine, melody, fragrance, the memory of happiness.	The town does not exist except where one black-haired tree slips up like a drowned woman into the hot sky. The town is silent. The night boils with eleven stars. Oh starry night! This is how I want to die. It moves. They are all alive. Even the moon bulges in its orange irons to push children, like a god, from its eye. The old unseen serpent swallows up the stars. Oh starry night! This is how I want to die: into that rushing beast of the night, sucked up by that great dragon, to split from my life with no flag, no belly, no cry.	Toe upon toe, a snowing flesh, A gold of lemon, root and rind, She sifts in sunlight down the stairs With nothing on. Nor on her mind. We spy beneath the banister A constant thresh of thigh on thigh– Her lips imprint the swinging air That parts to let her parts go by. One-woman waterfall, she wears Her slow descent like a long cape And pausing, on the final stair Collects her motions into shape.	White and red and yellow Rothko thought more mellow than mountains that are grandiose, or portraits. Figures can be gross and mountains overwhelm the view, but colors, when they're few and pure, achieve a balance of pure reason in every place, in every season.

Good poems produce sensational images in compressed and refined language. When you read a poem, the scene it describes seems to unfold before your eyes. Association is a phenomenon in which one idea evokes another. Five images found in poetry, for example, are as follows.

1. Visual: the light blue sky
2. Sound: the wind attenuating the acoustic image
3. Olfactory: The smell of grass flowers
4. Taste: sweet chocolate
5. Tactile: warm stars
6. Synesthesia: blue whistling sound

2.2. Review on Review on Color Association with Sound

Recently, the following works were devised to materialize multi-sensory appreciation for the color of artworks for the visually impaired [24]. An experiment comparing the cognitive capacity for color codes found that users could intuitively recognize 24 chromatic and 5 achromatic colors with tactile pictogram codes [25] and 18 chromatic and 5 achromatic colors with sound codes [26].

Cho et al. [25] presented a tactile color pictogram system to communicate the color information of visual artworks. The tactile color pictogram uses the shape of sky, earth, and human derived from the oriental philosophy of heaven, earth, and human as a metaphor. The tactile color pictogram used a slightly larger cell size compared to most tactile patterns currently used to indicate color, but code for more colors due to its simple structure. With user experience and identification tests, conducted with 23 visually impaired adults, the effectiveness of the tactile color pictogram suggested that they were helpful for the participants. Colors can thus be recognized easily and intuitively by touching the different patterns.

What the art teacher wanted to do most with her blind students was to have them imagine colors using a variety of senses—touch, scent, music, poetry, or literature. A comprehensive survey of associations between color and sound can be found in [24], including how different color properties such as value and hue are mapped onto acoustic properties such as pitch and loudness. Cho et al. [26] developed a sound coding color (Table 2) to express red, orange, yellow, green, blue, and purple using a melody (length: about 10 s) excerpted from a Classic music video with different musical instruments, intensity, and pitch of sound to express vivid, light, and dark colors. In Reference [26], musical instruments were classified for each color to ensure that they would be easily distinguished from one another. Red, a representative warm color, is expressed by a violin that plays a passionate and strong melody. A trumpet plays a high-pitched melody with energy, as if a bright light were expanding, to simulate yellow bursts. Orange is expressed by a viola playing a warm yet energetic melody. Green, which makes the eyes feel comfortable and psychologically stable, is expressed by a fresh oboe that plays a soft melody. Blue, a representative cold color, is expressed by a cello that plays a low, calm melody. Violet, where warm red and cold blue coexist, is expressed by a pipe organ that plays a magnificent yet solemn melody.

Vivid colors and bright and dark colors were distinguished through a combination of pitch, instrument tone, intensity, and tempo. High color lightness used a small, light, particle-like melody and high-pitch sounds, and a bright feeling was emphasized by using a melody of relatively fast and high notes. For low color lightness, a slow, dull melody with a relatively low range was used to create a sense of separation and movement away from the user.

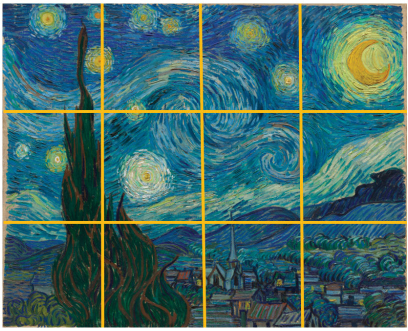
Table 2. Sound coding color [26] (the sound files developed in this research are provided separately as a Supplemental Materials).

Instrument and Music	L	V	D
R-Vn. High-frequency banded string instrument Tchaikovsky: Violin Concerto in D	1. R-L.wav	1. R-S.wav	1. R-D.wav
O-Va. Medium-frequency banded string instrument Stamitz: Viola Concerto in D	2.O-L.wav	2.O-S.wav	2.O-D.wav
Y-Trp. Brass instrument Haydn: Trumpet Concerto in E flat	3.Y-L.wav	3.Y-S.wav	3.Y.-D.wav
G-Ob. Woodwind instrument with reed Mozart: Oboe Concerto in C	4. G-L.wav	4. G-S.wav	4. G-D.wav
B-Vc. Low-frequency banded string instrument Bach: Cello Suite No. 1 in G	5.B-L.wav	5.B-S.wav	5.B-D.wav
P-Org. Keyboard instrument with simultaneous expressions Mozart: Eine Kleine Nachtmusik	6. P-L.wav	6. P-S.wav	6. P-D.wav

In Reference [26], the overall color composition of Van Gogh's "The Starry Night" was expressed as a single piece of music that accounted for color using the tone, key, tempo, and pitch of the instruments. To express the highly saturated (vivid) blue of the night sky, which dominates the overall hue of the picture, a strong, clear melody in the mid-range was excerpted from the Bach unaccompanied cello suite No. 1 to form the base of the whole song, and it is played repeatedly without interruption. To express the twinkling bright yellow of the stars, a light particle-like melody was extracted from Haydn's Trumpet Concerto and played as a strong, clear melody in the mid-range. The painting was divided into four lines, and worked with 16 bars per line, producing a total of 68 bars played in 3 min and 29 s. The user experience evaluation rate from nine blind people was 84%, and the user experience scores from eight sighted participants were 79% and 80% for the Classical and Vivaldi schemes, respectively. After one hour of practice, the cognitive success rate for three blind people was 100% for both the Classical and Vivaldi schemes. Here, we show another example of composing a single music piece representing colors in Van Gogh's "The starry night", as shown in Table 3.

The painting was divided into three rows and four columns, and played in 2 min and 10 s. The color sounds in each entry in the color map table were joined to form a single music piece. Therefore, while listening to music, the user can recall the color corresponding to each sound and the location of the color in the painting. These sound cues were prepared for the comparison purpose (as we shall explain in the system usability score test section).

Table 3. Color map table and single music coding colors in Van Gogh's "The starry night" (the sound files developed in this research are provided separately as a Supplemental Material).



starry
night-jdcho@skku.

star (Y-L) wind (B-L) tree (G-D)	star (Y-L) wind (B-L)	star (Y-L) wind (B-L)	moon (O-S)	a little dark (Snare drum)
star (Y-L) tree (G-D)	star (White) wind (B-L)	wind (B-L) stars (Y-L)	stars (Y-L) trees (G-D)	a little dark (Snare drum)
town (G-S)	tree (G-D)	town (G-S)	town (G-S)	dark (Timpani)

Bartolome et al. [27] expressed color and depth (advancing and retreating) as temperature and demonstrated the temperature-sensitive replica of Marc Rosco's work using a thermoelectric Peltier element and a control board. For the sighted people, at some point of the test, eight of the ten users agreed that the conceptual dichotomies warm-near and cold-far were correlated. This fact was proven during the test with persons with visual impairment, where 83% of the users (a total of five out of six) linked in both stages the warm temperature to the concept of being near something, and the cold temperature to the concept of being far [27].

Using an implicit associations test, Anikin et al. [28] confirmed the following cross-modal correspondences between visual and acoustic features. Pitch was associated with color lightness, whereas loudness mapped onto greater visual saliency. The hue of colors with the same luminance and saturation was not associated with any of the tested acoustic features, except for a weak preference to match higher pitch with blue (vs. yellow).

In the Doppler effect, the response of sound waves to moving bodies is illustrated in the example of the sounding of the locomotive whistle of a moving train. When the train blows its whistle while it is at rest in the station, stationary listeners who are either ahead of the engine or behind it will hear the same pitch made by the whistle, but as the train advances, those who are ahead will hear the sound of the whistle at a higher pitch. Listeners behind the train, as it pulls further away from them, hear the pitch of the whistle begin to fall [29]. Using such a principle, depth information of a color object located close to the viewer's gaze can be expressed by voice. In other words, if you increase the intensity of the voice, the color will feel close and intense. However, if you weaken the sound intensity, the color will feel pale and distant.

Jonas et al. [30] and Cogan et al. [31] found that a higher value in lightness is associated with higher pitch. Another way to match color and sound is to associate an instrument's tone with color, as in Kandinsky [32]. A low-pitched cello has a low-brightness dark blue color, a violin or trumpet-like instrument with a sharp tone feels red or yellow, and a high-pitched flute feels like a bright and saturated sky blue. As shown in Table 1, it is possible to compare the sensibility felt in each instrument tone with the sensibility felt in color. Chroma has a relationship with sound intensity [33,34]. This color encoding problem is the same as finding the entropy introduced in Claude Shannon's Theory of Information.

Entropy is the average (estimated) minimum resources required to provide information in an event.

Marks et al. [35] found that children of all ages and adults matched pitch to value and loudness to chroma. The value (i.e., lightness) is high and heavy, dependent on the light and dark levels of the color. Using the same concept in music, sound is divided into light and heavy feelings according to the high and low octaves of a scale. When the intensity of the sound is strong, the color sensed is close and sharp, whereas when the intensity of the sound is weak, the color becomes distant and muted [35]. Wilms and Oberfeld [36] observed that the increase of the arousal ratings from low to medium to high brightness is only present at the two higher saturation levels (among three levels). For the arousal dimension, it ranges from a relaxed, sleepy figure to an excited, wide-eyed figure. By Rowe [37], warm sound has a tilt towards the bass frequencies. The bass and vocals are more prominent, "bright" is the opposite of warm. Bright gear is better at reproducing high-pitched sounds.

Synesthete prevalence among fine-art students was estimated to be 23% [38]. Among various synesthesia due to color (N = 365), colored graphemes are the most common form occurring in two-thirds (66.8%) of a group of 365 synesthetes, and colored time units 19.2%, colored musical sounds 14.5%, and colored general sounds 12.1% [39]. Chromesthesia or sound-to-color synesthesia is a type of synesthesia in which sound involuntarily evokes an experience of color, shape, and movement [40]. Synesthetes that perceive color while listening to music experience the colors in addition to the normal auditory sensations. In Reference [41], synesthetes and non-synesthetes alike associate high-pitched sounds with lighter or brighter colors and low-pitched sounds with darker colors. For some individuals, chromesthesia is only triggered by speech sounds, while others' chromesthesia can be triggered by any auditory stimuli [42]. In a study investigating variability within categories of synesthesia, 40% of subjects with chromesthesia for spoken words reported that voice pitch, accent, and prosody influenced the synesthetic color [43].

2.3. Review on Color Association with Other Senses

Areas that are relatively high in chroma and lightness can seem to "come forward" in the sense of being visually more insistent than other areas, and orange-red, orange, and yellow paints attain higher chroma-lightness combinations than paints of other hues [44]. We can see through experience, that lighter, cooler colors seem to recede, thus making a room feel larger (giving it more "room"), while warmer, more saturated, and darker colors seem to advance, and take up more space in a room, thus making it appear smaller [45].

Ludwig and Simner [46] found that smooth, soft, and round stimuli tended to induce brighter colors, compared to rough, hard, and spiky stimuli.

In Slobodenyuk et al. [47], participants were asked to use colors to describe vibrotactile stimuli of varying frequencies and intensities, simulating variations in roughness–smoothness, heaviness–lightness, elasticity–inelasticity, and adhesiveness–non-adhesiveness. Analysis of the hue, chroma, and brightness of the chosen colors showed a bias towards the red, violet, and blue spectra of hue for the highest intensity haptic stimuli, and toward yellow and green for the lowest intensity, for which green colors were chosen the least. The least intense stimuli also had the lowest level of chroma and highest level of brightness, whereas the opposite was true for the most intense stimuli [47].

There are also indications of hedonic scores mediating cross-modal interactions of odors and colors. Namely, bright colors tend to be rated as pleasant, while darker colors tend to be found more unpleasant (Maric and Jacquot [48]). Strong fragrances are associated with dark colors (Kemp and Gilbert [49]), and there are research results that state that floral fragrances are associated with bright colors (Fiore [50]).

Kim [47] found that the floral and woody families showed more distinguishable opposite patterns in both hue and tone parameters: the floral family with brighter warm colors and the woody family with darker (or stronger) cool colors. The warm colors strongly evoked the floral family, while the cool colors the fresh family. The brighter (darker) their lightness values become, the more the floral (woody) scents are associated. Smoothness, softness, and roundness of stimuli positively correlated with luminance of the chosen color, and smoothness and softness also positively correlated with chroma [51]. These survey results are summarized in Tables 4 and 5.

Table 4. Semantic differential association between color warmness/coolness and various sensations.

The Semantic Differential Pair Relevant to Warmness and Coolness (Sensation)	Source
Strong~Week (sound)	Jonas et al. [30] Marks et al. [35]
Near~Far (thermal-tactile)	Bartolomé et al. [18]
Warm~Cool (vision)	N/A

Table 5. Semantic differential association between color lightness and various sensations.

The Semantic Differential Pair Relevant to Lightness and Darkness (Sensation)	Sources
Soft~Hard (haptic)	Ludwig and Simner [45] Slobodenyuk et al. [46]
Round (Bouba)~Pointed (Kiki) (haptic)	Ludwig and Simner [45]
Smooth~Rough (haptic)	Ludwig and Simner [45] Slobodenyuk et al. [46]
Light~Heavy (haptic)	Slobodenyuk et al. [46]
Light~Heavy (sound)	Marks et al. [35]
High~low (sound)	Ankins et al. [28] Jonas et al. [30] Cogan et al. [31] Kandinsky [32]
Aroused~Calm (vision)	Wilms and Oberfeld [36]
Pleasant~unpleasant (scent)	Maric and Jacquot [46]
Weak~Strong (scent)	Kemp and Gilbert [49]
Floral~Woody (scent)	Fiore [50], Kim [51]

3. Proposed ColorPoetry System

This section describes 'ColorPoetry', a systematic approach to color expression aids using poetry. The proposed system can explicitly model the color directivity of poems and effectively matches the color appearance dimension (vivid/light/dark/cool) of artwork and poem. Using a user study, we demonstrate that our color-coding scheme can effectively match colors in artwork with poems and provide a significantly higher user experience. We envision that this color-coding system utilizing poems will enable many useful applications in appreciating visual arts for the sighted people as well as people with visual impairment.

3.1. Color Selection

The perception of color is often described by referring to three dimensions of the color experiences: hue, intensity, and value. There are four visual perceptive terms that are used to describe color appearance, as defined below [44].

Hue: Hue could also be called "root" or "source" color. The hue is always one of the 12 key color places on the basic color wheel (Figure 1).

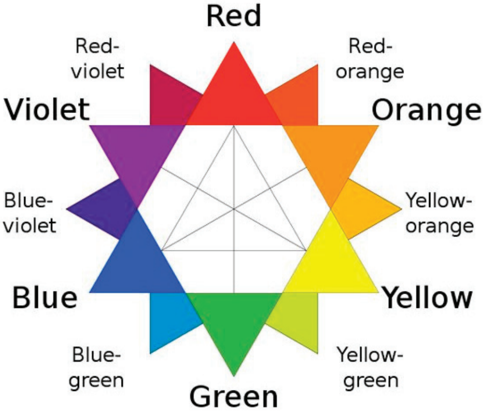


Figure 1. 12 RYB color wheel. These images are used under GNU Free Documentation License.

Value: The value of a color is always going to be compared against a basic palette of white, grey, or black. A color's value refers to the quality that distinguishes a light color from a dark one. In other words, it refers to the lightness or darkness of a hue.

Intensity: Intensity is more commonly called "saturation". To measure a brightness of any given color, it is first necessary to identify the root hue/color it is most closely related to. This is usually one of the two neighboring colors that sit next to it on the color wheel. Then, the new color's intensity can be characterized in terms of "brightness" or "dullness" as related to the predominant root hue.

Temperature: Temperature is perhaps the most subjective of the four basic color qualities outlined by Itten. It is also in this quality where it is easiest to see the emotional relationship Itten himself had with color. Temperature is expressed in terms of warm or cool.

Cho et al. [26] investigated that melodies with different pitch, tone, velocity, and tempo can be used as color sound codes to easily express the color lightness level. However, the number of colors that can be represented by sound was limited to 18 (three levels of brightness for each of six hues). In this paper, using poems, we extend the number of colors to be expressed up to 30, including warmer and cooler colors of each hue. In this way, the visually impaired can improve their color literacy to enjoy the colorful world in the painting. In this perspective, the extension can be done by using sounds to represent six color hues (red, orange, yellow, green, blue, violet) and poems to represent light/dark and warm/cool colors, simultaneously. In other words, the six color hues like red, orange, yellow, green, blue, and violet on the 12 RYB color wheel, Figure 1, are represented by sound codes. The six tertiary colors are warm and cool colors of red, yellow, and blue, such as red-orange (the warm color of R), red-violet (the cool color of R), yellow-orange (the warm color of Y), yellow-green (the cool color of Y), blue-green (the warm color of B), and blue-violet (the cool color of B), and are represented by sound code and poem (to represent the warmer or cooler colors of the color hue) together. Also, 12 lighter and darker colors from 6 color hues are also represented by sound (color hue) and poem (to represent the lighter or darker colors of the color hue) together. Also, there are five achromatic colors: white (WH), black (BK), and three levels of grays (dark gray, middle gray, light gray). In this system, the higher the lightness value, the closer the color is to white, and the lower the lightness, the closer it is to black. As shown in Figure 2, the light color (L) has a value of 7 and chroma of 8. The dark color (D) has a value of 3 and a chroma of 8. A saturated color (S) has high chroma (level 15) and colors with the lowest chroma (level 0) are achromatic.

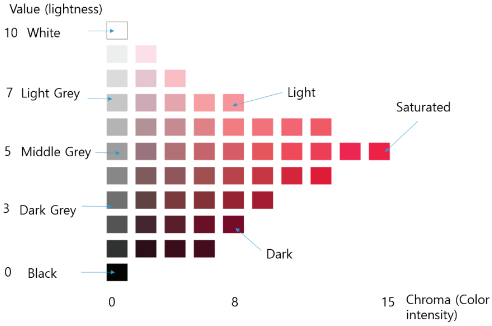


Figure 2. Saturated (S), light (L), and dark (D) for red.

3.2. *Algorithm of Poetry Coding Colors*

In this section, we explore the poem coding colors representing five color appearance dimensions. First, we need to identify the poems that feels "good" for the color appearance dimension used in the visual artwork. Next, we express such color appearance dimension using three levels of voice pitches respectively, while reciting a poem. Therefore, combining voice sounds and poems in this way makes it simpler and easier to identify color, including light, dark, warm, and cool colors, compared to the conventional way of using a single modality such as sound, tactile pattern, or temperature alone. Given an artwork image A and a set of poems DP, as illustrated in Table 6, the goal is to find the set of poems P ∈ DP that are mostly relevant to the color appearance dimension directivity of the image. Let S(A) (respectively S(P)) be denoted as a measure of the overall color appearance dimensions observed in artwork A (respectively P). We are now ready to describe the ColorPoetry Algorithm 1 as follows. The purpose of the algorithm is to find the set of "good" poems P with the maximum significance of Art–Poem similarity score denoted by sigmax S(A, P):

Algorithm 1 ColorPoetry (i.e., Poetry Coding Colors).

```
Input: a color C in an artwork A, a poem database DP.
Output: a set of poems P ∈ DP.
1: For every color in C
Identify the set of poems P, finding sigmax S (A, P), for all P ∈ DP.
2: Return P.
```

3.3. *Overall Solution Strategy*

There are three main steps to solving the problem.

(1) Determine poem database. Color poems are about a single color, using descriptions, nouns, and other key elements to express feelings about that color. An easy format for this type of poem is describing the color using the five senses: looks, sounds, tastes, feels, and smells. From Wikipedia, lyric poetry is a form of poetry that does not attempt to tell a story, as do epic poetry and dramatic poetry, but is of a more personal nature instead. Rather than portraying characters and actions, the lyric poet addresses the reader directly, portraying his or her own feelings, states of mind, and perceptions. In the poems, colors are used as a metaphor or to express the context in the poem, and we mainly collected them extensively through browsing relevant websites like "poemhunter.com" [48], and literature. The color database we built consists of 12–15 poems with different moods for each of the 6 hues: red, orange, yellow, green, blue, purple, including white, grey, and black. These poems are selectively analyzed and selected according to the appearance dimensions of the color in a given artwork.

(2) Determine color perceptual adjectives. The perceptual semantic difference method is used to establish a rigorous sensory analysis process to evaluate algorithms to find phrases in poetry that match the color of a given artwork. To reflect the user's perception

of color perception, it is necessary to use a set of color-related perceptual adjectives that have been validated in psychological research. The amount and fidelity of the collection of perceptual adjectives affects the quality of the research data. Adjectives that describe colors have been extensively collected through the existing literature.

(3) Data analysis and user experience test. The results of the evaluation of the correlation between the perceptual meanings of the colors that appeared in artwork and poetry were analyzed with statistical tools. The user experience test of the proposed color-coding system will be performed through user interviews.

3.4. Voice Modulation

Four color properties such as light/dark/warm/cool can be expressed by voice, so the brightness of the color is modulated according to the pitch of the sound, and the degree of warmth and coolness of the color can be expressed by the strength and reverberation of the voice. A voice was produced using adobe audition. To express the lighter color in voice, the pitch was raised by 2.5 semitones from the original voice. To express darker color in voice, the pitch was lowered by 5 semitones from the original voice. To express vivid (i.e., highly intense) color in voice, the speed was stretched to 75% from the original voice and made it faster. To express warmer color in voice, a stereo-reflection plate was set among presets at full reverb and amplified +2.5 dB from the original voice. Finally, to express cool color in voice, "deserted room" was set among presets in full reverb. Two experimental artworks are illustrated to evaluate the usability of the appreciation method for the colors of artworks for the visually impaired that have been confirmed through such a test. Two versions of voice modulation were made and used according to preference: (1) the first version (called "word") with the corresponding voice modulation applied for each color word, and (2) the second version (called "phrase") in which the words contained in the same line related to the color are applied with the same voice modulation corresponding to the color.

3.5. Use Cases

Here, Vincent van Gogh's 1889 work 'The Starry Night' (Modern museum of art, NY) and Henri Matisse's 1910 work 'Dance' (Hermitage Museum, Saint Petersburg) are illustrated as use cases. 'Starry Night' conceptually and dynamically expresses the fluctuating air of an invisible night sky. Matisse used only four colors: blue, green, orange, and red. These vivid hues create an intense contrast. Both of these works well match the conceptual elements of color, so they are suitable for testing the usefulness of how to appreciate works of art using multiple senses. One of two poems per color were selected from about 2000 color poems in the "poem hunter website" [52], as shown in Tables 6 and 7. The table also shows the vocal narration for coding each color in words and tones.

Table 6. Poetry for representing colors in Van Gogh's "The starry night" (the sound files developed in this research are provided separately as a Supplemental Material).

Color	Poet	Part of Poem Excerptfrom the Original Poem	Modulated Voice	
			Word	**Phrase**
Cool dark blue sky and night	*P1:* Dark Blue Sky Peter S. Quinn [53]	Dark blue sky's out there, Before daybreak is in; Falling stars here and there, Night and distant moon, Sleepless is my night;	MP3 P1.mp3	MP3 P1.mp3
	P2: Blue on Blue Sandra Feldman [54]	The skies so blue, The Sky, the Ocean, sadly Blue. Upon a blue-lit tingling Star, That Love can travel to the moon, And much that's Blue is a Mirage.	MP3 P2.mp3	MP3 P2.mp3

Table 6. *Cont.*

Color	Poet	Part of Poem Excerptfrom the Original Poem	Modulated Voice	
			Word	**Phrase**
Light warm yellow star White star	*P3:* Yellow Star Vijay Sai [55]	A yellow star In a sky Twinkling white ones It's a yellow bird My beautiful little yellow bird!	MP3 P3.mp3	MP3 P3.mp3
Light Cool blue wind	*P4:* The wind has a blue tail Arjen Duinker [56]	The wind has a blue tail The wind also has a white tail	MP3 P4.mp3	MP3 P4.mp3
	P5: Blue, Blue and Blue Karun Aker [57]	Blue smells like my mother's perfume On a windy day	MP3 P5.mp3	MP3 P5.mp3
Warm orange moon	*P6:* Crushed Orange Moon Robert Murray Smith [58]	Crushed orange moon, Hold your color for the delight of night. When you crush orange light for our sight.	MP3 P6.mp3	MP3 P6.mp3
Cool dark Green (or Cool brown) tree	*P7*: Beneath the Dark Green Grove Megan Cooper [59]	I think of us beneath the trees, Beneath the dark green grove.	MP3 P7.mp3	MP3 P7.mp3
	P8: Color Poems; Brown Alex Safford [60]	Brown in the color of nature itself. Brown is the color of an oakwood shelf. Brown is the trunks of the trees	MP3 P8.mp3	MP3 P8.mp3
Cool green town	*P9*: Green: The Color Bri Edwards [61]	My favorite colors are green and brown. Green-leafed trees polka-dot our town.	MP3 P9.mp3	MP3 P9.mp3

Table 7. Poetry for representing colors in Henry Matisse's "Dance".

Color	Poet	Part of Poem Excerpt from the Original Poem	Modulated Voice	
			Word	**Phrase**
Blue sky	*P10*: My Imagination of a Purple Sky Premila Patel [62]	Cool, deep, fathomless a perfect shade of blue the green of Sicilian olives, sunsets orange dazzle,	MP3 P10.mp3	MP3 P10.mp3
Cool Dark Green Grass	*P11*: Green Syed Ali Sagar [63]	Green the optimist and the eye opener Green means clean and hope Glory, Blossom and Life	MP3 P11.mp3	MP3 P11.mp3
Warm orange Body	*P12*: Orange (Color Poem) [64]	Orange is the sunset its fire, fire is orange also the most vivid color God dared to make.	MP3 P12.mp3	MP3 P12.mp3
	P13: Orange is Blue Summer Song [65]	The sky is green and the grass is blue the sky is green and the grass is blue blue is orange and orange is blue	MP3 P13.mp3	MP3 P13.mp3

3.6. Implicit Association Test to Find a Solution to the Problem of ColorPoetry

The purpose of this test is to find a solution to the algorithm described in Section 3.2 for the instance of Table 6. To do so, an implicit association test was performed to identify the intimacy of the color stimulus in the poems (Table 6) and the one in the artwork ("The starry night" by Van Gogh) through the intervention of a semantic differential adjective antonym in Tables 4 and 5. Fifteen students were recruited as participants of the experiment. The gender of the participants was 5 males and 10 females, and the average age was 22.3 years (minimum 20 years old, maximum 27 years old). The number of participants with music and literature experience is 9 and 5, and the average number of years they have experienced is 5.5 years and 3.5 years (Figure 3). While participating in the experiment, repeated auditory stimulation may cause side effects such as headaches, and if physical or mental discomfort is felt, participants are informed in advance that they can request to stop the experiment at any time.

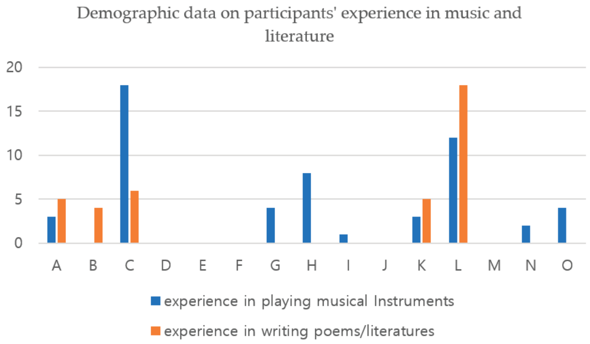


Figure 3. Demographic data on participants' experience in music and literature.

The participants first completed a training session that lasted approximately an hour, followed by an hour of evaluation. During the training session, the participants familiarized themselves with the concept behind the sound (modulated voices) and poem to be used to express various color appearance dimensions through tutorial materials, Tables 1–7, and Figures 1 and 2. The sound source and user testimonial questions to be used in the experiment were distributed just before the experiment began. After orientation, participants were presented with sound and voice clips shown in Tables 2, 3, 6 and 7. After that, the participants evaluated the implicit association and user experience.

Osgood [66] simplified the semantic space of relative adjectives into three aspects: evaluation, efficacy, and activity. The pair of semantic derivatives that were adopted in this experiment, including some adjectives in [66], are the pair of adjectives (Tables 4 and 5) people are familiar with for the light, dark, warm, and cold color of each hue collected through literature research [24]. In our experiments, participants are asked to experience multiple imagery sensations (see the definition of imagery in Section 2.1) like not only sight, but also, sound, touch, and smell from the artwork (Van Gogh's "The starry night") and poems in Table 6, respectively. Imagery with those four sensations is used for participants to appreciate the colors in artwork and poems. The sound is produced by means of voice narration of poems adjusting the intensity, pitch, and reverberation to distinguish warm/cool/vivid/light/dark dimensions of colors (see Section 3.4) so that the sound is matched with the characteristics of colors in artwork.

Test participants experienced two poem stimuli in Table 6 expressing each color stimulus in the artwork and the colors described in the poems they thought were associated with that stimulus. For each artwork or poem stimulus, there were provided to the participants a total of 14 concept adjectives pairs in Tables 4 and 5.

Table 8 illustrates an implicit association test questionnaire for the case of color stimuli "Blue sky and night" in artwork associated with Poem 1 and Poem 2. The following test guideline is given to the participants:

Table 8. An implicit association test questions for the case of color stimuli "Blue sky and night" in artwork associated with Poem 1 and Poem 2.

(a) **The Semantic Differential Pair Relevant to Warmer of Cooler Color of Each Hue**				
No	The Semantic Differential Pair	S(A)	S(P1)	S(P2)
1	Strong~Weak (sound)	-	-	-
2	Near~Far (thermal-tactile)	-	-	-
3	Warm~Cool (vision)	-	-	-
(b) **The Semantic Differential Pair Relevant to Lighter/Darker Color of Each Hue**				
1	Soft~Hard (haptic)	-	-	-
2	Round (Bouba)~Pointed (Kiki) (haptic)	-	-	-
3	Smooth~Rough (haptic)	-	-	-
4	Light~Heavy (haptic)	-	-	-
5	Light~Heavy (sound)	-	-	-
6	High~low (sound)	-	-	-
7	Aroused~Calm (vision)	-	-	-
8	Light~Dark (vision)	-	-	-
9	Pleasant~Unpleasant (scent)	-	-	-
10	Weak~Strong (scent)	-	-	-
11	Floral~Woody (scent)	-	-	-

"Based on your experience on each color in the artwork and two poems expressing the color today, check the box that reflects your immediate response to each adjective pair in the given table. Do not think too long to select a score among (2, 1, 0, −1, −2), and write it in each empty box. Make sure you respond to every adjective pair. Each of these 14 pairs of semantic differential adjectives, a visual feeling that is in context with the various sensations conveyed by the colors of art works or poem 1 and poem 2 is given 2 points for the closest one to the former, 1 point for the closer one to the former, −1 point for the closer to the latter, −2 points for the closest to the latter. Otherwise, if you don't know how to respond, simply score 0."

In this way, the similarity of Art–Poem Association with respect to colors denoted as S (A, P) for each of the two stimuli A and P is measured. Test results are shown in Figure 4. Then, the set of poems P having sigmax S (A, P) (as defined in Section 3.2) is selected to represent the color C in the artwork.

We measured cosine similarity between two vectors (like artwork A and poem P). The values of two vectors should be positive. Thus, the score range of (2, 1, 0, −1, −2) was replaced with (4, 3, 2, 1, 0). Since the similarity is measured by the cosine value of the angle formed by two vectors, the cosine similarity gets closer to 1 as the angle is smaller, and closer to 0 as the angle is larger. We obtained the similarity for Blue sky: S (A, P1) = 0.978 and S (A, P2) = 0.901, for Blue wind: S (A, P4) = 0.871 and S (A, P5) = 0.203, and for Green cypress: S (A, P7) = 0.501 and S (A, P8) = 0.569, respectively. According to the analysis based on cosine similarity, the three poems P1, P4, and P8 showed the most consistent dark blue, light blue, and dark green directivity respectively, among the poems used in the experiment settings. The poem depicting Van Gogh's Starry Night, created through the artwork–poetry matching process presented in this paper, is shown in Table 9.

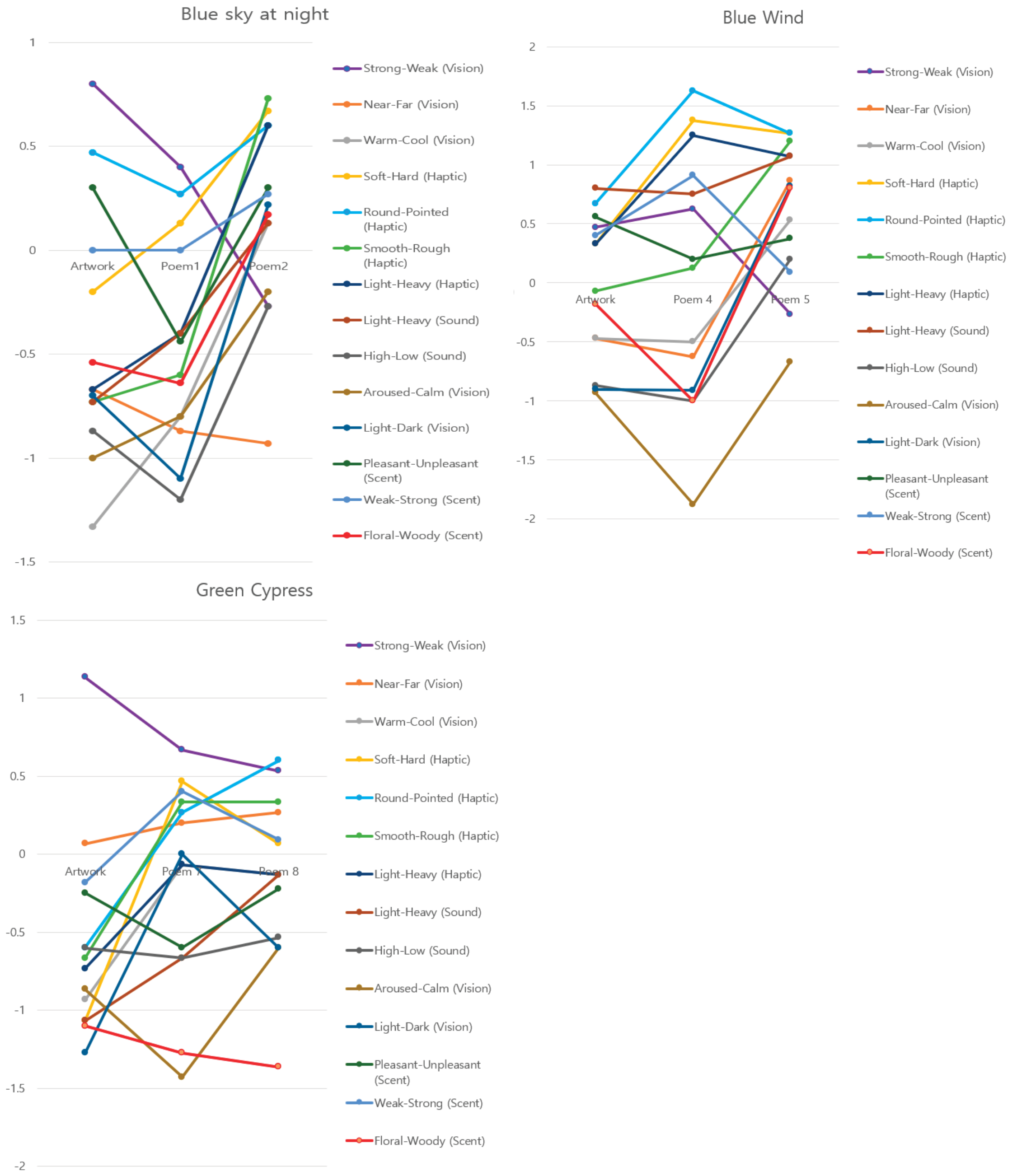


Figure 4. Comparison of response to semantic differential similarity between artwork and two poems with respect to various sensations (showing graphical representation to visualize the similarity between the artwork and two poems).

Table 9. Poetry for representing colors in Van Gogh's "The Starry Night" (the sound files developed in this research are provided separately as a Supplemental Material).

The Starry Night **Jun Dong Cho,** **2021**
poem-word.mp3 poem-phrase.mp3
Dark blue sky's out there, Before daybreak is in; Falling stars here and there, Night and distant moon, Sleepless is my night; [53]
A yellow star In a sky Twinkling white ones It's a yellow bird My beautiful little yellow bird! [55]
The wind has a blue tail The wind also has a white tail [56]
Crushed orange moon, Hold your color for the delight of night. When you crush orange light for our sight [58]
Brown in the color of nature itself. Brown is the color of an oakwood shelf. Brown is the trunks of the trees [59]
My favorite colors are green and brown. Green-leafed trees polka-dot our town [61]

Moreover, we performed a t-test with paired two sample for means. The average of adjective pairs was calculated for each experiment participant, and the difference between P1 and A and the difference between P2 and A was calculated, and the absolute value was taken. Respondents whose value is closer to 0 means that the difference between P and A is smaller. The differences between P1 and A and between P2 and A, which is the average of the larger one, were verified through a paired t-test. As a result of the verification, the verification statistics were $t\,(14) = -1.46$, $p = 0.17$ (two-tail verification), which could not be said to be statistically significant. The differences between P4 and A and between P5 and A, which is the average of the larger one, were verified through a paired t-test.

As a result of the verification, the verification statistics were $t\,(14) = -2.06$, $p = 0.03$ (two-tail verification), and there was a significant difference within the significance level of 5%. That is, P4 is statistically significant compared to P5 and is similar to A. The differences between P7 and A and between P8 and A, which was larger on average, were verified through a paired t-test. As a result of the verification, the verification statistics were $t\,(14) = -0.09$, $p = 0.46$ (two-tail verification), which could not be said to be statistically significant. According to these analyses, we reject P5 that showed the significant difference from P4. Therefore, finally, we could confirm that the set of poems (P1, P2, P4, P7, P8) are determined to be the set of "good" poems with the maximum significance of the Art–Poem similarity score with respect to each corresponding color stimulus.

As a result of the experiment conducted to find out the color directivity of each poem, the poem that showed the most consistent color directivity was P1. For P1, the adjectives chosen by the participants as they feel most suitable are cool (vision), calm (vision), far (vision), dark (vision), low (sound), strong (vision), rough (haptic), heavy (sound), woody (scent), round (haptic), and weak (scent), in the order of their preferences. Among them, the adjectives of the same directivity (positive or negative) in both artwork and P1 with having an intensity of 0.7 or more appear as cool (vision), calm (vision), far (vision), and dark (vision). It can be seen that the four adjectives matched well with the color characteristic of "blue sky at night". Therefore, we see that the poem P1 exhibits a wider variety of color appearance dimensions, including color temperature, color depth, and color emotion, as well as color lightness.

On the other hand, pleasant–unpleasant (scent) and soft–hard (haptic) were not such a case. Participants reacted "pleasantly" for artwork and unpleasantly for p1, showing the opposite directivity. In addition, participants responded with "hard" for artwork and "soft" for P1. However, their intensity of directivity was too weak, around 0.1~0.4. In addition, "weak–strong (scent)" was shown as "0" (random) in both artwork and P1.

4. Usability Test and Result

In the usability evaluation experiment, the SUS (System Usability Score) test was executed. During the SUS test, our proposed method "ColorPoetry" in this paper was compared with "ColorSound" [26]. Fifteen participants who also attended the implicit association test described in Section 3 were asked the following question: "Based on your experience on two color codes: ColorSound [26] and ColorPoetry *today*, score each of 10 items of SUS test set that reflects your immediate response to each statement. Don't think too long about each statement. Make sure you respond to every statement. If you strongly disagree then score 1, and if you strongly agree then score 5. Otherwise, if you don't know how to respond, simply score 3."

As a result of the analysis as shown in Figure 5, the ColorPoetry received good scores (74.5%), comparable to ColorSound (74.1%) in the user experience test (Figure 5). ColorSound was produced similarly to [17] for comparison purposes and was not discussed in detail during the training session, so it received a lower score than the score received in the previous study [26].

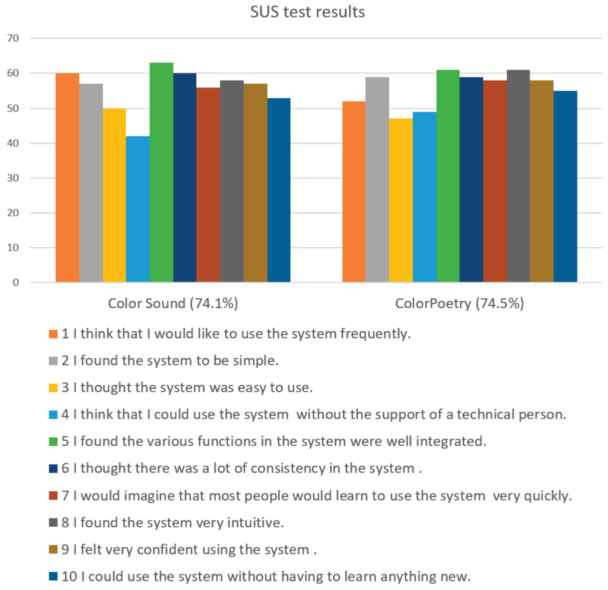


Figure 5. System Usability Scale test results.

Tables 10 and 11 list the participants' positive and negative feedback after reviewing the two color-coding systems, respectively. Most participants gave both positive and negative feedback. Two participants, C and L, who have more than 10 years of experience in playing musical instruments and 5 years of experiences in writing poems and literature, responded that "The system looks very efficient and very good to handle." On the contrary, in the negative evaluations other than the positive user feedback in Table 10, participants E and M, who do not have any experience in both music and literature, said, "It will be a little difficult if you don't understand the proper manual," and "If there is a simple explanation at first, it seems to be easy to use." The participant who provided negative feedback gave a low score on SUS statements 4 and 10 since they thought this system will need the support of a technical person and should learn while using this system.

Musical sound coding color is suitable for practical use because there are positive evaluations as a result of existing tests [26] when trying to convey colors by using instruments with characteristics that match the colors used in artworks with classical melodies. Representing the various color dimensions (e.g., distinguishing between warm, cool, light, and dark colors) itself into sound is a much more difficult and complex process than just using color names (with poetry, for example). The feeling that comes from a color is very subjective, so deciding which image to express with some effective sound or musical notes only is very difficult. Therefore, the current method of linking the characteristics of a musical instrument to the characteristics of color is objective and easy to access. Just as sighted people learn color names, blind people need to learn not only color names, but also musical codes that correspond to colors [26]. Table 12 shows conflicting user feedback and future works to resolve the conflict.

Table 10. Positive user feedbacks from the System Usability Scale test.

A	I felt the system's approach was very creative. I think two systems are easy to use and have a simple interface.
B	I think two systems are easy to use and have a simple interface. It does not require special technical knowledge, and I think it will be easy to use regardless of the individual characteristics of the user.
D	I think that it is a system that relies on human senses and feelings rather than going through thinking.
E	There seems to be no particular difficulty to use the system. I think they are common in everyday life, such as voices, colors, and poetry.
-	I think that it is a system that relies on human senses and feelings rather than going through thinking.
	I don't need any more knowledge to use the system, but I think that if you have background knowledge about music and poetry provided by the system, it will be of greater help in using the system.
F	It was easier to understand by making the sound different according to the color.
L	The system looks very efficient. This system was very good to handle.
M	It seemed simple to use because you can hear and feel it right away.
	At first, it was a little awkward to see how it works, but I think it will be useful afterwards

Table 11. Negative user feedback from the System Usability Scale test.

A	It was my first time using it, so I didn't feel confident about using the system because I was unfamiliar with it.
B	In the case of sound color, more diverse interpretations are possible depending on the user regardless of the matching between music and color. Therefore, it helps intuitive understanding that can be relative to each user. However, in the case of poetry, there may be various interpretations of the content for each user, and since the voice matched with poetry excludes such various possibilities, it is judged that it is difficult to feel intuitive from the user's point of view.
C	It seems that you should be able to learn how to use it.
E	It will be a little difficult if you don't understand the proper manual.
G	Poetry is okay because it is directly connected to color, but the voice with word is unnatural when savoring poetry. There is no big problem with both when it comes to integration.Poetry seems to be used a lot because the overall system is okay, but music is not. However, in the case of music, people who understand will use it a lot.
M	If there is a simple explanation at first, it seems to be easy to use.
O	When I first listened without any explanation, it was a little difficult to evaluate or use.

Table 12. Critical and conflicted user feedback and future works.

Conflicted User Feedbacks	Conflict Resolution (Future Works)
D: The sudden change of voice within a single paragraph seems to be unfamiliar yet. F: Having different voices with words could specifically indicate the meaning to be expressed. N: The voice with word units seems to have a more sophisticated feel. O: Applying the same voice to all words in a phrase that contains words related to color is more enjoyable and comfortable than applying a different voice to only words related to color.	These conflicted thoughts between participants can be mediated by using both methods of voice modulation, selectively considering the importance of delivering emotional and cognitive aspects of colors. Also, the voice delivery with modulation may be used limitedly, only for the key color words.
B: In the case of music provided in Color Sound, it provides a fairly structured service by varying the pitch and the instruments used, but I think the matching of poetry and color is quite random. G: The connection between musical instruments and colors is completely arbitrary without inevitable reasons. M: There was something I felt while listening to the song, but I don't think I felt that much about poetry.	Sound will be used for expressing color hues and poetry for expressing various color dimensions like warm, cool, light, dark, etc.

5. Discussion and Conclusions

Despite the availability of tactile graphics and audio guides, the visually impaired still face challenges in experiencing and understanding visual artworks. In previous works (as described in Section 2), musical melodies with different combinations of pitch, timbre, velocity, and tempo were used to distinguish vivid (i.e., saturated), light, and dark colors. However, it was rather difficult to distinguish among warm/cool/light/dark colors with using sound cues only. The way to use poems together when appreciating works has the advantage of enhancing the expressiveness of one work. These motivated us to develop a systematic algorithm to automate the generation of poetry that can be applied consistently to artworks, especially to help visually impaired users to perceive colors in the artworks. Therefore, in this paper, we presented a methodology to create poetry that matched well with a given artwork to ascertain whether a person with a visual impairment can interact with color through sound and poetry together without a complex learning process. Although several researchers have previously performed art–poetry matching, to the best of our knowledge, there is no art–poetry matching for the purpose of conveying different color dimensions such as warmth, coolness, lightness, and darkness. Moreover, there is no previous works that suggest a method of providing poems suitable for the various senses (sight, touch, hearing, smell) of the artwork.

In our experiments, an implicit association test was performed to identify the most suitable poem among the candidate poems to represent colors in artwork by finding the common semantic directivity between the given candidate poem with voice modulation and the artwork in terms of light/dark/warm/color dimensions. From the test, we found that the poem P1, for example, exhibited a wider variety of color appearance dimensions, including color temperature, color depth, and color emotion, as well as color lightness that matched well with "Blue sky and night" in Van Gogh's "The starry night".

Xu et al. [67] propose a memory-based neural model which exploits images to generate poems. Zhang et al. [68] presented a new image-driven poetry recommender system that takes a traveler's photo as input and recommends classical poems that can enrich the photo with aesthetically pleasing quotes from the poems. They developed a heterogeneous information network and neural embedding techniques. However, they did not take color matters with extensive implicit association tests into account. In this paper, a system usability test was also performed and user experience scores from 15 college student participants were 75.1%, which was comparable with the color–music coding system that received a user experience rating of 74.1%. After training three congenitally blind adults for about one hour, the recognition rate of 18 colors (6 hues and their 3 levels of lightness) using the color–music coding system [17] was 100%.

Even though the magic number 5 rule (Nielsen and Landauer, 1993) is vastly known and used for usability testing, the sample size is a long-running debate. Lamontagne et al. [69] investigated how many users are needed in usability testing to identify negative phenomena caused by a combination of the user interface and the usage context. They focused on identifying psychophysiological pain points (i.e., emotionally irritant experienced by the users) during a human–computer interaction. Fifteen subjects were tested in a new user training context and results show that out of the total psychophysiological pain points experienced by fifteen participants, 82% of them were experienced with nine participants.

Eye tracking studies take time. For qualitative eye tracking tests where recordings are manually reviewed, 5 users will suffice, but it is necessary to recruit at least 39 participants for meaningful heatmaps and other visualization that aggregate the actions of many users [70]. In the implicit association test done by Greenwald et al. [71], 32 (13 male and 19 female) students from introductory psychology courses at the University of Washington participated in exchange for an optional course credit.

Therefore, as a future work, we will further perform scaled experiments on people with visual impairment as a future work, along with experiments to find significant differences in perception of the various levels of the visually impaired for the proposed solution.

These studies can enhance the mental imagery experience of color using one or more modalities, such as sounds and poetry, presented in this paper. For practical application, we can identify a set of phrases in poems in a larger sized database (poemhunter.com) that best fit the color dimensions of a given piece of artwork using the same method presented in this paper. In addition, due to the nature of the auditory code, the usability of hearing is higher than the sense of touch and smell, so it can be used in art textbooks for the visually impaired, and it is easy to carry and distribute. When using the mobile phone's touch screen, colors (hue) can be expressed as sounds, and at the same time, other color dimensions like color temperature and lightness can be expressed by poems, vibrations, or odor, as described in the literature survey [24]. In other words, an integrated multi-sensory platform can convey color images effectively, taking advantage of temperature, vibration, and scent, as well as sound and poetry.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/electronics10091064/s1.

Author Contributions: Conceptualization, J.-D.C.; methodology, J.-D.C.; validation, J.-D.C.; formal analysis, J.-D.C.; investigation, J.-D.C.; resources, J.-D.C.; data curation, J.-D.C. and Y.L.; writing—original draft preparation, J.-D.C.; writing—review and editing, J.-D.C.; visualization, J.-D.C.; supervision, J.-D.C.; project administration and funding acquisition, J.-D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea, grant number 2018M3C1B6061353.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Sungkyunkwan University (protocol code: 2020-11-005-001, 22 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors are grateful to Young Jun Sah, who helped in statistical analysis.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Calvert, G.A.; Spence, C.; Stein, B.E. *The Handbook of Multisensory Processes*; MIT Press: London, UK, 2004.
2. Haverkamp, M. *Synesthetic Design: Handbook for a Multi-Sensory Approach*; Walter de Gruyter: Berlin, Germany, 2012; p. 139.
3. Cottin, M.; Faria, R.; Amado, E. *The Black Book of Colors*; Groundwood Books Ltd.: Toronto, ON, Canada, 2008.
4. Vinter, A.; Orlandi, O.; Morgan, P. Identification of Textured Tactile Pictures in Visually Impaired and Blindfolded Sighted Children. *Front. Psychol.* **2020**, *11*, 345. [CrossRef] [PubMed]
5. Damásio, A. *O Erro De Descartes: Emoção, Razão E O Cérebro Humano*; Editora Companhia das Letras: São Paulo, Brazil, 2012.
6. Damasio, A. Feeling & knowing: Making minds conscious. *Cogn. Neurosci.* **2020**, *12*, 1–2.
7. Neves, J. Multi-sensory approaches to (audio) describing the visual arts. MonTI. *Monogr. Traducción Interpret.* **2012**, *4*, 277–293. [CrossRef]
8. Rector, K.; Salmon, K.; Thornton, D.; Joshi, N.; Morris, M.R. Eyes-free art: Exploring proxemic audio interfaces for blind and low vision art engagement. *Proc. Acm Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–21. [CrossRef]
9. Schifferstein, H.N.J. Comparing Mental Imagery across the Sensory Modalities. *Imagin. Cogn. Pers.* **2009**, *28*, 371–388. [CrossRef]
10. Literary Devices. Imagery. 2021. Available online: https://literarydevices.net/imagery/ (accessed on 29 March 2021).
11. Trent, A. What Is Synesthesia in Poetry? Available online: https://penandthepad.com/synesthesia-poetry-1824.html (accessed on 29 March 2021).
12. Betts, J. Synesthesia Examples in Literature and Poetry. Available online: https://examples.yourdictionary.com/synesthesia-examples-in-literature-and-poetry.html (accessed on 29 March 2021).
13. Jansson, M. Ekphrasis and digital media. *Poet. Today* **2018**, *39*, 299–318. [CrossRef]
14. Wagner, P. Introduction: Ekphrasis, Iconotexts, and Intermediality–the State (s) of the Art (s). In *Icons-Texts-Iconotexts*; de Gruyter: Berlin, Germany, 2012; pp. 1–40.
15. De Armas, F.A. Visual Culture: Art and Ekphrasis in Early Modern Spain. In *A Companion to the Spanish Renaissance*; Brill: Leiden, The Netherlands, 2018; pp. 450–470.
16. Hirsch, E. *Transforming Vision: Writers on Art*; Little Brown and the Art Institute of Chicago: New York, NY, USA, 1994.
17. Yu, W.; Xiaojie, L. On the Expressions in Color in the Poems of William Carlos on the Expressions in Color in the Poems of William Carlos Williams. *Foreign Lit. Stud.* **2011**, *33*, 113–120.
18. Koch, K. *Rose, Where Did You Get That Red?* Random House: New York, NY, USA, 1973.
19. Routman, R. *Kids' Poems: Teaching Third and Fourth Graders to Love Writing Poetry*; Scholastic: New York, NY, USA, 2000.
20. Flanders, J. Van Gogh's Bed. 2015. Available online: https://www.poemhunter.com/poem/van-gogh-s-bed/ (accessed on 4 March 2021).
21. Sexton, A. *"The Starry Night" from the Complete Poems of Anne Sexton*; Houghton Mifflin: Boston, MA, USA, 1981; pp. 53–54.
22. Kennedy, X.J. *Nude Descending a Staircase*; Doubleday, Incorporated: New York, NY, USA, 1961.
23. Hepner, G. White, Red on Yellow. 2008. Available online: https://www.poemhunter.com/poem/white-red-on-yellow/ (accessed on 3 March 2021).

24. Cho, J.D. A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment. *Electronics* **2021**, *10*, 470. [CrossRef]
25. Cho, J.D.; Quero, L.C.; Bartolomé, J.I.; Lee, D.W.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2021**, *46*, 103–116. [CrossRef]
26. Cho, J.D.; Jeong, J.; Kim, J.H.; Lee, H. Sound Coding Color to Improve Artwork Appreciation by People with Visual Im-pairments. *Electronics* **2020**, *9*, 1981. [CrossRef]
27. Iranzo Bartolomé, J.; Cho, J.D.; Cavazos Quero, L.; Jo, S.; Cho, G. Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People. *Electronics* **2020**, *9*, 1939. [CrossRef]
28. Anikin, A.; Johansson, N. Implicit associations between individual properties of color and sound. *Atten. Percept. Psychophys.* **2019**, *81*, 764–777. [CrossRef] [PubMed]
29. Doppler Effect. Available online: https://science.jrank.org/pages/2147/Doppler-Effect-Doppler-effect-in-sound-waves.html#ixzz6qHpjElX0 (accessed on 29 March 2021).
30. Jonas, C.; Spiller, M.J.; Hibbard, P. Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychon. Bull. Rev.* **2017**, *24*, 1104–1112. [CrossRef] [PubMed]
31. Cogan, R.D. *Sonic Design: The Nature of Sound and Music*; Prentice Hall: Upper Saddle River, NJ, USA, 1976.
32. Kandinsky, V. *Concerning the Spiritual in Art*; Dover Publications: Mineola, NY, USA, 1977.
33. Hamilton-Fletcher, G.; Witzel, C.; Reby, D.; Ward, J. Sound properties associated with equiluminant colors. *Multisens. Res.* **2017**, *30*, 337–362. [CrossRef]
34. Giannakis, K. Sound Mosaics: A Graphical User Interface for Sound Synthesis Based on Audio-Visual Associations. Ph.D. Thesis, Middlesex University, London, UK, 2001.
35. Marks, L.E.; Hammeal, R.J.; Bornstein, H.M. Perceiving similarity and comprehending metaphor. *Monogr. Soc. Res. Child Dev.* **1987**, *52*, 1–102. [CrossRef] [PubMed]
36. Wilms, L.; Oberfeld, D. Color and emotion: Effects of hue, saturation, and brightness. *Psychol. Res.* **2018**, *82*, 896–914. [CrossRef]
37. Rowe, A. Audio Dictionary: "Warm" vs. "Neutral" vs. "Bright". Available online: https://xander51.medium.com/audio-dictionary-warm-vs-neutral-vs-bright-and-the-role-your-brain-plays-in-all-this-2ec761759f7b (accessed on 29 March 2021).
38. Domino, G. Synesthesia and creativity in fine arts students: An empirical look. *Creat. Res. J.* **1989**, *2*, 17–29. [CrossRef]
39. Cytowic, R.E. *Synesthesia: A Union of the Senses*; MIT Press: Cambridge, MA, USA, 2002.
40. Cytowic, R.E. *Synesthesia*; MIT Press: Cambridge, MA, USA, 2018; ISBN 9780262535090.
41. Maxwell, N.F. "Visualizing Sound: Effects of Pitch Height and Tonality on Luminance Matching. Inquiries J. Student Pulse 2013, 5(10). Available online: http://www.inquiriesjournal.com/a?id=767 (accessed on 28 April 2021).
42. Goller, A.I.; Otten, L.J.; Ward, J. Seeing sounds and hearing colors: An event-related potential study of auditory-visual syn-esthesia. *J. Cogn. Neurosci.* **2009**, *21*, 1869–1881. [CrossRef] [PubMed]
43. Niccolai, V.; Jennes, J.; Stoerig, P.; Van Leeuwen, T.M. Modality and variability of synesthetic experience. *Am. J. Psychol.* **2012**, *125*, 81–94. [CrossRef]
44. Brigg, D. The Dimension of Color. Available online: http://www.huevaluechroma.com/index.php (accessed on 29 March 2021).
45. Artfactory Studio. Two Steps Forward, One Step Back: Colors that Advance and Recede. 2021. Available online: https://artissima.wordpress.com/tag/temperature/ (accessed on 29 March 2021).
46. Ludwig, V.U.; Simner, J. What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex* **2013**, *49*, 1089–1099. [CrossRef] [PubMed]
47. Slobodenyuk, N.; Jraissati, Y.; Kanso, A.; Ghanem, L.; Elhajj, I. Cross-modal associations between color and haptics. *Atten. Percept. Psychophys.* **2015**, *77*, 1379–1395. [CrossRef] [PubMed]
48. Maric, Y.; Jacquot, M. Contribution to understanding odour–colour associations. *Food Qual. Prefer.* **2013**, *27*, 191–195. [CrossRef]
49. Kemp, S.E.; Gilbert, A.N. Odor intensity and color lightness are correlated sensory dimensions. *Am. J. Psychol.* **1997**, *110*, 35. [CrossRef] [PubMed]
50. Fior, A.M. Multisensory integration of visual, tactile, and olfactory aesthetic cues of appearance. *Cloth. Text. Res. J.* **1993**, *11*, 45–52. [CrossRef]
51. Kim, Y.J. Can eyes smell? Cross-modal correspondences between color hue-tone and fragrance family. *Color Res. Appl.* **2013**, *38*, 139–156. [CrossRef]
52. Hunter, P. Available online: https://www.poemhunter.com/poems/purple/ (accessed on 29 March 2021).
53. Quinn, P.S. Dark Blue Sky. 2007. Available online: https://www.poemhunter.com/poems/blue-sky/page-2/6741601/ (accessed on 3 March 2021).
54. Feldman, S. Blue on Blue. 2013. Available online: https://www.poemhunter.com/poems/Blue/page-2/36448079/ (accessed on 4 March 2021).
55. Vijay, S.R. Yellow Star 2010. Available online: https://www.poemhunter.com/poem/yellow-star/ (accessed on 4 March 2021).
56. Duinker, A. The Wind Has a Blue Tail. 1998. Available online: https://www.poetryinternational.org/pi/poem/4137/auto/0/0/Arjen-Duinker/The-wind-has-a-blue-tail/en/tile (accessed on 4 March 2021).
57. Aker, K. Blue, Blue and Blue. 2005. Available online: https://www.poemhunter.com/poem/blue-blue-and-blue/ (accessed on 4 March 2021).

58. Smith, R.M. Crushed Orange Moon. 2020. Available online: https://www.poemhunter.com/poem/crushed-orange-moon-3/ (accessed on 4 March 2021).
59. Cooper, M. Beneath the Dark Green Grove. 2008. Available online: https://www.poemhunter.com/poems/Green/page-2/10236520/ (accessed on 4 March 2021).
60. Safford, A. Color Poems; Brown. 1980. Available online: https://www.csmonitor.com/1980/0310/031059.html (accessed on 4 March 2021).
61. Edwards, B. Green: The Color 2013. Available online: https://www.poemhunter.com/poem/green-the-color-green-things/ (accessed on 4 March 2021).
62. Patel, P. My Imagination of a Purple Sky. 2009. Available online: https://www.poemhunter.com/poem/my-imagination-of-a-purple-sky/ (accessed on 4 March 2021).
63. Sagar, S.A. Green 2007. Available online: https://www.poemhunter.com/poem/green-21/ (accessed on 4 March 2021).
64. Pd. Is Here, Orange (Color Poem). 2006. Available online: https://www.poemhunter.com/poem/orange-colour-poem/ (accessed on 4 March 2021).
65. Song, S. Orange Is Blue. 2006. Available online: https://www.poemhunter.com/poems/Blue-7/page-7/4284787/ (accessed on 4 March 2021).
66. Osgood, C.E.; Suci, G.J.; Tannenbaum, P.H. *The Measurement of Meaning*; University of Illinois Press: Champaign, IL, USA, 1957.
67. Xu, L.; Jiang, L.; Qin, C.; Wang, Z.; Du, D. How images inspire poems: Generating classical Chinese poetry from images with memory networks. In Proceedings of the 32 AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
68. Zhang, D.Y.; Ni, B.; Zhi, Q.; Plummer, T.; Li, Q.; Zheng, H.; Wang, D. Through the eyes of a poet: Classical poetry recom-mendation with visual input on social media. In *2019 IEEE/ACM International Conference on Advances in Social Networks Anal-ysis and Mining (ASONAM)*; IEEE: Vancouver, BC, Canada, 2019; pp. 333–340.
69. Lamontagne, C.; Sénécal, S.; Fredette, M.; Chen, S.L.; Pourchon, R.; Gaumont, Y.; Léger, P.M. User Test: How Many Users Are Needed to Find the Psychophysiological Pain Points in a Journey Map. In *International Conference on Human Interaction and Emerging Technologies*; Springer: Cham, Switzerlands, 2019; pp. 136–142.
70. Dawson, L. Eye Tracking: What Is It for and When to Use It. Available online: https://usabilitygeek.com/what-is-eye-tracking-when-to-use-it/ (accessed on 9 April 2021).
71. Greenwald, A.G.; McGhee, D.E.; Schwartz, J.L. Measuring individual differences in implicit cognition: The implicit associa-tion test. *J. Personal. Soc. Psychol.* **1998**, *74*, 1464. [CrossRef]

electronics

Article

Construction of a Soundscape-Based Media Art Exhibition to Improve User Appreciation Experience by Using Deep Neural Networks

Youngjun Kim, Hayoung Jeong, Jun-Dong Cho and Jitae Shin *

College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, Korea; yjk931004@skku.edu (Y.K.); lmo3088@skku.edu (H.J.); jdcho@skku.edu (J.-D.C.)
* Correspondence: jtshin@skku.edu

Abstract: The objective of this study was to improve user experience when appreciating visual artworks with soundscape music chosen by a deep neural network based on weakly supervised learning. We also propose a multi-faceted approach to measuring ambiguous concepts, such as the subjective fitness, implicit senses, immersion, and availability. We showed improvements in appreciation experience, such as the metaphorical and psychological transferability, time distortion, and cognitive absorption, with in-depth experiments involving 70 participants. Our test results were similar to those of "Bunker de Lumières: van Gogh", which is an immersive media artwork directed by Gianfranco lannuzzi; the fitness scores of our system and "Bunker de Lumières: van Gogh" were 3.68/5 and 3.81/5, respectively. Moreover, the concordance of implicit senses between artworks and classical music was measured to be 0.88%, and the time distortion and cognitive absorption improved during the immersion. Finally, the proposed method obtained a subjective satisfaction score of 3.53/5 in the evaluation of its usability. Our proposed method can also help spread soundscape-based media art by supporting traditional soundscape design. Furthermore, we hope that our proposed method will help people with visual impairments to appreciate artworks through its application to a multi-modal media art guide platform.

Keywords: music recommendation system; multimedia data processing; weakly supervised learning; soundscape music; media art; exhibition environments; user experience

Citation: Kim, Y.; Jeong, H.; Cho, J.-D.; Shin, J. Construction of a Soundscape-Based Media Art Exhibition to Improve User Appreciation Experience by Using Deep Neural Networks. *Electronics* **2021**, *10*, 1170. https://doi.org/10.3390/electronics10101170

Academic Editors: Juan M. Corchado, George A. Tsihrintzis and Amir Mosavi

Received: 23 March 2021
Accepted: 7 May 2021
Published: 14 May 2021

1. Introduction

A soundscape is a contextual acoustic environment perceived by humans. Initially, this concept was used as a compositional approach in order to provide a sense of space via the recording and rearrangement of noise, including natural and environmental elements. This role of the soundscape is used to help an audience imagine its space [1,2]. Recently, the concept of the soundscape has been used in a variety of fields, such as architecture, art, and education [3–6]. In particular, some research [7] has proposed art-based spatial design by presenting the concept of shopping galleries through a combination of shopping spaces and soundscape music. The concept of the soundscape in art education has also been explored [8]. Recently, IT technology [9,10] has been used to improve user experience in map applications through data-based soundscape construction methods.

In this study, we applied the soundscape concept to an art exhibition environment to improve the artwork appreciation experience. At this point, music is an important component of soundscape-based exhibition environments because it can interfere with or help an audience appreciate the experience. In previous studies, soundscapes were designed by experts by selecting or composing music containing a message that the experts wanted to convey. However, this approach is expensive in terms of time and effort. In this study, we investigated the replacement of this traditional approach to soundscape construction with deep-neural-network-based methods. Our requirement was that the soundscape constructed using our method should provide as impactful of an experience as that of the expert's choice. If our requirements are met, our approach could spread the abundance soundscape-based media art at a lower cost than that which was previously possible.

The purpose of this study was to improve users' artwork appreciation experience through auditory cues, such as classical music, in order to provide an abundant media art exhibition environment. These multi-modal sense-based exhibits [11–14] not only provide the user with a more impressive, realistic, and immersive experience, but also have potential cognitive and emotional impacts on the appreciator [15–17]. Thus, we propose a soundscape-based methodology that uses deep neural networks to identify music associated with a given visual artwork through multi-modal data processing based on weakly supervised learning. With a multi-faceted approach, we measure whether our system can recommend music that can have an impact on the user. This soundscape-based media art will improve users' experience of appreciating artworks through metaphorical and psychological interactions as well as through the direct and material appreciation of media art. Our method of soundscape design using deep learning can help disseminate media art exhibitions by reinterpreting them as immersive media. Furthermore, we hope that our proposed method will help people with visual impairments to appreciate artworks through its application to a multi-modal media art platform. The contributions of this study are described below.

Contributions

1. We developed a system for matching classical music and paintings using the concept of the soundscape through an improved deep neural network based on weakly supervised learning.
2. We propose a multi-faceted approach to measuring ambiguous concepts, such as the subjective fitness, implicit senses, immersion, and availability; then, an interdisciplinary discussion is also provided.
3. We show the improvements in the appreciation experience, such as metaphorical and psychological transferability, time distortion, and cognitive absorption, with in-depth experiments involving 70 participants.

2. Related Studies

2.1. Multi-Modal Artwork Platform for People with Visual Impairments

Modern information systems rely on vision, resulting in differences in information gain between visually impaired people and non-visually impaired people. Technological developments have improved this situation [18,19], but access to cultural content beyond daily life remains challenging. In particular, persons with disabilities should have opportunities provided for cultural, physical, and artistic activities based on the 2008 Convention on the Rights of Persons with Disabilities [20]. Nevertheless, this has not been the case from the perspective of arts and culture, as pointed out by Kim Hyung-sik, a former member of the Committee on the Rights of Persons with Disabilities (CRPD) in the United Nations [21].

The blind touch project [22] was launched by ratifying Article 30 (2) of the Convention on the Rights of Persons with Disabilities, which states that "Parties shall take appropriate measures to enable persons with disabilities to have the opportunity to develop and utilize their creative, artistic, and intellectual potential, not only for their own benefit, but also for the enrichment of society". This project aimed to improve the artwork appreciation environment for people with visual impairments [23,24]. The blind touch project had two main objectives. The first objective was to allow blind people to experience, understand, and interpret art through various multi-modal senses, such as hearing, touch, temperature, and texture. The second objective was to develop a framework and technologies that would allow visually impaired people not only to experience art through their senses, but also to understand, interpret, and reflect upon it.

In a blind touch project, Cavazos et al. [25] developed a multi-modal artwork guide platform (see Table 1) that transformed an existing 2D visual artwork into a 2.5D (relief form) replica using 3D printing technology, making it accessible through touch, audio descriptions, and sound in order to provide a high level of user experience. Thus, visually impaired individuals could enjoy this artwork freely, independently, and comfortably through touch and sound without the need for a professional commentator. In addition, gamification concepts [26] were included to awaken various other non-visual senses and maximize enjoyment of an artwork. For example, vivid visual descriptions and sound effects were provided to maximize the sense of immersion in appreciating artworks. Such recreated artworks with multi-modal guides facilitated user-friendly interaction environments by sensing the event of tactile input on some part of an artwork and providing the related information. In addition, background music was created to elicit emotions similar to those of the work, taking into consideration the musical instrument's timbre, minor/major mode, tempo, and pitch. In this paper, we aimed to replace the background music with other classical music recommended by deep neural networks and soundscape concepts.

Table 1. Interactive multi-modal guideline for appreciating visual artworks and museum objects [25].

Type	Description
Sensing technology	Capacitive sensor connected to conductive ink-based sensors embedded under the surface of the model.
Input	Double-tap and triple-tap gestures on the surface.
Tactile presentation	Tactile bas-relief model
Output	Audio descriptions, sound effects, and background music
Objective	Improve visual artwork exploration

2.2. Soundscape Construction Using Deep Neural Networks

In this study, we constructed soundscapes based on music that matched well with a given artwork by using deep neural networks. We considered three technical approaches to constructing soundscapes. The first is a generative-model-based approach, in which generative model is used to translate a painting into music based on deep neural networks [27]. This proposed method is characterized by the use of consistent features that allow interconversion between music and painting. However, this method of using consistent and interchangeable features does not ensure well-matched music. Therefore, in this work, we did not adopt this method because we felt that it did not produce music that fits paintings well. Rather, we used the approach of finding music that matches the painting rather than using a generative models.

The second is an approach based on music recommendation systems. The recommendation systems were divided into user-based, content-based, and hybrid-based methods [28,29]. We focused on content-based recommendation systems [30] because our purpose was a recommendation between music and painting, not personalized recommendations. The key to this approach is the vectorization of content because it matches based on the similarities in vectorized content. Examples of vectorization methods include video and description summarization [31,32] and image captioning [33–35]. In particular, the authors of [36] matched poetry and images through captioning, the authors of [33] presented an automatic caption generation method for Impressionist artworks for people with visual impairments, and the authors of [28] used emotional features in music recommendations. However, when applying these methods to our research, not only was a capping module required for the music and images, but a user-based recommendation system was also needed.

Thus, we selected baseline networks as a third method [10,37–41] for feature matching by using kernel density estimation based on weakly supervised learning. Our training baseline was soundnet [37], in which a kernel mapped an audio sample space to an image sample space via weakly supervised learning for vectorization. In prior imaginary soundscape research [10], an application based on soundnet was used to improve users' experience with Google Maps. This application mapped vectors of images and audio into the same sample space by using soundnet. Our approach uses a similar framework. This approach has the limitation of not being able to use audio or image descriptions, but it has the advantage of being able to focus on natural features. In the following sections, we describe our audio feature extraction method and knowledge distillation method.

2.3. Audio Feature Extraction

Audio feature extraction is a major component of soundnet frameworks. There are three major methods. The first deals with audio data as a spectrogram image by using a short-term Fourier transform [42–45]. The second is an end-to-end method for dealing with raw audio data by using a shallow and wide raw feature extractor [46,47] rather than a Fourier transform. The third one uses improved learning techniques, such as data argumentation [47–49], pre-processing (or post-processing) [50,51], and other learning methods, such self (or weakly) supervised learning [10,37–41,47,51].

Spectrogram-based audio feature extraction depends on the hyper-parameters of the short-term Fourier transform. Thus, many state-of-the-art networks [37,47,51] have been studied by using raw feature extractors; however, we adopted a spectrogram-based approach because it has advantages in terms of knowledge distillation and its application to our domain. In particular, we selected the WaveMsNet [45] as our baseline feature extractor. The WaveMsNet fuses features of time and frequency domains to improve the dependence of the Fourier transformation on the hyper-parameter window size via multi-scale feature extraction in the time domain. However, the network still receives the gray channel as the input. In this paper, we propose a multi-time-scale transform [52,53] to convert audio data into an RGB image in order to receive RGB input instead of gray input. These techniques not only improve the receptive field, but also enable direct measurement of feature distance. In the next sections, we shall address knowledge distillation for mapping audio features and image features into the same sample space.

2.4. Knowledge Distillation

Knowledge distillation began from mimic model [54], and is a method to learn differences among distributions for model compression. This method has therefore been used in various fields related to distribution learning, such as domain adaptation [55] and knowledge transfer [56]. This method is also used in soundnet frameworks as a learning method. There are two main considerations when applying knowledge distillation. The first is the choice of the knowledge to learn, such as score maps [57], feature maps [58], attention maps [59], Jacobian matrixes [60], or decision boundaries [61]. The second is how the distilled knowledge is transferred, such as through mutual learning [62], knowledge projection [63], or teacher assistance [64].

In this study, we selected two research studies as baseline studies. The first was that of fitnet [58], which conducted knowledge distillation based on a feature map with abundant information. A previous fitnet study solved the problem of size mismatch between feature maps via a regressor, which problem is to transfer of knowledge from wide features to narrow features Later, an attention study [59] showed that learning without an additional regressor was possible by changing the learning structure from deep to shallow. In this paper, we conducted knowledge transfer through direct measurement of feature distance without an additional regressor by setting the same model structure for transferring from wide and deep features to wide and deep features. Our learning method is based on a deep mutual learning strategy [62] with symmetric Kullback–Leibler divergence. Furthermore, the strategy used by the learning method allows the audio feature extractors to be configured in the same way as the image feature extractors. We chose this learning method because it aims to reduce the difference between the two sample spaces rather than to increase the accuracy of specific tasks.

3. Proposed Architecture and Learning Method for Constructing a Soundscape

In this section, we deal with the proposed architecture of our application and present a learning method for constructing a soundscape in four subsections—namely, music–artwork matching for the soundscape, a training phase, a domain adaptation phase, and a multi-time-scale transform for audio feature extraction.

3.1. Music–Artwork Matching for the Soundscape

Figure 1 shows the architecture of our service application. In the preprocessing phase, music is converted into RGB images via a multi-time-scale transform. Multi-time-scale transform and audio feature extraction are performed, and the resulting data are stored in a database. Later, when a painting is entered through a service application, the application matches and recommends the nearest n music pieces that match well with the audio features stored in the database. The distance is measured by the cross-entropy. A feature extractor for the audio and images was constructed by using a wide resnet 101 with double width. For audio with a standard sample rate (44.1 kHz), the extracted features are stored as a JSON object. The stored JSON object had about 3.48 times more capacity than the existing audio because of the feature characteristics, such as the high resolution and multiple channels. We did not use any additional compression methods. Our music database consisted of 2000 items of classical music stored in the form of key values, and it took about 2.5 days for the database to be configured without parallel processing. In addition, a total inspection was conducted via cross-entropy without an additional search algorithm for the music–painting matching, which took about 3.2 h. Our device settings were as follows: CPU: Intel® Core™ i7-8700K processor, GPU: 2 RTX 2080 Ti. However, because we used unoptimized code, the technical issues described above are likely to be improved in real-world service via optimization. In the next section, we describe how the deep neural networks were trained.

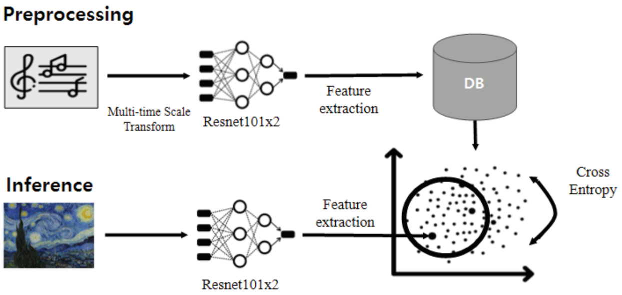


Figure 1. Our service application architecture, consisting of a preprocessing phase before the service and an inference phase at runtime. This figure shows a music–artwork matching method for soundscape construction.

3.2. Training Phase

Figure 2 shows our learning framework, which is similar to the soundnet framework. However, we additionally used methods such as a pre-trained model with shared features, an improved audio feature extractor, and improved weakly supervised learning. In the pre-training phase, a multi-domain convolutional neural network (CNN) was applied for sharing features. In the original soundnet framework, the feature extractor is based on two models—namely, the object and scene of the feature extractor for each pre-training phase and weakly supervised learning phase. However, the data cannot be utilized efficiently. In soundnet, the object feature extractor is learned in Imagenet, while the scene feature extractor is learned in Place365, which results in less data than Imagenet. Thus, it is difficult to determine if the feature extractor has learned enough. To address this issue, we applied a multi-domain CNN because the place and object can share features when using this method. The feature extractor used was the wide resnet101 with double width and with two classifier headers for the object and scene. The Imagenet and Place365 data were trained in one model so that the features could be shared. This method also had advantages in the training phase.

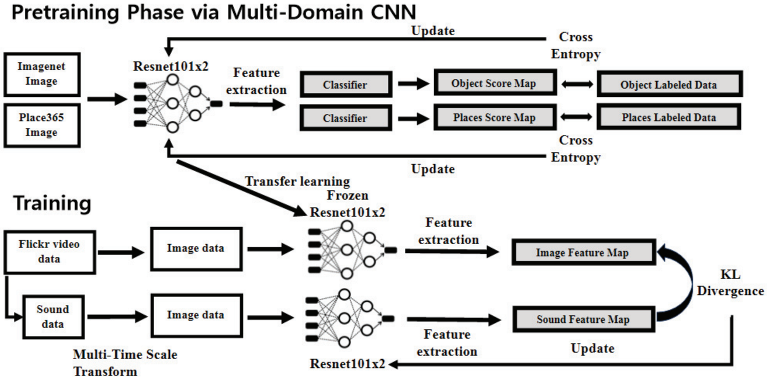


Figure 2. The training phase was divided into a pre-training phase and a training phase. This learning framework is similar to that of the soundnet framework; however, the pre-trained model with features shared via a multi-domain CNN, audio feature extraction via multi-time-scale transform, and weakly supervised learning via mutual learning are advantages of our framework.

In this training phase, we trained our network by using a Flickr video dataset [37] for cross-modal recognition. The video dataset was viewed as weakly labeled data with images assigned to the audio. Image features were extracted from a pre-trained model that was frozen. This frozen model meant that the results of learning were not reflected; they were used only for feature extraction. Audio data were then converted into RGB images via the multi-time-scale transform, which allowed feature extraction via the wide resnet with double width. These extracted features were used as the sources and targets for audio features and image features, respectively. The deep neural network was trained with a kernel function to map the source to a target, which meant that learning to extract features was similar for the audio and image data. Only the audio feature extractor reflected the learning results because our purpose was to approximate from the source to the target. Unlike soundnet, in the KL divergence, the source and target were configured as feature maps, not score maps. This is the same method as that used in the fitnet; however, our training frameworks conducted direct distribution training without an additional regressor through use of the same feature extractor structure. Furthermore, because the multi-domain CNN was applied in the pre-trained model, we also performed inter-distribution learning on one integrated model through only feature sharing.

3.3. Domain Adaptation

In the training phase, a network was trained to match audio with objects and scenes. In the domain adaptation phase, we developed a method to train our network to match music and a painting, which is challenging because of the absence of a related dataset. We therefore focused on appealing advertisements in our dataset. The purpose of emotionally appealing advertisements is to transfer emotional feelings to the customer rather than rational information. A new paradigm of emotionally appealing product advertisements has recently emerged, where the concept is not to focus on revealing production, but to convey a brand image and overall atmosphere. These advertisements convey a sense of artistry and atmosphere that transcend the boundaries of usability and beauty by using well-matched colors and atmospheric music. Thus, the video of an advertisement can be considered as an expert's well-matched labeling data. In this work, we matched paintings and music by using the color and the audio atmosphere rather than simply matching audio to objects or scenes. Therefore, we used advertisements for products. Examples of the emotionally appealing advertisements used are provided in Figure 3. In this work, data were collected manually. The collection criteria were as follows: The first was whether background music was included; the second was whether the advertisements emphasized color; the third was whether there was conversation. If conversations were frequent, the advertising was excluded from the collection. If several atmospheric parts existed in the collected advertisement, the video was divided into several videos according to the sections where the atmosphere changed. For example, in Figure 3, each row is from the same advertisement. This advertisement was divided based on the points at which the atmosphere changed.

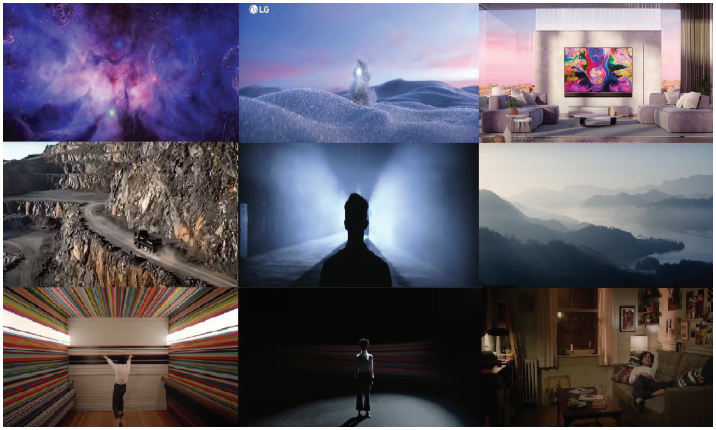


Figure 3. Dataset of emotionally appealing advertisements for domain adaption.

Figure 4 shows our domain adaptation method with its key features of mutual learning through advertisement data with symmetric *KL* divergence. The purpose of mutual learning was not to increase the accuracy of the task, but to train the two features to be the same. Therefore, we used advertising videos to fine-tune the network through mutual learning with symmetric *KL* divergence for domain adaptation. This domain adaptation had two main objectives. First, we wanted our system to learn a method of matching sounds and images based on considerations of the color and the atmosphere. Second, we focused on extracting these two features equally, rather than matching the sound to the space of the image via symmetric *KL* divergence.

$$D_{SymmetricKL}(P,Q) = D_{KL}(P \parallel Q) + D_{KL}(Q \parallel P) = \textstyle\sum_{x\in\chi} P(x)\ln\frac{P(x)}{Q(x)} + \sum_{x\in\chi} Q(x)\ln\frac{Q(x)}{P(x)}. \quad (1)$$

Equation (1) is the symmetric *KL* divergence for the mutual learning used in this study. The objective of mutual learning is to train model so that image and audio features can be extracted equally; the image is indicated by P and the audio by Q. Importantly, unlike in the training phase, a model freeze was not used in order to reduce the gap between the distributions of the two features.

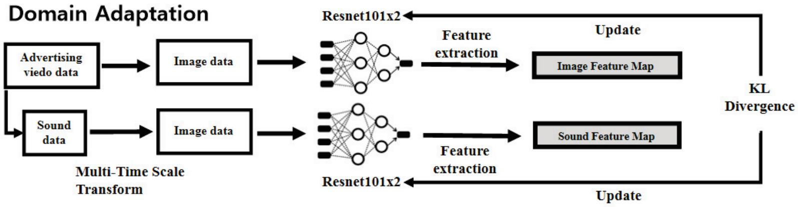


Figure 4. Domain adaptation phase with symmetric KL divergence based on mutual learning.

3.4. Audio Feature Extraction via Multi-Time-Scale Transform

Algorithm 1 is for audio feature extraction via the multi-time-scale transform in order to extract features from an RGB image rather than a conventional gray image. This method is an improvement over WaveMsNet [45,46], the Specaugment method [65], and time-wise multi-inference strategies. In the initialization section, the hyper-parameters of the FFT were experimentally obtained based on values that are commonly used in ESC-50, and the hyper-parameters of the MTST were obtained via a greedy search. The ranges of the greedy searches were as follows: The steps were [50, 100, 150, 200, 250, 300] and x_size was [224, 401, 501, 601, 701, 801]. M was a method of conversion from raw data into a mel-spectrogram with the conversion of power to decibels (dB). Audio features were then extracted in a similar manner, except that we used a multi-time-scale transform. The model that we used was the wide resnet101 with double width.

Figure 5 shows the key idea of the multi-time-scale transform. The two figures are the same graph, with the three-dimensional visualization shown on the left and the two-dimensional visualization on the right. The x-axis shows frequency, the y-axis shows time, and the *z*-axis shows power. A heatmap-based RGB image is shown on the left, while the multi-time-scale transform image is shown on the right. The heatmap-based RGB image was determined by the power. In other words, even when information from the three channels was combined, only the gray information was available because the gray information quantity was distributed among the three channels according to the power level. We distributed information into the RGB channel to increase the total information. The spectrogram was up-scaled to match input shapes via bilinear interpolation, then divided into three parts based on the time axis using multi-scale inference (see Algorithm 1).

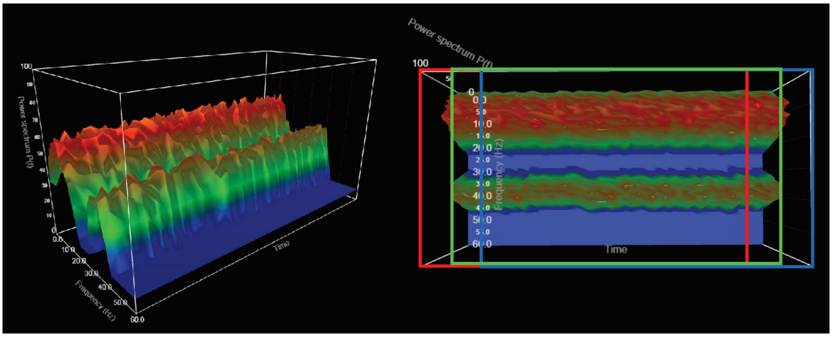


Figure 5. Lighting-chart-based simulation plots to help understand the key idea of the multi-time-scale transform.

Algorithm 1 Audio Feature Extraction via the Multi-Time-Scale Transform

```
Initialization;
[x]←audio_signal, [t]←time, [n_fft]←4096, [hop_length]←441, [n_mels]←224,
  [f_min]←20, [f_max]←20000, [top_db]←160, [x_size] ←701, [y_size] ←224, [step]
  ←200;
Audio Feature Extraction:
    spectrogram = |M(x,n_fft,hop_length,n_mels,f_min,f_max,top_db)|;
    image_rgb = $MTST(spectrogram, size = (x\_size, y\_size, step), p = 0.5)$
    features = model(image_rgb)
    return features;
MTST (spectrogram, size, p):
    x_size, y_size, step = size
    spectrogram = normalization(spectrogram)
    image_gray = $bilinear\_interpolation(spectrogram, (x\_size, y\_size))$
    r=spec_aug(image_gray$[; , 0 : max\_size - step], p$)
    g=image_gray$[; , step/2 : max\_size - step/2]$
    b=spec_aug(image_gray$[; , step : max\_size], p$)
    image_rgb = $concatenate(r, g, b)$
    return image_rgb
```

4. Experiments

We performed three main experiments. The first experiment was a quantitative evaluation in order to determine how effective the multi-time-scale transform was and if it improved the task of audio feature extraction. In the second experiment, we evaluated how reasonably music and paintings were matched by our system. In the third experiment, we evaluated whether the user experience improved while appreciating artworks and the soundscape music.

Hypothesis 1. *"If the soundscape music and painting are well matched, user appreciation of the experience will increase."*

The above hypothesis is the central focus of this study. However, assessing the match between the music and painting and the improvement of the user experience is challenging. We therefore developed strategies for measuring these factors, as described in the following subsections.

4.1. Audio Feature Extraction via the Multi-Time-Scale Transform

We performed an ablation and benchmark study to determine the effectiveness of the multi-time-scale transform. This experiment was conducted using the ESC-50 as an environmental sound classification dataset. The environmental settings were as follows: Pytorch 1.1 with Python 3.6. The batch size was 32, and a random shuffle was used. The optimizer used was Adam, and the initial learning rate was 1×10^{-4}; the scheduler used the cosine annealing learning rate scheduler. The hyper-parameter T value was 50. The hyper-parameters of training were obtained via a greedy search. The ranges of the greedy searches were as follows: The batch sizes were [8, 16, 32], and the initial learning rates were $[1 \times 10^{-4}, 2 \times 10^{-4}, \ldots, 5 \times 10^{-4}, 1 \times 10^{-5}, 2 \times 10^{-5} \ldots 5 \times 10^{-5}]$.

Table 2 shows the results of the ablation study. First, we conducted experiments using various backbone network settings, such as Mobilenet v2, Efficientnet, VGG, Resnet, and Densenet, with different widths and depths. However, networks other than Resnet and Densenet were omitted from Table 2 due to their poor results. We increased the depth of the network-provided setting and the width by one, two, and four times the depth. The width showed the best performance when it was doubled, and the depth tended to be better, but not in all cases. Resnet101 with double width showed the best performance. In

Table 2, the application of our multi-time-scale transform to resnet resulted in about 2.8% higher accuracy than that of the baseline network, WaveMsNet.

Table 2. Ablation study of the multi-time-scale transform.

Methods	Accuracy
Densenet with spectrogram	73.9%
Resnet with spectrogram	76.3%
WaveMsNet after first phase [45]	70.05%
WaveMsNet after second phase [45]	79.1%
Resnet with multi-time-scale transform	81.9%

Table 3 shows the sound classification results for various methods based on the ESC-50 benchmark. Methods without an asterisk in Table 3 are examples of supervised learning methods that use backbone and data argumentation methods, while methods with an asterisk were trained using various methods, such as weakly (or self) supervised learning, between-class learning, or some other method. Our network showed better performance than that of baseline, but lower performance than that of the state-of-the-art methods. AclNet uses raw feature extractors, such as Envnet v2. This is a weakness for mutual learning; therefore, we did not use these methods. Ensemble-fusing CNN is a state-of-the-art method, but is not a single-model method. Our model performed as well as the single-model methods, and had advantages for mutual learning because it is spectrogram-based. Therefore, in this work, we selected our network as the baseline. However, in our future research, various augmentations with ensemble-fusing CNNs should be explored for applications in our work. Our model's performance was about 12.2% poorer than that of WEANET. Therefore, work is needed to enhance the performance of our network. Nevertheless, our network has the advantage of being spectrogram-based and showed better performance than human performance. The validation results for each fold of our proposed method can be found in detail in Table 4.

Table 3. ESC-50 dataset benchmark.

Methods	Accuracy
EnvNet v2 + S.A. [47]	78.8%
WabeMsNet (our baseline) [45]	79.1%
Resnet with Multi-Time-Scale Transform	81.9%
AclNet [49]	85.65%
Ensemble-fusing CNN [48]	88.65%
Soundnet (our baseline) [37] *	74.2%
EnvNet v2 + S.A. + B.C.L. [47] *	83.9%
WEANET [66] *	94.1%
Human Performance [67]	81.3%

4.2. Relevance of Music–Artwork Matching

Definition. ***"Well-Matched Soundscape Music–Painting"***: In this work, we defined details factor of "Well-Matched Soundscape Music–Painting" for a measurement of the quality of a match. Blind touch is intended to provide direct and material interactions that go beyond metaphorical and psychological interactions between the art and the appreciator. However, this does not imply the absence of traditional art. Therefore, in this study, we wanted to measure the metaphorical and psychological relevance of matching between music and a painting. The metaphorical scale measured how much the same implicit multi-sensory experience was transferred because blind touch is a multi-sensory-based media art. The psychological scale was measured through subjective fitness, which is an individual's appreciation, because it is the most representative single scalar value of individual appreciation.

Table 4. Result of five-fold cross-validation. The F1 score is the weighted F1 score and the MCC is the Matthews correlation coefficient, which was obtained as the average of the MCC for each class. This metric tables were written by referring to [68,69].

Fold	Accuracy	Recall	Precision	F1	MCC	AUC
1	81.2%	81.2%	84.4%	81.1%	81.0%	90.4%
2	81.0%	81.0%	83.5%	80.4%	80.7%	90.3%
3	81.2%	81.2%	82.6%	80.7%	80.9%	90.4%
4	84.8%	84.8%	85.7%	84.2%	83.7%	91.8%
5	81.2%	81.2%	84.0%	81.1%	80.9%	90.4%

Experimental Setting: Twenty-four men and 16 women, all of whom are Korean, were evaluated in this study. Each group consisted of 10 people divided into four groups with a balanced gender ratio. The criteria for dividing the groups were as follows. First, the artistic culture of the subjects was evaluated via a simple test with the music and paintings used in this experiment. Second, the subjects were allocated to each group based on their test scores, which we averaged. The average music score was 2.3, the average painting score was 3.4, and a perfect score was 5. The roles of each group was as follows: Group 1 participated in subjective fitness experiments. Groups 2 and 3 participated in multi-sensory concordance experiments. Group 4 participated in multi-sensory concordance inverse measurement experiments. The groups were designed to avoid duplication of the experiments and to ensure that prior knowledge did not affect the experimental results. Furthermore, we did not provide any information other than information related to the experimental progress. This meant that the subjects did not know that the paintings and music were recommended via a deep neural network. Figure 6 shows the paintings used in our experiment, and the music allocated to the paintings by our system is indicated in Table 5.

(**a**) (**b**) (**c**) (**d**)

Figure 6. Paintings used in the experiments. (**a**) Starry Night; (**b**) Starry Night over the Rhône; (**c**) Café Terrace at Night; (**d**) Femme avec parasol dans un jardin. This figure is associated with Table 5.

Table 5. Music–painting matching scheme. The top three pieces of music selected by our system for each painting are shown. Rows 1, 2, and 3 correspond to musical choices 1, 2, and 3, respectively.

Painting	Music
(a)	Beethoven: Piano Sonata No. 3 in C, Op. 2 Prokofiev: 10 Pieces from Romeo and Juliet, Op. 75, No. 9, Dance of the Girls With Lilies Grofé: Grand Canyon Suite—4. Sunset
(b)	Edvard Grieg—Peer Gynt—Suite No. 1, Op. 46—III. Anitra's Dance Bolero—Maurice Ravel Ashkenazy: Sibelius—Valse Triste, Op. 44
(c)	Tchaikovsky—The Sleeping Beauty Suite, Op. 66a Joseph Haydn—Symphony No. 88 in Major, Largo Verdi: "Nabucco"—Schippers
(d)	Jean Sibelius—Valse Triste, Op. 44, No. 1 Edvard Grieg: "Peer Gynt—Morning Mood" Tchaikovsky—Swan Lake Op. 20, Act III

4.2.1. Measurements of Subjective Fitness

Experimental Methods: Subjective fitness was measured in group 1, which comprised five men and five women. First, we input the paintings into our system and extracted the top three music recommendations. Second, when both the paintings and music were experienced at the same time, we measured the subjective fitness using a five-point Likert scale. We changed the paintings in the painting–music pairs to ensure that previous and subsequent experiments did not affect the current experiments; for example, painting 1–music 1, painting 2–music 1, painting 3–music 1, painting 4–music 1, then painting 1–music 2, painting 2–music 2, etc.

Table 6 shows the results of the subjective fitness experiments. Columns (a), (b), (c), and (d) are associated with Figure 6, Music pieces 1, 2, and 3 are associated with Table 5, and Music 4 was used in the soundscape music exhibition [12] "Bunker de Lumières Van Gogh" held in Jeju, Korea. Therefore, Music 4 can be considered to be a type of ground truth labeled by experts. F-score is the fitness score, and the P-score is the preference ratio. The fitness score was measured on a five-point Likert scale, and the P-score was the proportion of people who scored three or more F-score points. Music 4 used the same values as in previous studies [70]. Therefore, the system matched the artwork and music well in terms of subjective fitness. However, the P-score was not stable for Music 2, and Music 3 had a low F-score and P-score. The average F-score for all music items was 3.24, and the average P-score for all music items was 75%. The average F-score of Music 1 was 3.68, and the average P-score of Music 1 was 97.5%. This result is similar to that reported in a previous study, which reported an average F-score of 3.16, an average F-score for Music 1 of 3.74, an average P-score of 87.6%, and an average P-score of Music 1 of 94%. Thus, our system matched music and artworks well, but did not perform better than previous systems, despite the improved feature representation. This is likely due to the small size of the music database. However, the reliability of the system was demonstrated by the attainment of results similar to those reported in previous studies. In future studies, we need to assess whether the stability and performance of the system increase in response to the expansion of the music database.

Table 6. Results of the subjective fitness experiment. The F-score is the fitness score and the P-score is the preference ratio. Fitness was measured on a five-point Likert scale, whereas the P-score reflects the proportion of individuals with three or more points.

Painting	Measure	Music 1	Music 2	Music 3	Music 4
(a)	F-score	**3.8**	3.2	3.1	3.77
	P-score	**100**%	80%	60%	85%
(b)	F-score	3.8	3.6	3.2	-
	P-score	100%	80%	70%	-
(c)	F-score	3.4	3	2.7	**3.85**
	P-score	**90**%	60%	40%	85%
(d)	F-score	3.7	3	2.4	-
	P-score	100%	70%	50%	-

4.2.2. Measurements of Implicit Multi-Sensory Concordance

Experimental Methods: This experiment was conducted in groups 2 and 3, each consisting of five men and five women. The members of group 2 first wrote a review after viewing the four paintings. Group 3 then wrote a review after listening to the 12 music items. At this time, the subjects were not provided any information about the content of the experiment; the only guidelines that they received were to use all five senses when assessing the artworks or music. Fourth, we measured multi-sensory concordance by comparing the sensory language similarity of the reviews. We mapped words in the review using a Korean sensory word classification table (Table [71]). Sensory word tables are classified as gustatory, tactile, and temperature sensations; examples of such words are bitter, sweet, salty, sour, nutty, astringent, spicy, plain, rough, smooth, soft (texture), soft (material), hard, moist, sharp, cold, cool, lukewarm, warm, hot, etc.

Table 7 shows the results of the implicit multi-sensory concordance experiments. The D-score refers to the Euclidean distance and the C-score to cosine similarity. Table 7 shows that similar sensory trends between the two groups were measured through the C-score. However, it can be confirmed that the C-score is not proportional to Table 6's P-score or F-score because the cosine similarity is advantageous for distance measurements in high-dimensional positive space, but the size of each dimension is not meaningful. In other words, only the trends in each dimension can be evaluated, while the size of the difference is difficult to assess. We used the D-score to overcome this limitation. The C-score in Table 7 indicates the sensory similarity between both the music matched by experts and the music matched by our system. The D-score was similarly measured. The results show that not only can media art provide similar multi-sensory appreciation, but our system can also provide similar results to those provided by experts. However, this result has the limitation that a criterion was not provided to determine the significance according to magnitude of the score. Future studies should develop a criterion to determine the significance of the score's magnitude.

4.2.3. Inverse Measurements of Implicit Multi-Sensory Concordance for Validation

Experimental Methods: This experiment was conducted in group 4, which comprised nine men and one woman. The purpose of this experiment was to verify the experiments in Section 4.2.2. A questionnaire was created based on a multi-sensory word table constructed in Section 4.2.2. The experiment was conducted in the same order as that described in Section 4.2.1, but the questionnaire was completed instead of a written review. The questionnaire evaluated the extent to which the appreciator agreed with the sensory table using a five-point Likert scale.

Table 8 presents the results of the inverse implicit multi-sensory concordance experiment. The A-score is the agreement scale, which was the average score obtained using the five-point Likert scale. The C-score is the cosine similarity of the A-scores calculated based on an active value (3 or higher). The purpose of this experiment was to measure how much the appreciator agreed with the implicit multi-sensory data. The A-scores and C-scores were low because of the failure of the experiment. Several problems were encountered during this experiment. The first problem was the mechanical marking phenomenon. When responding to the questionnaire, subjects gave low scores to senses that were mechanically opposed to the first high score. This was in contrast to the phenomenon where opposite senses were expressed together in the free reviews of appreciation. The second problem was the phenomenon of monotonous responses, which occurred when subjects became familiar with the experiment. This phenomenon demonstrated the tendency of subjects to exclude complex senses before appreciation. For example, the subjects gave low scores to combinations of options, such as the bitter taste of wine and the sour taste of candy, before even listening to music. Thus, in further studies, the validation of the experimental design should be improved.

Table 7. Results of the implicit multi-sensory concordance experiments. The D-score is the Euclidean distance and the C-score is the cosine similarity.

Painting	Measure	(a)	(b)	(c)	(d)
Music 1	D-score C-score	41.8 92%	46.9 89%	54.8 88%	74.2 75%
Music 2	D-score C-score	48.9 89%	50.9 87%	52.4 87%	55.7 87%
Music 3	D-score C-score	57.9 85%	56.6 82%	57.8 87%	60.0 84%
Music 4	D-score C-score	56.6 85%	- -	44.2 92%	- -

Table 8. Results of the inverse implicit multi-sensory concordance experiment. The A-score is the agreement scale and the C-score is the cosine similarity.

Painting	Measure	(a)	(b)	(c)	(d)
Music 1	A-score C-score	2.09 8%	2.34 10%	2.73 14%	2.68 27%
Music 2	A-score C-score	2.45 12%	2.55 12%	2.62 12%	1.46 15%
Music 3	A-score C-score	2.89 16%	2.83 15%	2.89 15%	1.62 16%
Music 4	A-score C-score	2.82 15%	- -	2.21 8%	- -

4.3. Improvement of the Appreciation Experience with the Soundscape

Definition. "Improvement of the appreciation experience": In this work, we defined the measurable factor of "Improvement of the appreciation experience" as the combination of appreciation and subjective satisfaction; however, this is an ambiguous concept. To address this, we proposed an evaluation method for the immersion based on the flow theory of Csikszentmihalyi. We differentiated flow and cognitive immersion in this study according to flow theory. Flow was indirectly measured via time distortion phenomena, and cognitive absorption was measured through a simple test of working memory and attention concentration. The subjective satisfaction score of the appreciation experience comprised experiences of the environment and appreciation. The subjective environmental satisfaction was not related to improvements in the appreciation experience; however, the environmental score was a useful indicator of if the environmental setting was appropriate.

Appreciation scores were measured with a questionnaire based on the SSID [72] and WHO-5 Well-Being indices. The SSID is an evaluation index for soundscapes, and the WHO-5 index is an evaluation index for quality of life. Our questionnaire was prepared based on the SSID and WHO-5 Well-Being indices.

Experimental Setting: The participants in this experiment included a total of 29 men and 1 woman. They were divided into three groups, with each group consisting of 10 participants. The roles of each group were as follows: Groups 5 and 6 participated in the immersion experiments. Group 7 participated in the subjective satisfaction experiments. The criteria for dividing the groups and other experimental conditions were similar to those described for the previous experiments.

4.3.1. Measurements of Immersion

Experimental Methods: This experiment was conducted in groups 5 and 6. The purpose of this experiment was to indirectly measure immersion. Time distortion was measured for the flow measurement experiments. Group 5 appreciated artworks without listening to music, and group 6 appreciated artworks while listening to music. Time distortion was measured as the subjective assessment of time between the two groups, and flow was evaluated indirectly from the time distortion measurements. Cognitive absorption was measured with a simple test of working memory and attention concentration. This test asked questions about the color, location, shape, and texture of an object or scene.

Table 9 presents the measurement results for the immersion experiment. Each row presents the subjective times that participants felt while appreciating the exhibition. The ground truth refers to the real length of the piece of music. The *p*-value is the result of the *t*-Test, which was performed under the assumption that the variances were different. The results of the time distortion experiment indicate that group 5 predicted a time closer to the ground truth than group 6. In particular, participants in group 6 felt that more time had elapsed than actually had. Two interesting phenomena were discovered during the analysis of the interviews. The first was the “sleepy” phenomenon. When music pieces 1 or 4 were played, participants in group 5 did not experience sleepiness, while more than 80% of the participants in group 6 experienced sleepiness. The subjective viewing time was 2 m 36 s on average, while the experimental time was longer than the appropriate viewing time, ranging from 3 m 37 s to 6 m 52 s. The second phenomenon was the phenomenon of ambiguous answers. Group 5 answered with specific times, such as 4 m 20 s and 5 m 30 s, while group 6 tended to answer with ambiguous numbers, such as “about 5 minutes” and “about 10 minutes”. In the experiments on cognitive absorption, the averages of the answers given by each group in our simple test were used. The perfect score for this test was 10. As shown in Table 8, groups 5 and 6 had an average score difference of 2.2. This can be attributed to the sleepy phenomenon. Thus, with the results obtained from this experiment, it was not possible to accurately confirm whether cognitive absorption was affected or not. However, music is qualitatively conducive to cognitive absorption. Examples of answers from participants included the following: **“The bouncy rhythm in the song reminded me of stars”; “The music felt like a young man in the country was leaving the village, and it helped me remember because there was a real village in the artwork”; “it wasn’t hard to find the location of the lover because I watched the couple carefully because of the sentimental music being played.”**

Table 9. Results of immersion experiments for assessing time distortion and cognitive absorption. A *t*-test was performed with the assumption of different variances.

Painting	Music 1	Music 2	Music 3	Music 4	Avg. Answer
Group 5	8 m 18 s	3 m 58 s	3 m 32 s	5 m 13 s	6.4
Group 6	10 m 22 s	4 m 43 s	5 m 18 s	7 m 20 s	4.2
Ground Truth	6 m 52 s	3 m 37 s	4 m 05 s	5 m 33 s	-
p-value	0.00137	0.00123	5.51×10^{-7}	1.03×10^{-10}	0.00511

4.3.2. Measurements of Subjective Satisfaction

Experimental Methods: This experiment was conducted in group 7. The purpose of this experiment was to measure the subjective satisfaction score. The subjects appreciated the paintings with the top music and filled out a questionnaire that was prepared based on the SSID and WHO-5 Well-Being indices.

The subjective satisfaction scores, which were based on the environment and the appreciation experience, are shown in Table 10. The top three questions were about environmental experience. The environmental satisfaction scores ranged from 3.3 to 3.6, which are fairly high scores. For the first question, discomfort was associated with 0 points and comfort with 5 points. Therefore, this experiment was appropriately constructed to assess the appreciation experience. The subjective satisfaction scores of appreciation ranged from 3.1 to 4.0. Therefore, soundscape music can aid in appreciation.

Table 10. Questionnaire based on the SSID and WHO-5 Well-Being indices.

Question	Avg. Score
Are you feeling uncomfortable with your appreciation?	3.3
Is the volume of the soundscape music appropriate for you?	3.5
Is the sound quality of the soundscape music being played good?	3.6
Did the soundscape music go well with the painting?	3.7
Did music help you to enjoy paintings more?	4.0
Did music help you to be more comfortable when appreciating the paintings?	3.1
Did music help you to take initiative and actively appreciate the paintings?	3.8
Did music make your appreciation of the paintings fresher?	3.3
Did music make your appreciation of the paintings more interesting?	3.3

5. Discussion

In this section, we would like to conduct an interdisciplinary analysis of how this metaphorical and psychological transfer could have occurred. This interdisciplinary approach consists of five parts: The first is the consideration of the inter-sensory transition phenomenon through the concept of synesthesia; the second is an artistic approach based on the characteristics of the paintings that we used in our experiments; the third is the consideration of the characteristics of deep neural networks in connection with the concept of synesthesia and the characteristics of our paintings; the fourth includes the limitations of this study and the future directions of development; the fifth is the overall blueprint of our research.

First, we deal with synesthesia by focusing on the transition of senses. References [15–17] show that media art can give potential cognitive and emotional impacts. We would like to discuss why this effect could occur. Synesthesia is a blending of the senses in which the stimulation of one modality simultaneously produces sensations in a different modality. Synesthesia is known to affect four percent of the population. However, via color–alphabet experiments with 400 people, studies [73] have shown that people without synesthesia perceive synesthesia but do not consciously recognize it. In other words, non-synesthesia implies that synesthesia is being perceived unconsciously. In particular, the study by [74] dealt with implicit associations between color and sound, particularly showing that various color properties can be mapped to acoustic properties, such as tone and volume. The study by [75] showed that the properties of sounds can be associated with colors. The study by [76] also demonstrated empirical investigations of the associations between auditory and visual perception. Based on these experimental results, transitions between implicit senses are possible; for example, value and hue can be mapped onto acoustic properties, such as pitch and loudness, pitch can be associated with color lightness, loudness can be mapped onto greater visual saliency, high loudness can be associated with orange/yellow rather than blue, and chroma can have a relationship with sound intensity. The following sections cover the features of our artworks.

Second, the artworks used in our experiments were Impressionistic paintings. Impressionism is a trend in art that focuses on colors, lighting, and textures, and it is characterized by the ability to describe nature in the changing colors of light and to accurately and objectively record the visible world using the momentary effects of colors and shades. In particular, three of our four paintings were by Van Gogh, and he is known as a post-impressionist. Van Gogh tried to thicken paint through the Impasto technique to express the maximum texture and create a three-dimensional effect. In addition, Van Gogh did not describe objects in the same way, but captured his emotion and feeling very strongly through the touch of his brush. This distorted form, intense color, and simplification of the form were sought to express emotions. The other painting was "Woman with a Parasol in a Garden". Renoir was an early impressionist who understood how to express the effects of light, and was known to use black to express shadows. Therefore, the paintings we used are from the early to the late Impressionist period, and they are characterized by their good representation of colors, light, and texture. Therefore, we selected our paintings in consideration of these features. The characteristics of the paintings were analyzed and the possibilities of sensory transitions were addressed.

Third, we discuss how a deep neural network was able to recommend music and paintings in the above consideration. The study by [77] dealt with the perspectives of beauty, sentiment, and remembrance of art in deep neural networks. The study by [77] quantitatively and qualitatively analyzed three subjective aspects of human consciousness: image features in relation to aesthetics, sentiment, and memorability. This study indicated that the CNN considered features of various aspects, such as color, intensity, harmony of colors, object emphasis, pattern, art style, semantics, genre, and content. This study also addressed how deep neural networks judge and remember emotional and artistic values by exploring predicted aesthetic and emotional memory scores in the context of art history. In particular, these experiments also showed that aesthetic and emotional scores and color correlations are consistent with common assumptions. These experimental results show that deep neural networks consider factors that can cause sensory transference. Therefore, these features make our sensory transition possible.

Fourth, we address the limitations of our experiment and the directions of future study. There is a limitation in that no global solution was shown. The limitations can be seen from three perspectives: The first is that our research only dealt with Impressionist paintings from an artistic point of view. Therefore, further study is needed as to whether such studies can be applied to different artistic styles beyond Impressionism. The second is the problem of generalization; this study is not representative of the general population because the experiments were conducted on people in their 20s in Korea. In addition, research should be conducted on whether the experimental results are universal or if they are due to social learning. If the results obtained are due to social learning, extended studies should also be carried out from various perspectives, including culture, age, and local areas. Future research will therefore need to extend these local solutions into a global solution. The third is the effectiveness of media art, which can have potential cognitive and emotional impacts. However, our research focused on the potential cognitive impacts of the experiments. Therefore, we measured emotional impact with the F-score, the most representative and implicit single scalar value. These representative features of F-score can be influenced by various factors, such as emotions and environments, as well as by potential cognitive factors. Therefore, more detailed measurements should be studied in the future.

Fifth, our experiment was intended to enhance the experience of appreciation of media art through the effect of visual and auditory transitions. The transitions of these senses can be expected not only in vision and hearing, but also between various other senses, such as vision and smell or vision and touch. In this study, we aimed to create a platform that guides blind people through sensory transition they appreciate media art. Finally, we hope that this multi-sensory experience will be harmonized and become abundant in media art. Furthermore, we hope that our proposed method will be applied to multi-modal media art platforms to help visually impaired people appreciate artworks.

6. Conclusions

This study described the construction of a soundscape-based exhibition environment using deep neural networks. The baseline was improved by different perspectives, such as modeling methods, learning methods, and domain adaptation methods. In addition, the soundscape music selected by our system was output via hyper-orientated speakers to improve the appreciation experience. To measure the improvements in user experience, we devised a soundscape music evaluation method and an appreciation experience evaluation method and conducted extensive experiments with 70 subjects. However, our research suffered from three major limitations. First, this was not state-of-the-art performance from the perspective of the deep neural network. In particular, models with low accuracy from the feature extraction perspective were selected for mutual learning. Therefore, the development of models with improved audio feature representation is required. Second, our current database comprised only 2000 music items. In addition, these 2000 music pieces were limited from the genre perspective. A wide spectrum of databases containing music from various cultures and times should be used in future studies. Third, we conducted experiments using 70 individuals, but each experiment was conducted on a group of only 10 people. Therefore, our experimental results cannot be generalized. Large-scale experiments need to be conducted to generalize these experimental results. This study shows the results of a pilot test with sighted test participants. We hope that the results of this study help people with visual impairments to appreciate art and that they help promote the cultural enjoyment rights of people with visual impairments.

Author Contributions: Conceptualization, methodology, and software, Y.K.; Data collection, experiments, and data curation, H.J.; Review and validation of UX experiments, J.-D.C.; Writing—review and editing, funding acquisition, and project administration, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2018-0-01798) supervised by the IITP (Institute for Information and Communications Technology Promotion).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Adams, M.; Cox, T.; Moore, G.; Croxford, B.; Refaee, M.; Sharples, S. Sustainable soundscapes: Noise policy and the urban experience. *Urban Stud.* **2006**, *43*, 2385–2398. [CrossRef]
2. Miller, N. Understanding soundscapes. *Buildings* **2013**, *3*, 728–738. [CrossRef]
3. Yang, W.; Kang, J. Soundscape and sound preferences in urban squares: A case study in Sheffield. *J. Urban Des.* **2005**, *10*, 61–80. [CrossRef]
4. Lavia, L.R.; Dixon, M.; Axelsson, O.; Witchel, H. Using a soundscape approach to develop an acoustic ecology plan for a city. *J. Acoust. Soc. Am.* **2011**, *130*, 2532. [CrossRef]
5. Aletta, F.; Margaritis, E.; Filipan, K.; Romero, V.P., Axelsson, Ö.; Kang, J. Characterization of the soundscape in Valley Gardens, Brighton, by a soundwalk prior to an urban design intervention. In Proceedings of the Euronoise Conference, Maastricht, The Netherlands, 31 May–3 June 2015.
6. Aletta, F.; Xiao, J. What are the current priorities and challenges for (urban) soundscape research? *Challenges* **2018**, *9*, 16. [CrossRef]
7. Hellström, B.; Sjösten, P.; Hultqvist, A.; Dyrssen, C.; Mossenmark, S. Modelling the shopping soundscape. *J. Sonic Stud.* **2011**, *1*. Available online: https://www.researchcatalogue.net/view/220808/220809 (accessed on 13 February 2021).
8. Gomes, J.A.; Hespanhol, N. Mini Mapa Sonoro Project: Soundscape, Technology, Art and Education. In Proceedings of the 9th International Conference on Digital and Interactive Arts, Braga, Portugal, 23–25 October 2019; pp. 1–6.

9. Casu, M.; Koutsomichalis, M.; Valle, A. Imaginary soundscapes: The SoDA project. In Proceedings of the 9th Audio Mostly: A Conference on Interaction with Sound, Aalborg, Denmark, 1–3 October 2014; pp. 1–8.
10. Kajihara, Y.; Dozono, S.; Tokui, N. Imaginary Soundscape: Cross-Modal Approach to Generate Pseudo Sound Environments. In Proceedings of the Workshop on Machine Learning for Creativity and Design (NIPS 2017), Long Beach, CA, USA, 8 December 2017; pp. 1–3.
11. Faustino, D.B.; Gabriele, S.; Ibrahim, R.; Theus, A.L.; Girouard, A. SensArt demo: A multisensory prototype for engaging with visual art. In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, Brighton, UK, 17–20 October 2017; pp. 462–465.
12. Bunker de Lumières. Available online: https://www.bunkerdelumieres.com (accessed on 13 February 2021).
13. Vi, C.T.; Ablart, D.; Gatti, E.; Velasco, C.; Obrist, M. Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition. *Int. J. Hum. Comput. Stud.* **2017**, *108*, 1–14. [CrossRef]
14. Cho, J.-D. Human ICT Convergence Exhibition: BlindTouch (Multisensory Painting Platform for the Blind), Siloam Center for the Blind S-Gallery, Sungkyunkwan University, Seoul, Korea, 2019. Available online: https://www.donga.com/news/Society/article/all/20200428/100840356/1 (accessed on 13 May 2021).
15. Wang, S. Museum as a Sensory Space: A Discussion of Communication Effect of Multi-Senses in Taizhou Museum. *Sustainability* **2020**, *12*, 3061. [CrossRef]
16. Feng, A.; Chen, Y.V.; Li, R.; Ding, L. How smell can help visually impaired in health and well-being—A cognitive experiment. *Des. J.* **2019**, *22*, 371–386. [CrossRef]
17. Baumgartner, T.; Esslen, M.; Jäncke, L. From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *Int. J. Psychophysiol.* **2006**, *60*, 34–43. [CrossRef] [PubMed]
18. 7 Smart Tech Developments for People Who Are Blind or Have Low Vision. Available online: https://news.microsoft.com/on-the-issues/2019/08/08/smart-tech-blind-low-vision (accessed on 13 February 2021).
19. With Lookout, Discover Your Surroundings with the Help of AI. Available online: https://www.blog.google/outreach-initiatives/accessibility/lookout-discover-your-surroundings-help-ai (accessed on 13 February 2021).
20. *Convention on the Rights of Persons with Disabilities (CRPD)*; United Nations, Department of Economic and Social Affairs Disability: New York, NY, USA, 2008.
21. Two Problems with Article 30 of the United Nations Convention on the Human Rights of Disabilities. Available online: http://abnews.kr/1QZK (accessed on 13 February 2021).
22. Blind Touch Project. Available online: http://blindtouch.org/ (accessed on 13 February 2021).
23. Lee, Y.H.; Kim, S.W.; Eom, M.S.; An, S.M.; Cho, J.D. An art exhibition needs assessment survey of persons with visual impairment. *Asia Pac. J. Multimed. Serv. Converg. Art Humanit. Sociol.* **2019**, *9*, 457–466.
24. The Life of People with Disabilities in 2020 Statistics. Available online: https://eiec.kdi.re.kr/policy/materialView.do?num=203276&topic= (accessed on 13 February 2021).
25. Cavazos Quero, L.; Bartolomé, L.C.; Cho, J.D. Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics. *Electronics* **2021**, *10*, 297. [CrossRef]
26. Cavazos Quero, L.; Bartolomé, L.C.; Cho, J.D. Gamification and accessibility. In Proceedings of the 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 11–14 January 2019; pp. 1–5.
27. Muller-Eberstein, M.; Van Noord, N. Translating Visual Art into Music. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, 27 October–2 November 2019.
28. Moscato, V.; Picariello, A.; Sperli, G. An emotional recommender system for music. *IEEE Intell. Syst.* **2020**, *1*, 026000. [CrossRef]
29. Katarya, R.; Verma, O.P. Efficient music recommender system using context graph and particle swarm. *Multimed. Tools Appl.* **2018**, *77*, 2673–2687. [CrossRef]
30. Thorat, P.B.; Goudar, R.M.; Barve, S. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *Int. J. Comput. Appl.* **2015**, *110*, 31–36.
31. Workie, A.; Sharma, R.; Chung, Y.K. Digital Video Summarization Techniques: A Survey. *Int. J. Eng. Technol.* **2020**, *9*. [CrossRef]
32. Basavarajaiah, M.; Sharma, P. Survey of Compressed Domain Video Summarization Techniques. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–29. [CrossRef]
33. Lee, D.; Hwang, H.; Jabbar, M.S.; Cho, J. Language of Gleam: Impressionism Artwork Automatic Caption Generation for People with Visual Impairments. In Proceedings of the 13th International Conference on Machine Vision (ICMV 2020), Rome, Italy, 2–6 November 2020.
34. Hossain, M.Z.; Sohel, F.; Shiratuddin, M.F.; Laga, H. A comprehensive survey of deep learning for image captioning. *ACM Comput. Surv. (CSUR)* **2019**, *51*, 1–36. [CrossRef]
35. Liu, X.; Xu, Q.; Wang, N. A survey on deep neural network-based image captioning. *Vis. Comput.* **2019**, *35*, 445–470. [CrossRef]
36. Zhang, D.Y.; Ni, B.; Zhi, Q.; Plummer, T.; Li, Q.; Zheng, H.; Wang, D. Through the eyes of a poet: Classical poetry recommendation with visual input on social media. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Vancouver, BC, Canada, 27–30 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 333–340.
37. Aytar, Y.; Vondrick, C.; Torralba, A. Soundnet: Learning sound representations from unlabeled video. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2016; pp. 892–900.

38. Kumar, A.; Raj, B. Audio event and scene recognition: A unified approach using strongly and weakly labeled data. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3475–3482.
39. Arandjelovic, R.; Zisserman, A. Objects that sound. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 435–451.
40. Hu, D.; Nie, F.; Li, X. Deep multimodal clustering for unsupervised audiovisual learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9248–9257.
41. Wilkinghoff, K. On open-set classification with L3-Net embeddings for machine listening applications. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 24–28 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 800–804.
42. Kumar, A.; Raj, B. Features and kernels for audio event recognition. *arXiv* **2016**, arXiv:1607.05765.
43. Baelde, M.; Biernacki, C.; Greff, R. A mixture model-based real-time audio sources classification method. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2427–2431.
44. Boddapati, V.; Petef, A.; Rasmusson, J.; Lundberg, L. Classifying environmental sounds using image recognition networks. *Procedia Comput. Sci.* **2017**, *112*, 2048–2056. [CrossRef]
45. Zhu, B.; Wang, C.; Liu, F.; Lei, J.; Lu, Z.; Peng, Y. Learning environmental sounds with multi-scale convolutional neural network. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8.
46. Tokozume, Y.; Harada, T. Learning environmental sounds with end-to-end convolutional neural network. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2721–2725.
47. Tokozume, Y.; Ushiku, Y.; Harada, T. Learning from between-class examples for deep sound recognition. *arXiv* **2018**, arXiv:1711.10282.
48. Nanni, L.; Maguolo, G.; Brahnam, S.; Paci, M. An Ensemble of Convolutional Neural Networks for Audio Classification. *arXiv* **2020**, arXiv:2007.07966.
49. Huang, J.J.; Leanos, J.J.A. Aclnet: Efficient end-to-end audio classification cnn. *arXiv* **2018**, arXiv:2007.07966.
50. Tak, R.N.; Agrawal, D.M.; Patil, H.A. Novel phase encoded mel filterbank energies for environmental sound classification. In Proceedings of the International Conference on Pattern Recognition and Machine Intelligence, Kolkata, India, 5–8 December 2017; pp. 317–325.
51. Sailor, H.B.; Agrawal, D.M.; Patil, H.A. Unsupervised Filterbank Learning Using Convolutional Restricted Boltzmann Machine for Environmental Sound Classification. In Proceedings of the InterSpeech (2017), Stockholm, Sweden, 20–24 August 2017; pp. 3107–3111. [CrossRef]
52. Kim, Y.; Jeong, H.; Shin, J. Sound Classification using 3D Image via Multi-Time Scale Transform. In Proceedings of the International Conference on Convergence Technology (ICCT), Jeju, Korea, 8–10 July 2020.
53. Available online: https://github.com/EzoBear/MTST (accessed on 13 February 2021).
54. Buciluǎ, C.; Rich, C.; Niculescu-Mizil, A. Model compression. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; ACM: New York, NY, USA, 2006.
55. Kouw, W.M.; Loog, M. A review of domain adaptation without target labels. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 766–786. [CrossRef]
56. Gou, J.; Yu, B.; Maybank, S.J.; Tao, D. Knowledge distillation: A survey. *arXiv* **2020**, arXiv:2006.05525.
57. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
58. Romero, A.; Ballas, N.; Kahou, S.E.; Chassang, A.; Gatta, C.; Bengio, Y. Fitnets: Hints for thin deep nets. *arXiv* **2014**, arXiv:1412.6550.
59. Zagoruyko, S.; Nikos, K. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *arXiv* **2016**, arXiv:1612.03928.
60. Srinivas, S.; François, F. Knowledge transfer with jacobian matching. *arXiv* **2018**, arXiv:1803.00443.
61. Heo, B.; Lee, M.; Yun, S.; Choi, J.Y. Knowledge transfer via distillation of activation boundaries formed by hidden neurons. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33.
62. Zhang, Y.; Xiang, T.; Hospedales, T.M.; Lu, H. Deep mutual learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
63. Zhang, Z.; Guanghan, N.; Zhihai, H. Knowledge projection for deep neural networks. *arXiv* **2017**, arXiv:1710.09505.
64. Mirzadeh, S.I.; Farajtabar, M.; Li, A.; Levine, N.; Matsukawa, A.; Ghasemzadeh, H. Improved knowledge distillation via teacher assistant: Bridging the gap between student and teacher. *arXiv* **2019**, arXiv:1902.03393.
65. Park, D.S.; Chan, W.; Zhang, Y.; Chiu, C.C.; Zoph, B.; Cubuk, E.D.; Le, Q.V. Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv* **2019**, arXiv:1904.08779.
66. Kumar, A.; Ithapu, V. A Sequential Self Teaching Approach for Improving Generalization in Sound Event Recognition. In Proceedings of the International Conference on Machine Learning—PMLR 2020, Virtual, 13–18 July 2020; pp. 5447–5457.

67. Piczak, K.J. ESC: Dataset for environmental sound classification. In Proceedings of the 23rd ACM International Conference on Multimedia, New York, NY, USA, 21–25 October 2015; pp. 1015–1018.
68. Le, N.Q.K.; Do, D.T.; Hung, T.N.K.; Lam, L.H.T.; Huynh, T.T.; Nguyen, N.T.K. A computational framework based on ensemble deep neural networks for essential genes identification. *Int. J. Mol. Sci.* **2020**, *21*, 9070. [CrossRef]
69. Le, N.Q.K.; Do, D.T.; Chiu, F.Y.; Yapp, E.K.Y.; Yeh, H.Y.; Chen, C.Y. XGBoost improves classification of MGMT promoter methylation status in IDH1 wildtype glioblastoma. *J. Pers. Med.* **2020**, *10*, 128. [CrossRef] [PubMed]
70. Jeong, H.; Kim, Y.; Cho, J. Convergence evaluation method using multisensory and matching painting and music using deep learning based on imaginary soundscape. *J. Korea Converg. Soc.* **2020**, *11*, 175–182.
71. Jennett, C.; Cox, A.L.; Cairns, P.; Dhoparee, S.; Epps, A.; Tijs, T.; Walton, A. Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput. Stud.* **2008**, *66*, 641–661. [CrossRef]
72. Mitchell, A.; Oberman, T.; Aletta, F.; Erfanian, M.; Kachlicka, M.; Lionello, M.; Kang, J. The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys—Questionnaires with Acoustical and Contextual Information. *Appl. Sci.* **2020**, *10*, 2397. [CrossRef]
73. Cytowic, R.E. *Synesthesia*; MIT Press: London, UK, 2018.
74. Anikin, A.; Johansson, N. Implicit associations between individual properties of color and sound. *Atten. Percept. Psychophys.* **2019**, *81*, 764–777. [CrossRef]
75. Hamilton-Fletcher, G.; Witzel, C.; Reby, D.; Ward, J. Sound properties associated with equiluminant colours. *Multisens. Res.* **2017**, *30*, 337–362. [CrossRef]
76. Giannakis, K. Sound Mosaics: A Graphical User Interface for Sound Synthesis Based on Audio-Visual Associations. Ph.D. Thesis, Middlesex University, London, UK, 2001.
77. Cetinic, E.; Lipic, T.; Grgic, S. A deep learning perspective on beauty, sentiment, and remembrance of art. *IEEE Access* **2019**, *7*, 73694–73710. [CrossRef]

electronics

MDPI

Article

Multi-Sensory Color Expression with Sound and Temperature in Visual Arts Appreciation for People with Visual Impairment

Jorge Iranzo Bartolome [1], Gilsang Cho [2] and Jun-Dong Cho [2,*]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Korea; jorgedavid@skku.edu
[2] Department of Human ICT Convergence, Sungkyunkwan University, Suwon 16419, Korea; jus5041@naver.com
* Correspondence: jdcho@skku.edu

Abstract: For years the HCI community's research has been focused on the hearing and sight senses. However, in recent times, there has been an increased interest in using other types of senses, such as smell or touch. Moreover, this has been accompanied with growing research related to sensory substitution techniques and multi-sensory systems. Similarly, contemporary art has also been influenced by this trend and the number of artists interested in creating novel multi-sensory works of art has increased substantially. As a result, the opportunities for visually impaired people to experience artworks in different ways are also expanding. In spite of all this, the research focusing on multimodal systems for experiencing visual arts is not large and user tests comparing different modalities and senses, particularly in the field of art, are insufficient. This paper attempts to design a multi-sensory mapping to convey color to visually impaired people employing musical sounds and temperature cues. Through user tests and surveys with a total of 18 participants, we show that this multi-sensory system is properly designed to allow the user to distinguish and experience a total of 24 colors. The tests consist of several semantic correlational adjective-based surveys for comparing the different modalities to find out the best way to express colors through musical sounds and temperature cues based on previously well-established sound-color and temperature-color coding algorithms. In addition, the resulting final algorithm is also tested with 12 more users.

Keywords: visually impaired people; accessibility; art appreciation; color; multi-sensory interaction

Citation: Bartolome, J.I.; Cho, G.; Cho, J.-D. Multi-Sensory Color Expression with Sound and Temperature in Visual Arts Appreciation for People with Visual Impairment. *Electronics* **2021**, *10*, 1336. https://doi.org/10.3390/electronics10111336

Academic Editor: George A. Tsihrintzis

Received: 3 March 2021
Accepted: 28 May 2021
Published: 2 June 2021

1. Introduction

Multi-sensory systems, or systems that combine several modes of interaction through different senses, have been gaining popularity in recent times in the field of HCI (Human–Computer Interaction). Each one of the five senses gives humans unique ways of experiencing the world around and, as a result, researchers have seen the potential of applying those natural ways of interacting with the world into different ways of interacting with computers. In addition, multi-sensory systems can also be useful to convey complex or abstract information since each sense can congruently aid the others. Moreover, recent research has shown that humans are used to interacting with the world in a multi-sensory way and that multi-sensory experiences tend to be more engaging and easier to remember when compared to unisensory ones [1].

One of the possible applications for multi-sensory systems is the creation of novel art which creates a novel experience for the spectator by means of its multi-sensory interfaces. Because of that, artists have been showing a growing interest in multi-sensory art exhibitions. This increase of multi-sensory experiences, which makes interaction drift from the common visual and auditory types to other types of interaction, can be really convenient for minority groups that lack one of those senses, such as the visually impaired people. People with visual impairments are interested not only in performing with ease the necessary activities of everyday life but also in visiting museums and enjoying visual arts, which is why many museums have been making some of their exhibitions and artworks accessible to visually impaired people by hosting specialized tours with access to tactile representations of the artworks [2–4]. Nevertheless, most of those tours and exhibitions are based solely on tactile and audio systems which do not make use of any other sensory modality for interacting with the artworks. Since visually impaired people lack only the sense of vision, we believe they could benefit greatly from multi-sensory experiences as they provide them not only with the possibility of experiencing the artwork without relying on single sense that they cannot have access to, but also with the increased engagement, gamification and artwork comprehension which multi-sensory experiences can facilitate.

In this work, we design a novel sound-temperature multi-sensory system that helps visually impaired people experience a total of 24 colors. Additionally, during the process, a throughout investigation about the relationship of the different sounds and temperatures with the warm/cold and bright/dark dimensions of colors was performed. This investigation included extensive surveys and tests with a total of 18 participants, and an additional 12 participants for the final multimodal test. The work is the continuation of a series of works whose aim is to develop a multi-modal system and algorithms that aid visually impaired people in experiencing pieces of art. In previous works, both a color-temperature mapping algorithm and a sound-color coding system were designed, implemented into physical prototypes, and tested [5–7]. This work introduces a method for finding the best way to turn those unisensory systems into an improved sound-temperature multi-sensory system for conveying colors. Both works will be explained in more detail in the following sections.

2. Background and Related Works

2.1. Multi-Sensory Systems

In recent times, multi-sensory systems have gained popularity among researchers thanks to strong evidence that multi-sensory experiences have some benefits over unisensory ones, especially when it pertains to experiences related to education and learning. Shams et al. [1] state that multi-sensory experiences aid in memorization by helping the brain retain the information faster and for longer times. Moreover, the authors argue that the human brain evolved to develop, adapt, learn and function optimally in multi-sensory environments. However, they also maintain that congruence between the multi-sensory cues seems to be a condition for the benefits to be optimum.

Sensory-substitution is the technique of representing the characteristics of one sensory modality into another sensory modality. The technique is strongly related to assistive multi-sensory systems for the visually impaired people since the main goal of those systems is to communicate through a different sense the information that cannot be acquired by the visually impaired user because of the lack of sight. The authors of [8] provide a tutorial guide on how to research sensory substitution techniques and multi-sensory systems in order to successfully design inclusive cross-modal displays. They state the three main guiding principles necessary for a multi-sensory process to occur are spatial coincidence, temporal coincidence, and inverse effectiveness. Our work takes care of following all those principles. Spatial and temporal coincidence is assured as long as the sound and the temperature are communicated to the user at the same time or in immediate succession. The inverse effectiveness guideline states that for multi-sensory information to be effective, none of the sensory modalities needs to overpower the other. Since both the musical sounds and the temperature cue are semantically congruent to what they represent (as will be proved during the tests in later sections) and offer the user different information about the color, there is no weak/strong bias towards any of them and both sensory modalities work united as one.

2.2. Multi-Sensory Art

The recent growing interest in multi-sensory experiences within the scientific and HCI community has also been accompanied by a growing interest in those technologies by artists and museums. For example, in [9], a prototype called SensArt, which uses music, vibration patterns, and temperature to translate descriptive and emotive qualities of the artwork to the user, is designed and tested. The tests with 12 participants showed that most participants preferred the multi-sensory experience over the normal one and commented that they would visit museums more often if those types of experiences were more available.

Tate Sensorium [10] is a multi-sensory display that was exhibited at the Tate Britain art gallery of London. Its authors set themselves to explore new ways of experiencing art while researching and gaining design insights related to multi-sensory systems. One of the most interesting results after real on-field testing with users was the conclusion that multi-sensory experiences, in general, do make art more engaging and stimulating.

It can be observed that in most multi-sensory art applications some type of sound of musical feedback is included. Research has shown that mixing music with visual art increases the emotional experience of the user [11]. Additionally, as can be inferred from previous works' results, multi-sensory systems seem to increase the engagement of the spectator and enhance the art exploration experience. Similarly, one of the goals of our sound-temperature multimodal system for conveying color is to enhance the artistic experience for visually impaired users. Particularly, we expect the art engagement to be raised and the experience and memorization of colors to be improved. This manuscript is the first iteration towards that goal: the definition and design of the multi-sensory color coding.

2.3. Multi-Sensory Assistive Devices

Multi-sensory systems have also been implemented as assistive devices for the visually impaired people. In [12] a multi-sensory picture book for children with visual impairment was presented. The work provided visually impaired children with a multi-sensory experience consisting of touch, sounds and smell, which was integrated with the storytelling. User tests with a total of 25 children showed the potential of multi-sensory experiences for increasing engagement and enhancing the learning and artistic experience of children. The use of smell as a sensory modality is particularly interesting because of its implementation and how uncommon smell interfaces are. The olfactory device is contained inside the book's page and the fragrance can be smelled as it is emitted from a small hole in the center of the device panel. Similarly, our work also investigates a less common interaction modality by implementing temperature cues.

Mapsense [13] is a multi-sensory interactive map for visually impaired children. The map consists of a colored tactile map overlay on top of a touchscreen, speakers, and conductive tangibles. The tactile map overlay has some point of interest which can detect the conductive tangibles when placed on top. Additionally, the system provides the user with audio feedback communicating the name of a city when a point of interest is tapped twice with the finger. Based on that premise, there are several modes available: guiding function, audio discovery, and navigation. The guiding function consists of vocal indications guiding the user to the destination (one of the points of interest). On audio discovery, sound effects (such as the song of water when sailing or religious chants when reaching a church) were trigger while navigating the map. Lastly, the navigation mode allows the user to navigate between "points of interest", "general directions" and "cities".

There are also multi-sensory art exhibitions that are designed specifically for visually impaired people, so they can be considered multi-sensory assistive art exhibitions. The "Feeling Vincent Van Gogh" exhibition is one of those types of exhibitions aimed at the visually impaired [14]. The artwork is communicated to the users through a large variety of interactive elements such as sounds, smells, and 3D versions of Van Gogh's most famous artworks. The highly detailed 3D reproductions allow the visually impaired spectator to appreciate the brush strokes of Van Gogh. In addition, the visual part of the exhibition is carefully taken care of so sighted people (either by themselves or accompanying a visually impaired person) are also able to enjoy the multi-sensory experience.

2.4. Sound-Color Cross-Modality

Color is a continuous spectrum for which there have been several representation models [15]. Munsell's color model is one of the earliest ones. In it, the color is organized into three dimensions: hue, chroma (or saturation), and value (or brightness). Hue refers to the color itself. Brightness is an indication of the amount of white or black of the color. The brighter the color, the closer it is to white, and vice versa. Saturation is an indicator of the vividness (clearness) of a color. Another common dimension is the warm-cold spectrum of colors. The closer a color is to the red end of the visible spectrum, the warmer it is. On the contrary, the closer a color is to the blue end of the spectrum, the colder it is [16].

Wang et al. [17] explored the putative existence of cross-modal correspondences between sound attributes and beverage temperature. The results, after an online pre-study and the main study itself, confirmed that the experience of drinking cold water is associated with significantly higher pitches and faster tempo. One possible explanation for this kind of effect is the formation of emotional associations [18].

Hamilton-Fletcher et al. [19] presented a color-sound sensory substitution device which consisted on a color image explored by the user on a tablet device. The color was then turned into sound, which the user was able to listen to while moving the stylus over the image. In addition, the device, together with other two sensory-substitution devices, was given to ten blind users which, among other things, addressed the importance of the sounds to be not only understandable but also aesthetically engaging.

Regarding aesthetics in sounds for color substitution, Cho et al. [6] investigated possibilities for creating beautiful sounds for representing colors by replicating the three main characteristics of color: hue, chroma, and value, by matching them to three features of sound: timbre, intensity, and pitch. Then, two sets of musical sounds for expressing colors were designed and tested: VIVALDI and CLASSIC. User tests were conducted with eight sighted adults and 12 visually impaired users. The results showed that both sound-color mappings were useful and engaging for the participants and that users were able to identify with high accuracy the different colors by hearing the musical sounds of both sets of audios. The present work is a continuation of that work. Here, those two sound-color mappings are implemented into a complete sound-temperature-color coding. In addition, the relative usefulness of each set (VIVALDI or CLASSIC) for the multi-sensory system is also investigated, in order to choose the best one among the two for the multi-sensory system.

2.5. Temperature-Color Cross-Modality

Regarding cross-modality between temperature and color, works like [5,20] showed that there is a color-temperature association with the color warm–cold spectrum. In [20], the authors designed and performed a method for conveying color information both through thermal intensity and thermal change rate. The method was tested with ten users and showed successful results. In [5], a method for discerning up to 50 colors by feeling three different temperatures was designed. It included extensive interviews and user tests with visually impaired users. Both works proved that temperature cues as a way of expressing color can be a successful interaction method and that visually impaired users can differentiate temperatures quite comfortably as long as there is, at least, a difference of 3 °C between them. This fact will be applied in our multi-sensory method for expressing colors.

3. Method

3.1. Previous Method

As we stated above, this paper addresses and aims to improve previous works in which unisensory color-coding was performed and tested. While in [5] a sound-color mapping algorithm was designed, in [6] the developed algorithm and system was a temperature–color one. It is our goal to mix both these methods for increasing the number of colors the user can experience and for making the sensory experience more stimulating and engaging. However, for developing a multi-sensory system mixing both of them, it is necessary to find the best way of implementing both methods in a multi-sensory way in order to increase the easiness of color recognition and the comfortability of the sensory experience.

The sound–color code designed in [6] for each of the colors and their saturated, bright, and dark variants can be seen in Table 1. There, each of the 18 colors is represented by a musical excerpt either from a sound set consisting of variations on a Vivaldi piece or from a more ample classical repertoire sound set. Each one of these sound sets received a name: VIVALDI, and CLASSIC. The melody, rhythm, harmony, instrument, and tessitura of the sound files of both sound sets were selected in order to make a comprehensible representation of the hue and quality of each color. While more information about the musical selection and processing method can be seen in [6], it is important to notice that, regardless of the method, each sound file was created so its musical excerpt represents both the hue of the color and the quality of the color. Therefore, all sounds belonging to the dark colors have differences, which allows the user to distinguish the hue, but also have similarities that represent the dark dimension of the color. It is important to remember that this set of sounds come in two different versions, CLASSIC and VIVALDI, since one of the research's goals will be finding out which one of those set of sounds is more convenient for being implemented in the multi-sensory system.

Table 1. Sound coding colors with saturated/bright/dark colors (Adapted from ref. [6]). The sound wav files are provided separately as a supplementary materials.

Color Hues	Saturated Colors	Bright Colors	Dark Colors
Red	SCC-R	SCC-Red-L	SCC-Red-D
Orange	SCC-O	SCC-Orange-L	SCC-Orange-D
Yellow	SCC-Y	SCC-Yellow-L	SCC-Yellow-D
Green	SCC-G	SCC-Green-L	SCC-Green-D
Blue	SCC-B	SCC-Blue-L	SCC-Blue-D
Violet	SCC-V	SCC-Violet-L	SCC-Violet-D

Similarly, a temperature–color mapping was designed based on the previous work from [5], where it was proved that users were able to discern with high accuracy temperatures with intervals larger than 3 °C. As a result, any division of the comfortable temperature range with more than 3 °C between temperatures can be used for discerning several colors or the other dimensions of the colors. Consequently, the following temperature-color codes were designed, one for expressing hues (Table 2) and the other for expressing other color dimensions (Table 3). The original main method, based on [5], would be to have the user feel two temperatures, one expressing hue and the second one expressing the color dimensions of that same hue.

Table 2. Temperature-Color coding for color hue.

Color Hue	Temperature (°C)
Red	38
Orange	34
Yellow	30
Purple	26
Green	22
Blue	14

Table 3. Temperature-Color coding for color warm/bright/dark and cold dimensions.

Color Dimension	Temperature (°C)
Warm	38
Bright	30
Dark	22
Cold	14

3.2. Multi-Sensory Improved Method

It was possible to mix both those methods to create a multi-sensory experience of colors if both sound–color and temperature–color mappings were simplified and mixed, dividing the roles each one had. For example, hue could be expressed only through sound while the color quality could be expressed only by means of the temperature cues, or vice versa. This created a multi-sensory experience of the colors more engaging and convenient than the original unisensory experiences. In addition, the separation of hue and the other color dimensions into different sensory modalities may make memorization of the mapping easier and faster for the user. Both novel multi-sensory sound-temperature-color systems can be seen in Tables 4 and 5. In Table 4, a coding where color hue is expressed as sound, and the other color dimensions are expressed as temperature, is presented. On the contrary, in Table 5 the color hue is expressed through temperature and the other dimensions are expressed by means of sound cues. We created a naming convention for these two possible mappings by following the order of the sensory experience: sound-temperature-color coding and temperature-sound-color coding, respectively.

Table 4. Designed sound–temperature–color coding.

Color Hues	Saturated Color Sound	Bright/Dark (°C)	Warm/Cold (°C)
Red	SCC-R sound	(Temperature) 30/22	(Temperature) 38/14
Orange	SCC-O sound		
Yellow	SCC-Y sound		
Green	SCC-G sound		
Blue	SCC-B sound		
Violet	SCC-V sound		

Table 5. Designed temperature–sound–color coding.

Color Hues	Saturated Color Temp (°C)	Bright/Dark Sound	Warm/Cold Sound
Red	38	Warmest/Coldest Color From VIVALDI or CLASSIC set	Darkest/Brightest Color From VIVALDI or CLASSIC set
Orange	34		
Yellow	30		
Purple	26		
Green	22		
Blue	14		

For the satisfactory development of a multi-sensory color-coding system based on these methods, three are the elements that needed to be figured out:

1. Finding out whether it was better to represent color hue with sounds and the color dimensions with temperature, or vice versa.
2. Knowing which one was those warmest and coldest colors (that appear in Table 5) from the sound set and the temperature set. In other words, setting the dark dimension as an example, with many sounds representing the quality of "dark" such as "dark green" or "dark red", it was important to figure out which one of all those dark color-sounds expressed the dark quality of color the best.

3. Finding out which sound set from the two options (VIVALDI and CLASSIC) is in general terms more suitable to express the different dimensions of color.

Tests were performed for answering these questions. These tests and their analysis and results will be explored in the following sections.

3.3. Hardware

For the temperature system, a thermal display that was developed for prior research was used [7]. The system consisted of an array of Petliers, which is a device that releases heat on one side and absorbs it through the other one when electric current passes through it. All the Petliers are driven by a dual H-bridge board controlled through an Arduino Mega microcontroller. Figure 1 shows the array of Petliers that was used during testing. An array of petliers was used in order to be able to set each petlier at a different temperature from the beginning, so the test could be perform faster. The temperature range of each petlier was from 10 °C to 40 °C.

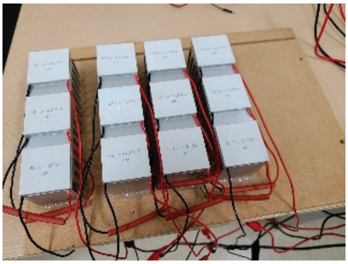


Figure 1. Peltier array temperature device and controller (Adapted from ref. [7]).

4. Experimentation

Inspired by the research related to music and color association [21], we chose to use a common semantic and emotive link to compare the different modalities. We used as a link a list of adjectives and the emotional-semantic space they created. Osgood [22] simplified the semantic space of adjectives into three aspects, which are (1) evaluation (like–dislike), (2) potency (strong–weak), and (3) activity (fast–slow). The adjectives we decided to adopt in this research are the pairs of adjectives that people are familiar with; emotion, shape, location, activity, texture, contrast, temperature, sound characteristics, and so on.

As a result, 18 pairs of adjectives, each set of six representing one of the three different Osgood semantic spaces, were chosen. In other words, there was a set of adjectives representative of the whole semantic space with which to create a comparison between modalities. That modified list of adjectives from Osgood [22] can be seen in Table 6, where the column indicates the Osgood adjective list.

The main idea for the test was the following: classifying the adjectives into warm/cold and dark/bright dimensions, correlating the different adjectives to the different sound and temperature cues, and, as a result, giving to each sound and temperature a warm/cold and dark/bright score to find out which cues expressed the warm, cold, dark or bright dimensions the best. Additionally, those scores would help us find out whether VIVALDI or CLASSIC was a better option for representing the color dimensions and whether it was better to express color hue with temperature and color dimension through sounds or vice versa.

Table 6. Selected semantic adjectives (modified from Osgood [22]) to be used as a common link between sound coding colors and color dimensions like warm/cold and bright/dark.

Evaluation	Potency	Activity
Bright~Dark	Strong~Weak	Fast/Agile~Slow/Dull
Clear~Cloudy	Hard~Soft (Pressure)	Noisy~Quiet
Joyful~Depressed	Rough~Smooth (Touch vibration)	Extroverted~introverted
Calm~Tense	Pointed (Kiki)~Round (Bouba) Sharp~Dull (Touch)	Centrifugal~Centritorial Dilated~Constricted
Comfortable~Anxious	Far~Near	Passionate~Depressed
Warm~Cold	High~Low (e.g., high-pitch~low-pitch)	Active~Inactive

In total, three different types of tests were performed: one for classifying the different adjectives into the warm/cold and dark/bright main dimensions, a second one for finding out the weights of each particular pair of adjectives with the dimensions they were classified into, and one last test which had the users selecting adjectives for each one of the temperatures and sound cues. The three test were analyzed and the results presented in the analysis section.

The number of participants was 18. They were college students who had normal eyesight and an average age of 21.5 years old. Since tests were performed on different days, not all of them were able to participate in all tests. 15 users participated in the first two tests and a total of 12 users in the third one. All the test sessions included an explanation of the test and its procedure before starting the surveys. Test duration varied, with the first and second one lasting together for about 25 min per person and the last test lasting for around 45 min per person. The testing procedure was the following:

(1) Introducing the context of the research, explaining about visually impaired people and art exploration, color, and multi-sensory systems.
(2) Introduction of the Petlier thermal display prototype (only during the third test).
(3) Lastly, the tester interacted with the different audio or temperature cues and was given the survey questions. Each audio cue was heard as many times as the user needed (the user was given the playback control for repeating it until he/she was ready to answer the survey question related to that cue). Similarly, the user was able to feel the different temperatures as many times as desired before answering to the survey question related to that temperature cue.

4.1. First Test: Classifying Pair of Adjectives

Users were asked to select which color dimension was related to each pair of adjectives the most: the bright/dark dimensions, the warm/cold dimensions, both of them or none. As an example, the answer of one of the users for the adjective pair “noisy–quiet” can be seen in Table 7. There, the user stated that the dimensions that is more correlated to the “noisy–quiet” adjective pair was the bright/dark dimension of the colors. Table 8 presents the results after summing up all the answers. The number indicates the total of users that ticked an option. Since some users selected the option “both”, it is possible that the sum of the bright/dark and warm/cold counts results in a number higher than that of the total number of users.

Table 7. Example of the dimension survey for the adjective pair "noisy-quiet".

Adjective Pair	Bright/Dark	Warm/Cold	Both	None
Noisy-Quite	×			

Table 8. Total count of answers for all adjectives after surveying 15 participants. L/D is the bright/dark dimension and W/C is the warm/cold dimension.

Adjective Pairs		L/D	W/C
Noisy	Quiet	15	1
Cloudy	Clear	10	8
Joy	Depressed	13	3
Round	Pointed	1	6
High	Low	8	5
Slow	Agile	4	3
Rough	Smooth	3	6
Hard	Soft	4	9
Strong	Weak	6	4
Active	Inactive	13	6
Calm	Tense	4	7
Near	Far	8	4
Introverted	Extroverted	11	7
Passionate	Depressed	13	8
Centritorial	Centrifugal	7	5

4.2. Second Test: Calculating Weights of Each Pair of Adjectives for Each Dimension

In this test, the users were asked to rate the correlation of each pair of adjectives with each individual dimension: warm, cold, bright, and dark. The scale used was from −2 to 2. Table 9 shows the answer of one of the users for the adjective pair "noisy-quiet" in relation to the color dimension "bright".

Table 9. Example of the weight score survey answered by one participant for the dimension bright and the adjective pair "noisy–quiet". Participants were asked to give a weight score of each pair of adjectives for each one of each color dimension.

	Bright					
	−2	**−1**	**0**	**1**	**2**	
Noisy		×				Quite

Once all the users rated all the adjective pairs in all the dimensions, four tables with the weights for each dimension and pair of adjectives, graded from −2 to +2, were made. Negative numbers indicate that the dimension is directly correlated with the adjective from the left column, while positive numbers indicate that it is directly correlated with the adjective from the right column. The number indicates how strongly the correlation is, with −2 and +2 being the strongest correlation and a zero meaning there is no correlation at all. As an example, in Table 10, the weights for the dimension of bright can be seen. Similar tables for warm, cold, and dark dimensions were also made but omitted here for brevity.

Table 10. All weight scores for each adjective pair or the "bright" dimension. Similar weight tables were acquired for the warm, cold and dark dimensions. The standard deviation is indicated next to the value in parenthesis.

	LIGHT	
Noisy	−1.13 (0,264)	Quiet
Cloudy	1.66 (0,261)	Clear
Joyful	1.86 (0,087)	Depressed
Round	0.06 (0,239)	Pointed
High	−0.73 (0,220)	Low
Slow	0.60 (0,157)	Agile
Rough	0.33 (0,203)	Smooth
Hard	0.80 (0,252)	Soft
Strong	−0.53 (0,247)	Wear
Active	−1.66 (0,121)	Inactive
Calm	−0.13 (0,228)	Tense
Near	−0.66 (0,203)	Far
Introverted	1.53 (0,185)	Extroverted
Passionate	−1.2 (0,285)	Apathetic
Centrifugal	0.93 (0,199)	Centritorial

4.3. Final Test: Linking Adjectives to Each Modality Cue

During the final test, all modalities and all cues (VIVALDI sound, CLASSIC sound, and temperatures) were given a score for each one of the pairs of adjectives stated above. For example, in the case of temperature modality, the process was the following.

First, the user would feel one of the temperatures seen in Table 2, and, for each one of those temperatures, a form sheet like the one shown in Table 11 would be filled up. As an example, Table 11 has been filled up with all the answers of one of the users after having felt the 38 °C temperature Peltier. This same process was performed not only with the rest of the temperatures but also after listening to each one of the sounds from the VIVALDI and CLASSIC sounds. As a result, there was an adjective-graded sheet for each one of the cues (each one of the temperatures and sounds) of the three different modalities contemplated.

Table 11. Adjective score table of the 38 °C temperature cue filled up with the answers of one of the participants. The users answered similar score tables after hearing to each musical sound and feeling each temperature.

			38 °C			
	1	**2**	**3**	**4**	**5**	
Noisy	×					Quiet
Clear					×	Cloudy
Joy		×				Depressed
Pointed				×		Round
High	×					Low
Agile					×	Slow
Rough		×				Smooth
Soft			×			Hard
Strong	×					Weak
Active	×					Inactive
Tense			×			Calm
Near	×					Far
Extroverted					×	Introverted
Passionate	×					Apathetic
Centrifugal					×	Centritorial

From all the testers' answers, an average score on the scale of [−2, 2] was calculated. The value of −2 would be the equivalent to all testers giving a score of 1 to the adjective during the survey. A score of +2 would be the result of all participants giving a score of 5. As an example, the results for the saturated red of the VIVALDI set of sounds can be seen in Table 12.

Table 12. Adjective score results for VIVALDI red saturated color. Similar adjective scores were acquired for all colors and temperatures from the third test.

RED-S	
Adjective Pair	**Score**
Noisy – Quiet	−0.750
Cloudy – Clear	1.083
Joy – Depressed	−1.333
Round – Pointed	0.667
High – Low	−1.167
Slow – Agile	0.917
Rough – Smooth	−0.083
Hard – Soft	0.417
Strong – Weak	−0.917
Active – Inactive	−1.333
Calm – Tense	0.167
Near – Far	−0.750
Introverted – Extroverted	0.833
Passionate – Apathetic	−1.167
Centritorial – Centrifugal	0.333

5. Analysis and Results

5.1. Analysis

Analysis of the results was performed for finding out the best design of the multi-sensory color-mapping system. The analysis process was the following.

First, by means of the results presented during the second test, we created a weight table indicating all the pairs of adjectives and their relative weights to each one of the four dimensions: bright, warm, cold, and dark (Table 13). However, only the adjectives that were selected during the first test as related to that particular dimension were taken into account when filling up the table. Therefore, empty weights are the set of adjectives that were uncorrelated to that particular dimension on the first test's results. Both the weight table and the adjective score (like the one shown in Table 12) for each one of the sounds and temperature cues were used to calculate a bright score, dark score, warm score, and cold score for each cue. The basic formula for each of those four scores was the following.

$$S_{cd} = \sum_{i} (W_{di} + S_{ci}) \tag{1}$$

where S_{cd} is the total score of the dimension "*d*" for the cue "*c*", W_{di} is the weight of the dimension "*d*" for the adjective pair "*i*", and S_{ci} is the score of the adjective pair "*i*" for the cue "*c*".

Table 13. Weight score summary of each adjective pair for each one of the four dimensions.

Adjective Pairs		**L/W/C/D Weights**			
−	+	**Light**	**Dark**	**Warm**	**Cool**
Noisy	Quite	−1,133	1,467		
Cloudy	Clear	1,667	−1,867	1,800	−1,200
Joyful	Depressed	−1,867	1,800		
Round	Pointed			−0,667	0,733
High	Low	−0,733	0,800		
Agile	Slow				
Rough	Smooth			1,067	−0,533
Hard	Soft			1,533	−0,867
Strong	Weak				
Active	Inactive	−1,667	1,533	−0,933	0,200
Calm	Tense			−0,600	−0,067
Near	Far	−0,667	0,867		
Introverted	Extroverted	1,533	−1,467	0,533	−0,333
Passionate	Apathetic	−1,200	1,600	−1,000	0,600
Centrifugal	Centritorial				

After applying the formula for all dimensions and each sound and temperature cue, three color dimension score tables were created, one for VIVALDI and all its sounds, other one for CLASSIC and all its sounds, and the last one for all the temperatures. An example of all the scores for the VIVALDI set can be seen in Table 14. Only the positive results are relevant, since it indicates cues that are directly correlated to the different dimensions. In Figures 2–5, bar plots graph indicating all the positive scores of all sets for each dimension can be seen. In Figure 6, the highest store per dimension for each set is presented. Lastly, in Figures 7–9, all the scores of all the cues (negative values included) are presented in a boxplot graph for each one of the methods and dimensions.

Table 14. Color dimension scores for VIVALDI. The scores are within the range [−100,100].

VIVALDI					
Color	**S/L/D**	**Light**	**Dark**	**Warm**	**Cool**
Red	S	35,635	−38,247	15,035	−7,448
	L	47,361	−51,163	12,587	−6,354
	D	−39,149	42,274	−15,382	7,014
Orange	S	42,882	−46,076	12,170	−6,337
	L	44,878	−48,628	19,028	−10,295
	D	−45,920	49,479	−9,392	5,104
Yellow	S	43,750	−47,049	17,014	−9,149
	L	39,792	−41,910	18,542	−10,486
	D	−41,215	44,427	−14,132	7,413
Purple	S	39,931	−43,090	17,795	−9,583
	L	40,278	−43,767	17,118	−9,670
	D	−35,295	36,892	−27,674	15,139
Green	S	28,021	−29,757	28,247	−16,528
	L	38,733	−41,580	26,024	−14,097
	D	−44,253	47,222	−10,729	5,590
Blue	S	39,913	−42,986	25,868	−14,045
	L	30,503	−32,431	22,326	−11,146
	D	−39,549	42,899	−8,785	4,271

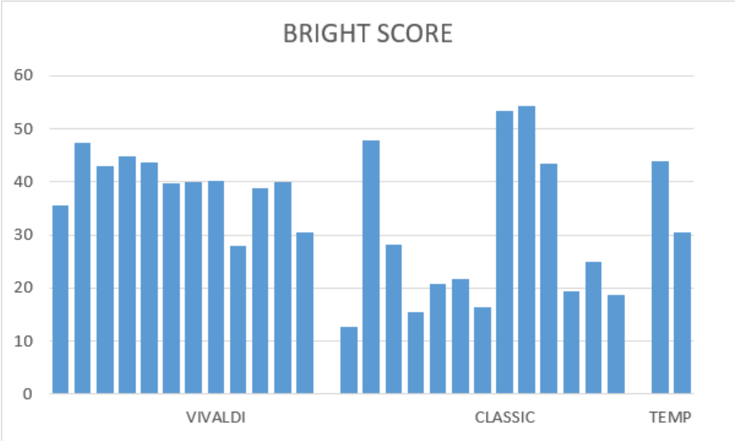


Figure 2. Positive scored cues in all sets for the bright dimension.

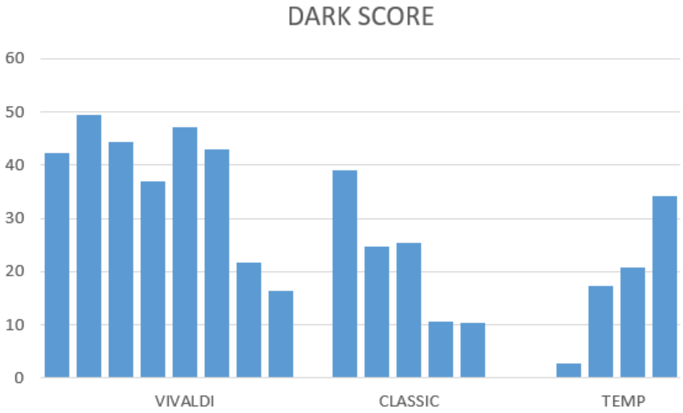


Figure 3. Positive scored cues in all sets for the dark dimension.

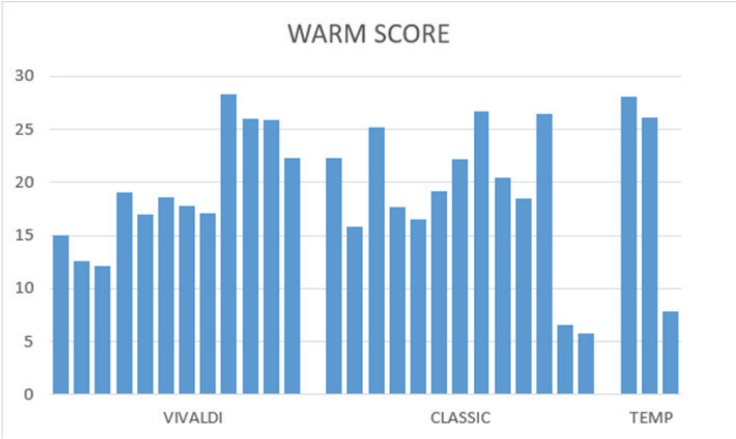


Figure 4. Positive scored cues in all sets for the warm dimension.

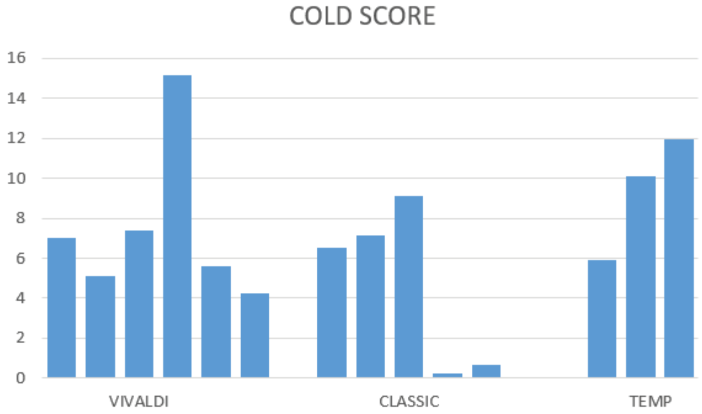


Figure 5. Positive scored cues in all sets for the cold dimension.

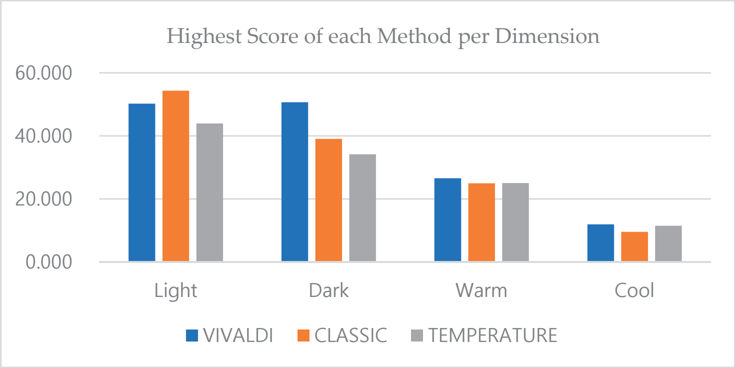


Figure 6. Highest score graph of the three methods for each of the four dimensions.

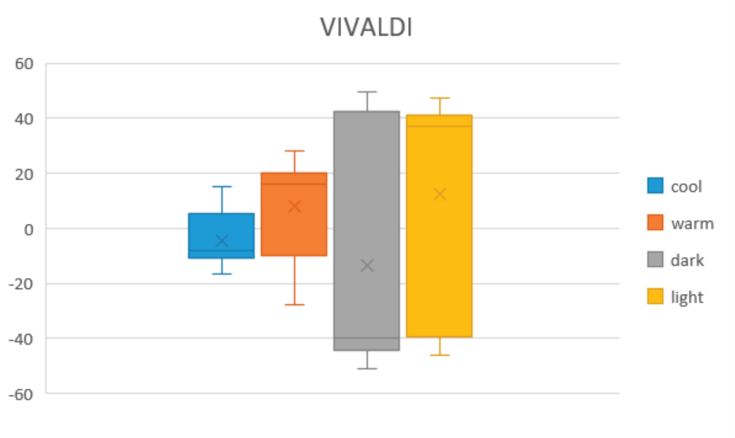


Figure 7. Boxplot of the four dimensions for all points for VIVALDI. The average, max., and min. values, together with the different percentiles, can be seen.

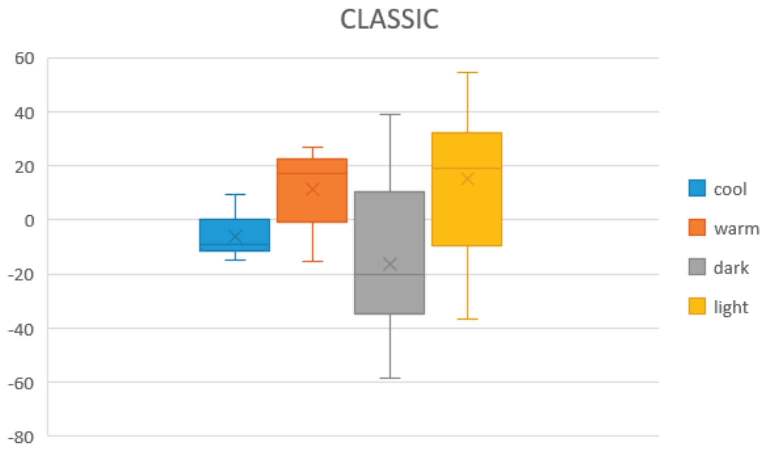


Figure 8. Boxplot of the four dimensions for all points for CLASSIC. The average, max., and min. values, together with the different percentiles, can be seen.

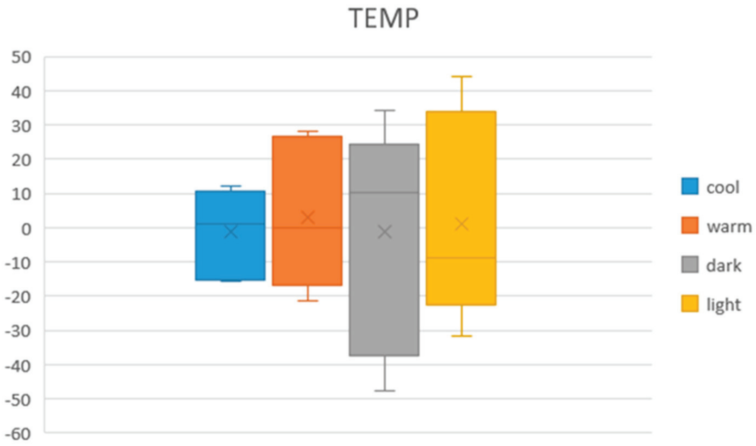


Figure 9. Boxplot of the four dimensions for all points for TEMPERATURE. The average, max., and min. value, together with the different percentiles, can be seen.

5.2. Results

These final scores can aid us in finding out which VIVALDI, CLASSIC and temperature cues are the ones that express the bright, dark, warm, and cold dimensions of colors the best. For finding that out, it is necessary to select the sound and temperature cues that scored the highest for each dimension. Table 15 presents the cues that score the highest for each dimension and method and the name of that cue.

Table 15. Highest scores attained by each method for each color dimension. The name of the cue which attained that highest score can be seen inside the parenthesis.

Method	Bright	Dark	Warm	Cold
VIVALDI	47.361 (RED-L)	49.479 (Orange-D)	28.247 (Green-S)	15.139 (Purple-D)
CLASSIC	54.323 (PURPLE-L)	39.063 (Red-D)	26.701 (Purple-L)	9.097 (Orange-D)
TEMP	43.935 (38 °C)	34.190 (14 °C)	28.056 (38 °C)	11.698 (14 °C)

Several interesting points can be observed from these results:

1. The CLASSIC sound set's cue which best expressed the bright and warm dimensions of color was the same sound. This was not convenient since users would not be able to discern whether the sound was expressing brightness or warmness. The cues for representing each dimension had be all different from each other so the user could relate each particular cue to a particular dimension.
2. Similarly, the highest temperature of 38 °C is the best one for representing both the warmness and brightness of the color. Additionally, the coldest temperature of 14 °C was the best temperature cue for conveying darkness and coldness. Like with the CLASSIC sound set, this reiteration of cues made temperature modality not suitable for representing the color dimensions.
3. On the contrary, none of the VIVALDI sounds is repeated. Each dimension is conveyed best by a different musical sound cue.
4. Except for conveying brightness, in which CLASSIC sound reached a higher score, VIVALDI sound reached higher scores for all the other three dimensions when compared to both the CLASSIC and temperature sets.

Taking into consideration these observations, it was possible to reach several conclusions. First of all, all the cues that expressed most accurately the different dimensions of color were clearly defined. Additionally, it is clear that, for expressing color dimensions, VIVALDI sounds were better than CLASSIC sounds since, except for one dimension, the scores reached by VIVALDI sound cues were higher. In addition, it was better to use temperatures as a way of expressing color hues and sounds to express color dimensions, since, as observed above, the warmest and the coldest temperature conveyed the best two dimensions each, but it was convenient for the user that each dimension was represented by one and only one cue. Therefore, temperature was not a good modality for representing color dimensions.

It can be concluded that the best multi-sensory algorithm was one where the temperature cues expressed color hue and VIVALDI sounds expressed the other color dimensions. In other words, the best multi-sensory method was the temperature–sound–color coding method presented in Table 4, which is shown now in its final complete state in Table 16. The algorithm of Table 16 is the temperature–sound–color designed through the tests and results that were observed.

Table 16. Final temperature–sound–color coding.

Color Hues	Saturated Color Temp (°C)	Bright	Dark	Warm	Cold
Red	38	VIVALDI RED-L	VIVALDI ORANGE-D	VIVALDI GREEN-S	VIVALDI PURPLE-D
Orange	34				
Yellow	30				
Purple	26				
Green	22				
Blue	14				

5.3. Final Temperature–Sound Coding Multimodal Test and Results

For assessing the final temperature–sound multimodal coding that was designed (Table 16, a final multimodal test with the final system was performed. The number of participants was 12. They were college students who had normal eyesight and an average age of 22 years. The test sessions included an explanation of the test and the method, a short training time so the user could familiarize himself/herself with the different temperatures and sounds, and a final test in which the user had to guess which color the multimodal system was representing through its sound and temperature. Test duration lasted around 25 min per person. The testing procedure was the following:

(1) Introducing the context of the research, the system and the method (5 min).
(2) Explanation of the 10 different cues and training time for both the six temperature cues and the four sound cues (10 min)
(3) After that the users were tested on 10 colors. The users were given the multimodal feedback related to each color and were asked to find out which color it represented. In other words, after feeling both the temperature cue and hearing sound at the same time, they were asked about the color. The users were given a printed page similar to Table 16 so they did not need to memorize what each temperature represented. All the users were given the same questions in the same order.

The accuracy during the test can be seen in Table 17, and a list with all the wrong and correct answers, divided into hue (represented through temperature) and dimension (represented through sound) during the test can be seen in Table 18. Additionally, in Tables 19 and 20, confusion matrices for both the hue and the dimension of the color (each one presented to the user as a temperature cue and as a sound, respectively) are presented. The total accuracy was 67.5%. Considering only the hue of the colors, the accuracy went up a little bit to 71.6%. On the other hand, the accuracy of guessing the dimension of the color through sound reached up to 92.5%. While 67.5% might not seem too high, it is important to consider the limited training time the users were given with. With a longer training time, the accuracy would likely increase. In addition, previous work [5] seems to suggest that the capacity for visually impaired people to discern temperatures might be higher than the one reflected here (with sighted users). On the other hand, the confusion matrices show clearly that the main recognition problem is caused by the colors yellow and orange, whose temperature cues (30 °C and 34 °C) are not easily discernible. Therefore, the temperatures cues for orange and yellow, and that particular temperature range of (30 °C, 34 °C) are things that will have to be modified and improved in the future in order to improve the system. Overall, the results are promising and the system seems to have the potential to be developed and improved on future iterations.

Table 17. Accuracy when discerning a color correctly (both hue and dimension) and same value when only taking hue or dimension into account.

Percentage of Correct Answers		
Total	**Only Hue**	**Only Dimension**
67.5%	71.6%	92.5%

Table 18. Final multimodal system test results for 12 users. "H" means the answer related to hue, "D" the one related to the dimension. As a result, if for the case of Dark Red a user answered Dark Blue, the "H" answer would be wrong while the "D" answer would be correct. "Tot_D" gives the total number of correct answers per user per dimension. "Tot" gives the total correct answers per user considering both hue and dimension together.

Col	**U1**		**U2**		**U3**		**U4**		**U5**		**U6**		**U7**		**U8**		**U9**		**U10**		**U11**		**U12**	
	H	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**	**H**	**D**
Dark Red	✗	✗	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓
Light Yellow	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓
Cool Green	✗	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Warm Orange	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗	✓
Dark Purple	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Warm Blue	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cool Red	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓
Light Yellow	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓
Cool Green	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Dark Orange	✗	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓
Tot_D	7	8	7	9	7	9	7	9	9	10	6	9	9	10	8	10	7	10	6	7	7	10	6	10
Tot.	6		6		6		7		9		5		9		8		7		5		7		6	

Table 19. Confusion matrix for the six color hues (presented through temperature cues). The main recognition problem is caused by the colors yellow and orange, whose temperature cues (30 °C and 34 °C) are not easily discernible.

	Red	**Orange**	**Yellow**	**Purple**	**Green**	**Blue**
Red	17	0	0	0	0	0
Orange	7	11	10	0	0	0
Yellow	0	13	14	1	0	0
Purple	0	0	0	11	0	0
Green	0	0	0	0	20	0
Blue	0	0	0	0	4	12

Table 20. Confusion matrix for the four dimensions of the colors (presented through sounds).

	Bright	**Dark**	**Warm**	**Cool**
Bright	24	0	0	0
Dark	0	31	0	5
Warm	0	2	24	0
Cool	0	3	0	31

6. Discussion

Throughout this work, three types of tests were performed to assess a multi-sensory method to convey color information by using temperature and musical sound cues. The results showed the intrinsic relationship of several sounds and temperature-based cues to the warm, cold, dark, and bright dimensions of colors. Based on these results, a promising multi-sensory system for expressing color through sound and temperature was designed. It was shown that, from both sound codes designed in [6], VIVALDI color coding was better for the task than the CLASSIC color coding. Similarly, different temperatures were also assessed and the best temperature cues to apply for each dimension were also found. Lastly, the most appropriate temperature–sound–color coding method was defined by defining the method and all its temperature and sound cues. In addition, this last coding method was tested with 12 different users. The final results show that the system is a promising solution for expressing color to users in a multimodal way.

The addition of a multi-sensory sound–temperature interaction as a way of communicating color can open the door to many new ways of experiencing art for the VIP. In addition, the extensive adjective survey for each one of the sounds and temperature cues can be handy for other investigations in the future. Finally, we expect this research to serve encouragement for many more multi-sensory application research and systems that could improve the way in which visually impaired people experience art.

7. Future Work

There are many ways in which this work could be continued or improved on such as:

1. The problem for differentiating the colors yellow and orange need to be addressed, since it is the main bottleneck for reaching high accuracy on the system.
2. Installing the multi-sensory system and using its method for performing complete on-field user tests with different artworks and on different exhibitions for the visually impaired people. This could be the beginning of research about the effect of multi-sensory systems on art engagement and artwork memorization by the visually impaired people.
3. Finding a way to increase the number of colors that can be expressed by the system, without increasing its complexity or making it less engaging.
4. Finding other applications based on the same sound-temperature concept or expressing other features of the artworks.

8. Conclusions

In this work, a multi-sensory sound-temperature cross-modal mapping method for conveying a total of 24 colors (six hues and four different color dimensions of dark, bright, warm and cold) was designed. The mapping was based on previous unisensory temperature-color and sound-color codes. They were adapted and tests were performed for finding out the best way to mix them for reaching a satisfactory multi-sensory cross-modal code. In addition, a semantic study of the musical sound cues and temperatures from those methods were acquired. The results showed that the musical sounds and temperatures can be used as a substitute for color hues and color dimensions. Additionally, the data from the results guided us into designing the optimum multi-sensory temperature-sound method based on those musical sounds and temperatures. We hope this work can encourage researchers to consider thermal and sound multi-sensory interaction both as a substitute for color and as a way to improve accessibility for the visually impaired people in visual artworks and color experience.

Supplementary Materials: The sound wav files on CLASSIC and VIVALDI sound color codes in Table 1 are available online at https://www.mdpi.com/article/10.3390/electronics10111336/s1.

Author Contributions: Conceptualization, J.-D.C.; methodology, J.-D.C.; software, G.C.; validation, J.-D.C. and G.C.; formal analysis, J.-D.C. and J.I.B.; investigation, J.-D.C., J.I.B. and G.C.; resources, J.-D.C.; data curation, G.C.; writing—original draft preparation, J.-D.C.; writing—review and editing, J.-D.C. and J.I.B.; supervision, J.-D.C.; project administration & funding acquisition, J.-D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea, grant number 2018M3C1B6061353.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Sungkyunkwan University (protocol code: 2020-11-005-001, 22 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The author declares that they have no conflicts of interest.

References

1. Shams, L.; Seitz, A.R. Benefits of multisensory learning. *Trends Cogn. Sci.* **2008**, *12*, 411–417. [CrossRef] [PubMed]
2. The Guggenheim Museums and Foundation. Mind's Eye Programs. Available online: https://www.guggenheim.org/event/event_series/minds-eye (accessed on 28 February 2021).
3. The Metropolitan Museum of Art. For Visitors Who Are Blind or Partially Sighted. Available online: https://www.metmuseum.org/events/programs/access/visitors-who-are-blind-or-partially-sighted (accessed on 28 February 2021).
4. Museum of Moden Art. Accessibility. Available online: https://www.moma.org/visit/accessibility/index#individuals-who-are-blind-or-have-low-vision (accessed on 28 February 2021).
5. Bartolome, J.D.I.; Quero, L.C.; Cho, J.; Jo, S. Exploring Thermal Interaction for Visual Art Color Appreciation for the Visually Impaired People. In Proceedings of the 2020 International Conference on Electronics, Information, and Communication (ICEIC), Barcelona, Spain, 19–22 January 2020; pp. 1–5.
6. Cho, J.D.; Jeong, J.; Kim, J.H.; Lee, H. Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments. *Electronics* **2020**, *9*, 1981. [CrossRef]
7. Bartolomé, J.I.; Cho, J.D.; Quero, L.C.; Jo, S.; Cho, G. Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People. *Electronics* **2020**, *9*, 1939. [CrossRef]
8. Lloyd-Esenkaya, T.; Lloyd-Esenkaya, V.; O'Neill, E.; Proulx, M.J. Multisensory inclusive design with sensory substitution. *Cogn. Res. Princ. Implic.* **2020**, *5*, 1–15. [CrossRef]
9. Faustino, D.B.; Gabriele, S.; Ibrahim, R.; Theus, A.-L.; Girouard, A. SensArt Demo: A Multi-sensory Prototype for Engaging with Visual Art. In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, New York, NY, USA, 17 October 2017; pp. 462–465. [CrossRef]
10. Pursey, T.; Lomas, D. Tate Sensorium: An experiment in multisensory immersive design. *Senses Soc.* **2018**, *13*, 354–366. [CrossRef]
11. Baumgartner, T.; Lutz, K.; Schmidt, C.F.; Jäncke, L. The emotional power of music: How music enhances the feeling of affective pictures. *Brain Res.* **2006**, *1075*, 151–164. [CrossRef] [PubMed]
12. Edirisinghe, C.; Podari, N.; Cheok, A.D. A multi-sensory interactive reading experience for visually impaired children; a user evaluation. *Pers. Ubiquitous Comput.* **2018**, 1–13. [CrossRef]
13. Brule, E.; Bailly, G.; Brock, A.; Valentin, F.; Denis, G.; Jouffrais, C. MapSense: Multi-Sensory Interactive Maps for Children Living with Visual Impairments. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 445–457.
14. Feeling Van Gogh, Van Gogh Museum, Amsterdam. Feel, Smell and Listen to the Sunflowers. Available online: https://www.vangoghmuseum.nl/en/visit/whats-on/feeling-van-gogh (accessed on 28 February 2021).
15. Ibraheem, N.A.; Hasan, M.A.; Khan, R.Z.; Mishra, P.K. Understanding color models: A review. *ARPN J. Sci. Technol.* **2012**, *2*, 265–275.
16. Bailey, R.J.; Grimm, C.M.; Davoli, C. The Real Effect of Warm-Cool Colors. Report Number: WUCSE-2006-17. All Computer Science and Engineering Research. 2006. Available online: https://openscholarship.wustl.edu/cse_research/166 (accessed on 31 May 2021).
17. Wang, Q.; Spence, C. The Role of Pitch and Tempo in Sound-Temperature Crossmodal Correspondences. *Multisens. Res.* **2017**, *30*, 307–320. [CrossRef]
18. Levitan, C.A.; Charney, S.; Schloss, K.B.; Palmer, S.E. The Smell of Jazz: Crossmodal Correspondences Between Music, Odor, and Emotion. In Proceedings of the CogSci, Pasadena, CA, USA, 22–25 July 2015; pp. 1326–1331.
19. Hamilton-Fletcher, G.; Obrist, M.; Watten, P.; Mengucci, M.; Ward, J. I Always Wanted to See the Night Sky. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; Association for Computing Machinery (ACM), San Jose, CA, USA, 7–12 May 2016; pp. 2162–2174.
20. Jia, Z.; Li, J.; Chen, C. Design and Evaluation of a Thermal Tactile Display for Colour Rendering. *Int. J. Adv. Robot. Syst.* **2015**, *12*, 162. [CrossRef]
21. Palmer, S.E.; Schloss, K.B.; Xu, Z.; Prado-Leon, L.R. Music-color associations are mediated by emotion. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8836–8841. [CrossRef] [PubMed]
22. Osgood, C.E.; Suci, G.; Tannenbaum, P.H. *The Measurement of Meaning*; University of Illinois Press: Urbana, IL, USA, 1957.

electronics

Article

Multi-Sensory Color Code Based on Sound and Scent for Visual Art Appreciation

Luis Cavazos Quero, Chung-Heon Lee and Jun-Dong Cho *

Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Korea;
luis@skku.edu (L.C.Q.); ch.lee@skku.edu (C.-H.L.)
* Correspondence: jdcho@skku.edu

Abstract: The development of assistive technologies is improving the independent access of blind and visually impaired people to visual artworks through non-visual channels. Current single modality tactile and auditory approaches to communicate color contents must compromise between conveying a broad color palette, ease of learning, and suffer from limited expressiveness. In this work, we propose a multi-sensory color code system that uses sound and scent to represent colors. Melodies express each color's hue and scents the saturated, light, and dark color dimensions for each hue. In collaboration with eighteen participants, we evaluated the color identification rate achieved when using the multi-sensory approach. Seven (39%) of the participants improved their identification rate, five (28%) remained the same, and six (33%) performed worse when compared to an audio-only color code alternative. The participants then evaluated and compared a color content exploration prototype that uses the proposed color code with a tactile graphic equivalent using the System Usability Scale. For a visual artwork color exploration task, the multi-sensory color code integrated prototype received a score of 78.61, while the tactile graphics equivalent received 61.53. User feedback indicates that the multi-sensory color code system improved the convenience and confidence of the participants.

Keywords: assistive technology; multi-sensory interface; auditory interface; scent interface; color

Citation: Cavazos Quero, L.; Lee, C.-H.; Cho, J.-D. Multi-Sensory Color Code Based on Sound and Scent for Visual Art Appreciation. *Electronics* **2021**, *10*, 1696. https://doi.org/10.3390/electronics10141696

Academic Editor: Mazdak Zamani

Received: 11 March 2021
Accepted: 12 July 2021
Published: 15 July 2021

1. Introduction

In 2010, the World Health Organization estimated that at least 285 million people worldwide have a form of visual impairment, 39 million of whom are blind. Moreover, the number will likely increase due to several factors, including an aging population. Driven by advances in technology and the need to improve the quality of life in visually impaired and blind people, the research community has recently experienced a growing interest in developing assistive technology [1]. Most of the focus has been on mobility, navigation, object recognition, and, today, in improving access to printed media and social interaction [2]. Although access to artistic culture prevents the isolation and fosters the proper functioning of visually impaired and blind people in society and communication [3], there is a limited amount of research regarding the development of assistive technology to access visual artworks contents. Traditionally, access to visual artworks by blind and visually impaired people has been made possible through accessible tours and workshops [4,5], audio guides [6], Braille leaflets with embossed tactile diagrams [7], tactile 3D models [8], and, more recently, by interactive interfaces that provide rich and location-based information through multiple sensory channels [9–13]. However, these methods fail to facilitate experiencing the artwork's color contents. Color is a fundamental artwork element that expresses depth, form, and movement. In addition, it strongly affects the mood, harmony, importance, and emotions expressed. Therefore, perceiving it is essential for a complete artwork experience.

In this work, we propose a multi-sensory color code for visual artworks that uses melodies and scents. We selected the audio and olfactory components following a cross-modal correspondence approach since there is evidence that it can facilitate learning [14–17]. Most of the literature has traditionally expressed the color in a uni-sensory way, but, as the number of encoded colors increases, the mental effort involved to learn and use the code also increases. We presume that using a hybrid approach eases the effort to recognize the encoded colors and improves the color identification success rate compared to the uni-sensory method. Additionally, the multi-sensory approach could be potentially more expressive as the different modalities could complement each other. In this work, we also explore using the proposed color code system and implement a sensory substitution device prototype for visual art color content exploration. Participants were requested to explore a visual artwork through the prototype and compare it with a tactile graphics alternative to elicit feedback and identify the advantages and challenges of the multi-sensory approach. The main contributions of this work are:

1. A formative study with five blind, one visually impaired, and one art professor participants to understand the importance and challenges of exploring and experiencing color in visual artworks.
2. A sound and scent-based multi-sensory color code system that can represent different colors across six different hues and the lightness (light-dark) color dimensions. We further extend the color dimensions to include the temperature (warm-cool) color dimensions.
3. A study with eighteen participants comparing the identification success rate achieved using a uni-sensory audio color code system and our multi-sensory approach.
4. A visual sensory substitution device prototype for visual artwork exploration that uses the proposed multi-sensory color code system.
5. A System Usability Scale survey with eighteen participants to compare the usability of the prototype and a tactile graphics alternative.

2. Related Work

2.1. Auditory Representation of Color

Visual Sensory Substitution Devices (SSDs) are systems that convey visual information through non-visual modalities. Typically, they translate visual features (vertical position, horizontal position, luminance, and color) into tactile or auditory stimuli (texture, shape, pitch, frequency, time, intensity) [18]. The translation often is based on the principle of cross modal correspondence. This principle refers to a compatibility effect between attributes or dimensions of a stimulus in different sensory modalities. For this work, the compatibility effect studied is between the visual stimulus and the sound and smell sensory modalities. The specific case of compatibility between color and audio is known as color sonification, and the set of associations or mappings made between sound and color is known as Sound Color Code (SCC).

Examples of SSDs that use color sonification are:

Soundview [19] represents the color of a point from an image selected with a stylus using a tablet. Color sonification maps the color hue, saturation, and brightness to a white noise sound. This sound is then modified using a low-pass filter. The color's brightness determines the filter's cutoff frequency. The color's hue determines which filter, out of 12 pitch-filters, modulates the signal. The color saturation affects the extent to which the pitch filter modifies the signal. Soundview has been evaluated for 2D geometric shape identification of black and white shapes with success [20].

Eyeborg [21] uses a head-mounted webcam. A single color is picked from in the center of the image using a weighted distribution of the pixel colors. The color's hue and saturation are used for color sonification. The color's hue is classified into 360 tones spanning an octave. The saturation modulates the volume of the sound.

See ColOr [22] horizontally samples 25 pixels from a video feed and simultaneously sonifies each pixel color based on its hue, saturation, and brightness. The color's hue is classified into seven colors. Each color is represented by the timbre of a musical instrument as follows: red (oboe), orange (viola), yellow (pizzicato), green (flute), cyan (trumpet), blue (piano), and purple (saxophone). Mixed color hues can be expressed by modulating the volume between them. Saturation is expressed by the pitch of the note of the hue's instrument in 4 steps. Zero to twenty-four percent is represented by a C note, 25–49% by a G note, 50–74% by a B lat note, and 75–100% by an E note. Luminance is mapped as follows: for values less than 50% a double bass sound is added, for values larger than 50% a signing voice is added. High luminance uses high-pitched notes, while low luminance uses low-pitched notes. See ColOr has been evaluated for color recognition [23], navigation, and object localization [24].

Kromophone [25] represents a specific color as the sum of focal colors and luminance. The focal colors are red, green, blue, and yellow. The luminance is divided into white, grey, and black. Each focal color is represented by a characteristic sound composed of pitch, timbre, and panning. A color representation is made by mixing the sound of the different focal colors and luminance. The contribution of each of the focal colors is expressed by its volume in the mix. Kromophone was evaluated for color identification, object discrimination, and navigation tasks.

ColEnViSon [26] converts the Red, Green, and Blue (RGB) space of an image into the CIELch color space, which characterizes colors into 267 centroids. The colors in the image are simplified into the color of the nearest centroid into ten color categories selected from hue and chroma. Each of the ten color categories is given a unique timbre for representation, and the luminance is represented using musical notes. Experiments show that users were able to interpret sequences of colored patterns and identify the number of colors in an image.

EyeMusic [14] conveys color information using a musical instrument's timbre for each of the white (choir), blue (brass), red (Reggae organ), green (Rapman's reed), and yellow (string) colors. Black is represented by silence.

EyeMusic resizes and clusters the colors in the image to produce a six-color 40 × 24-pixel image. Then, the image is divided into columns, which are processed from left to right to construct a soundscape. The musical note played in the instrument is determined by the y-axis coordinate of the pixel. The luminance determines the note's volume. All the sounds of each of the pixels in the column are combined. The combined audio is reproduced for each of the columns. Eye Music was evaluated for shape and color recognition achieving 91.5% and 85.6%, respectively.

Creole [27] is similar to Soundview in that it uses a tablet and stylus to explore two-dimensional colored images. The stylus is used to select a pixel in the image to translate its color information into sound. The pixel's RGB value is transformed into the CIE LUV color space, and the proportions of white, grey, black, red, green, yellow, and blue colors are calculated. Each of the colors is mapped to sound in the following way: white = 3520 Hz pure tone, grey = 100–3200 Hz noise, black = 110 Hz pure tone, red = 'u' vowel sound, green = 'i' vowel sound, blue = 262, 311, and 392 Hz tones, and yellow = 1047, 1319, and 1568 Hz tones. Creole was evaluated through color-sound associative memory tasks and object recognition. It was found that the Creole coding was easier to memorize and more effective to correctly identify the colors after less than 15 min of training.

Colorphone [28] is a wearable device that represents color by associating each RGB color component with a sine frequency. Specifically, red is associated with a 1600 Hz sine wave sound, green with 550 Hz, and blue with 150 Hz. The mixed signal is filtered with a low-pass white noise to represent the whiteness. This color system was evaluated for color recognition and navigation.

Most of the previous studies involving SSDs were made only for research purposes [17]. Like most assistive technologies, they mainly focus on functionality without any emphasis on aesthetics [27]. As shown in Table 1, in this paper, we explore a dual modality-based color code that utilizes the characteristics of hearing and smell to provide an improved user experience beyond the existing color codes that use single sensory modalities, such as touch, hearing, and smell.

Table 1. Comparison of existing color codes in terms of coding method and the number of colors coded.

Author (Modality)	Basic Pattern (Method)	No. of Colors
Taras et al. [29] (Touch)	Dots (Braille)	23 (6 hues + 2 levels of lightness for each hue + 5 levels of achromatic colors)
Ramsamy-Iranah et al. [30] (Touch)	Polygons (Children's Knowledge)	14 (6 hues + 5 other colors + 3 levels of achromatic colors)
Shin et al. [31] (Touch)	Lines (Orientation, Grating) The first eight colors are expressed in eight different angles of directionality by dividing a rainbow-shaped semicircle at intervals of 20 degrees	90 (8 hues + 4 levels of lightness, and 5 levels of saturation for each hue + 10 levels of brown and achromatic colors)
Cho et al. [32] (Touch)	Dots, lines, and curves (pictograms)	Simplified: 29 (6 hues + 2 levels of lightness, and 2 levels of saturation for each hue + 5 levels of achromatic colors) Extended: 53 (12 hues + 2 levels of lightness, and 2 levels of saturation for each hue + 5 levels of achromatic colors)
Cho et al. [33] (Hear)	Classical music melodies played on different instruments	23 (6 hues + 2 levels of lightness for each hue + 5 levels of achromatic colors)
Jabbar et al. [34] (Touch)	Embossed surface pattern by color wheel	Simplified: 24 (6 hues + 3 levels of lightness for each hue + 6 levels of achromatic colors) Extended: 32 (8 hues + 3 levels of lightness for each hue + 8 levels of achromatic colors)
Lee et al. [35] (Hear)	Spatial sound representation using binaural recoding in virtual environment	27 (8 hues + 3 levels of lightness for each hue + 3 levels of achromatic colors)
Bartolome et al. [36] (Hear & Touch)	Classical music melodies played on different instruments to code warmer/cooler/lighter/darker colors of each color hue and temperature to code color hue	30 (6 hues + 24 warmer/cooler/lighter/darker colors of each color hue)
Cho et al. [37] (Hear)	Classical music melodies played on different instruments to code 23 colors and Poetry to code colors with warm/cool/light/dark tones as supplements	35 (6 hues + 24 warmer/cooler/lighter/darker colors of each color hue + 5 levels of achromatic colors)
This work (Hear & Smell)	Classical music melodies played on different instruments to code color hues and scent to code warmer/cooler/lighter/darker colors of each color hue	30 (6 hues + 24 warmer/cooler/lighter/darker colors of each color hue)

Two recent works that propose color coding schemes designed for visual artwork color exploration with an emphasis on aesthetics are described in ColorPoetry [37] and Bartolome et al. [36]. ColorPoetry [37] proposes a color scheme that uses poem narrations with voice modulation to represent colors and their different shades. Bartolome et al. [36], instead, propose a multi-sensory approach involving the auditory and tactile channels that uses musical sounds and temperature cues to convey color. Compared to our work, ColorPoetry uses a uni-sensory approach for color representation, and was not evaluated for color identification performance. The approach followed by Bartolome et al. [36] describes a multi-sensory approach that uses sound and tact. Our work, in contrast, uses sound and smell. We use these sensory channels based on the results of the preliminary study were participants proposed the of use music and scent over other sensory channels. In addition, the temperature actuator proposed by Bartolome et al. [36] is limited in the number of colors it can represent spatially at the same time for tactile exploration. We compare the performance evaluations of the different methods in Table 2.

Table 2. Comparison of an existing multi-modal color codes in terms of user evaluation.

References	Color Substitution Modalities	Recognition Rate Evaluation	Usability Evaluation
Cho et al. [37]	Poetry	-	74.5
Bartolome et al. [36]	Sound & Temperature	67.5% (hue: 71.6%) (color dimensions: 92.5%)	-
Cavazos et al. [This work]	Sound & Scent	71.0% (hue: 75.6%) (color dimensions: 89.2%)	78.6

The sound color code study in Reference [33] confirmed that distinguishing light and dark colors using a sound (classical music melodies) approach is easier for most participants. However, when extending the palette to include warm and cool color variants, they started to experience difficulties. In this work, we propose to use scent in addition to sound to design a color code to serve two purposes. One is to easily differentiate the light-dark, warm-cool color variants from each other. The second one is to take advantage of using the additional sensory mode to convey the several sensorial properties of color. The proposed code decomposes a specific color into a hue and a set of color dimensions (saturated, light, and dark) for each hue. It uses the correspondence between musical instruments' timbre and color to facilitate hue identification. In addition, it employs pitch modulation and selected melodies for aesthetic color dimension representation. Besides sound, the proposed multi-sensory color code simultaneously integrates smell to represent each hue's saturated, light, and dark color dimensions. The advantages of this approach are several: most users seem to improve the correct identification of the hue's color dimensions and also report improved expressivity of the color and their artwork experience.

2.2. Olfactory Representation of Color

The correspondence between color, scent, and taste has been explored for the food and consumer industries by Frieling [38], who proposed the scent-hue mappings in Table 3. Li et al. [39] developed ColorOdor, a sensory substitution device that uses scents to help blind and visually impaired people identify colors. The device uses a camera to recognize the color of objects. Using a piezoelectric transducer, it vaporizes scents following the scent-hue mapping in Table 3. The scent-color mappings were designed through a survey with two visually impaired participants. ColorOdor intended use is for the color identification of everyday objects and as a learning tool for children with congenital blindness. The use of ColorOdor for visual artwork appreciation was not explored in the study. Lee and Cho [40] explored the implicit associations between color and concepts and described two relationships: color orientation and concept orientation. They use these two relationships to map the association between scent and color. They found that the orange scent represents a highly saturated orange color directionality, bright, extroverted, and the directionality of a strong stimulus. This property could also be applied when describing the characteristics of yellow and red. Considering that orange is a mixture of these two colors, orange has a universality that includes all three colors. The chocolate scent showed a brown color directionality with low brightness, and the concept directionality of round, low, warm, and introverted. Menthol and pine had similar turquoise color directivity, and the concept of coolness was considered to be about 22% higher in menthol than pine. Therefore, menthol was assigned a color associated with coolness, blue, and pine green, respectively. Using these mappings, they produce a Tactile Color Book [41] that conveys the color information of visual artworks to provide immersive and active exploration for blind and visually impaired people. The book is printed using a special ink impregnated with scent, that can be smelled when rubbed. Using this approach, blind and visually impaired students obtained a color identification accuracy of 94.3%. They expressed that the scent-color mapping was intuitive, easy to learn, and that it helps to understand the visual artwork content.

Table 3. Odor-hue color codes for olfactory representation of color.

Hue	**Frieling et al. [38]**	**Gilbert et al. [42]**	**Li et al. [39]**	**Lee and Cho [40]**
Pink	Sweet, mild	Aldehyde C-16, Methyl anthranile	-	-
Lavender	Sweet, unerotic	-	-	-
Purple	-	-	Lavender	-
Magenta	Heavy, narcotic, charmingly, sweet	-	-	-
Indigo	Scentless	-	-	-
Blue	Scentless	-	Blueberry	Menthol
Mint	Juicy, fresh, salty	-	-	-
Green	Fresh, green fragrance	Civet, 2-ethyl fenchol, galbanum, lavender, neroli, olibanum, pine	Camphor leaves	Pine
Olive	Musty	-	-	-
Lime Green	Sour, dry, fresh, bitter	-	-	-
Yellow	Perfume, flower	Bergamot oil	- Lemon	-
Orange	Hearty	-	-	Orange
Red	Sweet, hefty, hot	Cinnamic aldehyde	- Rose	-
Gold	Sweet, good, stunning	-	-	-
Ocher	Sourly, neutral	-	-	-
Brown	Aroma, musty	Caramel lactone, anise	-	Chocolate
White	Scentless	-	Lily	-
Grey	Bad	-	-	-
Black	-	-	Ink	-

The olfactory system is connected directly to the limbic system which is the section of the brain that processes emotions. When the olfactory receptors are stimulated by a scent, it often produces emotive responses on the subject which often trigger associated memories [43]. Thus, scents can be used to mediate the exploration of artworks through emotion [44]. Besides emotion, olfactory stimuli can trigger light or dark sensations [45]. Gilbert et al. [42] described that people tend to associate particular scents (Table 3) to a specific color in a non-random way. In addition, they associated the scents with the concepts of darkness and lightness. Civet was rated as the darkest, and Bergamot Oil, Aldehyde C-16, and Cinnamic aldehyde were rated as the lightest. Kemp et al. [46] described a correspondence between the strength-intensity of the scent with the lightness of a color. For example, strong scents were matched with darker colors. On the contrary, Fiore [47] described a correlation between floral scents and bright colors.

Olfactory stimuli can also trigger thermal sensations, like the feeling of coolness or warmness. Laska et al. [48] found that menthol (peppermint) and cineol (eucalyptus) consistently match the temperature conditions (cooling). Madzharov [49] pretested six essential oils, three of which we expected to be perceived as warm scents (warm vanilla sugar, cinnamon, pumpkin, and spice) and three as cool scents (eucalyptus- spearmint, peppermint, and winter wonderland). Thirty-three undergraduate students participated and evaluated each scent on perceived temperature and liking. Of the six scents, cinnamon and warm vanilla sugar were rated as the warmest, and peppermint was rated as the coolest. Cinnamon and peppermint were significantly different on the temperature dimension, as were warm vanilla sugar and peppermint. Adams et al. [50] identified lemon, apple, and peach scents as the brightest and lightest, whereas coffee, cinnamon, and chocolate scents as the dimmest and darkest odorants. Stevenson et al. [51] argued that the stronger cross-correspondence between scent and color occurs when the scent evokes a specific object (or context) producing a semantic match with a specific color. A summary of the correspondences between scent and color dimension described in previous works is shown in Table 4.

Table 4. Correspondence between scent and color dimensions described in previous literature.

Light	Dark	Warm	Cool
Bergamot [42]	Civet [42]	Cinnamon [49]	Menthol [48,49]
Strawberry [42]	Coffee [50]	Vanilla [49]	Eucalyptus [49]
Cinnamic aldehyde [42]	Cinnamon [50]	Chocolate [40]	Pine [40]
Lemon [50]	Chocolate [40]	-	-
Apple [50]	-	-	-
Peach [50]	-	-	-
Orange [40]	-	-	-
Rose [47]	-	-	-
Cherry blossom [47]	-	-	-
Acacia [47]	-	-	-

2.3. Multi-Sensory Representation of Color Based on Sound and Scent

While there are more developments to improve the access of blind and visually impaired to visual artworks that make use of the several human perception senses, it is still uncommon to provide access to color content through the use of multiple senses simultaneously. This work attempts to follow this multi-sensory approach which has also been explored in Reference [52]. Previous research has explored the relationship between scent and sound. Piesse [53] described correspondences between scent and sounds by using the musical notes in the diatonic scale. The match could also be made using variations on the sound pitch. Crisinel et al. [54] studied the correspondence between scent and musical features and found that the scent of orange and the iris flower could be mapped to higher-pitched sounds compared to the scent of musk and roasted coffee. They expand the correspondence study to include shapes and emotions. Scents judged as joyful, pleasant, and sweet were more frequently associated with a higher pitch and round-curved shapes. Scents judged as arousing were more frequently associated with the angular shapes, but no correlation was found with sound's pitch. Scents judged as brighter were associated with higher-pitch and round shapes. In Reference [55], Velasco et al. describe the emotional similarity between the olfactory and auditory information, which is potentially crucial for cross-modal correspondences and multi-sensory processing. Olfactory and sound multi-sensory representation are more frequently explored in media artworks. A couple of examples are the Tate Sensorium [56] and Perfumery Organ [57]. In the latter, a fragrance is scented when a piano is played using an "incense" that connects the fragrance and sound devised by Piesse [53]. Piesse matches the musical notes Do with rose, Re with violet, and Mi with acacia scents. Inspired by strong correspondences between color, sound, and scent, in this paper, we attempt to identify the correspondence between scent and four color dimensions warm-cool and light-dark using a semantic match mediation. Once the correspondence has been established, we will use them in conjunction with a timbre-hue auditory correspondence to create a multi-sensory color code.

2.4. Tactile Representation of Color

Tactile graphics refer to the use of raised lines and textures to convey images by touch. They serve as the basis for Tactile Color Patterns (TCPs), which are among the most common forms of accessible color representation for blind and visually impaired people. TCPs are a series of tactile pattern symbols that can be embossed along tactile reproductions of visual artworks to help identify the color of a specific object or area in the artwork. They follow a series of logical patterns to ease memorization and recollection. They are common on printed media as they can be embossed along with Braille and tactile graphics. Compared to other color representation methods, they have several advantages. The immediacy of feedback is one of them. A trained user can identify the color as soon as the pattern is touched. Additionally, it can also communicate other characteristics of color through the shape, size, and position of the pattern [32]. To ease the understanding and learning of the pattern, the designers base the tactile patterns of their color symbols on different properties or motifs. For example, Taras et al. [29] designed a TCP inspired for display on Braille display devices. It uses two dots in a Braille cell to represent the symbols for primary colors red, blue, and yellow. Secondary and tertiary color symbols are represented using a combination of the former color patterns. Ramsamy-Iranah et al. [30] based their tactile patterns on the children's knowledge of basic shapes and their surroundings. For example, the color red is represented by a circle which the children associate with the red 'bindi' dot used by Hindu women on their forehead. Blue is represented by the outline of a square as an analogy of the blue rectangular-shaped soap used in the laundry. The color yellow is represented by small dots reminiscent of the pollen in flowers. Shin et al. [31] developed a line pattern texture by decomposing color into three components: hue, saturation, and value (brightness). These components map into the line pattern texture following the convention: The hue affects the orientation of the lines, the saturation determines the width of the lines, and the value (brightness) dictates the interval (density) of the lines. Stonehouse [58] proposed a TCP based on common geometric shapes. Cho et al. [32] proposed three different TCPs: CHUNJIIN, CELESTIAL, and TRIANGLE. CHUNJIIN is inspired by the three basic components of the Korean alphabet. CELESTIAL is based on curved and straight lines, and the TRIANGLE TCP by Goethe's color triangle. TCPs are well suited for printed media and color learning. However, when embossed into the tactile graphic representation of an artwork, they obstruct exploration, making it hard to discern what is part of the visual artwork and what is part of the color pattern. A solution to this problem is producing two or three tactile graphic versions of the artwork, one without the TCP for easy shape recognition, and one with the TCP for color identification. The third one is a combined version of the first two [59]. This approach has the disadvantage that forces the reader to explore the different versions to build the complete mental image of the artwork [60]. More importantly, as the number of colors increases, so does the difficulty for correct identification and training required to be proficient. Multi-sensory color coding using sounds and scents, which we explore in this work, can be used to alleviate the complexity of using tactile graphics with tactile color patterns.

2.5. Visual Art Appreciation

Art appreciation is performed on the basic principles for exploration, technique examination, information analysis, and interpretation that enable the viewer to experience and understand an artwork. The community has long used art appreciation frameworks as a tool to establish a common ground and a defined process to appreciate visual artworks. Feldman [61] proposes a framework composed by the study of the artwork's information, analysis of its techniques, interpretation of the artwork's meaning, and value judgment. With the advent of modern art, the perception of art has moved from what was traditionally considered aesthetic or 'high art' into cognitive experiences such that viewing the artwork produces affective and self-rewarding aesthetic experiences. This has led to the development of new frameworks, such as the information-processing stage model of aesthetic processing (Leder et al. [62]), which considers art appreciation through both aesthetic experiences and judgments in a five-stage process that includes the perception of the artwork, explicit classification, implicit classification, cognitive mastering, and evaluation. Each stage exerts influence on each other, and, while the process follows an order, it also has feedback loops and can repeat its cycle. However, most of the frameworks include a perception phase in the early stage of the process. This stage involves acquiring different perceptual variables, such as complexity, contrast, symmetry, order, etc. This stage is particularly challenging for blind and visually impaired people since their vision might hinder their ability to assess the perceptual variables. The following stages will be influenced and produce distressed aesthetic judgment and emotional responses. This work focuses on improving the perception of the color contents of the visual artworks through non-visual channels. Facilitating the perceptual analysis of the color information will positively influence the following stages in the aesthetic processing model helping blind and visually impaired people reach better aesthetic judgment and experience richer aesthetic emotions.

3. Materials and Methods

3.1. Formative Study

We conducted a formative study with four blind, one visually impaired, and a school for the blind art teacher as participants to get a better understanding of the needs of blind and visually impaired people when exploring color in visual artworks.

3.1.1. Experiencing Colors in Visual Artwork by Blind and Visually Impaired People

The main focus of the study was on the current opportunities, tools available, and challenges faced by the participants to explore color information in visual artworks. The average age of the participants was 24.2 years (standard deviation of 1.6). Of the five participants in the study, three are women (60%), and two are men (40%). All of the participants are university students. More information on their characteristics is available in Table 5. All the participants gave signed informed consent for the study based on the procedures approved by the Sungkyunkwan University Institutional Review Board.

Table 5. Characteristics of blind and visually impaired participants in our formative study.

Participant	Sex	Age	Occupation	Sight
FP1	Female	23	University student	Total vision loss
FP2	Female	23	University student	Total vision loss
FP3	Female	24	University student	Total vision loss
FP4	Male	24	University student	Total vision loss
FP5	Male	27	University student	Profound vision loss

The formative study involved a semi-structured interview. We began by inquiring about the participant’s experience when exploring color from visual artworks. All the participants expressed the limited methods available at museums, art galleries, and even schools.

They commented that at museums and galleries, most color explanations are provided verbally by the staff. Some of them expressed being uncomfortable when receiving the information that way since they fear appearing *“dumb or being judged (FP2)”* if they repeatedly ask for further explanations. Participants without prior color experiences were very vocal about the challenges of receiving color properties verbally since *“without prior experience, learning such abstract concepts is very tough”*. After further inquiries, it seemed much of the interest in an artwork’s color information stemmed from peaking surrounding conversations of other people where they actively discussed or made expressions of wonder about the color contents. About the participants’ experience when exploring a visual artwork color contents using tactile means, only two of the participants expressed having explored an artwork through a tactile graphic at a museum-gallery setting. However, all five participants stated having experience tactually exploring colors during their early education years. Those participants that expressed having explored visual artwork color content by touch expressed their experience as challenging. *“(FP4) When you explore a paint or picture colors by touch, the first thing you need to do is find the key or legend to learn which texture patterns correspond to each color. If there are few uniform colors in the painting, this is not such a big problem, but, as soon as the number of colors grows, it becomes hard to remember all of them”*. The participants added that this approach does not work very well in paintings where colors transition gradually from one color to another as these transitions are difficult to pick up by tact. Another participant added that if the tactile graphic of the artwork and color textures are used together in one tactile graphic, it is challenging to identify which tactile features correspond to the painting shapes and which to color.

After inquiring about their current state, we proceeded to ask about their color experiences through sensory substitution. While none of the participants expressed having experienced color through a different channel (except for audio descriptions and tactile graphic experiences mentioned earlier) in a museum gallery setting, all the participants stated having experienced color through sensory substitution during their life. For example, one of the participants described his experience using color markers with fruit and flavor scents as a kid. The participant stated that using those markers was a pleasant experience that helped make common associations between colors and scents similar to sighted people. Other participants recollected their experience learning about color through temperature and brightness. When asked about their opinion about using musical sounds and scents for artwork color exploration, most users expressed their concern about the technical feasibility. Nevertheless, they showed interest in the experience as they believe that a proper mapping between color, music, and scent could help experience color in an enjoyable way other than just matching abstract semantic concepts that do not produce any reaction. *"(FP3) Even if I could touch the textures (tactile color representation) and immediately recognize the specific shades of many colors, what would be the artistic value of that? I want to touch or feel something (either by audio, smell, or other means) and feel awed by that just like a sighted person feels when they see the colors on the painting"*. Besides the time spent with the participants, we had the opportunity to interview an art teacher from a school for the blind. This participant emphasized the importance of a multi-sensory approach for color education by stating that color education is more memorable using the multi-sensory approach. In addition, the abstract concepts of color are easier to grasp through analogies with other sensory experiences. The art teacher also emphasized the importance of color education from an artistic and aesthetic perspective.

3.1.2. Design Requirements

Stemmed from the feedback received from the formative study and our objective, we list the following design requirements for a multi-sensory color code. Our priority is the communication of a richer visual artwork color exploration experience through multiple senses.

1. The multi-sensory code should be simple to learn and use while offering an expanded palette of colors.
2. The color code system should not only focus on expanding the color palette but also improve the aesthetic exploration of the artwork.
3. Because previous tactile approaches require extensive training and are obtrusive for tactile artwork exploration, it should focus on using music and scent to provide an extended palette. We also choose to focus on music and scent because of two additional reasons. The first is that music can be aesthetic and expressive, while smell can evoke emotions and memories, contributing to enhancing appreciation of works of art. The second reason is that we have the future goal to integrate the multi-sensory color code system to an interactive interface for artwork appreciation previously described in References [13,63], which already uses the tactile channel for artwork spatial exploration and auditory (verbal) channel for general information.

3.2. Color

For our color palette, we selected six unique hues (red, green, yellow, blue, orange, and purple), based on the color system described by Munsell [64], in which color is composed of hue, value, and chroma. Our proposed palette for this work expresses each of the six hues across five dimensions, saturated, light, dark, warm, and cool. The light and dark dimensions are obtained by adjusting the value in the Munsell color system. A light color has a value of 7 and a chroma of 8, while a dark color has a value of 3 and a chroma of 8. Warm and cool colors follow an approach derived from the RYB color wheel proposed by Issac Newton in Opticks [65]. Warm colors result from the combination of red-orange, yellow-orange, and blue-green hues. Cool colors result from the combination of red-violet, yellow-green, and blue-violet hues. Following this scheme, our proposed palette can express 30 colors across 6 hues and 5 dimensions and is shown in Figure 1.

3.3. Multi-Sensory Color Code

Our proposed multi-sensory color code system is composed of two components. A sound color code component (Section 3.3.1) that maps the auditory channel in the form of melodies to the six different hues shown in Figure 1. In addition, a scent color code component (Section 3.3.2) maps the olfactory channel through different scents to the five dimensions shown in Figure 1. To determine a specific color of the palette, the user must listen to the melodies to identify the hue while matching a scent to identify the saturated, light, dark, warm, or cool dimension of the color. In the following sections, we elaborate on each of the components.

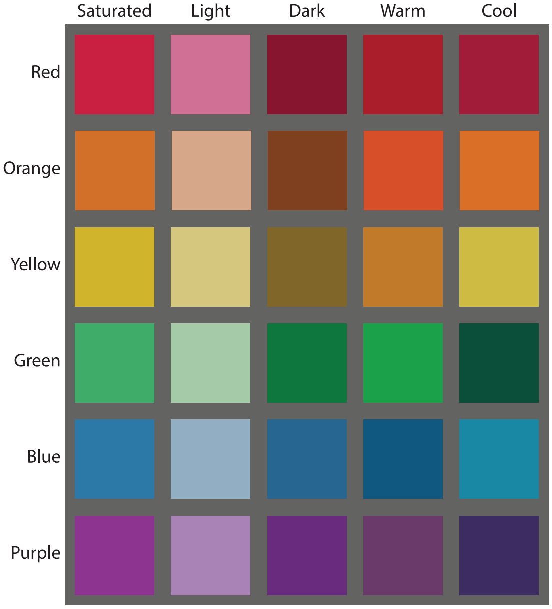


Figure 1. The 30-color palette used for the multi-sensory color code experimentation. The six hues on the left of the figure are described through the saturated, light, dark, warm, and cool dimensions. The saturated, light, and dark dimension are based on the BCP-37 [66].

3.3.1. Sound Color Code Component

The first component in the design of our multi-sensory color code is the sound color code component. While there are many different sound color codes based on several properties of sound, such as pitch and tempo, we use the VIVALDI sound color code previously developed by Cho et al. [33] as the sound color component of our multi-sensory color code. The VIVALDI color code was also designed by decomposing a color into hue and the saturated, light, and dark color dimensions. VIVALDI expresses six different color hues: red, orange, yellow, green, blue, and purple using several musical instruments. Red and orange are represented by string instruments (violin + cello and guitar); yellow and green by brass instruments (trumpet + trombone and clarinet + bassoon); blue and purple by percussion instruments (piano and organ). The pairing between the color and musical instruments was made following the correspondence between the instruments' timbre and the color hue. To express the saturated, light, and dark color dimensions, VIVALDI uses a different set of pitches for each dimension and fragments of Vivaldi's Four Seasons

Spring, Autumn, and Summer, respectively. Regarding the pitch, the saturated dimension is represented using an A major chord (medium pitch), the light dimension using F major (high-pitch), and Dark using E minor (low-pitch). The marked difference in pitch helps the listener to identify the color dimension variations more easily. The complete sound color code is resumed in Table 6.

Table 6. VIVALDI Sound Color Code.

Hue	Instrument	Saturated	Light	Dark
Red	Violin & Cello	Vivaldi's Four Seasons Spring [A major]	Vivaldi's Four Seasons Autumn [F major]	Vivaldi's Four Seasons Summer [E minor]
Orange	Guitar	Vivaldi's Four Seasons Spring [A major]	Vivaldi's Four Seasons Autumn [F major]	Vivaldi's Four Seasons Summer [E minor]
Yellow	Trumpet & Trombone	Vivaldi's Four Seasons Spring [A major]	Vivaldi's Four Seasons Autumn [F major]	Vivaldi's Four Seasons Summer [E minor]
Green	Clarinet & Bassoon	Vivaldi's Four Seasons Spring [A major]	Vivaldi's Four Seasons Autumn [F major]	Vivaldi's Four Seasons Summer [E minor]
Blue	Piano	Vivaldi's Four Seasons Spring [A major]	Vivaldi's Four Seasons Autumn [F major]	Vivaldi's Four Seasons Summer [E minor]
Purple	Organ	Vivaldi's Four Seasons Spring [A major]	Vivaldi's Four Seasons Autumn [F major]	Vivaldi's Four Seasons Summer [E minor]

Sound files are provided in the Supplementary Material.

3.3.2. Scent Color Code Component

In this work, we propose that using a sound and scent-based multi-sensory approach to design a color code will ease the effort to learn the code and improve the identification of the encoded colors. The previous section centered on the definition of the sound component; this section focuses on the scent component. Scent can be used to either identify the hue or the color dimension that compose a specific color. We decided to use scent only to identify the color dimension because it is already simple to identify the hue using sound (it is just necessary to identify the type of instrument, as seen in Table 6. In addition, it reduces the number of scents required to represent the complete palette. The color palette for the multi-sensory color code has five color dimensions: saturated, light, dark, warm, and cool. Since the saturated color dimension is the neutral shade of the color, we determined that we could represent the saturated dimension by either not using any smell or using a distinctive neutral one, like coffee. For the remaining light-dark and warm-cool color dimension pairs, the next step is to find a scent for each dimension. Instead of pairing each one to an arbitrary scent, we decided to select the scents based on their correspondence to the color dimensions. This can improve and establish the user's mental association between scents and colors.

For the color dimension-scent pair matching, instead of making an arbitrary match, we performed a semantic differential survey based on adjectives to exclude the participants' biases as much as possible. This kind of survey is valid for visually impaired and non-visually impaired people. For example, the latter tend to match scents to hue as follows: lemon-yellow, apple-red, and cinnamon-brown. Even though blind people have never seen colors, they also similarly match colors because of their education and social interaction with non-visually impaired people. During the survey, the participants smell different scents and choose which semantic adjectives relate most to that particular scent. Using those adjectives, we match them with the color dimension. Making the color dimension-scent pairs based on the survey results provides a more empathic solution for blind people than simply connecting colors and scents directly.

We use the semantic differential survey proposed by Reference [67] to determine the scent-color dimension correspondence. We establish the association between the semantic adjectives in Table 7 and each of the candidate scents in Table 8 through the survey. The selection of semantic adjective in Table 7 and their association with the light-dark and warm-cool pairs was made through a literature survey. The candidate scents in Table 8 were selected from the scent-color correspondence literature discussed in Section 2.2. Therefore, using the results of the survey, we determine the association between scents and color dimension through the semantic adjectives. The survey was performed following the next steps:

1. Each test participant is presented with one of the scent candidates from Table 8.
2. After the participant receives the olfactory stimuli, for each of the semantic adjective pairs, he evaluates the closeness (or directivity) between the stimuli and either of the adjectives using a scale from −2 to +2 in increments of 1. A value of 0 indicates no closeness (or directivity) to either of the adjectives. The evaluation is recorded in a document similar to Table 9.
3. Steps 1 and 2 are performed for each of the scents in Table 8, and the results are aggregated by each scent.
4. Once all the participants have evaluated all the scents across all the semantic adjectives pairs, it is possible to establish the closeness (or directivity) magnitude of each scent. All the evaluation scores for a specific scent are aggregated into two groups, the light-dark dimensions group, and the warm-cool dimensions group. The magnitude sign indicates the directivity between the two dimensions in the group.
5. Obtaining the overall directivity of the scent is done by identifying the maximum value among the directivity magnitudes of the light-dark and warm-cool dimensions obtained in step 4.
6. Once the overall directivity of each of the scents is obtained, the scent with the maximum directivity magnitude is selected as the match for each dimension. If a dimension does not have a maximum value, then the scent with the minimum directivity value of the opposite dimension can be selected as the maximum value. For the case where a specific scent is the top candidate for several dimensions, the match is done between the scent and the dimension where the scent has the maximum magnitude. The second to top candidate is then matched to the remaining dimension.

Table 7. Semantic adjectives for scent and color dimensions (light-dark and warm-cool) mediation.

Light-Dark	Warm-Cool
Noisy-Quiet [68]	Strong-Weak [69,70]
Joyful-Depressed [71]	Pointed-Round [72]
Sharp-Dull [73]	Active-Inactive [74]
High-Low [75]	Tense-Calm [76]
Rough-Smooth [73,77,78]	Near-Far [79]
Soft-Hard [73,77,78]	Expand-Contract [80]

Table 8. Olfactory stimuli used during our implicit association test.

Caramel	Lemon
Cinnamon	Chocolate
Eucalyptus	Vanilla
Apple	Civet
Cool mint	Pine

Following the implicit test procedure described before, we performed a test with the help of eighteen participants (average of 24.7 years with a standard deviation of 2.9, eight are women (44%), and ten are men (56%)) to identify the best four scents to pair to the light, dark, warm, and cool dimensions. The final candidate list after processing the test results is shown in Table 10. The final scent-color dimension pair is as follows, apple is matched with the light dimension, chocolate with the dark dimension, lemon to the warm dimension, and caramel to the cool dimension.

Table 9. Semantic adjectives association test example.

Participant: 1, Olfactory stimuli: Apple						
	−2	−1	0	+1	+2	
Light-Dark						
Noisy	X					Quiet
Joyful		X				Depressed
Warm-Cool						
Strong			X			Weak
Pointed				X		Round
Participant: 2, Olfactory stimuli: Apple						
	−2	−1	0	+1	+2	
Light-Dark						
Noisy		X				Quiet
Joyful					X	Depressed
Warm-Cool						
Strong		X				Weak
Pointed				X		Round
Directivity of Apple						
Directivity (Light-Dark): P1{−2 −1} + P2{ −1 + 2} = −2 = Light [2]						
Directivity (Warm-Cool): P1{ 0 + 1} + P2{−1 + 1}= 1 = Cool [1]						
Apple directivity: max(Light-Dark, Warm-Cool) = **Light [2]**						

The directivity of a scent is calculated by aggregating the evaluations for each of the semantic adjectives of the test users grouped by dimension and choosing the dimension with the largest magnitude.

Table 10. Scent candidates for scent—color dimension correspondence.

Light	**Dark**	**Warm**	**Cool**
Lemon [89]	* Chocolate [49]	* Lemon [105]	* Caramel [45]
* Apple [67]	Caramel [39]	Eucalyptus [73]	Vanilla [26]
Eucalyptus [49]	Vanilla [17]	Apple [64]	Chocolate [9]

* Selected candidate to represent the corresponding color dimension. Each cell describes the scent candidate and directivity magnitude.

3.4. Multi-Sensory Color Code Based Visual Sensory Substitution Prototype

To study the feasibility of our proposed multi-sensory color code to experience color contents, we decided to integrate it into a sensory substitution device prototype. The prototype is based on and extends our integrated multimodal guide platform described in Reference [13]. It uses a combination of tactile, audio, and olfactory modalities to communicate the color contents of an artwork. For artwork tactile exploration, the prototype presents the user with a 2.5D relief model of the artwork. The relief model is 3D printed and embedded with touch capacitive sensors on the model surface using conductive ink spread over the main features in the artwork to create interactive areas. A colored paint coat is then applied. The surface of the model can be explored by touch to determine the features of the artwork. The embedded sensors are connected to a control board composed of an Arduino Uno microcontroller (Arduino, Somerville, MA, USA), a WAV Trigger polyphonic audio player board (SparkFun Electronics, Boulder, CO, USA), an MPR121 proximity capacitive touch sensor controller (Adafruit Industries, New York, NY, USA), and four water atomization modules(Seeed, Shenzhen, China). Touch-capacitive data sensed during the user's tactile exploration is handled through the MPR121 controller, which communicates touch and releases events to the microcontroller through an I2C interface. The microcontroller processes the events to determine the user's touch gesture (touch, tap, and double-tap). Depending on the gesture, the microcontroller can send commands through its UART port to the audio board to play audio tracks. The microcontroller can also send digital signals through its general purpose IO ports to the atomization modules to produce fine water mist. The four atomization modules are installed in small bottles containing a dilution of scent and water. The modules are housed in an enclosure and placed above the relief model installed on a vertical exhibitor, as shown in Figure 2b.

Interaction with the prototype is simple. The user only needs to approach the prototype, wear a pair of headphones located at the side, and touch the surface to start an exploration session. While the original platform described in Reference [13] had several functions, like audio explanations and sound effects of the artwork features, these were disabled to avoid bias during the evaluation. For this prototype, the user can freely explore the artwork by touch to identify the features and double-tap in any of them to trigger the audio and olfactory stimuli. The audio and olfactory stimuli are based on the proposed multi-sensory color code. The audio files reproduced correspond to the audio tracks in Table 6 and the olfactory stimuli to the scents in Table 10. The user can then hear and smell the stimuli to determine the color of the double-tapped feature.

4. Evaluation

The following sections describe two evaluations of the multi-sensory color code. The first evaluation compares the effectiveness in terms of color identification success rate of the multi-sensory color code and a uni-sensory color code. The second evaluation examines the usability of the multi-sensory color code when used to experience the color contents in a painting. The evaluation involves using a visual sensory substitution device prototype that uses the proposed multi-sensory color code and a tactile graphic version of the artwork.

4.1. Participants

The eighteen participants from the implicit association test also took part in the evaluations. The participants' age ranged from 22 to 35 years, with an average of 24.7 years (standard deviation of 2.9). Of the eighteen participants recruited, eight are female (44%), and ten are men (56%). All of the participants had previous experience with accessible visual artworks and expressed interest in the visual arts. None of the participants took part in the formative study. All the participants or their legal guardians gave signed informed consent. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Sungkyunkwan University (protocol code: 2020-11-005-001, 22 February 2021).

4.2. Multi-Sensory Color Code Effectiveness

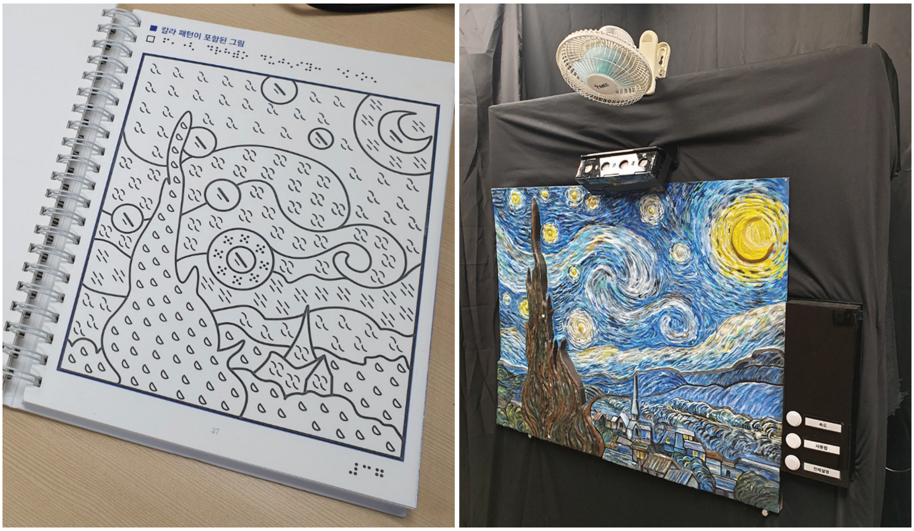
In this section, we investigate the effectiveness of our multi-sensory approach and a uni-sensory alternative color codes by comparing the color identification success rate achieved by the participants when using them. We performed this evaluation to determine if the proposed method improves the color identification success rate and explore its causes.

4.2.1. Materials

The effectiveness evaluation was performed using two sets of color codes. We used the VIVALDI color code described in Section 3.3.1 as the baseline for our experiments. The VIVALDI color code is entirely based on sound. It uses different musical instrument timbres to express the red, orange, yellow, green, blue, and purple hues shown in Table 6. The second set is our proposed multi-sensory color code described in Section 3.3. The multi-sensory color code extends the VIVALDI color code by adding different scents to facilitate the identification of the saturated, light, and dark dimensions. The scent selection was performed following the cross-modal correspondence approach through a semantic differential scale survey performed with the help of the participants. The selection method is described in detail in Section 3.3.2. The scents of coffee, apple, and chocolate were selected to represent the saturated, light, and dark color dimensions, respectively. The sound color code component of the color codes was provided by reproducing pre-recorded audio tracks on a set of high-fidelity speakers at a comfortable volume level. The audio files are provided separately in the supplement materials of this work. The scent color code component was provided in bottled scent samples made available during the evaluation.

4.2.2. Methodology

The evaluation was performed at our usability laboratory located within Sungkyunkwan University. Participants were invited into an isolated room and sat in front of a desk. The materials were provided to the participants by the test facilitator. A set of speakers was set at a comfortable volume and placed facing the participants. Capped bottles containing the scents were placed on one side of the table. When necessary, the test facilitator opened the scent bottles and provided them for a couple of seconds to the test participant for olfactory stimuli. The facilitator, observer, and test participants used sanitary face masks during the session. The participants could temporarily remove their masks to facilitate the olfactory stimuli. Between each participant evaluation, the area and scent bottles were sanitized.



(**a**) (**b**)

Figure 2. (**a**) Tactile graphic version of Van Gogh's Starry Night featuring the tactile color patterns described in Reference [32] used during the evaluation. (**b**) Visual sensory substitution prototype that expresses color contents through sound and scent.

The evaluation consisted of two parts. The first was a training session that lasted for 25 min. In this session, the facilitator introduced the purpose and methodology of the evaluation. Then, he explained the theory of color, its dimensions, and the terminology used. He emphasized the relationship between the hue, light, dark, warm, and cool color dimensions for the composition of a determined color. Participants were encouraged to ask any questions regarding the contents and explanations they received to ensure proper understanding. The second part of the evaluation lasted for about 50 min. There, the facilitator introduced and evaluated the VIVALDI and multi-sensory color codes. The color codes introduction and evaluation order followed a 2 × 2 Latin Square test design to counterbalance practice, fatigue, or order effects. For the VIVALDI color code, the facilitator explained the relationship between the color hue and the different musical instruments' timbre. He also described the relationship between the spring, autumn, and summer melodies selected from Vivaldi's Four Seasons and the saturated, light, and dark color dimensions. During the explanation, the facilitator reproduced the audio files of the VIVALDI color code and highlighted their characteristic features. Then, the facilitator played some of the audio files and asked the participant to identify some of the colors as a warming-up exercise and to verify the participant's understanding of the evaluation procedure. The participant was given ten min to review any of the colors of the VIVALDI color code in preparation for the color identification evaluation. For the multi-sensory color code explanation, the facilitator provided a similar explanation to that of the VIVALDI color code. In addition, he introduced the scents from the multi-sensory color code selected in this work and their relationship with the saturated, light, and dark color dimensions. The participant experienced the audio and scents stimuli during the introduction. A color identification warm-up exercise was also performed. After which, the participant had ten min to review any of the colors in the multi-sensory color code. After each of the color code explanations and reviews, the participant performed a color identification test. The test consisted of eighteen identification tasks. The tasks consisted of experiencing audio or audio and scent stimuli depending on the color code being evaluated and identifying the color it represents. Each task corresponded to a random color from the color code palette. Results were recorded for further analysis, and a short questionnaire was performed after the participant evaluated both color code systems to identify perceived preferences, learnability, and memorability. The results of this evaluation are described in detail in Section 5.1.

4.3. Visual Artwork Color Content Exploration Using a Visual Sensory Substitution Prototype

This section describes the second evaluation where participants use the visual sensory substitution prototype described in Section 3.4 to explore the color contents of a visual artwork. The participants also explored a tactile graphic equivalent using tactile color patterns. A System Usability Survey was performed for each exploration method to investigate their feasibility, usability, and elicit feedback of the systems.

4.3.1. Materials

The visual artwork exploration evaluation was performed after the multi-sensory color code effectiveness evaluation. The evaluation consisted of two visual artwork color content exploration tasks. Two different methods were used for the exploration tasks. One of them consisted in exploring a tactile graphic version of Van Gogh's Starry Night shown in Figure 2a. This tactile graphic was designed to highlight the contours of the features of the artwork and uses a series of different tactile color patterns to express the simplified color contents. A legend with the tactile color patterns and the corresponding name of the color written in Braille was also available for the participants to review the tactile pattern-color pairs. The second method involved the use of the sensory substitution device for visual art color content exploration described in Section 3.4. This prototype makes use of the multi-sensory color code proposed in this work. The prototype was inside the usability laboratory where the other evaluations took place but located in a semi-isolated exhibition area set up as a small art gallery. The prototype has a touch interactive 2.5D surface that also depicts Van Gogh's Starry Night, as shown in Figure 2b. To prevent bias and inconsistency between the exploration methods, both provide feedback with the same color representations at similar regions. For example, the moon's color is represented by a yellow-saturated (color composed of yellow hue and saturated dimension). Thus, the tactile pattern in the tactile graphics method corresponds to the yellow-saturated pattern. Similarly, in the sensory substitution device prototype, the audio feedback is Vivaldi's Four Seasons spring melody executed with a trumpet and trombone. In addition, the olfactory feedback is the scent of coffee. Together, audio and scent, also represent the yellow-saturated color. We assigned the following colors to the following features: the tree and mountains are green-dark, the wind and church blue-light, and the shine from the moon and haze above the mountains are blue-light.

4.3.2. Methodology

The second evaluation was performed in a semi-isolated temporal gallery that we set up within our usability laboratory. The sensory substitution device for visual art color content exploration (Section 3.4) was placed in the gallery. Participants had a ten-minute break between the two evaluations. The test facilitator guided the participants to the gallery. There, following a 2 × 2 Latin Square test design, presented one of the two artwork exploration methods. No color identification tasks were assigned, instead, the participants had ten min to freely explore the visual artwork color contents. Before the exploration task started, the test facilitator provided a short explanation of the exploration method. For the tactile graphic, he also presented a tactile legend with the color-tactile pattern pairs and explained their design. For the sensory substitution prototype, the facilitator explained its function, use, and revised that the participants could trigger the audio and olfactory stimuli using touch gestures. Then, he pointed out the location of the headphones and let the participant explore the artwork independently. When the exploration time was over, the test facilitator conducted a System Usability Survey with the participant. Upon completion, the participant was presented with the other exploration method and performed a similar procedure. Once the participant completed exploration with the two methods, a short unstructured interview was conducted to expand on the information gathered from the survey and obtain feedback for improvements and any personal thoughts or remarks about their exploration experience. The results of this evaluation are described in detail in Section 5.2.

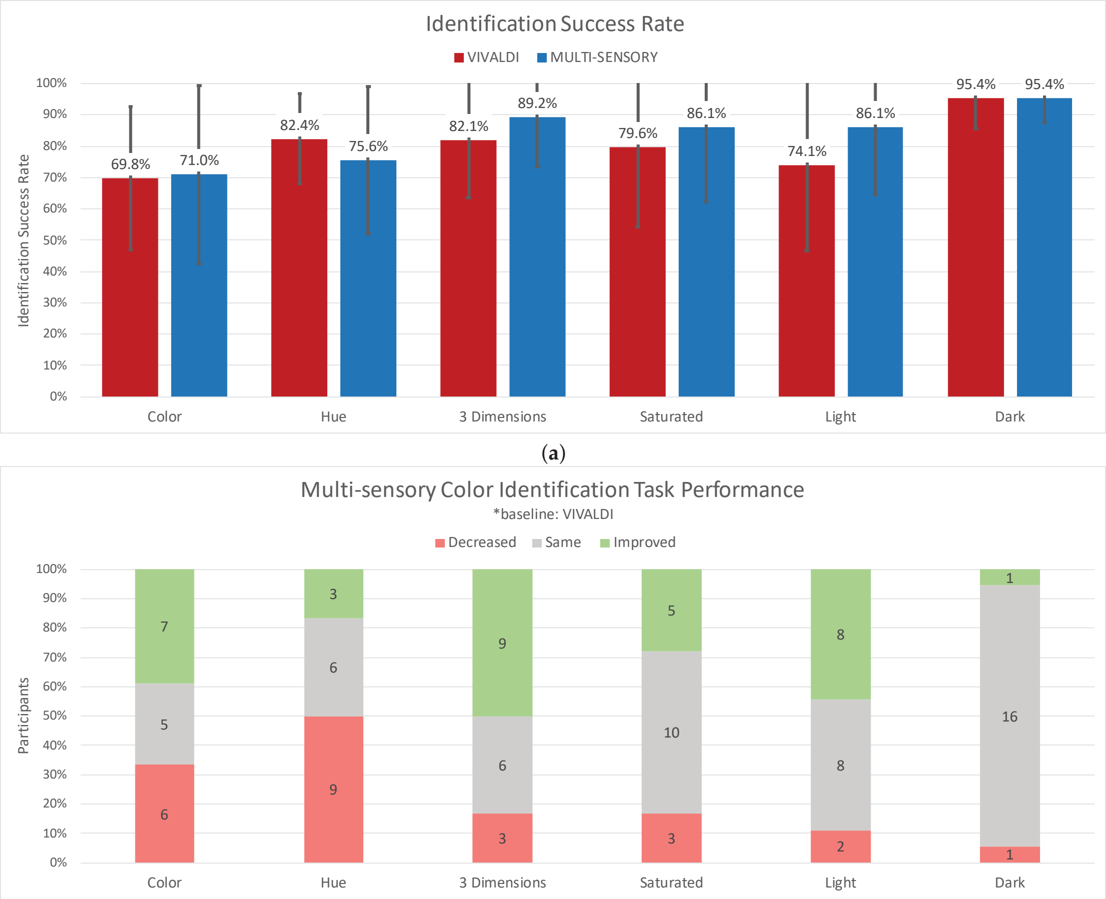
5. Results & Discussion

5.1. Multi-Sensory Color Code Effectiveness

We define the effectiveness of a color code as the color identification success rate achieved by the participants. The color identification success rate is simply the ratio between the number of color identification tasks completed correctly and the total number of identification tasks (eighteen). We computed the rates for both color codes to determine whether the users performed better in color identification tasks using the uni-sensory or multi-sensory approach. We also aggregated the results in different granularities to understand the effect of the several sensory stimuli on the participants' performance. The aggregated results are shown in Figure 3a. Participants were able to identify the color (hue and color dimensions) with a success rate of 69.8% and 71.0% when using the VIVALDI and multi-sensory color codes. In addition, the standard deviation increased from 0.23 to 0.29, which indicates that the multi-sensory color code improved some of the participants' performance, while it hurt others. Thus, from a group perspective, the multi-sensory approach did not perform better in terms of color identification. However, the results for the color identification tasks in Figure 3b show that twelve of the participants (67%) performed equal or better when using the multi-sensory color code. This result is relevant since it suggests that the majority of the participants benefited from the multi-sensory approach. Yet, it also emphasizes the possibility that it might not be an effective solution for everyone. We also calculated the identification success rate for the hue and color dimensions separately. For the hue, the VIVALDI color code had better performance at 82.4% compared to 75.6% of the multi-sensory. Half of the participants experienced a decreased performance when using the multi-sensory approach. Upon further inquiry, we found that some participants reported having experienced cognitive overload during the hue identification task. For example, one of the participants stated that *"it is important to note that there are some parts where it is difficult to concentrate on music while focusing on scent"*. Another participant also revealed an interesting point that we did not consider during the training session, *"it was difficult to understand the connection between scent and sound"*. Our training session covered the relationship between color-sound and color-scent. However, it did not cover the relationship between sound-scent. After the interview with the participants, we found that they mainly followed two strategies to identify the color during the multi-sensory evaluation. Some of them used a separate approach where they focused on one sensory stimulus at a time and processed the hue and color dimensions identification separately, while others processed the sensory stimuli and identification simultaneously.

For the color dimension identification, the multi-sensory color code performed better with an 89.2% success rate compared to 82.1% achieved with the VIVALDI code. Nine of the participants improved their color dimension performance. Based on the participants' performance, the combined audio and olfactory stimuli of the multi-sensory approach helped some of the participants recognize the color dimensions more easily. One of the participants reported that *"as scent was added to the sound, I (the participant) could feel in more detail and understand more deeply in the process of imagining the color. In addition, it is possible to describe the atmosphere and environment that the music tries to represent in more detail by using the music and scent together"*. Most participants reported that the difference between the color dimensions based on the audio was more subtle than those based on scent. The pronounced difference helped them to identify more easily the color dimension. Regarding the identification success rate for each of the saturated, light, and dark color dimensions, the performance of the multi-sensory approach was higher than that of the VIVALDI code for the first two cases and similar for the last one. For the saturated color dimension, the VIVALDI color code performance was 79.6%, while the multi-sensory achieved 86.1%. In the case of the light dimension, the VIVALDI performance was 74.1%, while the multi-sensory maintained 86.1%. Nevertheless, there was no evidence of statistical difference. However, revising the color identification task performance for the saturated and light color dimensions in Figure 3b reveals that most of the participants improved or maintained their performance using the multi-sensory color code. The participants achieved a 95.4% success rate for the dark dimension for both color codes. Upon hearing the participants' feedback, we found that the high performance achieved when identifying the dark color dimension was because the participants strongly associated the olfactory stimuli of the chocolate scent. While the associations between apple-light and coffee-saturated were not as *"strong"* or *"close"* as reported by the participants. For the VIVALDI color code, the participants stated that both the slow pace of the melody and the low pitch of the notes used also were easy to link to the dark dimensions, while the differences between the pace of the melodies and pitch selected for the saturated and light dimensions where more *"ambiguous"*.

We conducted a short questionnaire to learn how the participants' perceived the color codes' learnability, memorability, and their preference. Ten (55.6%) of the participants expressed that the multi-sensory color code was simpler to learn compared to the VIVALDI color code. Since the multi-sensory color code builds on top of the VIVALDI code, we expected it to be perceived as more complicated to learn. However, the participants expressed that the simple and few associations between the scents and the color dimensions are simple enough. For the memorability of the color codes, eleven participants (61.2%) stated that it was easier to recognize the color from the olfactory stimuli compared to the audio-only stimuli from Vivaldi. One participant considered that *"the relationship (with color) will be remembered longer with the scented one (color code)"* because the scent triggers fond memories. Another participant expressed that it was also the synergy between the audio, smell, and color that allowed her to more easily recognize the color. The last question was about the preference between the two color codes. Thirteen participants (72.2%) preferred the multi-sensory color code over VIVALDI. Among the reasons, one was that they felt the difference between the color dimensions was clearer, and another was that the scents evoked feelings beyond the color space. For example, one of the participants described feeling *"fresh, springy, and like summer"* during the evaluation. Another participant described the experience as feeling the *"lightness and heaviness"* of the colors. In general, the questionnaire results suggest that the memorability and the experience created by the olfactory stimuli tilted the preference towards the multi-sensory color code.



(**a**)

(**b**)

Figure 3. (**a**) Success rate for the different color and color dimensions identification tasks during the VIVALDI and multi-sensory color codes evaluation. (**b**) Comparison of the participants performance achieved during the evaluation of the VIVALDI and multi-sensory color codes.

5.2. Visual Artwork Color Content Exploration Using a Visual Sensory Substitution Prototype

The results of the System Usability Survey for the tactile graphic and visual sensory substitution prototype are described in Table 11. In general, most participants (fifteen) gave a better usability score to the multi-sensory-based visual sensory substitution prototype. According to the participants' survey results, the visual sensory substitution prototype scored 78.61, while the tactile graphic method received a 61.53 out of 100. These scores can be interpreted as "acceptable" and "marginal-low" based on the acceptability ranges scores proposed by Bangor et al. [81]. It is important to note that two participants gave a lower overall score to the multi-sensory-based prototype. One of them felt overloaded by the multi-sensory approach. Influenced by this, he obtained the lowest color identification success rate among the group (17% compared to the group average of 71%). Despite the low performance, the user stated preferring the multi-sensory approach because it considered it more *"aesthetic and supportive to encourage artwork exploration"*. On the other hand, the other participant that gave a lower score had a 100% identification success rate on both exploration methods. This participant's evaluation only focused on the ease of color identification. In her opinion, using the tactile pattern was better *"Since it's more immediate, you don't need to wait for the audio or the smell and as soon as you touch the tactile graphic, you can identify the pattern"*. Thus, there is evidence of the different motivations to explore a visual artwork and how a single approach might be insufficient to cover diverse user needs. Nevertheless, most participants considered that the multi-sensory prototype facilitated the exploration of color contents in visual artworks mainly through its convenient interface, which improved their confidence in identifying the colors and stimulated their imagination experience through its expressive audio and scents.

Table 11. Tactile Graphics and Visual Sensory Substitution Prototype System Usability Survey report.

1 *	2	3	4	5	M	SD
S1. I think that I would like to use this system frequently.						
	3	4	8	3	3.61	0.98
1		1	8	8	4.22	1.00
S2. I found the system unnecessarily complex.						
3	8	6	1		2.28	0.83
8	6	1	3		1.94	1.11
S3. I thought the system was easy to use.						
	2	7	7	2	3.50	0.86
	2		9	7	4.17	0.92
S4. I think that I would need the support of a technical person to be able to use this system.						
2	4	2	6	4	3.33	1.37
2	11	1	2	2	2.50	1.20
S5. I found the various functions in this system were well integrated.						
	5	2	8	3	3.50	1.10
		2	4	12	4.56	0.70
S6. I thought there was too much inconsistency in this system.						
15	3				1.17	0.38
9	5	2	2		1.83	1.04
S7. I would imagine that most people would learn to use this system very quickly.						
	4	1	8	5	3.78	1.11
			11	7	4.39	0.50
S8. I found the system very cumbersome to use.						
1	6	1	10		3.11	1.08
7	8	1	2		1.89	0.96
S9. I felt very confident using the system.						
2	7	3	6		2.72	1.07
	1	1	6	10	4.39	0.85
S10. I needed to learn a lot of things before I could get going with this system.						
1	10	2	5		2.61	0.98
5	9	1	3		2.11	1.02
⬛	Tactile Graphics					
🟦	Visual Sensory Substitution Prototype					

* SUS score ranges from 1 ("Strongly disagree") to 5 ("Strongly agree").

Participants expressed their inclination to use both methods frequently. Several stated that it would be better if both were available. They could use the visual sensory substitution prototype to explore thoroughly those artworks in which they have an interest and use the tactile graphic to skim through those artworks they do not. Participants did not consider either of the methods to be complex. They said that both were easy to approach, start using, and operate. They deemed the tactile graphics method slightly more complicated because the pairing between the tactile patterns and color is not intuitive without previous training. The audio and scents were perceived more intuitive, *"(Using the multi-sensory approach) I think even if the color is not evident, I can still feel and experience some sensations from the music like brightness and darkness, or freshness from the smell"*. This characteristic also led the participants to express that they could explore the color contents more independently with the multi-sensory prototype. The opinion was divided in the case of the tactile graphic. Those participants with high tactile proficiency reported a high level of independence were those with low stated the contrary.

Regarding the function integration, the participants gave high scores to both methods. In particular, the multi-sensory prototype was considered well-integrated because of the combined localized audio and olfactory stimuli triggered by touch gestures. Most participants expressed surprise for being able to experience each of the features in the artwork separately through different audio and smell in a similar way as to how visual exploration occurs. However, some touch gesture sensing implementation challenges prevented some of the participants' gestures to trigger the correct stimuli, which lead them to rate the system with some inconsistency. On the other hand, the participants rated the tactile graphic approach as very consistent due to its simplicity. However, we identified that many participants assume responsibility for any problems during the exploration. For example, one of the participants that experienced confusion when touching the patterns said: *"I guess I should practice more (touching) the patterns, I keep forgetting what they mean and the little differences between them"*. The participants described both methods as very easy and straightforward to learn, principally due to the method's simple interface.

The two main advantages of the multi-sensory prototype compared to the tactile graphics concerning usability were its convenience and the confidence felt by the participants. In terms of convenience, the participants' opinion was divided for the tactile graphic approach. Participants that gave a high rating expressed the immediacy of the touch stimuli, the convenience of not having to wear any device. The participants that gave a low rating described the effort to distinguish between the features contour lines from the color patterns and constantly having to review the color pattern legend as very cumbersome. They also pointed out that, while the mental effort to identify the color from the patterns might be the same as for the multi-sensory method, it feels less *"pleasant and artistic and more mechanic"*. The other advantage reported by the participants was the confidence they felt using the multi-sensory prototype in comparison to the tactile graphics. When using the tactile approach, the participants felt difficulty identifying the colors correctly as this approach does not provide other means to confirm the correct identification. The participants felt that the multi-sensory prototype presented them with two alternatives (sound and scent) to perceive the color, making them feel more confident in the identification.

Beyond the identification of color and the usability of the system, the participants communicated that the multi-sensory prototype was better suited to explore and experience visual artworks as the combination of sound and olfactory stimuli encourage exploration and reflection of the artwork. They could not confirm whether their reactions towards the art piece are the same as those intended by the artist, but, in general, they agreed that they had a stronger reaction compared to that experienced when using the tactile graphic.

5.3. Limitations

We identified several limitations in our study. The proposed color code is based on the results of a semantic differential survey which should be extended by including more participants with diverse characteristics to make it more robust. For example, colors have different meanings and symbolize different concepts across cultures. Therefore, instead of proposing a fixed set of color-sound-scent pairs, we believe a selection should be tailor-made for the audience. Similarly, if fixed, some color-sound-scent pairs can cause incoherence when used for different artworks. Another limitation is the different needs of the audience. Some people can find the audio and scent stimuli bothersome if used over extended periods.

Regarding the number of test participants required for usability testing, problem-discovery studies typically require between three and twenty participants. For comparative studies, such as the A/B test that compares two designs against each other to determine which one is better, group sizes from eight to 25 participants typically provide valid results [82]. For psychophysiological pain tests, a problem-discovery test, Lamontagne et al. [83] found that 82% of the total paint points experienced by 15 participants were experienced by 9 participants. Therefore, the number of usability test participants in this paper of 18 is considered sufficient. Greenwald et al. [84] performed an implicit association test on 32 psychology course students (13 males and 19 females). Therefore, in future work, we will further investigate scaled implicit association tests with more participants.

6. Conclusions and Future Work

In this work, we presented a multi-sensory color code system that can represent up to 30 colors. It uses melodies to express each color's hue and scents to express the saturated, lightness (light-dark), and temperature (warm-cool) color dimensions. The scent selection and pairing were made through a semantic correspondence survey in collaboration with eighteen participants. We evaluated the color code system to determine if using a multi-sensory approach eases the effort to recognize the encoded colors and help improve the color identification compared to the commonly used uni-sensory method. The results from the evaluation suggest that the multi-sensory approach does improve color identification, however not for everyone. The cause of this seems to be the extra cognitive effort and sensory overload experienced by some. In addition, we integrated the color code into a sensory substitution device prototype to determine if the color code could be more suitable and expressive when exploring visual art color content compared to a tactile graphics alternative. The results from this evaluation indicate that the multi-sensory-based prototype is more convenient and improves the confidence for visual artwork color content exploration. Moreover, the results also suggest suitability for artwork exploration since the multi-sensory stimuli improve the experience and reaction to the artwork. In the future, we would like to expand the color code to include more color-audio-scent pairs to study the applicability across different styles of visual artworks. In this work, two non-visual sensory reproductions of artworks were used to evaluate the proposed color code system. As future work, experiments can be are carried out using more diverse styles of non-visual sensory reproductions to further support the results proposed in this paper. In addition, we believe it is relevant to explore the effect of the semantic incoherence that could happen from the color code usage and its influence on the artwork experience and the interpretation. While experiencing color is just a fraction of the art appreciation process, we believe our work contributes towards designing and studying accessible art appreciation frameworks for all.

Supplementary Materials: The following are available at https://www.mdpi.com/2079-9292/10/14/1696/s1.

Author Contributions: Conceptualization, methodology, J.-D.C.; formal analysis, data curation, L.C.Q., C.-H.L.; writing—original draft preparation, visualization, L.C.Q.; project administration, J.-D.C.; Investigation, L.C.Q., J.-D.C., C.-H.L.; Resources, writing—review and editing, supervision, funding acquisition, J.-D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2018 Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea (2018M3C1B6061353).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Sungkyunkwan University (protocol code: 2020-11-005-001, 22 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available within the article.

Acknowledgments: We would like to thank all volunteers for their participation and the reviewers for their insights and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Abbreviations

The following abbreviations are used in this manuscript:

TCP	Tactile Color Pattern
SSD	Sensory Substitution Device
RGB	Red Green, and Blue
SCC	Sound Color Code

References

1. Bhowmick, A.; Hazarika, S.M. An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends. *J. Multimodal User Interfaces* **2017**, *11*, 149–172. [CrossRef]
2. Terven, J.R.; Salas, J.; Raducanu, B. New Opportunities for Computer Vision-Based Assistive Technology Systems for the Visually Impaired. *Computer* **2014**, *47*, 52–58. [CrossRef]
3. Pawłowska, A.; Wendorff, A. Extra-visual perception of works of art in the context of the audio description of the Neoplastic Room at the Museum of Art in Łódź. *Quart* **2018**, *47*, 18–27.
4. Hoyt, B. Emphasizing Observation in a Gallery Program for Blind and Low-Vision Visitors: Art Beyond Sight at the Museum of Fine Arts, Houston. *Disabil. Stud. Q.* **2013**, *33*. [CrossRef]
5. Krantz, G. Leveling the Participatory Field: The Mind's Eye Program at the Guggenheim Museum. *Disabil. Stud. Q.* **2013**, *33*. [CrossRef]
6. Neves, J. Multi-sensory approaches to (audio) describing the visual arts. In *MonTI. Monografías de Traducción e Interpretación*; Redalyc: Alicante, Spain, 2012; pp. 277–293
7. Hinton, R. Use of tactile pictures to communicate the work of visual artists to blind people. *J. Vis. Impair. Blind.* **1991**, *85*, 174–175. [CrossRef]
8. Holloway, L.; Marriott, K.; Butler, M.; Borning, A. Making sense of art: Access for gallery visitors with vision impairments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–12
9. Reichinger, A.; Carrizosa, H.G.; Wood, J.; Schröder, S.; Löw, C.; Luidolt, L.R.; Schimkowitsch, M.; Fuhrmann, A.; Maierhofer, S.; Purgathofer, W. Pictures in Your Mind: Using Interactive Gesture-Controlled Reliefs to Explore Art. *ACM Trans. Access. Comput.* **2018**, *11*. [CrossRef]
10. Leporini, B.; Rossetti, V.; Furfari, F.; Pelagatti, S.; Quarta, A. Design Guidelines for an Interactive 3D Model as a Supporting Tool for Exploring a Cultural Site by Visually Impaired and Sighted People. *ACM Trans. Access. Comput. TACCESS* **2020**, *13*, 1–39. [CrossRef]
11. D'Agnano, F.; Balletti, C.; Guerra, F.; Vernier, P. Tooteko: A case study of augmented reality for an accessible cultural heritage. Digitization, 3D printing and sensors for an audio-tactile experience. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *40*, 207 [CrossRef]
12. Iranzo Bartolome, J.; Cavazos Quero, L.; Kim, S.; Um, M.Y.; Cho, J. Exploring art with a voice controlled multimodal guide for blind people. In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, Tempe, Arizona, 17–20 March 2019; pp. 383–390
13. Cavazos Quero, L.; Iranzo Bartolomé, J.; Cho, J. Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics. *Electronics* **2021**, *10*, 297. [CrossRef]
14. Abboud, S.; Hanassy, S.; Levy-Tzedek, S.; Maidenbaum, S.; Amedi, A. EyeMusic: Introducing a visual colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.* **2014**, *32*, 247–257. [CrossRef] [PubMed]
15. Stiles, N.R.; Shimojo, S. Auditory sensory substitution is intuitive and automatic with texture stimuli. *Sci. Rep.* **2015**, *5*, 1–14. [CrossRef] [PubMed]
16. Hamilton-Fletcher, G.; Wright, T.D.; Ward, J. Cross-modal correspondences enhance performance on a colour-to-sound sensory substitution device. *Multisensory Res.* **2016**, *29*, 337–363. [CrossRef]
17. Buchs, G.; Haimler, B.; Kerem, M.; Maidenbaum, S.; Braun, L.; Amedi, A. A self-training program for sensory substitution devices. *PLoS ONE* **2021**, *16*, e0250281. [CrossRef]
18. Hamilton-Fletcher, G.; Ward, J. Representing colour through hearing and touch in sensory substitution devices. *Multisensory Res.* **2013**, *26*, 503–532. [CrossRef]
19. van de Doel, K. *Soundview: Sensing Color Images by Kinesthetic Audio*; Georgia Institute of Technology: Atlanta, GA, USA, 2003
20. van den Doel, K.; Smilek, D.; Bodnar, A.; Chita, C.; Corbett, R.; Nekrasovski, D.; McGrenere, J. *Geometric Shape Detection with Soundview*; Georgia Institute of Technology: Atlanta, GA, USA, 2004
21. Else, L. A cyborg makes art using seventh sense. *New Sci.* **2012**, *215*, 50. [CrossRef]

22. Bologna, G.; Deville, B.; Pun, T.; Vinckenbosch, M. Transforming 3D coloured pixels into musical instrument notes for vision substitution applications. *EURASIP J. Image Video Process.* **2007**, *2007*, 76204. [CrossRef]
23. Bologna, G.; Deville, B.; Vinckenbosch, M.; Pun, T. A perceptual interface for vision substitution in a color matching experiment. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–6 June 2008; pp. 1621–1628
24. Bologna, G.; Deville, B.; Pun, T. Blind navigation along a sinuous path by means of the See ColOr interface. In Proceedings of the International Work-Conference on the Interplay Between Natural and Artificial Computation, Santiago de Compostela, Spain, 22–26 June 2009; pp. 235–243.
25. Capalbo, Z.; Glenney, B. Hearing color: Radical pluralistic realism and SSDs. In Proceedings of the AP-CAP, Tokyo, Japan, 1–2 October 2009; pp. 135–140
26. Ancuti, C.O.; Ancuti, C.; Bekaert, P. ColEnViSon: Color Enhanced Visual Sonifier-A Polyphonic Audio Texture and Salient Scene Analysis. In Proceedings of the International Conference on Computer Vision Theory and Applications, Lisboa, Portugal, 5–8 February 2009; Volume 1, pp. 566–572
27. Hamilton-Fletcher, G.; Obrist, M.; Watten, P.; Mengucci, M.; Ward, J. "I Always Wanted to See the Night Sky" Blind User Preferences for Sensory Substitution Devices. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 2162–2174
28. Osiński, D.; Bizoń, P.; Midtfjord, H.; Wierzchoń, M.; Hjelme, D.R. Designing auditory color space for color sonification systems. In Proceedings of the International Conference on Auditory Display, Newcastle upon Tyne, UK, 23 June 2019; pp. 331–334
29. Taras, C.; Ertl, T. Interaction with Colored Graphical Representations on Braille Devices. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, San Diego, CA, USA, 19–24 July 2009; Springer: Berlin, Germany, 2009; pp. 164–173.
30. Ramsamy-Iranah, S.; Rosunee, S.; Kistamah, N. Application of assistive tactile symbols in a 'tactile' book on colour and shapes for children with visual impairments. *Int. J. Arts Sci.* **2017**, *10*, 575–590.
31. Shin, J.; Cho, J.; Lee, S. Please Touch Color: Tactile-Color Texture Design for The Visually Impaired. In Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–7
32. Cho, J.D.; Quero, L.C.; Bartolomé, J.I.; Lee, D.W.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2021**, *46*, 103–116. [CrossRef]
33. Cho, J.D.; Jeong, J.; Kim, J.H.; Lee, H. Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments. *Electronics* **2020**, *9*, 1981. [CrossRef]
34. Jabbar, M.S.; Lee, C.H.; Cho, J.D. ColorWatch: Color Perceptual Spatial Tactile Interface for People with Visual Impairments. *Electronics* **2021**, *10*, 596. [CrossRef]
35. Lee, Y.; Lee, C.H.; Cho, J.D. 3D Sound Coding Color for the Visually Impaired. *Electronics* **2021**, *10*, 1037. [CrossRef]
36. Bartolome, J.I.; Cho, G.; Cho, J.D. Multi-Sensory Color Expression with Sound and Temperature in Visual Arts Appreciation for People with Visual Impairment. *Electronics* **2021**, *10*, 1336. [CrossRef]
37. Cho, J.D.; Lee, Y. ColorPoetry: Multi-Sensory Experience of Color with Poetry in Visual Arts Appreciation of Persons with Visual Impairment. *Electronics* **2021**, *10*, 1064. [CrossRef]
38. Frieling, H. *Farbe Hilft Verkaufen*; Musterschmidt-Verlag: Gottingen, Germany, 1967
39. Li, S.; Chen, J.; Li, M.; Lin, J.; Wang, G. ColorOdor: Odor Broadens The Color Identification of The Blind. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 2746–2751
40. Lee, H.; Cho, J. A Research on Using of Color-Concept Directed Scent for Visually Impaired Individuals to Appreciate Paintings. *Sci. Emot. Sensib* **2020**, *23*, 73–92. [CrossRef]
41. Cho, J.D. *Tactile Color Book*; Book Publishing Braille: Seoul, Korea, 2020.
42. Gilbert, A.N.; Martin, R.; Kemp, S.E. Cross-modal correspondence between vision and olfaction: The color of smells. *Am. J. Psychol.* **1996**, *109*, 335–351. [CrossRef] [PubMed]
43. Delplanque, S.; Coppin, G.; Sander, D. Odor and Emotion. In *Handbook of Odor*; Springer: Cham, Switzerland, 2017; pp. 101–102
44. Vi, C.T.; Ablart, D.; Gatti, E.; Velasco, C.; Obrist, M. Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition. *Int. J. Hum. Comput. Stud.* **2017**, *108*, 1–14. [CrossRef]
45. Spence, C. Olfactory-colour crossmodal correspondences in art, science, and design. *Cogn. Res. Princ. Implic.* **2020**, *5*, 1–21. [CrossRef] [PubMed]
46. Kemp, S.E.; Gilbert, A.N. Odor intensity and color lightness are correlated sensory dimensions. *Am. J. Psychol.* **1997**, *110*, 35. [CrossRef]
47. Fiore, A.M. Multisensory integration of visual, tactile, and olfactory aesthetic cues of appearance. *Cloth. Text. Res. J.* **1993**, *11*, 45–52. [CrossRef]
48. Laska, M. Perception of trigeminal chemosensory qualities in the elderly. *Chem. Senses* **2001**, *26*, 681–689. [CrossRef]
49. Madzharov, A.V.; Block, L.G.; Morrin, M. The cool scent of power: Effects of ambient scent on consumer preferences and choice behavior. *J. Mark.* **2015**, *79*, 83–96. [CrossRef]

50. Adams, C.; Doucé, L. What's in a scent? Meaning, shape, and sensorial concepts elicited by scents. *J. Sens. Stud.* **2017**, *32*, e12256. [CrossRef]
51. Stevenson, R.J.; Rich, A.; Russell, A. The nature and origin of cross-modal associations to odours. *Perception* **2012**, *41*, 606–619. [CrossRef]
52. Cho, J.D. A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment. *Electronics* **2021**, *10*, 470. [CrossRef]
53. Piesse, G.W.S. *The Art of Perfumery and the Methods of Obtaining the Odours of Plants*; Longmans: London, UK, 1879
54. Crisinel, A.S.; Jacquier, C.; Deroy, O.; Spence, C. Composing with cross-modal correspondences: Music and odors in concert. *Chemosens. Percept.* **2013**, *6*, 45–52. [CrossRef]
55. Velasco, C.; Balboa, D.; Marmolejo-Ramos, F.; Spence, C. Crossmodal effect of music and odor pleasantness on olfactory quality perception. *Front. Psychol.* **2014**, *5*, 1352. [CrossRef] [PubMed]
56. Ablart, D.; Velasco, C.; Vi, C.T.; Gatti, E.; Obrist, M. The how and why behind a multisensory art display. *Interactions* **2017**, *24*, 38–43. [CrossRef]
57. Inc., T. Perfumery Organ. Available online: https://www.nticc.or.jp/en/archive/works/perfumery-organ/ (accessed on 1 May 2021).
58. Stonehouse, P. Tactile Colours. Available online: http://cpwstonehouse.com/tactile-colours/ (accessed on 1 May 2021).
59. Kardoulias, T. Guidelines for making tactile diagrams and accompanying narratives. In *Art Beyond Sight. A resource on Art Creativity and Visual Impairment*; American Foundation for the Blind: Arlington, VA, USA, 2003; pp. 267–296
60. JunDong, C.; Jeon, H.R.; Jeong, J.; Jo, S.; Lee, D.; Cho, K.S.; Lee, H.; hye Kim, J.; Sah, Y.J. Multi-sensory Visual Art Experience Exhibition for People with Visual Impairment. *J. Korean Soc. Exhib. Des.* **2020**, *34*, 1–13.
61. Feldman, E.B. *Varieties of Visual Experience*; Harry N. Abrams. Inc.: New York, NY, USA, 1992.
62. Leder, H.; Belke, B.; Oeberst, A.; Augustin, D. A model of aesthetic appreciation and aesthetic judgments. *Br. J. Psychol.* **2004**, *95*, 489–508. [CrossRef] [PubMed]
63. Cavazos Quero, L.; Iranzo Bartolomé, J.; Lee, S.; Han, E.; Kim, S.; Cho, J. An Interactive Multimodal Guide to Improve Art Accessibility for Blind People. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '18, Galway, Ireland, 22–24 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 346–348. [CrossRef]
64. Nickerson, D. History of the Munsell color system and its scientific application. *JOSA* **1940**, *30*, 575–586. [CrossRef]
65. Newton, I. *Opticks*; Prabhat Prakashan: New Delhi, India, 1730
66. Palmer, S.E.; Schloss, K.B. An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8877–8882. [CrossRef]
67. Osgood, C.E.; Suci, G.J.; Tannenbaum, P.H. *The Measurement of Meaning*; Number 47; University of Illinois Press: Champaign, IL, USA, 1957
68. Taft, C. Color meaning and context: Comparisons of semantic ratings of colors on samples and objects. *Color Res. Appl.* **1997**, *22*, 40–50.:1<40::AID-COL7>3.0.CO;2-4. [CrossRef]
69. Jonas, C.; Spiller, M.J.; Hibbard, P. Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychon. Bull. Rev.* **2017**, *24*, 1104–1112. [CrossRef] [PubMed]
70. Marks, L.E.; Hammeal, R.J.; Bornstein, M.H.; Smith, L.B. Perceiving similarity and comprehending metaphor. *Monogr. Soc. Res. Child Dev.* **1987**, *52*, i-100 [CrossRef]
71. Maric, Y.; Jacquot, M. Contribution to understanding odour–colour associations. *Food Qual. Prefer.* **2013**, *27*, 191–195. [CrossRef]
72. Kandinsky, W.; Rebay, H. *Point and Line to Plane*; Courier Corporation: North Chelmsford, MA, USA, 1979
73. Ludwig, V.U.; Simner, J. What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex* **2013**, *49*, 1089–1099. [CrossRef]
74. Levy, B.I. Research into the psychological meaning of color. *Am. J. Art Ther.* **1984, 23, *58–62***
75. Anikin, A.; Johansson, N. Implicit associations between individual properties of color and sound. *Atten. Percept. Psychophys* **2019**, *81*, 764–777. [CrossRef]
76. Torres, A.; Serra, J.; Llopis, J.; Delcampo, A. Color preference cool versus warm in nursing homes depends on the expected activity for interior spaces. *Front. Archit. Res.* **2020**, *9*, 739–750. [CrossRef]
77. Slobodenyuk, N.; Jraissati, Y.; Kanso, A.; Ghanem, L.; Elhajj, I. Cross-modal associations between color and haptics. *Atten. Percept. Psychophys.* **2015**, *77*, 1379–1395. [CrossRef]
78. Jraissati, Y.; Slobodenyuk, N.; Kanso, A.; Ghanem, L.; Elhajj, I. Haptic and tactile adjectives are consistently mapped onto color space. *Multisensory Res.* **2016**, *29*, 253–278. [CrossRef] [PubMed]
79. Iranzo Bartolomé, J.; Cho, J.D.; Cavazos Quero, L.; Jo, S.; Cho, G. Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People. *Electronics* **2020**, *9*, 1939. [CrossRef]
80. Kandinsky, W. *Concerning the Spiritual in Art*; Courier Corporation: Chelmsford, MA, USA, 2012.
81. Bangor, A.; Kortum, P.T.; Miller, J.T. An empirical evaluation of the system usability scale. *Intl. J. Hum. Interact.* **2008**, *24*, 574–594. [CrossRef]

82. Six, J.M.; Macefield, R. How to determine the right number of participants for usability studies. *San Francisco (CA): UXmatters*. 2016. Available online: https://www.uxmatters.com/mt/archives/2016/01/how-to-determine-the-right-number-of-participants-for-usability-studies.php (accessed on 1 May 2021).
83. Lamontagne, C.; Sénécal, S.; Fredette, M.; Chen, S.L.; Pourchon, R.; Gaumont, Y.; De Grandpré, D.; Léger, P.M. User Test: How Many Users Are Needed to Find the Psychophysiological Pain Points in a Journey Map? In Proceedings of the International Conference on Human Interaction and Emerging Technologies, Strasbourg, France, 28–30 April 2021; Springer: Berlin/Heidelberg, Germany, 2019; pp. 136–142.
84. Greenwald, A.G.; McGhee, D.E.; Schwartz, J.L. Measuring individual differences in implicit cognition: The implicit association test. *J. Personal. Soc. Psychol.* **1998**, *74*, 1464. [CrossRef]

electronics

MDPI

Review

Image Accessibility for Screen Reader Users: A Systematic Review and a Road Map

Uran Oh *, Hwayeon Joh and YunJung Lee

Department of Computer Science and Engineering, Ewha Womans University, Seoul 03760, Korea;
e.jhy@ewhain.net (H.J.); yunjung_lee20@ewhain.net (Y.L.)
* Correspondence: uran.oh@ewha.ac.kr

Abstract: A number of studies have been conducted to improve the accessibility of images using touchscreen devices for screen reader users. In this study, we conducted a systematic review of 33 papers to get a holistic understanding of existing approaches and to suggest a research road map given identified gaps. As a result, we identified types of images, visual information, input device and feedback modalities that were studied for improving image accessibility using touchscreen devices. Findings also revealed that there is little study how the generation of image-related information can be automated. Moreover, we confirmed that the involvement of screen reader users is mostly limited to evaluations, while input from target users during the design process is particularly important for the development of assistive technologies. Then we introduce two of our recent studies on the accessibility of artwork and comics, AccessArt and AccessComics, respectively. Based on the identified key challenges, we suggest a research agenda for improving image accessibility for screen reader users.

Keywords: image accessibility; touchscreen; nonvisual feedback; blind; visual impairment; systematic review

Citation: Oh, U.; Joh, H.; Lee, Y. Image Accessibility for Screen Reader Users: A Systematic Review and a Road Map. *Electronics* **2021**, *10*, 953. https://doi.org/10.3390/electronics10080953

Academic Editor: Juan M. Corchado

Received: 8 March 2021
Accepted: 5 April 2021
Published: 16 April 2021

1. Introduction

According to the World Health Organization, at least 2.2 billion people have a visual impairment, and the number is likely to increase with population growth and aging [1]. For them, understanding visual information is one of the main challenges.

To improve the accessibility of images for people who are blind or have low vision (BLV), a number of studies have been conducted to assess the effectiveness of custom-made tactile versions of images [2–8]. Cavazos et al. [5], for instance, proposed a 2.5D tactile representation of the artwork where blind users can feel the artwork by touch while listening to audio feedback. Holloway et al. [7] also investigated tactile graphics and 3D models to deliver map information such as the number of entrances, location and direction of certain landmarks. This approach with extra tactile feedback is found to be effective as it can deepen one's spatial understanding of images by touch [9–11]. However, it requires additional equipment, which potential users have limited access to (e.g., 3D printer, custom devices). Moreover, tactile representations need to be designed and built for each image, and thus it is not ideal for supporting a number of different images in terms of time and cost.

Meanwhile, others have relied on digital devices that are commercially available (e.g., PC, tablets, smartphones) for conveying image descriptions (also known as alternative text or alt text) on the web in particular [12–14]. For instance, Zhong et al. [13] generated alt text for images on the web that are identified as important using crowdsourcing. In addition, Stangl et al. [12] used natural language processing and computer vision techniques to automatically extract visual descriptions (alt text) on online shopping websites for clothes. Unlike tactile approaches, this software-based approach is more scalable, especially with the help of crowds or advanced machine learning techniques. However, listening to a set of verbal descriptions of an image may not be sufficient for understanding its spatial layout of content or objects within each image.

To leverage the issues of two different approaches above, researchers have worked on touchscreen-based image accessibility that enables users to explore different regions on images by touch to help them have a better spatial understanding. In this paper, to gain a more holistic perspective of this approach by examining the current states and identifying the challenges to be solved, we conducted a systematic literature review of 33 papers, following PRISMA guidelines [15]. To be specific, our goal is to identify the following: supported image types, provided information, collection and the delivery method of the information, and the involvement of screen reader users during the design and development process.

As a result, we found that research studies on touchscreen-based image accessibility have been mostly focused on maps (e.g., directions, distance), graphs (e.g., graph type, values) and geometric shapes (e.g., shape, size, length) using audio and haptic feedback. Moreover, it revealed that the majority of them manually generated image-related information or assumed that the information was given. We also confirmed that while most user studies are conducted with participants who are blind or have low vision for user evaluation, a few studies involved target users during the design process.

In addition, to demonstrate how other types of images can be made accessible using touchscreen devices, we introduce two of our systems: AccessArt [16–18] for artwork and AccessComics [19] for digital comics.

Based on the challenges and limitations identified by conducting systematic review and from our own experience of improving image accessibility for screen reader users, we suggest a road map for future studies in this field of research. The following are the the contributions of our work:

- A systematic review of touchscreen-based image accessibility for screen reader users.
- A summary of the systematic review in terms of image type, information type, methods for collecting and delivering information, and the involvement of screen reader users.
- The identifications of key challenges and limitations of studying image accessibility of screen reader users using touchscreen devices.
- Recommendations for future research directions.

The rest of the content covers a summary of prior studies on image accessibility and touchscreen accessibility for BLV people (Section 2), followed by a description of how we conducted a systematic review (Section 3), and the results (Section 4), demonstrations of two systems for improving the accessibility of artwork and digital comics (Section 5), discussions on the current limitations and potentials of existing work and suggestions on future work (Section 6), and conclusions (Section 7).

2. Related Work

Our work is inspired by prior work on image accessibility and touchscreen accessibility for people who are blind or who have low vision.

2.1. Image Accessibility

Screen readers cannot describe an image unless its metadata such as alt text are present. To improve the accessibility of images, various solutions have been proposed to provide accurate descriptions for individual images on the web or on mobile devices [12–14,20–22]. Winters et al., for instance (Reference [14]) proposed an auditory display for social media that can automatically detect the overall mood of an image and gender and emotion of any faces using Microsoft's computer vision and optical character recognition (OCR) APIs. Similarly, Stangl et al. [12] developed computer vision (CV) and natural language processing (NLP) modules to extract information about clothing images on an online shopping mall. To be specific, The CV module automatically generates a description of the entire outfit shown in a product image, while the NLP module is responsible for extracting price, material, and description from the web page. Goncu and Marriott [22], on the other hand, demonstrated the idea of creating accessible images by the general public using a web-based tool. In addition, Morris et al. [20] proposed a mobile interface that provides screen reader users with rich information of visual contents prepared using real-time crowdsourcing and friend-sourcing rather than using machine learning techniques. It allowed users to listen to the alt text of a photograph and ask questions using voice input while touching specific regions with their fingers.

While most of the studies for improving the accessibility of images that can be accessed with digital devices tend to focus on how to collect the metadata that can be read out to screen reader users, others investigated how to deliver image-related information with tactile feedback [2,4,5,23,24]. Götzelmann et al. [23], for instance, presented a 3D-printed map to convey geographic information by touch. However, some worked on using computational methods to automatically generate tactile representations [3,25]. For example, Rodrigues et al. [25] proposed an algorithm for creating tactile representations of objects presented in an artwork varying in shape and depth. While it is promising, we focused on improving image accessibility on a touchscreen, which is widely adopted in personal devices such as smartphones and tablets, since it does not require additional hardware.

2.2. Touchscreen Accessibility

While we chose to focus on touchscreen-based image accessibility, touchscreen devices are innately inaccessible as they require accurate hand–eye coordination [26]. Thus, various studies have been conducted to improve touchscreen accessibility by providing tactile feedback using additional hardware devices [27–29]. *TouchCam*, for example, designed and implemented a camera-based wearable device that can be worn on a finger, which is used to access one's personal touchscreen devices by interacting with their skin surface to provide extra tactile and proprioceptive feedback. Physical overlays that can be placed on the top of a touchscreen were also investigated [6,30]. For instance, *TouchPlates* [6] allows people with visual impairments to interact with touchscreen devices by placing tactile overlays on the top of the touch display. Meanwhile, software-based approaches have been proposed as well such as supporting touchscreen gestures that can be performed anywhere on the screen [26,31–34]. *BrailleTouch* [32] and No-Look Notes [34], for example, proposed software solutions for supporting eyes-free text entry for blind users by using multi-touch gestures. Similarly, smartphones on the market also offer screen reader modules with location-insensitive gestures: iOS's VoiceOver (https://www.apple.com/accessibility/vision/,accessedon15April2021) and Android's Talkback (https://support.google.com/accessibility/android/answer/6283677?hl=en). These screen readers read out the contents on the screen if focused, and users can navigate different items by directional swipes (i.e., left-to-right and right-to-left swipe gestures) or by exploration-by-touch.

Again, we are interested in how touchscreen devices can be used to improve image accessibility mainly because they are readily available to a large number of end-users including BLV people as they have their own personal devices with touchscreens. In addition, as touchscreen devices offer screen reader functionality, they are accessible.

3. Method

To identify the road map of future research directions on image accessibility for people with visual impairments using touchscreen devices, we conducted a systematic review following PRISMA guidelines [15]. The process is shown in Figure 1.

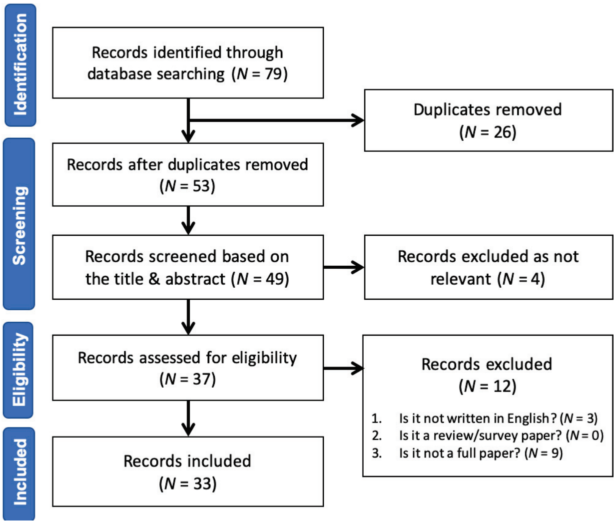


Figure 1. A PRISMA flow diagram that shows the process of identifying eligible papers on touchscreen-based image accessibility for people who are blind or have visual impairments.

3.1. Research Questions

We had five specific research questions for this systematic review:

- *RQ1. What types of images have been studied for image accessibility?*
- *RQ2. What types of image-related information has been supported for BLV people?*
- *RQ3. How has image-related information been collected?*
- *RQ4. How has image-related information been delivered?*
- *RQ5. How have BLV people been involved in the design and evaluation process?*

3.2. Identification

To identify research papers related to touchscreen-based image accessibility for BLV, we checked if at least one of the following search keywords from each category—target user (*User*), target object (*Object*), supported feature (*Feature*), supported device (*Device*)—is included either in the title, the abstract or authors' keywords:

- *User: "blind", "visual impairment", "visually impaired", "low vision", "vision loss"*
- *Object: "image", "picture", "photo", "figure", "drawing", "painting", "graphic", "map", "diagram"*
- *Feature: "description", "feedback"*
- *Device: "touchscreens", "touch screens"*

As a result, a total of 79 papers were identified from three databases: Scopus (*N* = 50), ACM digital library (*N* = 25), and IEEE Xplore (*N* = 4). We then removed 26 duplicates.

3.3. Screening

Then, we examined the titles and abstracts of the rest 53 unique papers and excluded four papers that are not relevant, which were all conference reviews.

3.4. Eligibility

Of the 49 remaining papers, we excluded 12 papers that met the following exclusion criteria:

1. Not written in English (*N* = 3).
2. A survey or review paper (*N* = 0).
3. Not a full paper such as posters or workshop and case study papers (*N* = 9).

Then, papers were included if and only if the goal of the paper was improving the accessibility of any type of images for people with visual impairments.

4. Results

As the result of the systematic review, 33 papers were considered as eligible for the analysis. We summarize the papers mainly in terms of the research questions specified in Section 3.1.

4.1. Overview

As shown in Figure 2, the first paper was published in 2008; note that the first generation of iPhone was released in 2007. While the topic of this area is not that active, at least three papers have been published each year since 2013. As for the country of authors' affiliations, United States had the dominant number of papers, which was 17. It was followed by Germany ($N = 4$), Canada ($N = 3$), Australia ($N = 2$), and Japan ($N = 2$). Other countries had only one paper each: Austria, Brazil, China, France, Italy, Lebanon, South Korea and Spain.

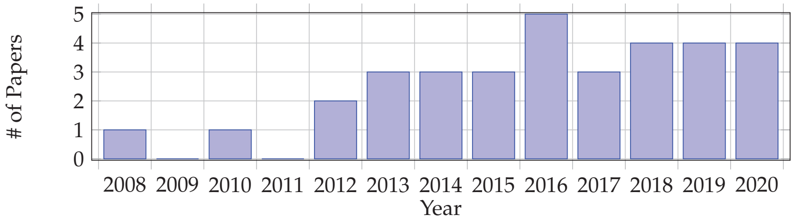


Figure 2. The number papers related to touchscreen-based image accessibility.

4.2. Supported Image Types

The types of images that have been studied in prior works for providing better accessibility for screen reader users (*RQ1*) are summarized in Table 1. While three papers were designed to support any type of images in general [35–37], most of the papers focused on specific types of images. To be specific, approximately half of the papers studied images of maps ($N = 10$) and graphs ($N = 6$). Interestingly, while the accessibility of photographs for BLV was largely investigated in terms of web accessibility [21,38], only three out of 33 papers aimed to support photographs using touchscreen devices in particular. In addition, as touchscreen devices themselves have accessibility issues for people with visual impairments, requiring accurate hand–eye coordination [26], four papers focused on improving the accessibility of the touchscreen-based interface itself such as soft buttons [39–41] and gestures [42].

Table 1. Types of supported images for people who are blind or have low vision (BLV).

Image Type	Count	Papers
Map	10	[23,24,37,43–49]
Graph	6	[37,50–54]
Geometric Shapes	4	[50,52,55,56]
Images in General	3	[35–37]
Lines	3	[57–59]
Photographs	3	[37,60,61]
Touchscreen UI	3	[39–41]
Floor plan	3	[62–64]
Diagram	2	[37,65]
Artwork	1	[66]
Touchscreen Gestures	1	[42]

4.3. Information Type

We have identified types of information provided to BLV to improve the accessibility of images (*RQ2*), which is shown in Table 2. The name of the object, mainly the object that is touched, was provided the most (*N* = 10), followed by spatial layout of various objects in each image (*N* = 8). As expected, types of provided information differ depending on the image type. For instance, direction/orientation, distance, and other geographic information was provided for map images. On the other hand, shape, size, boundary of objects, and length were mostly offered for images related to geometric figures. Scene descriptions (*N* = 5), textures (*N* = 4), graph values (*N* = 2), and texts written in images (*N* = 2) were also present. Note that only two papers supported color information for photographs [35] and artwork [66]. Other information types include types of graph [53] and weather [24]. Five papers did not specify the information they provided.

Table 2. Types of information provided to improve image accessibility.

Info Type	Count	Papers
Object Names	10	[23,24,43,46,49,55,62–64,66]
Spatial Layout	8	[46,47,49,54,55,60–62]
Direction/Orientation	8	[43–45,47,51,59,61,65]
Distance	5	[43,44,46,47,49]
Geographic information	5	[23,46–49]
Scene	5	[36,37,53,63,66]
Shape	4	[50–52,55]
Texture	4	[35,40,56,66]
UI Element	3	[40,41,61]
Area/Size	2	[56,59]
Object Boundary	2	[35,58]
Stroke	2	[42,54]
Length	2	[51,65]
Values in Graphs	2	[51,53]
Texts in Images	2	[53,61]
Color	2	[35,66]
Not specified	5	[39,50,55,57,64]

4.4. Information Preparation

In addition to types of information supported to improve image accessibility (*RQ3*), we have found that most studies have not specified how the visual information is collected or created (see Table 3). This suggests that the aim of many of these studies is “delivering” visual information that is inaccessible to BLV as is while assuming that the information is given rather than “retrieving” the information. Meanwhile, close to one-third of the studies seem to manually create the data they need to provide (*N* = 9) or use metadata such as alternative text (alt text) or textual descriptions that are paired with the images (*N* = 4). Others relied on automatic approaches to extract visual information of images: image processing, optical character recognition (OCR), and computer vision (*N* = 5). Meanwhile, two papers proposed a system where the image descriptions are provided by crowdworkers.

Table 3. Data preparation/collection methods for providing image-related information.

Data Collection	Count	Papers
Manually Created	9	[39,40,45,53,54,56,58,59,66]
Metadata (e.g., alt text)	4	[41,43,64,66]
Image Processing	2	[61,62]
Object Segmentation	2	[35,36]
Crowdsourcing	2	[60,61]
OCR	1	[61]
Not specified	18	[23,24,37,41,42,44–52,55,57,63,65]

4.5. Interaction Types

We were also interested in how image-related information is delivered using touchscreen devices (*RQ4*). We have identified the interaction type in terms of input types and output modalities as follows:

Input types. As shown in Table 4, the major input type is touch, as expected; most of the studies allowed users to explore images by touch with their bare hands ($N = 28$ out of 33). Moreover, touchscreen gestures were also used as input ($N = 5$). On the other hand, physical input devices ($N = 4$ for *keyboard*, $N = 3$ for *stylus*, and $N = 2$ for *mouse*) were used in addition to touchscreen devices. While it is known that aiming a camera towards a target direction is difficult for BLV [67], a camera was also used as a type of input, where users were allowed to share image feeds from cameras with others so that they could get information about their surrounding physical objects such as touch panels on a microwave [60,61].

Table 4. Types of input used for improving images on touchscreen devices.

Input	Count	Papers
Touch	28	[24,35–37,39,42–46,48–59,61–64,64–66]
Gesture	5	[37,42,46,47,49]
Keyboard	4	[24,35–37]
Stylus	3	[37,40,61]
Mouse	2	[35,37]
Voice command	2	[46,49]
Camera	2	[60,61]
Physical UI	1	[41]

Output Modalities. As for output modalities, various types of feedback techniques were used (see Table 5). Approximately half of the studies used a single modality: audio only ($N = 14$; including both speech and non-speech audio) or vibration only ($N = 3$). On the other hand, others used multimodal feedback, where the combination of audio and vibration was most frequent ($N = 6$), followed by audio with tactile feedback ($N = 5$). The most widely used output was speech feedback that verbally describes images to BLV users using an audio channel as a screen reader reads out what is on the screen using text-to-speech (e.g., Apple's VoiceOver). On the other hand, non-speech audio feedback (e.g., sonification) was also used. For instance, different pitches of sound [24,35,36,42] or rhythms [55] were used to convey image-related information. Meanwhile, vibration was as popular as non-speech audio feedback, while some used tactile feedback to convey information. For example, Gotzelmann et al. [23] used a 3D-printed tactile map. Zhang et al. [41] also made user interface elements (e.g., buttons, sliders) with a 3D printer to improve the accessibility of touchscreen-based interfaces in general by replacing virtual elements on a touchscreen with physical ones. Moreover, Hausberger et al. [56] proposed an interesting approach using kinesthetic feedback along with frictions. Their system dynamically changes the position and the orientation of a touchscreen device in a 3D space for BLV to explore shapes and textures of images on a touchscreen device.

Table 5. Output modalities used for improving images on touchscreen devices.

Output	Count	Papers
Speech	20	[24,36,41,42,45–49,51,53,54,58,60–63,65,66]
Non-speech Audio	14	[23,24,24,35–37,42,43,48,51,52,55,55,65]
Vibration	14	[39,40,43–45,47,50–55,57,59]
Tactile Feedback	9	[23,24,39,41,44,47,56,63,64]
Force Feedback	3	[39,40,56]

4.6. Involvement of BLV

Finally, we checked if BLV, the target users, were involved in the system development and evaluation processes; see Table 6. We first examined if user evaluation was conducted regardless of whether target users were involved or not. As a result, we found that all studies but two had tested their system with human subjects. Most of them had a controlled lab study, where metrics related to task performance were collected for evaluation such as the number of correct responses and completion time. However, close to half of the studies had subjective assessments such as easiness and satisfaction in a Likert scales, or open-ended comments about their experience after using the systems.

Of the remaining 31 papers, three papers had user studies but with no BLV participants. The rest of the 28 papers had evaluated their system with participants from the target user group. In addition to user evaluation, seven studies used participatory design approaches during their design process. Moreover, some papers had BLV participate in their formative qualitative studies at an early stage of their system development to make their ideas concrete (i.e., survey, interview).

Table 6. Methodologies used in the studies and BLV's involvement in system design and evaluation. Note that the following three studies conducted user studies with blind-folded sighted participants [54–56].

Process	Methodology	Count	Papers
Evaluation	Controlled Lab Study	28	[36,37,39–44,46–63,66]
	Subjective Assessment	15	[24,36,41–43,47,50,51,53,58,61–65,65,66]
Design	Participatory Design	7	[24,50,61–65]
Ideation	Interview	3	[24,50,65]
	Survey	2	[24,65]
N/A	No user study	2	[23,35]

5. AccessArt and AccessComics

Based on our systematic review results, we have confirmed that various types of images were studied to improve their accessibility for BLV people. However, most of the studies focused on providing knowledge or information based on facts (e.g., maps, graphs) to users rather than offering improved user experience that BLV users can enjoy allowing subjective interpretations. Thus, we focused on supporting two types of images in particular that are rarely studied for screen reader users: artwork and comics. Here, we demonstrate how these two types of images can be supported and appreciated with improved accessibility: AccessArt [16–18] and AccessComics [19] (see Table 7 as well).

Table 7. A summary of AccessArt and AccessComics following identifying factors used in our systematic review.

	AccessArt	**AccessComics**
Image Type	2D paintings	Digital comics
Information Type	Object names, spatial layout, shape, size, color	Panel number, background, character appearance and actions
Data Preparation	Manual, crowdsourced	Manual (visual descriptions) open dataset (panel, balloon, script)
Interaction	touch/gesture input, speech output	touch/gesture input, speech output
User Study with BLV	semi-structured interview, design probe study	online survey, semi-structured interview, design probe study

5.1. AccessArt for Artwork Accessibility

BLV people are interested in visiting museums and wish to know about artwork [68–70]. However, a number of accessibility issues exist when visiting and navigating inside a museum [71]. While audio guide services are in operation for some exhibition sites [72–75], it can still be difficult for BLV people to understand the spatial arrangement of objects within each painting. Tactile versions of artwork, on the other hand, allows BLV people to learn the spatial layout of objects in the scene by touch [2–5]. However, it is not feasible to make these replicas for every exhibited artwork. Thus, we began to design and implement touchscreen-based artwork exploration tool called *AccessArt*.

AccessArt Ver1. The very first version of AccessArt is shown in Figure 3, in which it had four paintings with varying genres: landscape, portrait, abstract, and still life [17]. As for the object-level labels, we segmented each object along with descriptions. Then we developed a web application that allowed BLV users to (1) select one of the four paintings they wish to explore and (2) scan objects within each painting by touch with its corresponding verbal description including object-level information such as the name, color and position of the object. For example, if a user touches the moon on "The Starry Night", then the system reads out the following: *"Moon, shining. Its color is yellow and it's located at the top right corner"*. Users can either use swipe gestures to go through a list of objects or freely explore objects in a painting by touch to better understand objects' location within an image. In addition, users can also specify objects and attributes they wish to explore using filtering options. Eight participants with visual impairments were recruited for a semi-structured interview study using our prototype and provided positive feedback.

AccessArt Ver2. The major problem with the first version of AccessArt was the object segmentation process, which was not scalable as it was all manually done by a couple of researchers. Thus, we investigated the feasibility of relying on crowdworkers who were not expected have expertise in art [16]. We used Amazon Mechanical Turk (https://www.mturk.com/) for collecting object-label metadata for eight different paintings from an anonymous crowd. Then we assessed the effectiveness of the descriptions generated by crowd with nine participants with visual impairments, where they were asked to go through four steps of the *Felman Model of Criticism* [76]: *description*, *analysis*, *interpretation*, and *judgment*). Findings showed that object-level descriptions provided by anonymous crowds were sufficient for supporting BLV's artwork appreciation.

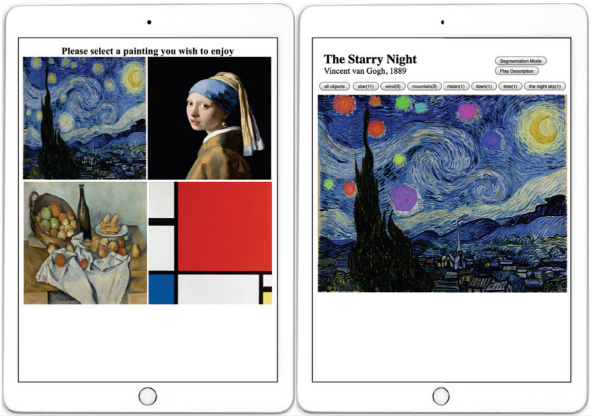


Figure 3. User interface prototype with two interaction modes: *Overview* (**left**) and *Exploration* (**right**). As for the *Exploration Mode* example, *star* is selected as object of interest, highlighted in various colors.

AccessArt Ver3. As a final step, we implemented an online platform (https://artwiki-hci2020.vercel.app/) as shown in Figure 4. It is designed to allow anonymous users to freely volunteer to provide object segmentation and description, inspired by Wikipedia (www.wikipedia.org). While no user evaluation has been conducted with the final version yet, we expect this platform to serve as an accessible online art gallery for BLV people where the metadata are collected and maintained by crowd to support a greater number of artwork, which can be accessed anywhere using one's personal device.

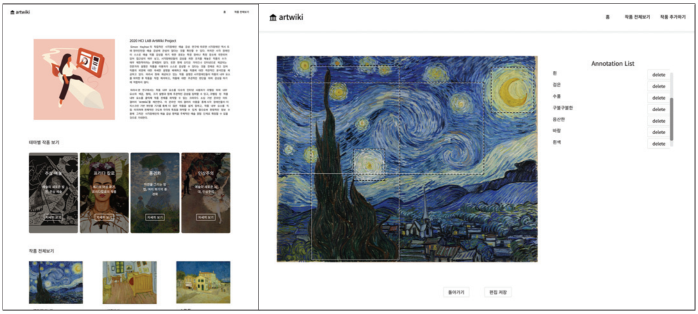
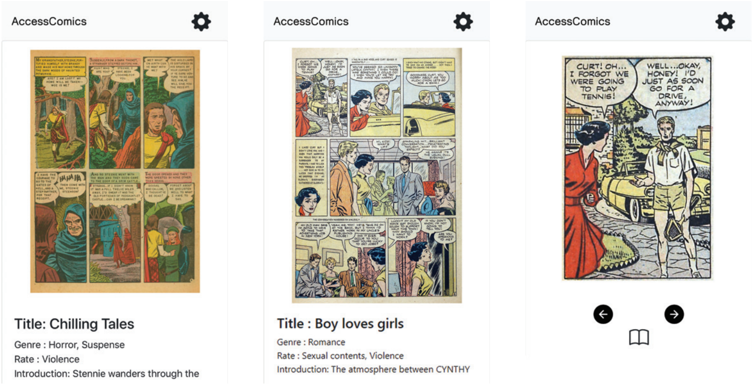


Figure 4. Screenshot examples of the main page (**left**) and edit page (**right**) of AccessArt Ver3.

5.2. AccessComics for Comics Accessibility

Compared to artwork accessibility, fewer studies have been conducted to improve the accessibility of comics. For instance, Ponsard et al. [77] proposed a system for people who have low vision or have motor impairments, which can automatically retrieve necessary information (e.g., panel detection and ordering) from images of digital comics and reads out the content on a desktop computer controlled with a TV remote. *ALCOVE* [78] is another web-based digital comic book reader for people who have low vision. The authors conducted a user study with 11 people who have low vision, and most of them preferred their system over the .pdf version of digital comics. Inspired by this study, our system, AccessComics [19], is designed to provide BLV users with overview (as shown in Figure 5a,b), various reading units (i.e., page, strip, panel), magnifier, text-to-speech, and autoplay. Moreover, we mapped different voices with different characters to offer a high sense of immersion in addition to improved accessibility similar to how Wang et al. [79] used a voice synthesis technique that can express various emotional states given scripts. Here, we briefly describe how the system is implemented.



Title: Chilling Tales
Genre : Horror, Suspense
Rate : Violence
Introduction: Stennie wanders through the

Title : Boy loves girls
Genre : Romance
Rate : Sexual contents, Violence
Introduction: The atmosphere between CYNTHY

(**a**) Introduction page ex1 (**b**) Introduction page ex2 (**c**) Reading page Ex

Figure 5. Solution and error progression curve for different times and ε values, with $n = m = 40$ and $t_{max} = 1$.

Data Preparations. As for the information we wish to provide about comics, we used *eBDtheque* [80], which is a dataset of comics consisting of pairs of an image file (.jpg) and a metadata file (.svg). The metadata has segmented information of *panel*, *character*, and *balloon* of a comic page. In addition, we manually added visual descriptions such as background and appearance and actions of characters.

Interaction. Similar to AccessArt, AccessComics allow users to select a comic book they wish to read and navigate to different elements in each panel, panels themselves as well as pages and listen to displayed content. For example, as for Figure 5c, the following would be played using the audio channel starting with the panel number, followed by background and character-related visual descriptions.

6. Discussion

Here, we discuss the current state of research on touchscreen-based image accessibility and missing gaps to be investigated in the future based on the findings from systematic review and our own experience of designing two systems for artwork and comics accessibility.

6.1. From Static Images to Dynamic Images

Various types of images displayed on touchscreen devices were studied in terms of accessibility over a decade since the year of 2008. However, all but one [42] have supported still images without motion. However, dynamically changing images such as animations and videos (e.g., movies, TV programs, games, video conferences) has rarely been explored in terms of accessibility for touchscreen devices for BLV users. Considering the rapid growth of YouTube [81] and its use for gaining knowledge [82], videos are another type of images (a series of images) that have various accessibility issues. While the area has been explored as well regardless of the medium [83–85], it would be interesting to examine how it can be supported for touchscreen devices.

6.2. Types of Information Supported for Different Image Types

As it has been found in prior work that BLV wish to get different types of information depending on the context [21], different types of information was provided for different image types. For example, geographic information such as building locations, direction, and distance were offered for map and graph images. On the other hand, shape, size, and line-length information were conveyed to users for geometric objects. However, little study has been done about other types of images, although specific locations or spatial relationships of objects within an image such as photographs and touchscreen user interface are considered important [20,30]. To identify types of information that users are interested in for each type of image, adopting recommendation techniques [86,87] can be a solution for providing user-specific content based on users' preference, interests, and needs.

6.3. Limited Room for Subjective Interpretations

The majority of the studies have prioritized images that contain useful information (e.g., facts, knowledge) over images that can be interpreted subjectively, differently from one person to another, such as artwork, using touchscreen devices. Even for artwork images, many studies have focused on delivering encyclopedia-style explanations (i.e., title, artist, painting styles) [2–5,88]. AccessArt [16] was an exception, where they demonstrated if their artwork appreciation system can enable BLV people to make their own judgements and criticism about artwork they explored. We believe that more investigations are needed to improve the experience of enjoying the content of the images or of making decisions based on subjective judgements for BLV people (e.g., providing a summary of product reviews of others as a reference).

6.4. Automatic Retrieval of Metadata of Images for Scalability

The greatest number of studies that we have identified in our systematic review, all but seven out of 33 papers, assumed that image-related information is given. If not, researchers manually created the information. However, a number of images on the web do not have alt text, although it is recommended by Web Content Accessibility Guidelines (WCAG) (https://www.w3.org/TR/WCAG21/). Moreover, it is not feasible for a couple of researchers to generate metadata for individual images. Thus, automatic approaches such as machine learning techniques have been studied [12,13,89,90]. However, since the accuracy of descriptions produced by humans is not as high, we recommend the crowdsourcing approach for generating descriptions [18,20] if precise annotations are needed. This can serve as a human–AI collaboration for validating auto-generated annotations [13]. Eventually, these data can be used to train machine learning models for implementing a fully automated image description generation system [91–93].

6.5. Limited Input and Output Modalities of Touchscreen Devices

Unlike other assistive systems that require BLV to physically visit certain locations (e.g., [75,88]) or that require special hardware devices with tactile cues (e.g., [3,4]), touchscreen devices benefit from being portable, where a variety of images can be accessed using a personal device with less physical and time constraints. However, the input and output modalities that touchscreen devices can offer are limited to audio and vibration feedback. To provide more intuitive and rich feedback, more in-depth studies on how to ease the design of 2.5D or 3D models and how the cost and time for producing tactile representations of images can be reduced should be conducted. One way to do so is open-sourcing the process, as in Instructables (https://www.instructables.com/), which is an online community where people explore and share instructions for do-it-yourself projects. Meanwhile, we also recommend touchscreen-based approaches to make a larger number of images accessible to a greater number of BLV people.

6.6. Limited Involvement of BLV People during Design Process

The findings of our systematic review revealed that most studies had user evaluation of proposed systems with BLV participants after the design and implementation. However, it is important to have target users participate in the design process at an early state when developing a new technology [94]. A formative study with surveys or semi-structured interviews is recommended to understand the current needs and challenges of BLV people before making design decisions. Iterative participatory design process is also great way to reflect BLV participants' opinions into the design, especially for users with disabilities [95–97].

7. Conclusions

To have complete understanding of existing approaches and identify challenges to be solved as the next step, we conducted a systematic review of 33 papers on touchscreen-based image accessibility for screen reader users. The results revealed that image types other than maps, graphs and geometric shapes such as artwork and comics are rarely studied. Furthermore, we found that only about one-third of the papers provide multi-modal feedback of audio and haptic. Moreover, our findings show that ways to collect image descriptions was out of the scope of interest for most studies, suggesting that automatic retrievals of image-related information is one of the bottlenecks for making images accessible on a large scale. Finally, while the majority of studies did not involve people who are blind or have low vision during the system design process, future studies should consider inviting target users early in advance and reflect their comments for making design decisions.

Author Contributions: Conceptualization, U.O.; methodology, U.O.; validation, U.O.; formal analysis, H.J. and Y.L.; investigation, U.O., H.J., and Y.L.; data curation, U.O., H.J., and Y.L.; writing—original draft preparation, U.O.; writing—review and editing, U.O.; visualization, U.O.; supervision, U.O.; project administration, U.O.; funding acquisition, U.O. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the Science Technology and Humanity Converging Research Program of National Research Foundation of Korea (2018M3C1B6061353).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blindness and Vision Impairment. Available online: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (accessed on 29 March 2021).
2. Gyoshev, S.; Karastoyanov, D.; Stoimenov, N.; Cantoni, V.; Lombardi, L.; Setti, A. Exploiting a graphical Braille display for art masterpieces. In Proceedings of the International Conference on Computers Helping People with Special Needs, Linz, Austria, 11–13 July 2018; pp. 237–245.
3. Cantoni, V.; Lombardi, L.; Setti, A.; Gyoshev, S.; Karastoyanov, D.; Stoimenov, N. Art Masterpieces Accessibility for Blind and Visually Impaired People. In Proceedings of the International Conference on Computers Helping People with Special Needs, Linz, Austria, 11–13 July 2018; pp. 267–274.
4. Iranzo Bartolome, J.; Cavazos Quero, L.; Kim, S.; Um, M.Y.; Cho, J. Exploring Art with a Voice Controlled Multimodal Guide for Blind People. In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, Tempe, AZ, USA, 17–20 March 2019; pp. 383–390.
5. Cavazos Quero, L.; Iranzo Bartolomé, J.; Lee, S.; Han, E.; Kim, S.; Cho, J. An Interactive Multimodal Guide to Improve Art Accessibility for Blind People. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; pp. 346–348.
6. Kane, S.K.; Morris, M.R.; Wobbrock, J.O. Touchplates: Low-cost tactile overlays for visually impaired touch screen users. In Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, Bellevue, DC, USA, 21–23 October 2013, pp. 1–8.
7. Holloway, L.; Marriott, K.; Butler, M. Accessible maps for the blind: Comparing 3D printed models with tactile graphics. In Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13.
8. Panotopoulou, A.; Zhang, X.; Qiu, T.; Yang, X.D.; Whiting, E. Tactile line drawings for improved shape understanding in blind and visually impaired users. *ACM Trans. Graph. (TOG)* **2020**, *39*, 89. [CrossRef]
9. Paterson, M. 'Seeing with the hands': Blindness, touch and the Enlightenment spatial imaginary. *Br. J. Vis. Impair.* **2006**, *24*, 52–59. [CrossRef]
10. Heller, M.A. *The Psychology of Touch*; Psychology Press: East Sussex, UK, 2013.
11. Watanabe, T.; Yamaguchi, T.; Nakagawa, M. Development of software for automatic creation of embossed graphs. In Proceedings of the International Conference on Computers for Handicapped Persons, Linz, Austria, 11–13 July 2012; pp. 174–181.
12. Stangl, A.J.; Kothari, E.; Jain, S.D.; Yeh, T.; Grauman, K.; Gurari, D. BrowseWithMe: An Online Clothes Shopping Assistant for People with Visual Impairments. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; pp. 107–118.
13. Zhong, Y.; Matsubara, M.; Morishima, A. Identification of Important Images for Understanding Web Pages. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 3568–3574.

14. Winters, R.M.; Joshi, N.; Cutrell, E.; Morris, M.R. Strategies for auditory display of Social Media. *Ergon. Des.* **2019**, *27*, 11–15. [CrossRef]
15. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Prisma Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* **2010**, *8*, 336–341. [CrossRef] [PubMed]
16. Nahyun Kwon, Y.L.; Oh, U. Supporting a Crowd-powered Accessible Online Art Gallery for People with Visual Impairments: A Feasibility Study. *Univers. Access Inf. Soc.* **2021**, accepted.
17. Kwon, N.; Koh, Y.; Oh, U. Supporting object-level exploration of artworks by touch for people with visual impairments. In Proceedings of the Conference of 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2019), Pittsburgh, PA, USA, 28–30 October 2019; pp. 600–602. [CrossRef]
18. Soohyun Yeo, S.C.; Oh, U. Crowdsourcing Based Online Art Gallery for Visual Impairment People. *HCI Korea Annu. Conf.* **2021**, *33*, 531–535.
19. Lee, Y.; Joh, H.; Yoo, S.; Oh, U. AccessComics: An Accessible Digital Comic Book Reader for People with Visual Impairments. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 12–22 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; ISBN 978-1-4503-5650-3.
20. Morris, M.R.; Johnson, J.; Bennett, C.L.; Cutrell, E. Rich representations of visual content for screen reader users. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; p. 59.
21. Stangl, A.; Morris, M.R.; Gurari, D. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–13.
22. Goncu, C.; Marriott, K. Creating ebooks with accessible graphics content. In Proceedings of the 2015 ACM Symposium on Document Engineering, Lausanne, Switzerland, 8–11 September 2015; pp. 89–92.
23. Götzelmann, T. CapMaps: Capacitive sensing 3D printed audio-tactile maps. In Proceedings of the Conference of 15th International Conference on Computers Helping People with Special Needs (ICCHP 2016), Linz, Austria, 13–15 July 2016; Volume 9759, pp. 146–152. [CrossRef]
24. Lazar, J.; Chakraborty, S.; Carroll, D.; Weir, R.; Sizemore, B.; Henderson, H. Development and Evaluation of Two Prototypes for Providing Weather Map Data to Blind Users through Sonification. *J. Usability Stud.* **2013**, *8*, 93–110.
25. Rodrigues, J.B.; Ferreira, A.V.M.; Maia, I.M.O.; Junior, G.B.; de Almeida, J.D.S.; de Paiva, A.C. Image Processing of Artworks for Construction of 3D Models Accessible to the Visually Impaired. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Orlando, FL, USA, 21–25 July 2018; pp. 243–253.
26. McGookin, D.; Brewster, S.; Jiang, W. Investigating touchscreen accessibility for people with visual impairments. In Proceedings of the 5th Nordic conference on Human–computer interaction: Building bridges, Lund, Sweden, 20–22 October 2008; pp. 298–307.
27. Vanderheiden, G.C. Use of audio-haptic interface techniques to allow nonvisual access to touchscreen appliances. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; SAGE Publications Sage CA: Los Angeles, CA, USA, 1996; Volume 40, p. 1266.
28. Landau, S.; Wells, L. Merging tactile sensory input and audio data by means of the Talking Tactile Tablet. In Proceedings of the EuroHaptics, Dublin, Ireland, 6–9 July 2003; Volume 3, pp. 414–418.
29. Stearns, L.; Oh, U.; Findlater, L.; Froehlich, J.E. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *1*, 1–23. [CrossRef]
30. Kane, S.K.; Morris, M.R.; Perkins, A.Z.; Wigdor, D.; Ladner, R.E.; Wobbrock, J.O. Access overlays: Improving non-visual access to large touch screens for blind users. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA, USA, 16–19 October 2011; pp. 273–282.
31. Kane, S.K.; Bigham, J.P.; Wobbrock, J.O. Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. In Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, Halifax, NS, Canada, 13–15 October 2008; pp. 73–80.
32. Frey, B.; Southern, C.; Romero, M. Brailletouch: Mobile texting for the visually impaired. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, Orlando, FL, USA, 9–14 July 2011; pp. 19–25.
33. Guerreiro, T.; Lagoa, P.; Nicolau, H.; Goncalves, D.; Jorge, J.A. From Tapping to Touching: Making Touch Screens Accessible to Blind Users (vol 15, pg 48, 2008). *IEEE MULTIMEDIA* **2009**, *16*, 13.
34. Bonner, M.N.; Brudvik, J.T.; Abowd, G.D.; Edwards, W.K. No-look notes: Accessible eyes-free multi-touch text entry. In Proceedings of the International Conference on Pervasive Computing, Helsinki, Finland, 17–20 May 2010; pp. 409–426.
35. Toffa, O.; Mignotte, M. A Hierarchical Visual Feature-Based Approach for Image Sonification. *IEEE Trans. Multimed.* **2021**, *23*, 706–715. [CrossRef]
36. Banf, M.; Blanz, V. Sonification of images for the visually impaired using a multi-level approach. In Proceedings of the Conference of 4th Augmented Human International Conference, Stuttgart, Germany, 7–8 March 2013; pp. 162–169. [CrossRef]
37. Neto, J.; Freire, A.; Souto, S.; Abílio, R. Usability evaluation of a web system for spatially oriented audio descriptions of images addressed to visually impaired people. In Proceedings of the Conference of 8th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2014—Held as Part of 16th International Conference on Human-Computer Interaction, 8514 LNCS, HCI International 2014, Crete, Greece, 22–27 June 2014; pp. 154–165. [CrossRef]

38. Zhao, Y.; Wu, S.; Reynolds, L.; Azenkot, S. The Effect of Computer-Generated Descriptions on Photo-Sharing Experiences of People with Visual Impairments. *Proc. ACM Hum.-Comput. Interact.* **2017**, *1*. [CrossRef]
39. Onishi, J.; Sakajiri, M.; Miura, T.; Ono, T. Fundamental study on tactile cognition through haptic feedback touchscreen. In Proceedings of the Conference of 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013, Manchester, UK, 13–16 October 2013; pp. 4207–4212. [CrossRef]
40. Kyung, K.U.; Lee, J.Y.; Park, J. Haptic stylus and empirical studies on braille, button, and texture display. *J. Biomed. Biotechnol.* **2008**, *2008*. [CrossRef] [PubMed]
41. Zhang, X.; Tran, T.; Sun, Y.; Culhane, I.; Jain, S.; Fogarty, J.; Mankoff, J. Interactiles: 3D printed tactile interfaces to enhance mobile touchscreen accessibility. In Proceedings of the Conference of 20th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2018, Galway, Ireland, 22–24 October 2018; pp. 131–142. [CrossRef]
42. Oh, U.; Branham, S.; Findlater, L.; Kane, S. Audio-based feedback techniques for teaching touchscreen gestures. *ACM Trans. Access. Comput. (TACCESS)* **2015**, *7*. [CrossRef]
43. Rodriguez-Sanchez, M.; Moreno-Alvarez, M.; Martin, E.; Borromeo, S.; Hernandez-Tamames, J. Accessible smartphones for blind users: A case study for a wayfinding system. *Expert Syst. Appl.* **2014**, *41*, 7210–7222. [CrossRef]
44. Watanabe, T.; Kaga, H.; Yagi, T. Evaluation of virtual tactile dots on touchscreens in map reading: Perception of distance and direction. *J. Adv. Comput. Intell. Intell. Inform.* **2017**, *21*, 79–86. [CrossRef]
45. Grussenmeyer, W.; Garcia, J.; Jiang, F. Feasibility of using haptic directions through maps with a tablet and smart watch for people who are blind and visually impaired. In Proceedings of the Conference of 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016, Florence, Italy, 6–9 September 2016; pp. 83–89. [CrossRef]
46. Götzelmann, T. LucentMaps: 3D printed audiovisual tactile maps for blind and visually impaired people. In Proceedings of the Conference of 18th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2016, Reno, NA, USA, 24–26 October 2016; pp. 81–90. [CrossRef]
47. Yatani, K.; Banovic, N.; Truong, K. SpaceSense: Representing geographical information to visually impaired people using spatial tactile feedback. In Proceedings of the Conference of 30th ACM Conference on Human Factors in Computing Systems, CHI 2012, Austin, TX, USA, 5–10 May 2012; pp. 415–424. [CrossRef]
48. Su, J.; Rosenzweig, A.; Goel, A.; De Lara, E.; Truong, K. Timbremap: Enabling the visually-impaired to use maps on touch-enabled devices. In Proceedings of the Conference of 12th International Conference on Human-Computer Interaction with Mobile Devices and Services, Mobile HCI2010, Lisbon, Portugal, 7–10 September 2010; pp. 17–26. [CrossRef]
49. Götzelmann, T. Visually augmented audio-tactile graphics for visually impaired people. *ACM Trans. Access. Comput. (TACCESS)* **2018**, *11*, 1–31. [CrossRef]
50. Bateman, A.; Zhao, O.; Bajcsy, A.; Jennings, M.; Toth, B.; Cohen, A.; Horton, E.; Khattar, A.; Kuo, R.; Lee, F.; et al. A user-centered design and analysis of an electrostatic haptic touchscreen system for students with visual impairments. *Int. J. -Hum.-Comput. Stud.* **2018**, *109*, 102–111. [CrossRef]
51. Gorlewicz, J.; Tennison, J.; Uesbeck, P.; Richard, M.; Palani, H.; Stefik, A.; Smith, D.; Giudice, N. Design Guidelines and Recommendations for Multimodal, Touchscreen-based Graphics. *ACM Trans. Access. Comput. (TACCESS)* **2020**, *13*. [CrossRef]
52. Tekli, J.; Issa, Y.; Chbeir, R. Evaluating touch-screen vibration modality for blind users to access simple shapes and graphics. *Int. J. Hum. Comput. Stud.* **2018**. *110*, 115–133. [CrossRef]
53. Hahn, M.; Mueller, C.; Gorlewicz, J. The Comprehension of STEM Graphics via a Multisensory Tablet Electronic Device by Students with Visual Impairments. *J. Vis. Impair. Blind.* **2019**, *113*, 404–418. [CrossRef]
54. Klatzky, R.; Giudice, N.; Bennett, C.; Loomis, J. Touch-screen technology for the dynamic display of 2d spatial information without vision: Promise and progress. *Multisensory Res.* **2014**, *27*, 359–378. [CrossRef]
55. Silva, P.; Pappas, T.; Atkins, J.; West, J.; Hartmann, W. Acoustic-tactile rendering of visual information. In Proceedings of the Conference of Human Vision and Electronic Imaging XVII, Burlingame, CA, USA, 23–26 January 2012; Volume 8291. [CrossRef]
56. Hausberger, T.; Terzer, M.; Enneking, F.; Jonas, Z.; Kim, Y. SurfTics-Kinesthetic and tactile feedback on a touchscreen device. In Proceedings of the Conference of 7th IEEE World Haptics Conference (WHC 2017), Munich, Germany, 6–9 June 2017; pp. 472–477. [CrossRef]
57. Palani, H.; Fink, P.; Giudice, N. Design Guidelines for Schematizing and Rendering Haptically Perceivable Graphical Elements on Touchscreen Devices. *Int. J. Hum.-Comput. Interact.* **2020**, *36*, 1393–1414. [CrossRef]
58. Tennison, J.; Gorlewicz, J. Non-visual Perception of Lines on a Multimodal Touchscreen Tablet. *ACM Trans. Appl. Percept. (TAP)* **2019**, *16*. [CrossRef]
59. Palani, H.; Tennison, J.; Giudice, G.; Giudice, N. Touchscreen-based haptic information access for assisting blind and visually-impaired users: Perceptual parameters and design guidelines. In Proceedings of the Conference of AHFE International Conferences on Usability and User Experience and Human Factors and Assistive Technology, Orlando, FL, USA, 21–25 July 2018; Volume 794, pp. 837–847. [CrossRef]
60. Zhong, Y.; Lasecki, W.; Brady, E.; Bigham, J. Regionspeak: Quick comprehensive spatial descriptionsof complex images for blind users. In Proceedings of the Conference of 33rd Annual CHI Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Korea, 18–23 April 2015; Volume 2015, pp. 2353–2362. [CrossRef]

61. Guo, A.; Kong, J.; Rivera, M.; Xu, F.; Bigham, J. StateLens: A reverse engineering solution for making existing dynamic touchscreens accessible. In Proceedings of the Conference of 32nd Annual ACM Symposium on User Interface Software and Technology, UIST 2019, New Orleans, LA, USA, 20–23 October 2019; pp. 371–385. [CrossRef]
62. Goncu, C.; Madugalla, A.; Marinai, S.; Marriott, K. Accessible on-line floor plans. In Proceedings of the Conference of 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, 18–22 May 2015; pp. 388–398. [CrossRef]
63. Madugalla, A.; Marriott, K.; Marinai, S.; Capobianco, S.; Goncu, C. Creating Accessible Online Floor Plans for Visually Impaired Readers. *ACM Trans. Access. Comput. (TACCESS)* **2020**, *13*. [CrossRef]
64. Taylor, B.; Dey, A.; Siewiorek, D.; Smailagic, A. Customizable 3D printed tactile maps as interactive overlays. In Proceedings of the Conference of 18th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2016, Reno, NA, USA, 24–26 October 2016; pp. 71–79. [CrossRef]
65. Grussenmeyer, W.; Folmer, E. AudioDraw: User preferences in non-visual diagram drawing for touchscreens. In Proceedings of the Conference of 13th Web for All Conference, Montreal, QC, Canada, 11–13 April 2016. [CrossRef]
66. Reinholt, K.; Guinness, D.; Kane, S. Eyedescribe: Combining eye gaze and speech to automatically create accessible touch screen artwork. In Proceedings of the 14th ACM International Conference on Interactive Surfaces and Spaces, ISS 2019, Daejeon, Korea, 10–13 November 2019; pp. 101–112. [CrossRef]
67. Jayant, C.; Ji, H.; White, S.; Bigham, J.P. Supporting blind photography. In Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, Dundee, UK, 24–26 October 2011; pp. 203–210.
68. Hayhoe, S. Expanding our vision of museum education and perception: An analysis of three case studies of independent blind arts learners. *Harv. Educ. Rev.* **2013**, *83*, 67–86. [CrossRef]
69. Handa, K.; Dairoku, H.; Toriyama, Y. Investigation of priority needs in terms of museum service accessibility for visually impaired visitors. *Br. J. Vis. Impair.* **2010**, *28*, 221–234. [CrossRef]
70. Hayhoe, S. Why Do We Think That People who are Visually Impaired Don't Want to Know About the Visual Arts II? In Proceedings of the Workshop: Creative Sandpit@ the (Virtual) American Museum, Bath, UK, 9–10 November 2020; Creative Sandpit: Bath, UK, 2020.
71. Asakawa, S.; Guerreiro, J.; Ahmetovic, D.; Kitani, K.M.; Asakawa, C. The Present and Future of Museum Accessibility for People with Visual Impairments. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; pp. 382–384.
72. The Audio Guide by Metropolitan Museum of Art. Available online: https://www.metmuseum.org/visit/audio-guide (accessed on 20 April 2019).
73. Accessibility at Museum of Modern Art. Available online: http://www.moma.org/learn/disabilities/sight (accessed on 20 April 2019).
74. Art Beyond Sight Tours at Seattle Art Museum. Available online: http://www.seattleartmuseum.org/visit/accessibility#tou (accessed on 20 April 2019).
75. Out Loud, The Andy Warhol Museum's Inclusive Audio Guide. Available online: https://itunes.apple.com/us/app/the-warhol-out-loud/id1103407119 (accessed on 15 April 2021).
76. Feldman, E.B. *Practical Art Criticism*; Pearson: Leondon, UK, 1994.
77. Ponsard, C.; Fries, V. An accessible viewer for digital comic books. In Proceedings of the International Conference on Computers for Handicapped Persons, Linz, Austria, 9–11 July 2008; pp. 569–577.
78. Rayar, F.; Oriola, B.; Jouffrais, C. ALCOVE: An accessible comic reader for people with low vision. In Proceedings of the 25th International Conference on Intelligent User Interfaces, Cagliari, Italy, 17 March 2020; pp. 410–418.
79. Wang, Y.; Wang, W.; Liang, W.; Yu, L.F. Comic-guided speech synthesis. *ACM Trans. Graph. (TOG)* **2019**, *38*, 1–14. [CrossRef]
80. Guérin, C.; Rigaud, C.; Mercier, A.; Ammar-Boudjelal, F.; Bertet, K.; Bouju, A.; Burie, J.C.; Louis, G.; Ogier, J.M.; Revel, A. eBDtheque: A representative database of comics. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 25–28 August 2013; pp. 1145–1149.
81. Figueiredo, F.; Benevenuto, F.; Almeida, J.M. The tube over time: Characterizing popularity growth of youtube videos. In Proceedings of the fourth ACM International Conference on Web Search and Data Mining, Hong Kong, China, 9–12 February 2011; pp. 745–754.
82. Hartley, J. Uses of YouTube: Digital literacy and the growth of knowledge. In *YouTube: Online Video and Participatory Culture*; Digital Media and Society: Cambridge, UK, 2009; pp. 126–143.
83. Yuksel, B.F.; Fazli, P.; Mathur, U.; Bisht, V.; Kim, S.J.; Lee, J.J.; Jin, S.J.; Siu, Y.T.; Miele, J.A.; Yoon, I. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. In Proceedings of the 2020 ACM Designing Interactive Systems Conference, Eindhoven, The Netherlands, 6–10 July 2020; ACM: New York, NY, USA, 2020; pp. 47–60.
84. Encelle, B.; Ollagnier-Beldame, M.; Pouchot, S.; Prié, Y. Annotation-based video enrichment for blind people: A pilot study on the use of earcons and speech synthesis. In Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, Dundee, UK, 24–26 October 2011; pp. 123–130.
85. Maneesaeng, N.; Punyabukkana, P.; Suchato, A. Accessible video-call application on android for the blind. *Lect. Notes Softw. Eng.* **2016**, *4*, 95. [CrossRef]
86. Stai, E.; Kafetzoglou, S.; Tsiropoulou, E.E.; Papavassiliou, S. A holistic approach for personalization, relevance feedback & recommendation in enriched multimedia content. *Multimed. Tools Appl.* **2018**, *77*, 283–326.

87. Pouli, V.; Kafetzoglou, S.; Tsiropoulou, E.E.; Dimitriou, A.; Papavassiliou, S. Personalized multimedia content retrieval through relevance feedback techniques for enhanced user experience. In Proceedings of the 2015 13th International Conference on Telecommunications (ConTEL), Graz, Austria, 13–15 July 2015; pp. 1–8.
88. Rector, K.; Salmon, K.; Thornton, D.; Joshi, N.; Morris, M.R. Eyes-free art: Exploring proxemic audio interfaces for blind and low vision art engagement. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 93. [CrossRef]
89. Wu, S.; Wieland, J.; Farivar, O.; Schiller, J. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, ON, USA, 25 February–1 March 2017; pp. 1180–1192.
90. Huang, X.L.; Ma, X.; Hu, F. Machine learning and intelligent communications. *Mob. Netw. Appl.* **2018**, *23*, 68–70. [CrossRef]
91. Vinyals, O.; Toshev, A.; Bengio, S.; Erhan, D. Show and tell: A neural image caption generator. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3156–3164.
92. Dai, B.; Fidler, S.; Urtasun, R.; Lin, D. Towards diverse and natural image descriptions via a conditional gan. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2970–2979.
93. Li, Y.; Ouyang, W.; Zhou, B.; Wang, K.; Wang, X. Scene graph generation from objects, phrases and region captions. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1261–1270.
94. Fischer, B.; Peine, A.; Östlund, B. The importance of user involvement: A systematic review of involving older users in technology design. *Gerontologist* **2020**, *60*, e513–e523. [CrossRef] [PubMed]
95. Schuler, D.; Namioka, A. *Participatory Design: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 1993.
96. Sanders, E.B. From user-centered to participatory design approaches. *Des. Soc. Sci. Mak. Connect.* **2002**, *1*, 1.
97. Sahib, N.G.; Stockman, T.; Tombros, A.; Metatla, O. Participatory design with blind users: A scenario-based approach. In *IFIP Conference on Human-Computer Interaction*; Springer: Berlin, Germany, 2013; pp. 685–701.

electronics

Review

A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment

Jun Dong Cho

Department of Human Information and Cognition Technology Convergence, SungkyunKwan University, Seoul 16419, Korea; jdcho@skku.edu; Tel.: +82-10-4332-7127

Abstract: Visually impaired visitors experience many limitations when visiting museum exhibits, such as a lack of cognitive and sensory access to exhibits or replicas. Contemporary art is evolving in the direction of appreciation beyond simply looking at works, and the development of various sensory technologies has had a great influence on culture and art. Thus, opportunities for people with visual impairments to appreciate visual artworks through various senses such as hearing, touch, and smell are expanding. However, it is uncommon to provide an interactive interface for color recognition, such as applying patterns, sounds, temperature, or scents. This review aims to convey the visual elements of the work to the visually impaired through various sensory elements. In addition, to open a new perspective on appreciation of the works, the technique of expressing the color coded by integrating patterns, temperature, scent, music, and vibration was explored, and future research topics were presented.

Keywords: visual impairment; accessibility; aesthetics; color; multi-sensory; museum exhibits

Citation: Cho, J.D. A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment. *Electronics* **2021**, *10*, 470. https://doi.org/10.3390/electronics10040470

Academic Editor: Jun Liu and Osvaldo Gervasi

Received: 19 December 2020
Accepted: 9 February 2021
Published: 15 February 2021

1. Introduction

Around 1.3 billion people worldwide live with some form of blindness, and their limited access to artwork cannot be ignored in a world of increasing inclusion. Museums are obligated to accommodate people with varying needs, including people with visual impairments [1]. Art is arguably one of the most intriguing creations of humanity, and as such should be available to every person; accordingly, making visual art available to the visually impaired has become a priority. However, for the visually impaired, visiting a museum can feel alienating.

Therefore, it is necessary to expand research on universal exhibition art appreciation assistance, exhibition contents, and art exhibition environment for the visually impaired. In other words, the development of technology to interpret the context of artwork using non-visual sensory forms such as sound, color, texture, and temperature is positive, and through this, the visually impaired can open a new way to enjoy art and culture in social and psychological aspects. For this reason, many studies have been conducted.

However, the use of vision and sound for interaction has dominated the field of human–computer interaction for decades, even though humans have many more senses for perceiving and interacting with the world. Recently, researchers have started trying to capitalize on touch, taste, and smell when designing interactive tasks, especially in gaming, multimedia, and art environments. The concept of multimodality, or communicating information by means of several sensations, has been of vital importance in the field of human–computer interaction. More innovative approaches can be used, such as multisensory displays that appeal to sight, hearing, touch, and smell [2]. The combination of the strengths of several interfaces allows for a more efficient user–machine communication, which cannot be accomplished by means of a single interaction mode alone. The combination of the strengths of various modes can make up for the lack of sense of vision for the visually impaired. One way to cultivate social, cognitive, and emotional empathy is to appreciate artworks through multiple senses (sight, hearing, touch, smell, etc.) [3].

Based on such thoughts, multiple senses can work together to increase the experience of the visually impaired allowing indirect sensing of the colors and images of the exhibits through media such as sound, texture, temperature, and scent. These technologies not only help the visually impaired enjoy the museum experience, but also allow sighted people to view museum exhibits in a new way.

Museums are evolving to provide enjoyable experiences for everyone, moving beyond audio guides to tactile exhibitions [4,5]. A previous study [6,7] reviewed the extent and nature of participatory research and accessibility in the context of assistive technologies developed for use in museums by people with sensory impairments or learning disabilities. Some museums have successfully produced art replicas that can be tactilely experienced. For example, the Metropolitan Museum of Art in New York has displayed replicas of the artworks exhibited in the museum [8,9]. The American Foundation for the Blind offered guidelines and resources for the use of tactile graphics for the specific case of artworks [10]. The Art Institute of Chicago also uses 3D-printed copies of its collection to support its curriculum for design students. Converting artworks into 2.5D or 3D allows the visually impaired to enjoy them using touch, with audio descriptions and sound effects provided to enhance the experience. In 2.5D printing, a relief model, a tactile diagram of a computer-edited drawing, is printed onto microcapsule paper called swell paper that enables the visually impaired to easily distinguish the texture and thickness of lines [10]. Bas-relief tactile painting is a sculptural technique that produces specific shapes that protrude on a plane [10]. The quality of the relief is measured using the perceived quality of the represented 3D shape. Three-dimensional (3D)-printed artworks are effective learning tools to allow people to experience objects from various perspectives, improving the accessibility of art appreciation and visual descriptive skills of the visually impaired by providing an interactive learning environment [11]. Such 3D printing technology improves access to art by allowing the visually impaired to touch and imagine an artwork. For example, the Belvedere Museum in Vienna used 3D printing technology to create a 3D version of Gustav Klimt's "The Kiss" [12] and the Andy Warhol Museum [13] released a comprehensive audio guide that allows visitors to touch 3D replicas of artwork during an audio tour.

Furthermore, colors should not be forgotten, because they retain a symbolic meaning, even for children without sight. Color is an absolute element that gives depth, form, and motion to a painting. Colors are expressed in such a way that different feelings can pop out of objects. Layers of color can provide an infinite variety of sensory feelings and show multi-layered diversity, liberating objects from ideas. According to the perception theorem, viewers give meaning to a work according to their experiences, and therefore, color is not an objective attribute, but a matter of perception that exists in the mind of the perceiver. Therefore, this review also attempts to convey color to the visually impaired through multiple sensory elements.

This review is organized as follows. In Section 2, some examples of multisensory art reproduction in museums and the museum's multi-sensory experiences of touch, smell, and hearing will be addressed. In Section 3, we look at how to express colors through sound, pictograms, and temperature. Section 4 will be dedicated to non-visual multi-sensory integration. Finally, conclusions will be drawn in Section 5.

2. Multi-Sensory Experiences in Museums

Multi-sensory interaction aids learning, inclusion, and collaboration, because it accommodates the diverse cognitive and perceptual needs of the visually impaired. Multiple sensory systems are needed to successfully convey artistic images. However, few studies have analyzed the application of assistive technologies in multisensory exhibit designs and related them to visitors' experiences. The museums of providing multiple senses that are possible in the method of delivering art works appreciation considering the visually impaired are Birmingham Museum of Art, Cummer Museum of Art and Garden, Finnish National Gallery, The Jewish Museum, Metropolitan Museum of Art, Omero Museum,

Museum of Fine Art, Museum of Modern Art, Cooper Hewitt Smithsonian Design Museum, The National Gallery, Philadelphia Museum of Art, Queens Museum of Art, Tate Modern, Smithsonian American Art Museum, and van Abbe Museum. They operate a variety of tours and programs to allow the visually impaired to experience art. The monthly tour provides an opportunity to touch the exhibits by providing sensory explanations and tactile aids through audio. There are also braille printers, 3D printers, voice information technology, tactile image-to-speech technology, color change technology, etc. in the form of helping the audience to transform the work or appreciating the exhibition by carrying auxiliary tools.

The "Eyes of the Mind" series at the Guggenheim Museum in New York also offered a "sensory experience workshop" for museum visitors with visual impairment or low vision. In addition to describing the artworks, it used the senses of touch and smell. In the visually impaired, the sense of touch can stimulate neurons that are usually reserved for vision. Neuroscience suggests that, with the right tools, the visually impaired can appreciate the visual arts, because the essence of a picture is not vision but a meaningful connection between the artist and the audience.

2.1. Tactile Perception of the Visually Impaired

The sense of touch is an important source of information when sight is absent. According to many studies, tactile spatial acuity is enhanced in blindness. Already in 1964, scientists demonstrated that seven days of visual deprivation resulted in tactile acuity enhancement. There are two competing hypotheses on how blindness improves the sense of touch. According to the tactile experience hypothesis, reliance on the sense of touch drives tactile-acuity enhancement. The visual deprivation hypothesis, on the other hand, posits that the absence of vision itself drives tactile-acuity enhancement. Wong et al. [14] tested the participants' ability to discern the orientations of grooved surfaces applied to the distal pads of the stationary index, middle, and ring fingers of each hand, and then to the two sides of the lower lip. A study comparing those hypotheses demonstrated that proficient Braille readers—those who spend hours a day reading with their fingertips—have much more sensitive fingers than sighted people, confirming the tactile experience hypothesis. In contrast, blind and sighted participants performed equivalently on the lips. If the visual deprivation hypothesis were true, blind participants would outperform sighted people in all body areas [14].

Heller [15] reported two experiments on the contribution of visual experience to tactile perception. In the first experiment, sighted, congenitally blind, and late blind individuals made tactual matches to tangible embossed shapes. In the second experiment, the same subjects attempted tactile identification of raised-line drawings. The three groups did not differ in the accuracy of their shape matching, but both groups of blind subjects were much faster than the sighted. Late (acquired) blind observers were far better than the sighted or congenitally blind participants at tactile picture identification. Four of the twelve pictures were correctly identified by most of the late blind subjects. The sighted and congenitally blind participants performed at comparable levels in picture naming. There was no evidence that visual experience alone aided the sighted in the tactile task under investigation, because they performed no better than the early blind. The superiority of the late blind suggests that visual exposure to drawings and the rules of pictorial representation could help in tactile picture identification when combined with a history of tactual experience [15].

2.2. Tactile Graphics and Overlays

Tactile graphics are made using raised lines and textures to convey drawings and images by touch. Advances in low-cost prototyping and 3D printing technologies aim to tackle the complexity of expressing complex images without exploration obstruction by adding interactivity to tactile graphics. Thus, the combination of tactile graphics, interactive interfaces, and audio descriptions can improve accessibility and understanding of visual art works for the visually impaired. Taylor et al. [16] presented a gesture-controlled interactive audio guide based on low-cost depth cameras that can track hand gestures on relief surfaces during tactile exploration of artworks. Conductive filament was used to provide touchscreen overlays. LucentMaps developed by Götzelmann et al. [17] uses 3D-printed tactile maps with embedded capacitive material that, when overlaid on a touchscreen device, can generate audio in response to touch. They also provided the results of a survey done with 19 visually impaired participants to identify their previous experiences, motivations, and accessibility challenges in museums. An interactive multimodal guide uses both touch and audio to take advantage of the strengths of each mode and provide localized verbal descriptions.

While mobile screen readers have improved access to touchscreen devices for people with visual impairments, graphical forms of information such as maps, charts, and images are still difficult to convey and understand. The Talking Tactile Tablet developed by Landau et al. [18] allows users to place tactile sheets on top of a tablet that can then sense a user's touches. The Talking Tactile Tablet holds tactile graphic sheets motionless against a high-resolution touch-sensitive surface. A user's finger pressure is transmitted through a variety of flexible tactile graphic overlays to this surface, which is a standard hardened-glass touch screen, typically used in conjunction with a video monitor for ATMs and other applications. The computer interprets the user's presses on the tactile graphic overlay sheet in the same way that it does when a mouse is clicked while the cursor is over a particular region, icon, or object on a video screen [18]. Table 1 summarizes a list of these projects and their interaction technologies.

Table 1. Interactive Tactile Graphics and Multimodal Guide for educations and map explorations.

Author–Name	Input	Output	Focus
Taylor et al. [16]	Touch (Touchscreen)	Tactile overlay Verbal descriptions	Map Exploration Geography Education
Gotzelmann et al. [17] LucentMaps		Voice, Tactile overlay Visual augmentation	
Brule et al. [19] MapSense		Tokens (Capacitive) Tactile overlay Smell and taste infused tangible tokens Verbal descriptions	
Landau et al. [18] The Talking Tactile Tablet	Touch (Surface)	Tactile overlay Verbal descriptions	Map Exploration Education and Scientific Diagrams
Shen et al. [20] CamIO	Touch (Mounted camera)	Tactile graph Tactile 3D Map Tactile Object Verbal descriptions	Access to 3D objects Map Exploration Access to appliances Access to documents
Baker et al. [21] Tactile Graphics with a Voice	Touch (Wearable Camera)	Voice, Tactile graph Verbal descriptions	STEM Education Map Exploration
Fusco et al. [22] The Tactile Graphics Helper	Touch (Mobile Camera)		
Holloway et al. [23]	Touch (Embedded capacitive sensors)	Tactile 3D Map Verbal descriptions	Map Exploration

2.3. *Interactive Tactile Graphics and 2.5D Models*

In the last decades, researchers have explored the improvement of tactile graphics accessibility by adding interactivity through diverse technologies. Despite the availability of many research findings on tactile graphics and audio guides focused on map exploration and STEM education, as described earlier, visually impaired people still struggle to experience and understand visual art. Artists are still more concerned with accessibility of reasoning, interpretation, and experience than providing access to visual information. The visually impaired wish to be able to explore art by themselves at their own pace. With these in mind, artists and designers can change the creative process to make their work more inclusive. The San Diego Museum of Art Talking Tactile Exhibit Panel [24] allows visitors to touch Juan Sánchez Cotán's master still-life, *Quince, Cabbage, Melon, and Cucumber*, painted in Toledo, Spain in 1602 [25]. If you touch one of these panels with your bare hands or wearing light gloves, you can hear information about the touched part. This is like tapping on an iPad to make something happen; but instead of a smooth, flat touch screen, these exhibit panels can include textures, bas-relief, raised lines and other tactile surface treatments. As you poke, pinch or prod the surface, the location and pressure of your finger-touches are sensed, triggering audio description about the part that was touched [25].

Volpe et al. [26] explored semi-automatic generation of 3D models from digital images of paintings, and classified four classes of 2.5D models (tactile outline, textured tactile, flat-layered bas-relief, and bas-relief) for visual artwork representation. An evaluation with 14 blind participants indicated that audio guides are required to make the models understandable. Holloway et al. [27] evaluated three techniques for visual artwork representation: tactile graphics, 3D printing (sculpture model), and laser cut. Among them 3D printing and laser cut were preferred by most participants to explore visual artworks. There are projects that add interactivity to visual artwork representations and museum objects. Anagnostakis et al. [28] used proximity and touch sensors to provide voice guidance on museum exhibits through mobile devices. Reichinger et al. [29,30] introduced the concept of a gesture-controlled interactive audio guide for visual artworks that uses depth-sensing cameras to sense the location and gestures of the user's hands during tactile exploration of a bas-relief artwork model. The guide provides location-dependent audio descriptions based on user hand positions and gestures. Vaz et al. [31] developed an accessible geological sample exhibitor that reproduces audio descriptions of the samples when picked up. The on-site use evaluation revealed that blind and visually impaired people felt more motivated and improved their mental conceptualization. D'Agnano et al. [32] developed a smart ring that allows to navigate any 3D surface with fingertips and get in return an audio content that is relevant in relation to the part of the surface while touching in that moment. The system is made of three elements: a high-tech ring, a tactile surface tagged with NFC sensors, and an app for tablet or smartphone. The ring detects and reads the NFC tags and communicates in wireless mode with the smart device. During the tactile navigation of the surface, when the finger reaches a hotspot, the ring identifies the NFC tag and activates, through the app, the audio track that is related to that specific hotspot. Thus, a relevant audio content relates to each hotspot.

Quero et al. [33–35] designed and implemented an interactive multimodal guide prototype based on the needs found through our preliminary user study [33] and inspired mainly in the related works Holloway et al. [27] and Reichinger et al. [30]. Table 2 compares the main technical differences among the related works and their approach. The prototype identifies tactile gestures that trigger audio descriptions and sounds during exploration of a 2.5D tactile representation of the artwork placed on top of the prototype. The body of work on interactive multimodal guide focused on artwork exploration is summarized in Table 2.

Table 2. Interactive multimodal guide for appreciating visual artwork and museum objects.

Author–Name	Input	Output	Focus
Talking Tactile Exhibit Panel [24,25]	Touch	Audio Descriptions	Museum Object Exploration
Halloway et al. [27]	Capacitive sensor board connected to discrete copper interaction points placed on the surface of the model Double tap and long tap gestures on the surface	Audio Descriptions	Tactile 3D map model. Improve Mobility and Orientation
Anagnostakis et al. [28]	Touch (PIR and touch sensors)	Tactile Objects Verbal descriptions	Museum Object Exploration
Reichinger et al. [29,30]	Touch (Camera) Hand gestures (Camera)	Tactile bas-relief Artwork Model Verbal descriptions	Artwork exploration
Vaz et al. [31]	Touch (Embedded capacitive sensors)	Tactile Objects Verbal descriptions	Museum Object Exploration
D'Agnano et al. [32]	Touch (Ring NFC reader)	Tactile 3D model Verbal descriptions	Archeological site exploration Artwork exploration
Cavazos et al. [33–35]	Capacitive sensor connected to conductive ink-based sensors embedded under the surface of the 2.5D model Double tap and triple tap gestures on the surface	Tactile bas-relief model Audio Descriptions Sound effects and Background music	Artwork exploration

2.4. An Example of Interactive Multimodal Guide for Appreciating Visual Artwork

Cavazos et al. [33–35] developed an interactive multimodal guide that transformed an existing flat painting into a 2.5D (relief form) using 3D printing technology that used touch, audio description, and sound to provide a high level of user experience. Thus, the visually impaired can enjoy it freely, independently, and comfortably through touch to feel the artwork shapes and textures and to listen and explore the explanation of objects of their interest without the need for a professional curator. The interactive multimodal guide [35] complies with the following processes: (1) Create a 2.5D (relief) model of a painting using image processing and 3D printing technologies (Figure 1); (2) 3D-print the model. (3) use conductive paint (Figure 1) applied to objects in the artwork to create the touch sensors such that a microcontroller can detect touch gestures that trigger audio responses with different layers of information about the painting; (4) add color layers to the model to the replicate original work; (5) place the interactive model into an exhibition stand; (6) connect the model to a control board (Arduino and capacitive sensor MPR121); (7) connect headphones to the control board; (8) touch to use; (9) engage in independent tactile exploration while listening to mood-setting background music; (10) tap anywhere on the artwork to listen to localized information, such as the name of an object, its color or shape, its meaning, and so on; (11) double tap anywhere on the artwork to listen to localized audio, such as the sound of leaves on a tree in autumn or the noise of a rural town at night; (12) touch a physical button to hear a recorded track containing use instructions and general information about the artwork, such as the painter's historical and social context, which is an essential part of understanding any work.

(**a**) Paul Cezanne “Compotier, Glass, and Apples”, 1880
(Left: Painted with normal color; Right: Conductive paint coated)

(**b**) Wookjin Jang (Korean Artist) “Magpie”, 1986
(Left: Painted with normal color; Right: Conductive paint coated)

Figure 1. Tactile 3-D printed reproductions in 2.5D model of works (courtesy of Luis Cavazos Quero, Ph.D. student, Dept. of Electronic, Electronic, and Computer Engineering, SungkyunKwan University).

In BlindTouch project [35,36], gamification [37] is included to awaken other senses and maximize enjoyment of the artwork. Through vivid visual descriptions, including sound effects, viewers can maximize their sense of immersion in the space of the painting. Each artwork was reproduced using materials that can recognize the timing of tactile input, so that when a person taps part of the artwork with a fingertip once, they can hear audio description about that part; if they tap twice, they can hear a sound effect about that part. It directly informs the sound of objects expressed in the work of art and informs viewers of what is depicted in the work with natural sound at the same time, while the emotional side is transferred to the background music with feelings similar to the emotion of the work, taking into consideration the musical instrument’s timbre, minor/major, tempo, and pitch. Two-dimensional speaker placement provides more detailed information on key objects such as perspective and directionality so that the visually impaired can maintain the direction of sound through hearing to awaken the real sense of space and to recognize sensory information about their direction. Furthermore, the voice interactive multimodal guide prototype developed by Bartolome et al. [34] identifies tactile gestures and voice commands that trigger audio descriptions and sounds while a person with visual impairment explores a 2.5D tactile representation of the artwork placed on the top surface of the prototype. The prototype is easy and intuitive, allowing users to access only the information they want, reducing user fatigue.

As a preliminary study, BlindTouch [36] was intended to represent various visual elements in the work such as ambient sounds that reflect periodic, seasonal, temporal, and regional information about the work as realistically as possible. In that first study, the auditory interaction was applied to Vincent van Gogh's 1889 work "The Starry Night". When users touch the BlindTouch painting three times, they hear a sound representing the starlight and the sound of a tall cypress swaying in the wind. The sound of the wind was played through two speakers to express the swirling movement of the wind, and the moonlight and starlight in the sky at the top of the work were expressed as a twinkling ringtone. The sounds of shaking leaves and grass bugs on a summer night were added. To those sounds, background music was added with an atmosphere similar to the emotions inspired by "The Starry Night". To express the warmth coming from the village, an oboe played a major scale, and a slightly fast, lyrical melody in the high pitch range of the piano provided a cold feeling of dawn. The completed exhibition environment used six-channel speakers arranged on flat plates, and the wind sounds were swirled between two speakers to arouse a sense of space and enhance appreciation of the artwork through a sense of three-dimensional sound. The blind user who experienced the exhibited work left the following words. "I'm so happy that I can now tell my friends that I understand Starry Night better through the blind touch. Thank you for making art more enjoyable. Especially when I'm older, it's so interesting because I can remember it in a different way."

BlindTouch works were exhibited for three weeks at St. Mary's School (a special school for the visually impaired operated by a 65-year-old Catholic institution located in Cheongju, Korea) for three weeks from October 12 to 30, 2020 [36]. A student, Geon Tak (Figure 2), who especially liked "The Starry Night", looks very satisfied.

Figure 2. A student viewing the "Starry Night" by Vincent van Gogh reproduced in 2.5D at the BlindTouch exhibition (Cheongju St. Mary's School) [36].

Hye-ryeon Jeon, a school art teacher who participated in this exhibition, left the following word. "This BlindTouch study is amazing, especially by conducting research focused on multi-sensory color coding, this barren realm that no one cares about. It seems that visually impaired people will enjoy the richness of life with a very delicate study on appreciation of the artworks." Participants in the experiment responded to the sound that expressed the wind in the work, "It feels like the wind is fighting with each other"; "The sound is played from side to side to express the feeling of wind blowing"; "There is a lot of wind and it feels cool and cold in the air." They replied that the applied wind sound was similar to the actual wind blowing and helped to remind them of it. In addition, the use of two speakers to effectively express the swirling wind in the work received positive reviews from participants. Additionally, while listening to the ambient sound of the stars in the sky and the moon in the work, the participants of the experiment recalled the stars, saying, “I feel like a sparkling in a quiet place” and “There is a sound of something shining.” However, there was an opinion that “when I first heard it, I didn’t know what it was, but later I knew that it was the sound of a star,” and the sound was artificial and the expression was insufficient to be reminiscent of a starry night sky. The participant in the experiment who heard the sound of the cypress tree on the left side of the work replied, “The landscape of grass bugs are crying and the tree next to it is swaying in the wind,” and responded that it was a lonely atmosphere. He said that it helped to remind him of trees.

In addition, interviews were conducted with sighted people after appreciating the work through sight and hearing. There were four participants in the experiment, two males and two females in their 20s. The average age was 22.25 years (SD = 1.92). There was an opinion that the ambient sounds of various objects in the work were reproduced, and that they aroused interest, and the ability to appreciate the interaction through multiple senses rather than a single sense led to a positive evaluation. When the background and appropriate sound of the work were applied to a work of art, it helped both the visually impaired and sighted people to appreciate the work, and users said that it is possible to imagine the appearance, space, and situation of the work and induce a deeper atmosphere and sensibility in it. Participants believed that appreciating works of art using multiple senses can communicate deeply and provide a rich aesthetic experience. To understand how educators perceived tactile art books and/or 3D printed replicas as a new experience for children with visual impairments, their interactive experiences was evaluated with those children. In this study, a high level of participation was observed from both teachers and children. They admitted that they had not experienced any attempts to include multisensory interactions. The visually impaired students enjoyed the BlindTouch works displayed under the guidance of art teachers and returned to the art room to express their feelings without hesitation. When the teacher talked about the atmosphere of the paintings that the students completed by themselves, various reactions emerged.

Here are the works of three students who participated in the exhibition and art classes. The basic information for the students who participated in the art activities is shown in Table 3. Drawings using paints can be difficult for visually impaired students, but teachers tried to induce pictorial expressions from visually impaired students by using wheat flour paste. Students created works with a similar arrangement and composition to the works they enjoyed, because their appreciation of the exhibition works was tactile, and detailed information on the objects in those works could thus be obtained. The students heard a story about Vincent van Gogh’s life and the characteristics of his paintings, which let them express their appreciation in flour paste and paints as vivid as Vincent van Gogh’s brushstrokes. The work below expresses the feeling of appreciation for “The Starry Night” in paints and clay. The artworks use various expressions of the material in “The Starry Night” to express the students’ experiences of touch and hearing with the BlindTouch exhibition. The visually impaired students touched the objects in the exhibits with their hands, and they received auditory information. Then they used their memories to express their feelings. The three students who participated in the BlindTouch exhibition were actively stimulated to express their emotions through the multi-sensory exhibition experience, and during the art class activity, their emotions and feelings toward the subject exhibition became abundant, giving them an opportunity to naturally express their feelings.

Table 3. Students' works that reflects personal impressions of the BlindTouch exhibit [36].

Name	Year/Gender	Disability and Expressive Ability Characteristics	Artwork	Description
Seok Kang-hee	High school/Female	- Level 1 blind - She thinks a lot and is prudent in choosing the subject matter or expressing		It is expressed while slowly thinking about "Starry Night" with clay. Trees, clouds, stars, and even the moon are expressed similarly to actual works.
HaEugene	Elementary school/Female	- Low vision - With mature thoughts, the subject matter and stories to be expressed are very diverse, and the pictorial expression drawn without clogging is excellent.		The "Starry Night" is reproduced using wheat flour, and expressed in accordance with the liveliness of Gogh's work. Gold powder is sprinkled on yellow stars to emphasize the sparkling feeling.
Lee Seah	Elementary school/Female	- Level 1 blind - Strong willingness to express her thoughts with strong inner energy. Very enjoyable and likes to play with paint.		She said she wants to see "Starry Night" and express the cypress tree in red. The red trees were covered up in the night sky. It was a time to experience the joy of creation.

2.5. Immersive Interaction through Haptic Feedback

Tactile feedback can be classified into contact tactile and non-contact tactile. The sunburn, snow, wind, and sensation of heat and humidity can be contact or non-contact. Haptic experiences for improving immersive interaction through haptic feedback are diverse and complex, and humans can perceive a variety of tactile sensations, including the kinematic sensations of objects and skin feedback when users manipulate them.

Only few assistive technologies rely on tangible interaction (e.g., the use of physical objects to interact with digital information [38]). For instance, McGookin et al. [39] used tangible interaction for the construction of graphical diagrams: non-figurative tangibles were tracked to construct graphs on a grid, using audio cues. Manshad et al. [40] proposed audio and haptic tangibles for the creation of line graphs. Pielot et al. [41] used a toy duck to explore an auditory map.

As digital interaction tools for introducing museum exhibits, Petrelli et al. [42] introduced "Museum Mobile App", "Touchable Replicas", and "NFC Smart Cards with Art Drawings". Here, when the replica (tangible) is placed on the NFC reader on the exhibition table, an introduction to the work is played on the multimedia screen. The NFC smart card on which the artwork is drawn works likewise. As a result of surveying visitor preferences for these three, it was found that they most preferred the use of replicas and smart cards. Among the three modes, the proportion of participants who did not prefer to use mobile apps was the highest. It is very noteworthy that this is because it interferes with the enjoyment of participating in the exhibition (55%, N = 31) [42].

Information is typically integrated across sensory modalities when the sensory inputs share certain common features. Cross-modality refers to the interaction between two different sensory channels. Although many studies have been conducted on cross sensation between sight and other senses, there are not many studies on cross sensation between non-visual senses [43]. The Haptic Wave [44] allows audio engineers with visual impairments to "feel" the amplitude of sound, gaining salient information that sighted engineers get through visual waveforms. If cross-modal mapping allows us to substitute one sensory modality for another, we could map the visual aspects of digital audio editing to another sensory modality. For example, if visual waveform displays allow sighted users to "see the sound", we could build an alternative interface for visually impaired users to "feel the sound". The demo will allow visitors, sighted or visually-impaired, to sweep backwards and forwards through audio recordings (snippets of pop songs and voice recordings), feeling sound amplitude through haptic feedback delivered by a motorized fader [44]. Gardner et al. [45] developed a waist belt with built-in sound, temperature, and vibration patterns to provide a multisensory experience of specific artworks.

Brule et al. [19] created a raised-line overlaying multisensory interactive map on a capacitive projected touch screen for visually impaired children after a five-week field study in a specialized institute. Their map consists of several multisensory tangibles that can be explored in a tactile way but can also be smelled or tasted, allowing users to interact with them using touch, taste, and smell together. A sliding gesture in the dedicated menu filters geographical information (e.g., cities, seas, etc.). Conductive tangibles with food and/or scents are used to follow an itinerary. Double tapping on an element of the map provides audio cues. Maps can be navigated in a tactile way, but consist of several different sensory types that can smell or taste, allowing users to interact with the system through three senses: tactile, taste, and smell. Multi-sensory interaction supports learning, inclusion, and collaboration, because it accommodates the diverse cognitive and perceptual needs of the blind. To analyze the data, the Grounded Theory [46] method was followed with open-coded interviews transcriptions and observations. One observation of children using a kinesthetic approach for learning and feedback from the teachers led to multi-sensory tangible artefacts to increase the number of possible use cases and improve inclusivity. Mapsense [19] consists of a touchscreen, a colored tactile map overlay, a loudspeaker, and conductive tangibles. These conductive tangibles are detected by the screen as the tangible's touch events. Users could navigate between "points of interest", "general directions", and "cities". Once one of this type of information is selected (e.g., cities for example), MapSense gives the city name through text-to-speech when it detects a double tap on a point of interest. Children could also choose "audio discovery", which triggered ludic sounds (e.g., the sound of a sword battles in the castle, of flowing waters where they were going to take a boat, of religious songs for the abbey, etc.). Finally, when users activate the guiding function, vocal indications ("left/right/top/bottom") help the users move the tangibles to their target. 3D printer PLA filament was used as material, and aluminum was added around tangibles, as it is conductive, and could be detected when it touches a point of interest in tactile map overlay [19].

Empathy is a communication skill by which one person can share another person's personal perceptions and experiences. A similar concept is rapport, which refers to understanding other people's feelings and situations and forming a consensus (or trust) with them. Empathy is an essential virtue in segmented modern society. Ambi, Figure 3, created by Daniel Cho, RISD, Province, RI, USA, 2015, is a nonverbal (visual, tactile, or sound) telepresence and communication tool used for promoting empathy and rapport between family members and couples. Sensors recognize intuitive and non-verbal (visual, tactile, or sound) signals and exchange empathetic emotions. Ambi's proximity sensor recognizes a person's presence and emotional state and communicates it to another person. One way to express affection is to wrap your hand around the Ambi's waist. In addition, through the non-visual sense of touch or sound, the visually impaired can share empathetic emotions with the others. The constant and immediate tactile feedback of another's presence and nudges allows the visually impaired to have more intimate connection and non-verbal communication with others that the video chat applications alone cannot provide.

Figure 3. A soft robot that conveys telepathy. Ambi (Courtesy of Daniel Cho, RISD, Province, RI, USA, 2015).

2.6. Smell

A tactile interaction created in 3D can be communicated through the touch of a brush and an olfactory stimulus that matches the space in the work, allowing the visually impaired to experience works of art through several senses [47]. Although many people have considered the effects of adding scent to art and museum exhibits, the addition of this normally unstimulated sense will not necessarily enhance the multisensory experience of those who are exposed to it [48]. Nina Levent and Alvaro Pascual-Leone in their book "The Multi-Sense Museum" [49] emphasized the use of forms such as smell, sound, and touch, providing visual and other impaired customers with a more immersive experience and a variety of sensory engagement. Although the use of congruent scents has been shown to enhance people's self-reported willingness to return to a museum [50], the appropriate distribution of scent in/through a space faces significant challenges [49]. More than any other sensory modality, olfaction contributes a positive (appetitive) or negative (aversive) valence to an environment. Certain odors reproducibly induce emotional states [51]. Odor-evoked memories carry more emotional and evocative recollections than memories triggered by any other cue [52].

Dobbelstein et al. [53] introduced a mobile scent operated device that connects to a 3.5 mm audio jack and contains only one scent. Scent actuators that trigger mobile notifications by touch screen input or incoming text message. The scent was less reliable than the traditional vibrations or sound, but it was also perceived as less disruptive and more pleasant. Individual scents can add anticipation and emotion to the moment of being notified and entail a very personal meaning. For this reason, scent should not replace other output modalities, but rather complement them to convey additional meaning (e.g., amplifying notifications). Scent can also be used to express a unique identity.

For Sound Perfume [54], a personal sound and perfume are emitted during interpersonal face-to-face interactions, whereas for light perfume [55], the idea was to stimulate two users with the same visual and olfactory output to strengthen their empathic connection.

Additionally, picture books are also considered beneficial to children because they provide a rich experience [56,57]. Some picture books offer multisensory experiences to enrich learning and gratitude. For example, the Dorling Kindersley publishing house (https://www.dk.com/uk/ accessed on 30 November 2020) has introduced a variety of books that children can touch, feel, scratch, and smell. These books have tactile textures in the pictures and contain a variety of smells [56–58]. The MIT Media Lab has developed an interactive pop-up book that combines material experimentation, artistic design, and engineering [59]. To improve the expression of movement, a study introduced continuous acoustic interaction to augmented pop-up books to provide a different experience of storytelling. The mental image of a blind person is a product of touch, taste, smell, and sound.

Edirisinghe et al. [60] introduced a picture book with multisensory interactions for children with visual impairments and it was found to provide an exciting and novel experience. It emits a specific odor through the olfactory device, which uses a commercially available Scentee (https://scentee.com/ accessed on 30 November 2020) device to respond to sounds. Children with visual impairments can smell and imagine broken objects. The olfactory device is contained inside the page, and the fragrance is emitted from a small hole in the center of the panel [60].

At the Cooper Hewitt Smithsonian Museum, chemist and artist Sissel Tolaas designed a touch-activated map with fragrant paint. After analyzing the scent molecules of different elements from within Central Park, Tolaas reproduced them as closely as possible, using a "microencapsulation" process, containing them inside tiny capsules. She then mixed them with a latex-based binder, creating a special paint that was applied to the wall of the Cooper Hewitt, which can be activated by touch. When visitors go to the wall that has been painted with the special paint, just by touching the wall they are able to break the capsules open and release the scent: a scientifically advanced scratch-and-sniff sticker [61]. Using powdered scents, incense, and spices, Ezgi Ucar stamped fragrances on different photos that form part of a painting. She took inspiration from scratch-and-sniff stickers and used the same method, allowing visitors to scratch and sniff some of the photographed parts of the painting. The human sense of smell has been called the "poet of sensory systems", because it is deeply connected to structures in our brain that relate to our emotions, memories, and awareness of the environment, which can be exploited to enhance user experiences.

Given the ability of smell to influence human experiences, multimodal interfaces are increasingly integrating olfactory signals to create emotionally engaging experiences [62]. Sense of Agency [63] can be defined as "the sense that I am the one who is causing or generating an action". The sense of agency is of utmost importance when a person is controlling an external device, because it influences their affect toward the technology and thus their commitment to the task and its performance. Research into human–computer interactions has recently studied agency with visual, auditory, haptic, and olfactory interfaces [64]. Jacobs et al. [65] showed that humans can define an arbitrary location in space as a coordinate location on an odor grid.

2.7. Hearing (Sound)

Hopkin [66] confirmed that congenital blind or who lost their sight during the first two years of life do indeed recognize changes in pitch more precisely than sighted people. However, there were no significant differences in performance between sighted people and people who had lost their sight after their first two years of life. These findings reveal the brain's capacity to reorganize itself early in life. At birth, the brain's centers for vision, hearing, and other senses are all connected. Those connections are gradually eliminated during normal development, but they might be preserved and used in the early blind to process sounds.

The Metropolitan Museum of Art in New York introduced the reproductions of sound-sensitive art objects by attaching sound switches [67]. The switch plates were cut into shapes based on the form of the major elements of the painting. When someone touches a particular element, an ambient sound related to that element of the painting is produced.

The sense of immersion is improved for the viewer when an artwork is experienced using more than one sense [68,69]. Visual images affect the sensibility of the viewer, conveying meaning, and sound affects the sensibility of the listener. Thus, the effect of a visual image can be maximized by harmonizing the sensibility of the visual image with the ambient sounds. Research has shown that pairing music and visual art enhances the emotional experience of the participant [70]. In the "Feeling Vincent Van Gogh" exhibition [71], a variety of interactive elements were used to communicate artworks to viewers, who could see, hear, and touch Van Gogh's works and thus appreciate them through multiple senses. Visitors could feel Van Gogh's brush strokes on 3D reproductions of Sunflowers and listen to a fragment of background sound through an audio guide [71]. The experience was intended to stimulate a deep understanding of the work and provide a rich imaginative experience. "Carrières de Lumières" in Levod Provence, France, and "Bunker de Lumières" in Jeju, Korea [72], are immersive media art exhibitions that allow visitors to appreciate works through light and music, providing an experience of immersing in art beyond sight.

Every moment of seeing, hearing, and feeling an object or environment generates emotions, which appear intuitively and immediately upon receiving sensory stimulation. Sensibility is thus closely related to the five senses, of which the visual and auditory are most important. Among sighted, hearing people, information from the outside is accepted in the proportions of 60% visual; 20% auditory; and 20% touch, taste, and smell together [73].

Sound can work together with sight to create emotion, allowing viewers to immerse themselves in a space. Therefore, Jeong et al. [74] designed a soundscape of visual and auditory interactions using music that matches paintings to induce interest and imagination in visitors who are appreciating the artwork. That study connected painting and music using a deep-learning matching solution to improve the accessibility of art appreciation and construct a soundscape of auditory interactions that promote appreciation of a painting. The multimodal evaluation provided an evaluation index to measure new user experiences when designing other multisensory artworks. The evaluation results showed that the background music previously used in the exhibit and the music selected by the deep-learning algorithm were somewhat equal. Using deep-learning technology to match paintings and music offers direction and guidelines for soundscape design, and it provides a new, rich aesthetic experience beyond vision. In addition, the technical results of that study were applied to a 3D-printed tactile picture, and then 10 visually impaired test participants were evaluated in their appreciation of the artworks [74].

3. Coding Colors through Sound, Pictograms, Temperature, and Vibration

According to Merleau-Ponty (1945/2002), color was not originally used to show the properties of known objects, but to express different feelings suddenly emerging from objects. According to Jean-Paul Sartre's aesthetics of absence, art is to lead to the world of imagination through self-realization and de-realization of the world. Aesthetic pleasure is caused by hidden impractical objects. What is real is the result of brushing, the thick layer of paint on the canvas, the roughness of the surface, and the varnish rubbed over the paint, which is not subject to aesthetic evaluation. The reason for feeling beauty is not mimesis, color, or form. What is real is never beautiful, and beauty is a value that can only be applied to the imaginary. Absence is a subject that transcends the world toward the imaginary. When reading the artist's work, the viewer feels superior freedom and subjectivity. According to the theory of perception, viewers give meaning to the work according to their experiences. Color is not an objective attribute, but a matter of perception that exists in the mind of the perceiver. It is also known that emotions related to color are highly dependent on individual preferences for the color and past experiences. Therefore, color has historically and socially formed images, and these symbols are imprinted in our minds, and when we see a color, we naturally associate the image and symbol of that color. For example, we can look at such embodied images [75] of color in Vincent van Gogh's work. Vincent Van Gogh went to Arles in February 1888 in search of sunlight. There he gradually fell in love with the yellow color. His signature yellow color is evident in his vase with fourteen sunflowers. Gogh was drawn to the yellow color of the sunflower, which represents warmth, friendship, and sunlight. He said to himself, "I try to draw myself by using various colors at will, rather than trying to draw exactly what I see with my eyes." On the other hand, according to Goethe's Color theory, yellow has a bright nature from purity, giving a pleasant, cheerful, colorful, and soft feeling.

Synesthesia is a transition between senses in which one sense triggers another. When one sensation is lost, the other sensations not only compensate for the loss, but the two sensations are synergistic by adding another sensation to one [76]. For example, sight and sound intermingle. Music causes a brilliant vision of shapes, numbers and letters appear as colors. Weak synesthesia refers to the recognition of similarities or correspondences across different domains of sensory, affective, or cognitive experience–for example, the similarity between increasingly high-pitched sounds and increasingly bright lights (auditory pitch-visual color lightness). Strong synesthesia, in contrast, refers to the actual arousal of experiences in another domain, as when musical notes evoke colors [76]. Synesthesia artists paint their multi-sensory experiences. Vincent van Gogh's work is known for being full of lively and expressive movements, but his unique style must have a reason. Many art historians believe that Vincent van Gogh has a form of synesthesia, the sense of color. This is a sensational experience in which a person associates sound with color. This is evident in the various letters Van Gogh wrote to his brother. He said, "Some artists have tense hands in their paintings, which makes them sound peculiar to violins", he said. Van Gogh also started playing the piano in 1885, but he had a hard time holding the instrument. He declared that the playing experience was overwhelming, as each note evokes a different color.

The core of an artwork is its spirit, but grasping that spirit requires a medium that can be perceived not only by the one sense intended, but also through various senses. In other words, the human brain creates an image by integrating multiple nonvisual senses and using a matching process with previously stored images to find and store new things through association. So-called intuition thus appears mostly in synesthesia. To understand as much reality as possible, it is necessary to experience reality in as many forms as possible, so synesthesia offers a richer reality experience than the separate senses, and that can generate unusually strong memories. For example, a method for expressing colors through multiple senses could be developed.

The painter Wassily Kandinsky was also ruled by synesthesia throughout his life. Kandinsky literally saw colors when he heard music, and heard music when he painted. Kandinsky said that when observing colors, all the senses (taste, sound, touch, and smell) are experienced together. Kandinsky believed abstract painting was the best way to replicate the melodic, spiritual, and poetic power found in music. He spent his career applying the symphonic principles of music to the arrangement of color notes and chords [77].

The art philosopher Nikolai Hartmann, in his book *Aesthetics* (1953), considered auditory–visual–touch synesthesia in art. Taggart et al. [78] found thar synesthesia is seven times more common among artists, novelists, poets, and creative people. Artists often connect unconnected realms and blend the power of metaphors with reality. Synesthetic metaphors are linguistic expressions in which a term belonging to a sensory domain is extended to name a state or event belonging to a different perceptual domain. The origin of synesthetic experience can be found in painting, poetry, and music (visual, literary, musical). Synesthesia appears in all forms of art and provides a multisensory form of knowledge and communication. It is not subordinated but can expand the aesthetic through science and technology. Science and technology could thus function as a true multidisciplinary fusion project that expands the practical possibilities of theory through art. Synesthesia is divided into strong synesthesia and weak synesthesia [78].

Martino et al. [79] reviewed the effects of synesthesia and differentiated between strong and weak synesthesia. Strong synesthesia is characterized by a vivid image in one sensory modality in response to the stimulation of another sense. Weak synesthesia, on the other hand, is characterized by cross-sensory correspondences expressed through language or by perceptual similarities or interactions. Weak synesthesia is common, easily identified, remembered, and can be manifested by learning. Therefore, weak synesthesia could be a new educational method using multisensory techniques. Since synesthetic experience is the result of unified sense of mind, all experiences are synesthetic to some extent. The most prevalent form of synesthesia is the conversion of sound into color. In art, synesthesia and metaphor are combined [79].

To some extent, all forms of art are co-sensory. Through art, the co-sensory experience becomes communicative. The origin of co-sensory experience can be found in painting, poetry, and music (visual, literary, musical) [80].

Today, the ultimate synesthetic art form is cinema. Regarding the senses, Marks [81] wrote: In a movie, sight (or tactile vision) can be tactile. It is "like touching a movie with the eye", and further, "the eye itself functions like a tactile organ".

Colors can be expressed as embossed tactile patterns for recognition by finger touches, and incorporated temperature, texture, and smell to provide a rich art experience to person with visual impairments. *The Black Book of Colors* by Cottin [82] describes the experience of a fictional blind child named Thomas, who describes color through association with certain elements in his environment. This book highlights the fact that blind people can gain experience through multisensory interactions: "Thomas loves all colors because he can hear, touch, and taste them." An accompanying audio explanation provides a complementary way to explore the overall color composition of an artwork. When tactile patterns are used for color transmission, the image can be comprehensively grasped by delivering graphic patterns, painted image patterns, and color patterns simultaneously [83–85].

This suggests to us the possibility and the justification for developing a new way of appreciating works in which the colors used in works are subjectively explored through the non-visual senses. In other words, it can be inferred that certain senses will be perceived as being correlated with certain colors and concepts through unconscious associations constructed with the concepts. The following works were designed to prove this assumption and to materialize it as a system for multisensory appreciation of artworks for the visually impaired.

An experiment comparing the cognitive capacity for color codes named ColorPictogram, ColorSound, ColorTemp, ColorScent, and ColorVibrotactile found that users could intuitively recognize 24 chromatic and five achromatic colors with tactile pictogram codes [86], 18 chromatic and five achromatic colors with sound codes [87], six colors with temperature codes [88], five chromatic and two achromatic colors with scent codes [89], and 10 chromatic and three achromatic colors with vibration codes [90].

For example, Cho et al. [86] presented a tactile color pictogram system to communicate the color information of visual artworks. The tactile color pictogram [86] uses the shape of sky, earth, and people derived from thoughts of heaven, earth, and people as a metaphor. Colors can thus be recognized easily and intuitively by touching the different patterns. What the art teacher wanted to do most with her blind students was to have them imagine colors using a variety of senses—touch, scent, music, poetry, or literature.

Cho et al. [87] expresses color using part of Vivaldi's *Four Seasons* with different musical instruments, intensity, and pitch of sound to express hue, color lightness, and saturation. The overall color composition of Van Gogh's "The Starry Night" was expressed as a single piece of music that accounted for color using the tone, key, tempo, and pitch of the instruments. Bartolome et al. [88] expresses color and depth (advancing and retreating) as temperature in Marc Rosco's work using a thermoelectric Peltier element and a control board. It also incorporates sound. For example, tapping on yellow twice produces a yellow-like sound expressed by a trumpet. Lee et al. [89] applied orange, menthol, and pine to recognize orange, blue, and green as fragrances.

3.1. ColorPictogram

With tactile sense, visually impaired people can access their works more independently, and have better tactile perception than non-visually impaired people. Baumgartner et al. [91] found that visual experience is not necessary to shape the haptic perceptual representation of materials. Color patterns are easy to understand and can be used even among people who do not share a language and culture. Braille-type color codes have been created for use on Braille devices [85]. Another method of expressing colors uses an embossed tactile pattern that is recognized by touching it with a finger [86,92–95].

Using that method, it is possible to express the shape of an object through the color pattern without deliberately creating the outline of a shape. The tactile color pictogram, which is a protruding geometric pattern, is an ideogram designed to help person with visual impairment to identify colors and interpret information through touch. Tactile sensations, together with or as an alternative to auditory sensations, enable users to approach artworks in a self-directed and attractive way that is difficult to achieve with auditory stimulation alone.

Raised geometric patterns called tactile color pictograms are ideographic characters designed to enable the visually impaired to interpret visual information through touch. Cho et al. [86] developed three tactile color pictograms to code colors in the Munsell color system; each color pattern consists of a basic cell size of 10 mm × 10 mm. In each tactile color pictogram, these basic geometric patterns are repeated and combined to create primary, secondary, and tertiary color pictograms of shapes indicating color hue, intensity, and color lightness. Each tactile color pictogram represents 29 colors including six hues, and these can be further expanded to represent 53 colors. For each of six colors (red, orange, yellow, green, blue, and purple), vivid, light, muted, and dark colors can also be expressed, along with five levels of achromatic color. These tactile color pictograms have a slightly larger cell size compared to most currently used tactile patterns but have the advantage of coding for more colors. Application tests conducted with 23 visually impaired adult volunteers confirm the effectiveness of these tactile color pictograms.

As shown in Figure 4, the graphic of colors floating into the colorless paper with two kinds of tactile color pictograms [86] represents the fact that although the artwork with the patterns looks "colorless" to a visually unimpaired person, a person with visual imparity can experience the full range and diversity of colors in the artworks.

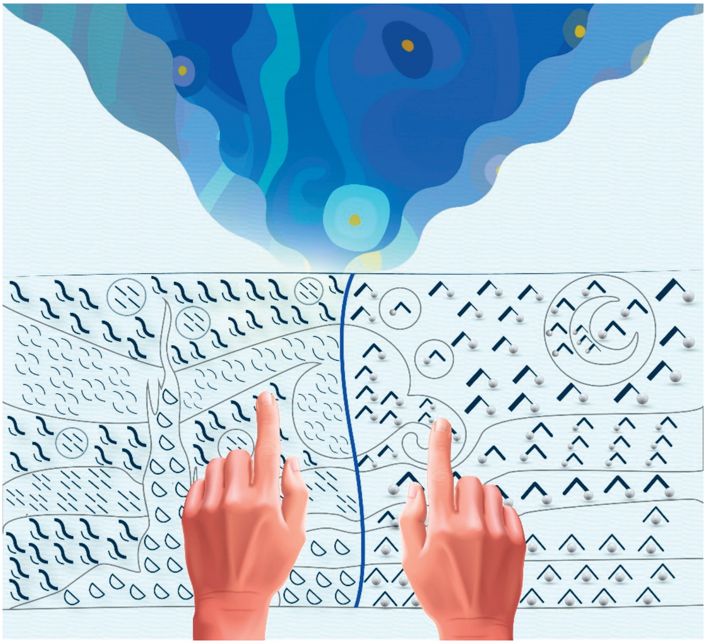


Figure 4. Two tactile color pictograms, Cover image of [86], Editing Services, Wiley. Dec. 2020.

3.2. ColorSound

When using sound to depict color, touching a relief-shaped embossed outline area transforms the color of that area into the sound of an orchestra instrument [96]. Palmer et al. [97] explored the relationship between color and music as a cross-modal correlation based on emotion. In "Barbiere et al. (2007), The color of music", college students listened to four song clips. Following each clip, the students indicated which color(s) corresponded with the clip by distributing five points among eleven basic color names. Each song had previously been identified as either a happy or sad song. Each participant listened to two happy and two sad songs in random order. There was more agreement in color choice for the songs eliciting the same emotions than for songs eliciting different emotions.

Brighter colors such as yellow, red, green, and blue were usually assigned to the happy songs, and gray was usually assigned to the sad songs. It was concluded that music–color correspondences occur via the underlying emotion common to the two stimuli.

Color sound synthesis [98] starts with a single (monotonous) sine wave for gray, changing in pitch according to color lightness. With red, a tremolo is created adding a second sine wave, just a few Hertz apart. A beat of two very close frequencies (diff. < 5 Hz) creates a tremolo effect. The more reds the color turns, the smaller the gap is tuned between both frequencies, increasing in speed of the perceived tremolo. To simulate the visual perception of warmth with yellow, the volume of bass is increased as well as the number of additional sine waves (tuned to the frequencies of only the even harmonics of the fundamental sine wave). The bass as well as the even harmonics are acoustically perceived to be warm. The result sounds like an organ. The coldness of blue was originally planned to be sonified, adding the odd harmonics, which would lead to a square wave, creating a cold and mechanical sound. However, the sound so produced is too annoying to be used, so we applied one of the Synthesis Toolkit's pre-defined instrument models that can synthesize a sound of a rough flute or wind. An increase in blue is represented by an increase of the wind instrument's loudness. Finally, to create an opponent sound characteristic to vibrant red, we represent green, as a calm motion of sound in time using an additional sine wave tuned to a classical third to the fundamental sine wave, forming a third chord, as well as two further sine waves, one tuned almost like the fundamental sine, the other like the second sine, far enough apart to create not the vibrant tremolo effect but a smooth pattern of beats, moving slowly through time [98].

Cavaco et al. [99] mapped the hue value into the fundamental frequency, f0, of the synthesized sound (which gives the perception of pitch). There is an inverse correspondence between the sound's pitch and color frequencies: when the color's frequency decreases from violet to red, the sound's pitch increases (by increasing the f0) [100,101]. The synthesized waveform starts off as a sinusoidal wave (i.e., a pure tone), but the final waveform can be different from a pure tone, because the signal's spectral envelope can be modified by the other attributes (saturation and value). The signal's spectral envelope (which is related to the perception of timbre) is controlled by the attribute saturation. The shape of the waveform can vary from a sinusoid (for the lowest saturation value) to a square wave with energy only in the odd frequency partials (for the highest saturation value). Finally, the attribute value (ranging from 0 to 1) is used to determine the intensity of the signal (which gives the perception of loudness). All frequency partials are affected in the same way, as the signal is multiplied by value [99].

Cho et al. [87] developed two sound codes (Table 4) to express vivid, bright, and dark colors for red, orange, yellow, green, blue, and purple. Fast notes in a major key are yellow or orange, and slow notes in a major key are blue and gray. Codes expressing vivid, bright, and dark colors for each color (red, orange, yellow, green, blue, and purple) were used in [86]. In this system, the shape of the work can only be distinguished by touching it with a hand, but the overall color composition is conveyed as a single piece of music, thereby reducing the effort required to recognize color from that needed to touch each pattern one by one. Vivid colors and bright and dark colors were distinguished through a combination of pitch, instrument tone, intensity, and tempo. High color lightness used a small, light, particle-like melody and high-pitch sounds, and a bright feeling was emphasized by using a melody of relatively fast and high notes. For low color lightness, a slow, dull melody with a relatively low range was used to create a sense of separation and movement away from the user. Beginning with Vivaldi's *Four Seasons*, a melody that matches the color lightness/saturation characteristics of each color was extracted from the theme melody of each season. In the excerpts, the composition was changed, and the speed and semblance were adjusted to clarify the distinction between saturation and color lightness. From the classical music, a melody that fits the characteristics of each color (length: about 10 to 15 s) has been excerpted [87].

Table 4. Two sound coding colors with using instruments and classical melodies [86].

Colors	Sound Coding Color: Excerpts from Classical Music	Sound Coding Color: Excerpts from Vivaldi: Four Seasons
Red	Violin: High frequency banded string instrument Tchaikovsky: Violin Concerto in D	Violin + Cello Vivaldi: Four Seasons
Orange	Viola: Stamitz: Viola Concerto in D	Guitar Vivaldi: Four Seasons
Yellow	Trumpet: Haydn: Trumpet Concerto in E flat	Trumpet + Trombone Vivaldi: Four Seasons
Green	Oboe: Woodwind instrument with reed Rossini Variations for oboe	Clarinet + Bassoon Vivaldi: Four Seasons
Blue	Cello: Bach Cello Suite No. 1 in G	Piano Vivaldi Four Seasons
Purple	Organ: Keyboard instrument with simultaneous expressions Mozart: Eine Kleine Nachtmusik	Organ Vivaldi: Four Seasons

3.2.1. Sound Color Code: Vivaldi Four Seasons

Each hue in [87] has its own unique tone using brass, woodwind, string, and keyboard instruments, so it is classified by designating groups of instruments that are easy to distinguish from one another. The characteristics of each instrument group's unique tone matched the color characteristics as much as possible. Red, a representative warm color, is a string instrument group with a passionate tone (violin + cello). A group of brass instruments with energy, as if bright light were expanding, is used to simulate yellow bursts (trumpet + trombone). Orange is an acoustic guitar with a warm yet energetic tone. Green is a woodwind instrument with a soft and stable tone to produce a comfortable and psychologically stable feeling (clarinet + bassoon). Blue, a representative cold color, is a piano, which has a dense and solid tone while feeling refreshing. Purple, which contains both warm red and cold blue, is a pipe organ using brass tones.

3.2.2. Color Sound Code: Classical

In [87], musical instruments were classified for each color to ensure that they would be easily distinguished from one another. Red, a representative warm color, is a violin that plays a passionate and strong melody. A trumpet plays a high-pitched melody with energy, as if a bright light were expanding, to simulate yellow bursts. Orange is a viola playing a warm yet energetic melody. Green, which makes the eyes feel comfortable and psychologically stable, is a fresh oboe that plays a soft melody. Blue, a representative cold color, is a cello that plays a low, calm melody. Violet, where warm red and cold blue coexist, is a pipe organ that plays a magnificent yet solemn melody. Each color of Marc Roscoe's works, Orange and Yellow (1956) and No. 6 Violet Green and Red (1951), is expressed with these sound codes. Vivid, bright, and dark colors were distinguished using a combination of pitch, instrument tone, intensity, and tempo.

3.2.3. ColorSound: The Starry Night

In [87], Vincent Van Gogh's work "The Starry Night" was transformed into a single song using the classical sound code just described. To express the highly saturated blue of the night sky, which dominates the overall hue of the picture, a strong, clear melody in the mid-range was excerpted from the Bach unaccompanied cello suite No. 1 to form the base of the whole song; it is played repeatedly without interruption. To express the twinkling bright yellow of the stars, a light particle-like melody was extracted from Haydn's Trumpet Concerto and played as a strong, clear melody in the midrange.

The painting was divided into four lines and worked with 16 bars per line, producing a total of 68 bars played in 3 min and 29 s. The user experience evaluation rate from nine blind people was 84%, and the user experience scores from eight sighted participants were 79% and 80% for the classical and Vivaldi schemes, respectively. After about 1 h of practice, the cognitive success rate for three blind people was 100% for both the classical and Vivaldi schemes.

3.3. ColorTemp

Recently, visual artworks have been reconstructed using 3D printers and various 3D transformation technologies to help the visually impaired rely on their sense of touch to appreciate works of art. However, while there is work in HCI on multimodality and cross-modal association based on haptic interfaces, such as vibrotactile actuators, thermal cues have not been researched to that extent. In that context [88,102], explored a way to use temperature sensation to enhance the appreciation of artwork.

Bartolome et al. [88] designed, developed, and implemented a color-temperature mapping algorithm to allow the visually impaired to experience colors through a different sense. The algorithm was implemented in a tactile artwork system that allowed users to touch the artwork. Temperature stimulation has some influence on image appreciation and recognition. An image presented along with an appropriate temperature is perceived as an augmented image by the viewer [103]. One VR device uses a small Peltier device to provide a temperature stimulus to maximize the sense of the field [104]. Lee et al. [105] explored the modal relationship between temperature and color focused on the spectrum of warm and cold colors. Bartolome et al. [102] expresses color and depth (advancing and retreating) as temperature in Marc Rosco's work using a thermoelectric Peltier element and a control board. An obvious way of conveying the warm or cold feeling of color to the visually impaired is to have a finger touch the temperature generating device (e.g., Peltier devices that are used in dehumidifiers and coolers) to identify the color. Temperature stimulation has some influence on image appreciation and recognition. An image presented along with an appropriate temperature is perceived as an augmented image by the viewer. Like sound, temperature is a modality that the visually impaired can use to enhance their appreciation of visual artwork. The modal relationship between temperature and color focused on the spectrum of warm and cold colors, including 3D printing techniques, interactive narration, tactile graphic patterns, and color-sensibility delivery through temperature. A temperature generator using Peltier element allows the visually impaired to perceive the color and depth in an artwork. The control unit, which controls the Peltier element, used an Arduino mega board, and a motor driver controlled the forward and reverse currents to manage the endothermic and heat dissipation of the Peltier element. Because constant voltage and current are important in maintaining the temperature of a Peltier device, a multi-power supply was used for stability. Twelve Peltier elements were densely placed in a 4 × 3 array, on top of which a thick paper coated with conductive ink was coated inside each cell except for the boundary of each cell. On top of that, relief-shaped artwork using swell paper was placed. The artwork is divided into 4 × 3 cells, providing the matched temperature for the color. The visually impaired feel a sense of temperature by touching the cell with their fingers, and they can obtain color and depth information corresponding to the temperature. Russian-born painter Marc Rothko is known for his color field works, Figure 5. If you touch a part of the artwork twice, the color of that part can be recognized more clearly through temperature and sound coding colors [36,88].

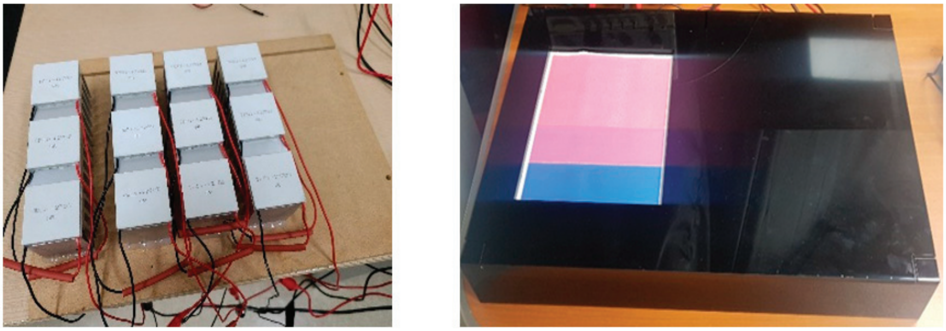


Figure 5. Peltier array and finished prototype with an artwork on top [88].

Chromostereopsis [106] is a visual illusion whereby the impression of depth is conveyed in two-dimensional color images. In this system, red and yellow are conveyed with warmth, and blue and green are conveyed with cold. Warm colors feel close, and cold colors feel far away, so music can be used to reinforce the cross-modal correlation between color and temperature. The thermoelectric element produced temperatures from 38 to 15 °C in 4° intervals, enabling six different colors or depths to be distinguished [102]. It is difficult to distinguish between bright and dark colors with temperature, so musical notes with different combinations of pitch, timbre, velocity, and tempo can be used to distinguish vivid, light, and dark colors. Mapping the depth of field and color-depth to temperature can help the visually impaired comprehend the depth dimension of an artwork through touch. Iranzo et al. [102] developed an algorithm to map color-depth to temperature in two different contexts: (1) artwork depth and (2) color-depth.

First, the temperature range was selected within a comfortable range. In general, the visual perceived distance between the extreme depth levels in a piece will be assessed, and the extreme temperatures will then be selected accordingly (higher temperature difference for larger distances). However, to simplify the algorithm, the extreme depth levels can consistently be linked to the extreme temperatures of 14 and 38 °C, regardless of their perceived relative distances. Second, the total number of perceived depth levels is counted. For example, an image with two people, one in front of the other, contains two depth levels, front and back. Third, the temperature is equally divided into as many segments as needed to assign a temperature to each depth level. The highest and lowest temperatures are always assigned to the nearest and farthest depth levels, respectively.

Fourth, if the difference between the temperatures of two consecutive depth levels is less than 3 °C, some of the levels can be clustered to make the temperature distinctions between levels easier to feel. The prototype was designed, developed, and implemented using an array of Peltier devices with relief-printed artwork on top. Tests with 18 sighted users and six visually impaired users revealed an existing correlation between depth and temperature and indicated that mapping based on that correlation is an appropriate way to convey depth during tactile exploration of an artwork [102].

3.4. *ColorScent*

As seen in previous studies, the amount of color that can be expressed is very limited, because the intensity of fragrance perception is poor. Scent acts as a good trigger for memory and emotion, because it can mediate the exploration of works of art in terms of general memory or emotion. Because smells and memories are connected in the brain, memories can be recalled by smells, and smells can sometimes be evoked by memories.

De Valk et al. [107] conducted a study of odor–color associations in three distinct cultures: the Maniq, Thai, and Dutch. These groups represent a spectrum in terms of how important olfaction is in the culture and language. For example, the Maniq and Thai have elaborate vocabularies of abstract smell terms, whereas the Dutch have a relatively impoverished language for olfaction that often refers to the source of an odor instead of the scent itself (e.g., it smells like banana). Participants were tested with a range of odors and asked to associate each with a color. They also found that across cultures, when participants used source-based terms (i.e., words naming odor objects, such as "banana"), their color choices reflected the color of the source more often than when they used abstract smell terms such as "musty". This suggests that language plays an important mediating role in odor–color associations [107].

Gilbert et al. [108] confirmed that humans have a mechanism that unconsciously associates specific scents with specific colors. For example, aldehyde C-16 and methyl anthranilate are pink; bergamot oil is yellow; caramel lactone and star anise oil are brown; cinnamic aldehyde is red; and civet artificial, 2-ethyl fenchol, galbanum oil, lavender oil, neroli oil, olibanum oil, and pine oil are reminiscent of green. Repeated experiments that produced similar results demonstrated that those results were not random. Also, of the 13 scents, civet was rated as the darkest, and bergamot oil, aldehyde c-16, and cinnamic aldehyde were rated as the lightest [108].

Kemp et al. [109] found that the color lightness of a color was perceived to correlate with the density of a fragrance and that strong scents are associated with dark colors.

Li et al. [110] developed the ColorOdor, an interactive device that helps the visually impaired identify colors. In this method, the camera attached to the glasses worn by the user recognizes the color, and the Arduino controls the piezoelectric transducer system through Bluetooth to vaporize the liquid scent associated with the color. Although culture plays a role in color–odor connection, user research showed color–odor mappings are (white, lily), (black, ink), (red, rose), (yellow, lemon), (green, camphor leaves), (blue, blueberry), and (purple, lavender). When a blind person touches the "white" part of the picture with a finger, it sends a signal to the fragrance generator so that "lily" is emitted. The visually impaired who knows that white and lily scent are related can know that the part is white through lily scent. However, this study was not intended to allow the visually impaired to appreciate works of art, and there is a limitation that the association of used fragrance and color was not based on scientific experiments and results, but mostly due to subjective selection of researchers. Nevertheless, the attempts of those who used scent to convey color to the visually impaired offers us many implications [110].

Lee et al. [89] assumed that each scent has its own unconscious relationship with color and concept, which the researchers called color directivity and concept directivity, respectively. Through experiments, they found specific scents with color directivity and concept directivity and then used those scents to successfully deliver information about the colors used in artworks to the visually impaired. Another study on the transmission of color information using scent found a scent with consistent color and concept orientation and applied it to tactile paper, allowing the visually impaired to actively explore and be immersed in the form and color of an artwork. Instead of understanding art appreciation as an educational technique that unilaterally conveys the authority and interpretation of third parties, such projects invite the visually impaired to directly experience artwork through their own senses. In this case, a special ink scent was applied to the surface of swell paper. After visually impaired students were exposed to the work, their degree of comprehension was measured in terms of the accuracy of their scent–color recognition, a usability evaluation, and an impression interview. In [89], scent is released when a user rubs their finger over the area where the scent was applied to the painting in Tactile Color Book. Unlike people with congenital blindness, people with an acquired visual impairment retain the concept of color. When they smell something, they naturally associate it with color through the memories associated with it. Sight and hearing do not have the same powerful recall ability as smell. For the prototype, the scents of menthol, orange, and pine were chosen to express blue, orange, and green, and the scents of rose, lemon, grape, and chocolate were chosen to express red, yellow, purple and brown, respectively. When smelling a particular scent, sensitivity decreases by 2.5% every second, with 70%

disappearing within one minute. However, even under adaptive fatigue conditions, other odors can be identified. On the paintings, the scents were arranged at intervals of 3 cm or more to prevent mixing. The paintings in Tactile Color Book use the same perfumes used in aromatherapy. In their experiments, color-concept directivities were found for the scent of oranges (orange color), chocolate (dark brown), mint (blue), and pine (green). Orange scent shows high saturation orange color directivity, and it shows concept directivity of bright, extroverted, and strong stimulus. Chocolate scent showed brownish color directivity with low color lightness, and showed concept directivity of roundness, lowness, warmth, and introversion. Pine scent and menthol scent appeared in turquoise color in terms of color directivity, but concept directivity menthol was found to have a greater association with the concept of coolness. The rest of the scents used in the experiment did not show distinctly significant and consistent characteristics in color directivity and concept directivity. Based on the results of the experiment, the scent of orange is associated with orange, menthol is blue, pine is green, and chocolate is brown colors, respectively, as shown in Table 5. Orange has been shown to be bright, extroverted, and associated with strong stimuli (angled form, high notes). This property is also generally applicable when describing the characteristics of yellow and red. Considering that orange is a blend of these two colors, one can give orange a universality that encompasses all three colors. Menthol and pine were similar in color directivity; the concept of coolness was considered to be about 22% higher in menthol than pine. Therefore, menthol was designated as blue, the color associated with coolness, and pine was designated as green, respectively. Using this scent–color association, the scent coding color is used in tactile textbooks for the visually impaired to appreciate artworks. The test result showed a high color recognition rate, and a positive result was obtained that induces subjective immersion in the artwork appreciation experience of the visually impaired. Among visually impaired students, the color recognition accuracy was 94.3%, and the usability evaluation score averaged 70 points. In the interviews, students said that the system was intuitive and easy to learn. In addition, the students said that the scents allowed them to better understand the content of the artwork. In the other scents, except for the above four scents, rose, phoenix, apricot, strawberry, lemon, and apple, as a result of the experiment, no significant color orientation or conceptual orientation could be observed visually or numerically [89].

Table 5. Color directivity, concept directivity, and color matching of each scent [89].

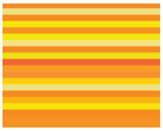
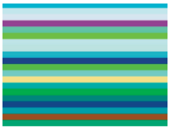
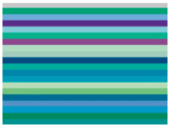
Scent	Color Directivity	Concept Directivity	Color Matching
Orange		Angular, Very bright, Very extravert, Very high note	Orange (Red, Yellow)
Chocolate		Rounded, Low, Warm, Introvert	Brown
Menthol		Angular, Very cool, Bright	Blue

Table 5. *Cont.*

Scent	Color Directivity	Concept Directivity	Color Matching
Pine		Angular, High (position), Cool	Green

Nehmé et al. [111] investigated odor–color associations and based on the three tested populations like French–Lebanese–Taiwanese, odor–color associations could be affected by the function of odors in different countries. Culture induced experiences influence the perception of odors familiarity, which will affect the prevalence of either perceptive (intensity, irritancy, and hedonics) or semantic processing of these associations. According to Stevenson et al. [112], color brightness correlates with perceptual attributes of odors (odors that are more irritating, intense, and unpleasant are associated with brighter colors) and semantic attributes (more familiar and identifiable odors are associated with more saturated colors).

Maric et al. [113] presented their French participants with 16 odorants including caramel, cucumber, lavender, lemon, lime, mint chlorophyll, mirabelle plum (at low and high intensity), orange blossom, peppermint, rose, pineapple, shallot, smoked, violet, and wild strawberry. Two pairs of similar odorants (namely lemon/lime and peppermint/mint chlorophyll) were included in the hope of teasing out the finer nuances of odor–color matching. The participants had to pick one of 24 color patches varying in terms of their hue, saturation, and brightness for each odor. They also rated the odorants on 11-point scales in terms of their intensity, pleasantness, familiarity, and edibility. Significantly non-random color matches (with two to six of the colors) were reported for all of the odorants. Of interest, subsequent data analysis highlighted a significant positive relationship between the rated pleasantness of the odorants and both the lightness and saturation (chroma) of the color chosen, with more pleasant odors matched to colors with higher lightness and saturation. They [113] were chosen to cover of the four basic groups: flower (floral odors), sweet (fruits and candies odors), bad (smoked odors), and nature (plant odors). Fifteen food and floral natural aromas were selected as olfactory stimuli: lavender, orange blossom, rose, and violet as floral odors; caramel, mirabelle plum, pineapple, and wild strawberry as sweet odors; smoked as bad odor; cucumber, lemon, lime, mint chlorophyll, peppermint, and shallot as nature odors. Fourteen odors had the same aromatic intensity (fixed by our supplier). The same odor of mirabelle plum (a small yellow plum, specialty of the French region of Lorraine) was presented twice; that is to say with two different aromatic intensities (at low and high intensity). Sixteen olfactory stimuli were thus prepared by injecting 1 mL of each odorant into a small piece of carded cotton that was previously placed into a small opaque glass bottle. No salient visual cues were therefore available to participants [113].

Kim [114] showed that compared with the oriental and fresh families, the floral and woody families showed more distinguishable opposite patterns in both hue and tone parameters: the floral family with brighter warm colors, and the woody family with darker (or stronger) cool colors. The warm colors strongly evoked the floral family, while the cool colors did the fresh family. The brighter (darker) their lightness values become, the more the floral (woody) scents are associated.

Adams [115] showed that the odorants were assigned significantly different values on these line scales. Therefore, for example, the brightest and lightest odorants were lemon, apple, and peach, whereas the dimmest and darkest odorants were coffee, cinnamon, and chocolate (making one wonder whether these mappings may have been driven by the color of the source objects). As has been documented previously, robust shape associations were also documented in response to the odorants.

Russian composer Scriabin talked about fragrances synchronized with the lighting score that he had designed for his tone poem Prometheus: Poem of Fire [116].

3.5. ColorVibrotactile

Research on diversifying cognitive patterns is being actively conducted to deliver meaningful information using vibrational tactile sensation [117,118]. Among the vibrational and tactile studies, Cappelletti et al. [90] compared two solutions. A firstly devised solution consisted in the most similar setting to the one of color: the superposition of three signals (vibrations) at different frequencies. Single frequencies—well in the 50 to 300 Hz range of skin sensibility—were not distinguished. It turned out that amplitude could be much better discriminated than frequency. The second solution was for all to have the same frequency, but each can independently be modulated in amplitude. Just three levels of amplitude are admitted at the present stage of the research: Low (L), Medium (M), and High (H), corresponding to signals of 0 v, 0 mA; 0.7 v, 30 mA; and 1.4 v, 60 mA, respectively, all at 75 Hz. The RGB color model is an additive color model in which red, green, and blue light are added together in various ways to reproduce a broad array of colors. Thus, with the RGB color model, color is encoded as a triple of vibrations. The corresponding color codes are black (L, L, L); dark red (M, L, L); red (H, L, L); orange (H, M, L); yellow (H, H, L), dark green (L, M, L); green (L, H, H); sky-blue (L, H, H); blue (L, L, H); dark-blue (L, L, M); violet (H, L, H); grey (M, M, M); white (H, H, H). The choice was made in order to have a palette of well-distinguishable colors with definite names. The experiments, performed on subjects with different sight histories, are satisfying [90].

Brewster's Tacton [119] links specific tactile signal patterns and specific meanings to deliver specific information to users by playing the corresponding tactile pattern during interaction. The more the recognition rate of feeling, the change of vibration is getting better. In the range of 100–160 Hz, the change of vibration was most sensitively recognized, and in the range of 160–315 Hz, it was found that the perception of change gradually became dull. However, simply distinguishing colors by varying the intensity of vibrations is likely to prevent users from finding a correlation between colors and vibrations. All of the subjects answered that it was difficult to semantically connect color and vibration. If instead of converting colors into emotions, they were converted in other ways to create vibrating tactile sensations, it could be possible to realize more clear tactile sensations of colors [119].

Maclean and Enriquez [120] studied semantic tactile messages, called haptic icons, created by changing signals in the dimensions of frequency, amplitude, and waveform (sine, square, and triangle). Subjects were able to consistently distinguish between the two dimensions of the data: frequency and waveform. The frequency range of 10–20 Hz was optimal for the user to recognize the signal. The initial experiment used 36 stimuli, combining three wave shapes (sine, square, and sawtooth), four frequencies (0.5, 5, 20, and 100 Hz) and three amplitudes (12.3, 19.6, and 29.4 mNm), each with duration of 2 s. All force magnitudes are scaled as torque values in peak-to-peak mNm. Effect of shape–smooth vs. jerky: there is a clear separation between the sine and the square/sawtooth wave shapes. This is likely due to the discontinuity of the square and sawtooth waves relative to the smooth derivatives of the sine wave. However, separations between smooth and discontinuous shapes diminish with frequency; experience with the stimuli confirms that these shape differences were indeed less perceptible at higher frequencies. While still most important, frequency does not dominate other parameters to the same extent [120].

Another potential issue is the extent to which this kind of coding is intuitive, given the perceptual tendency to link vibrotactile frequency to luminance rather than hue in sighted individuals [121].

Saket et al. [122] conducted an experiment to understand how mobile phone users perceive the urgency of ten simple vibration alerts that were created from four basic signals: short on, short off, long on, and long off. The short and long signals correspond to 200 and 600 ms, respectively. To convey the level of urgency of notifications and help users prioritize them, the design of mobile phone vibration alerts should consider that the gap length preceding or succeeding a signal, the number of gaps in the vibration pattern, and the vibration's duration affect an alert's perceived level of urgency. Their study specifically shows that shorter gap lengths between vibrations (200 vs. 600 ms), a vibration pattern with one gap instead of two, and shorter vibration all contribute to making the user perceive the alert as more urgent. A vibration pattern is defined as an arrangement of the simplest repeatable alternating sequence of an actuator's on and off state, with specific lengths (short and long) assigned to each state. They limited the variable of short to 200 ms and long to 600 ms, without any median values due to its susceptibility to detection errors. Participants could differentiate vibrotactile signals with extreme values well but were less able to do so with median values. We used four basic types of signal to form distinguishable vibration patterns: short on, short off, long on, and long off. In order to produce unique repeatable sequences, an equal number of on and off signals must be alternated in arrangements that do not replicate any other sequence. Two pairs of such signals form additional 16 patterns. An initial test confirmed that the ten patterns were distinguishable, while those that were comprised of three pairs of on–off signals were hard to distinguish. Thus, they did not consider patterns consisting of three or more pairs of on–off signals in the study. Three underlying factors contribute to users' perceived urgency of vibration alerts: gap length is the strongest factor, followed by number of gaps, and finally vibration length. Experiment results and qualitative analysis reveal that the short on signal is highly susceptible to varying perceptions of its level of urgency, depending on the length of the gaps that precede and succeed it. A gap length of 200 ms between short on signals heightens perceived urgency, because the short and sharp pulse is delivered in a stronger manner. On the other hand, a 600 ms gap length preceding or succeeding a short on signal diminishes its strength, making the pulse feel weaker [122].

3.6. ColorPoetry

In poetry, synesthesia refers specifically to figurative language that includes a mixing of senses. For example, saying "he wore a loud yellow shirt" is an example of synesthesia, as it mixes visual imagery (yellow) with auditory imagery (loud). Here is another example "The Loudness of Color by Jennifer Betts":

> The music of white dances softly around
> The soft silence and blue are bound
> Purple is calm, the sound soft and sweet
> The color lightness of a rainbow is a hypnotic beat
> In yellow, the silence is loud
> While red is a yell, robust and proud

A poem is a piece of writing that has features of both speech and song, whereas the poetry is the art of creating these poems. Throughout the poem, seeing and hearing are used to understand color. Red does not actually yell, but many would describe it as loud. Using sound to describe color makes it fun and interesting for the reader. The art of writing poetry about paintings is known as ekphrasis, which means a verbal description of a visual art. Cho [123] introduced a style of painting incorporating the meaning of poetry motivated by "poetry-based paintings". Poems can be created from paintings by applying the same techniques in the opposite direction. Poets have been inspired by works of art. Korean artist Lee Jing (1581–1674 or after) had an outstanding capacity to make "poetry-based paintings", demonstrating the patterns and trends of paintings in the period [124]. The characteristics of Lee's poetry-based paintings can be summarized as follows. At first, it is clear that the main objects of his poetry-based paintings were the poems of the scholars. Secondly, the poems portrayed the social aspect as well as describing the romantic expressions of individual emotions. Thirdly, the poetry-based paintings were mainly ordered by the royal family and power elites of the society. Fourth, many paintings are divisional in the composition of picture and poem. Last, the main themes of the poetry-based painting are landscape and four honorable plants.

Audio or verbal description on an artwork generally attempts to explain the painting without expressing the individual subjectivity. When making visual art accessible to the visually impaired, it is not enough to describe the colors and situations portrayed, because that objective information does not attach to anything in their experience. Using expressions that contain the sensibility of poems, color, and situation can be matched with the parts of each painting.

4. Multisensory Integration

Multisensory (or multimodal) integration is an essential part of information processing by which various forms of sensory information, such as sight, hearing, touch, and proprioception (also called kinesthesia, the sense of self-movement and body position), are combined into a single experience [125,126].

Information is typically integrated across sensory modalities when the sensory inputs share certain common features. Cross-modality refers to the interaction between two different sensory channels. Cross-modal correspondence is defined as the surprising associations that people experience between seemingly unrelated features, attributes, or dimensions of experience in different sensory modalities. Although many studies have been conducted on the cross sensation between the sight and other senses, there are not many studies on the cross sensation between the non-visual senses.

Through the investigation of weak synesthesia that is perceived at the same time by intersecting various senses such as hearing, touch, smell, etc., it explores other sensory information that can be connected with specific form and color information in sight. In this task so far, efforts have been made to create different single-mode sensory perceptions for color. A mapping technology that can be easily recognized by expressing colors as temperature, sound, and scent was designed and tested. However, while this helps visually impaired users perceive and understand colors in completely different ways, it becomes necessary to integrate all perceptual sensations into a single multi-sensory system, where all the senses are perfectly connected and interchanged.

Making art accessible to the visually impaired requires the ability to convey explicit and implicit visual images through non-visual forms. It argues that a multi-sensory system is needed to successfully convey artistic images. It also designed a poetic text audio guideline that blends sounds and effects called “sound painting” to translate the work into an ambiguous artistic sound text [127].

Testing was conducted to extend the results from previous work into a complete multisensory color experience system. The relationships between color–temperature/color–sound have been studied through existing research, and through this, visually impaired people feel or perceive color through temperature or sound. However, the connection between temperature and sound is unclear. When appreciating works using both temperature and sound, we need to explore whether temperature and sound interfere with each other, causing confusion in color perception, or synergize with each other to positively affect color perception.

4.1. Temperature and Sound

Wang et al. [128] explored the putative existence of cross-modal correspondences between sound attributes and beverage temperature. An online pre-study was conducted first to determine whether people would associate the auditory parameters of pitch and tempo with different imagined beverage temperatures. The same melody was manipulated to create a matrix of 25 variants with five different levels of both pitch and tempo. The participants were instructed to imagine consuming hot, room-temperature, or cold water and then to choose the melody that best matched their imagined drinking experience. The results revealed that imagining drinking cold water was associated with significantly higher pitches than drinking both room-temperature and hot water and with a significantly faster tempo than drinking room-temperature water. Next, the online study was replicated with participants in a lab tasting samples of hot, room-temperature, and cold water while choosing a melody that best matched their actual drinking experience. Those results confirmed that, compared with room-temperature and hot water, the experience of drinking cold water was associated with significantly higher pitches and a faster tempo [128]. One potential explanation for those results is emotional associations [129]. Evidence already exists for cross-modal correspondences between sound and smell [130] and between sound and taste [131] that are mediated by emotion, and both fast tempo and high pitch [132] are associated with increased arousal. The experience of drinking cold water might therefore be associated with a fast tempo and high pitches because it is deemed arousing and refreshing. Hot water, on the other hand, could be associated with soothing, calming warm beverages such as tea. This was especially true in the main study, where the hot water was served at 45 °C, a comfortable drinking temperature. It would be interesting to ask participants in an online study to associate pitch and tempo with both extremely hot water (around boiling, at 100 °C) and a comfortable 45 °C. One might expect the very hot (hence arousing) water to be associated with a faster tempo and higher pitches than the comfortably warm water. Of course, to truly verify the emotional association hypothesis, a future study would need gather information about the emotions that participants associate with each beverage sample [128].

Brunstrom et al. [133] explored oral temperature (e.g., of a beverage), a multi-sensory structure that includes odor and sound in addition to tactile and oral sensations. Successful mappings between temperature and color and then sound and color have been designed and tested. However, although such mapping does help the visually impaired to appreciate and understand colors in a new way, it is important to integrate those perceptual mappings into single multisensory system in which all those perceptual sensations can be perfectly linked and interchanged. In other words, a system is needed to enable the interactions of color, temperature, and sound to work together and give both sighted and visually impaired users the chance to experience "colors" through different perceptual sensations, thereby expanding their experience of colors and what they involve [133].

Cho et al. [87] investigated what color-directed sound have. Melodies with different pitch, tone, velocity, and tempo can be used as color sound codes to easily express the color lightness level. It is necessary to explore the cross-mode relationship between sound and temperature in order to express the color more concisely and perceptibly by integrating sound and temperature simultaneously. Two different coding color schemes with combining temperature and sound are suggested, for example:

(1) Celsius temperature represents six colors (e.g., 38—red, 34—orange, 30—yellow, 26—purple, 22—green, 14—blue), and three sound codes with different pitch and tempo (Table 4) represent three levels of color lightness.

(2) Isaac Newton's RYB color model consists of red, orange, yellow, green, blue, violet, red-orange (warmer red), red-violet (cooler red), yellow-orange (warmer yellow), yellow-green (cooler yellow), blue-violet (warmer blue), and blue-green (cooler blue). The color hue is expressed as sound (like in Table 4) and the warm and cold colors as two temperatures (e.g., 34 and 22 degrees Celsius). High color lightness (light), medium color lightness (muted) and low color lightness (dark) can be coded with temperatures like 14, 30, and 38 degrees Celsius, respectively. Combining temperature and sound in this way makes it simpler and easier to identify more colors, including warm/cool colors.

4.2. Temperature and Scent

Wnuk et al. [134] investigated the bases of those cross-modal associations, suggesting several possibilities, including universal forces (e.g., perception), and culture-specific forces (e.g., language and cultural beliefs). They examined odor–temperature associations in three cultures—Maniq, Thai, and Dutch—that differ with respect to their cultural preoccupation with odors, their odor lexicons, and their beliefs about the relationship between odors (and odor objects) and temperature. Their analysis revealed cross-modal associations that could not be explained by language but could be the result of cultural beliefs. Another possibility is that odor–temperature associations do not depend on cultural beliefs but are universal, perhaps due to shared physiology. It is often assumed that odors are associated with hot and cold temperatures because odor processing can trigger thermal sensations, such as the connection between coolness and mint. They found that menthol (peppermint) and cineole (eucalyptus) were consistently matched with the temperature term "cool". Laska et al. [135] also found that menthol (peppermint) and cineol (eucalyptus) consistently match the temperature conditions (cooling) [134].

Madzharov et al. [136] pretested six essential oils, three of which we expected to be perceived as warm scents (warm vanilla sugar, cinnamon pumpkin, and spice) and three as cool scents (eucalyptus–spearmint, peppermint, and winter wonderland). Following an established procedure (see Krishna, Elder, and Caldara 2010), 33 participants evaluated each scent on perceived temperature and liking ("smells like a cool/warm scent", seven-point scales). Of the six scents, cinnamon and vanilla were rated as the warmest, and peppermint was rated as the coolest. Cinnamon and peppermint were significantly different on the temperature dimension, as were vanilla and peppermint. According to Mackenzie [137], cinnamon and vanilla not only taste good to many people, but the scent of cinnamon or vanilla can invoke a warm, comforting feeling [136].

As shown in Table 5, orange and chocolate were used to easily express the color lightness level, and chocolate and menthol to express the temperature "warm/cool", respectively. It is necessary to explore the cross-mode relationships between scent and temperature, and scent and color lightness to express the color in terms of warm/cool and light/muted/dark. The following color-coding scheme with integrating temperature and scent is suggested.

Celsius (°C) temperature represents six colors (e.g., 38—red, 34—orange, 30—yellow, 26—purple, 22—green, 14—blue), and scents like orange and pine can convey two levels of color lightness (light/dark) (Table 5). Finally, warm and cold colors can be expressed by scents like chocolate and menthol.

Note that the six color hues cannot be expressed as scent since only four colors are associated with scent, as shown in Table 5. Combining temperature and scent in this way makes it simpler and easier to convey more colors, including warm/cool colors.

4.3. Scent and Sound

Researchers have started to document the existence of cross-modal correspondences between olfactory and auditory stimuli. For instance, Belkin [138] and Piesse [139] showed that people matched a series of different odors with sounds that differed in pitch. Piesse [139] introduces the idea that Olfaction can be described in ways that correlate to the musical notes on a diatonic scale. Those results were extended by [140,141], who found that people tended to match certain odors with the timbres of musical instruments.

Crisinel et al. [141] found that odors were preferentially matched to musical features: for example, the odors of candied orange and iris flower were matched to significantly higher pitches than the odors of musk and roasted coffee. Meanwhile, the odor of crème brûlée was associated with a more rounded shape than the musk odor. Moreover, by simultaneously testing cross-modal correspondences between olfactory stimuli and matches in two other modalities, they were able to compare the ratings associated with each correspondence. Stimuli judged as happier, more pleasant, and sweeter tended to be associated to both higher pitch and a more rounded shape, whereas other ratings seemed to be more specifically correlated with the choice of either pitch or shape. Odors rated as more arousing tended to be associated with the angular shape, but not with a particular pitch; odors judged as brighter were associated with higher pitch and, to a lesser extent, rounder shapes [141]. The emotional (hedonic) similarity between olfactory and auditory information could be crucial to both cross-modal correspondences and multisensory information processing [142].

Currently, Touch the Sound [143] and Perfumery Organ [144] are cross-sensory media works, but research on color expression is extremely rare. In Perfumery Organ [144], the fragrance is scented when played on the piano using the "incense" that connects the fragrance and sound devised by the perfumer Septimus Piesse [139], who matches "do (C4)" with rose, "le (C4)" with violet, and "mi (C4)" with acacia.

In Table 5, orange and chocolate can be used to easily express the color lightness level, and chocolate and menthol to express the temperature "warm/cool", respectively. It is necessary to explore the cross-modal relationships between scent and sound, and between scent and color lightness to express the color in terms of "warm/cool" and "light/muted/dark" and make it more easily perceptible. For example, the following color-coding scheme with integrating scent and sound is suggested.

Six colors can be expressed with sounds (Table 4). Scents of orange and pine convey two levels of brightness (brightness/darkness) (Table 5). Finally, scents like chocolate and menthol convey warm and cold colors (Table 5). Combining scent and sound in this way can make it simpler and easier to convey more colors, including warm/cool colors.

4.4. Scent and Shape

Odors rated as more arousing tended to be associated with the angular shape, but not with a particular pitch; odors judged as brighter were associated with higher pitch and, to a lesser extent, rounder shapes [140].

Humans do not arbitrarily attach sounds to shapes, as can be seen in the Kiki/Bouba effect [145,146]. Köhler [145] found that 95–98% assigned the name "bouba" to the rounded shape and "kiki" to the jagged shape, Figure 6.

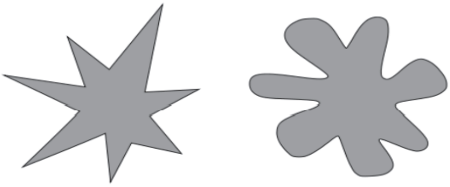


Figure 6. (**Left**) kiki: angular, jagged shapes; (**Right**) bouba: smooth, rounded shapes [146].

Adeli et al. [147] investigated the cross-modal correspondences between musical timbre and shapes. One hundred and nineteen subjects (31 females and 88 males) participated in the online experiment. Subjects included 36 claimed professional musicians, 47 claimed amateur musicians, and 36 claimed non-musicians. Thirty-one subjects have also claimed to have synesthesia-like experiences. Subjects have strongly associated soft timbres with blue, green or light gray rounded shapes, harsh timbres with red, yellow or dark gray sharp angular shapes. This is consistent with Kiki–Bouba experiment where subjects mostly chose a jagged shape for Kiki and a rounded shape for Bouba [145–148]. It is also consistent with Parise's findings [148] where subjects associated sine waves (soft sounds) with a rounded shape and square waves with a sharp angular shape [147].

Hanson-Vaux et al. [149] investigated how children relate emotions to smells and 3D shapes. Fourteen participants (ages 10–17 years) performed a cross-modal association task that gave emotional character to the transformation of the "kiki"/"bouba" stimulus presented as a 3D type model with lemon and vanilla flavors. The results of the study confirmed the association between the combination of the angular shape ("kiki") and the stimulating lemon scent, the round shape ("bouba") and the soothing vanilla scent. This expands the new results for the cross-mode response in terms of stimuli (3D rather than 2D shapes), samples (children), and delivered content compared to previous studies. We explored how these findings could contribute to the design of more comprehensive interactive multi-sensory technology.

Metatla et al. [150] investigated cross-modal associations between 20 odors (a selection of those commonly found in wine) and visual shape stimuli in a sample of 25 participants (mean age of 21 years). Two of the odors were found to be significantly associated with an angular shape (lemon and pepper) and two others with a rounded shape (raspberry and vanilla). Principal component analysis indicated that the hedonic value and intensity of odors are important in this cross-modality association, with more unpleasant and intense smells associated with more angular forms.

Lee et al. [89] investigated cross-modal associations between scents and visual shape stimuli like "kiki" and "bouba". The participants of the experiment were visually presented at the same time a paper with an angular shape and a rounded shape that corresponds to the words "kiki" and "bouba", respectively. The results of the study (Table 5) confirmed the association of the angular shape ("kiki") with the stimulating menthol, pine, and orange scents (associated with blue, green, and orange colors), and the round shape ("bouba") with the soothing chocolate scent (associated with brown). There was no significant difference in sharpness between menthol, pine and orange.

In summary, red and yellow are associated with "kiki" and blue and green with "bouba" [149]. Lemon is associated with an angular shape and vanilla with a rounded shape [149,150]. From [147–150], we can conjecture red and yellow are associated with lemon scent, and blue and green are associated with vanilla. From [89], orange, menthol, and pine scents correspond to orange, blue, and green that are associated with an angular shape. Additionally, chocolate scent corresponds to brown, which is associated with a rounded shape. Therefore, the results of research on the relationship between fragrance and shape might differ according to the cultural background of the participants.

5. Conclusions

In this review, a holistic experience using synesthesia acquired by people with visual impairment was provided to convey the meaning and contents of the work through rich multi-sensory appreciation. In addition, pictograms, temperatures, scents, music, and new forms incorporating them were explored to find a new way of conveying colors in artworks to the visually impaired. A method that allows people with visual impairments to engage in artwork using a variety of senses, including touch and sound, helps them to appreciate artwork at a deeper level than can be achieved with hearing or touch alone. The development of such art appreciation aids for the visually impaired will ultimately improve their cultural enjoyment and strengthen their access to culture and the arts.

The development of this new concept of post-visual art appreciation aids ultimately expands opportunities for the non-visually impaired as well as the visually impaired to enjoy works of art at the level of weak synesthesia and breaks down the boundaries between the disabled and the non-disabled in the field of culture and arts. It is made through continuous efforts to enhance accessibility. In addition, the developed multi-sensory expression and delivery tool can be used as an educational tool to increase product and artwork accessibility and usability through multi-modal interaction. Schifferstein [151] observed that vivid images occur in all sensory modalities. The quality of some types of sensory images tends to be better (e.g., vision, audition) than of others (e.g., smell and taste) for sighted people. The quality of visual and auditory images did not differ significantly. Therefore, training these multi-sensory experiences introduced in this paper may lead to more vivid visual imageries or seeing with the mind's eye.

Author Contributions: Conceptualization: J.D.C.; methodology: J.D.C.; validation: J.D.C.; formal analysis: J.D.C.; investigation: J.D.C.; resources: J.D.C.; data curation: J.D.C.; writing—original draft preparation: J.D.C.; writing—review and editing: J.D.C.; visualization: J.D.C.; supervision: J.D.C.; project administration and funding acquisition: J.D.C. The author has read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Technology and Humanity Converging Research Program of the National Research Foundation of Korea, grant number 2018M3C1B6061353.

Conflicts of Interest: The author declares that he has no conflict of interest.

References

1. Making Museums Accessible to Those with Disabilities. January 2020. Available online: https://www.museumnext.com/article/making-museums-accessible-to-those-with-disabilities (accessed on 30 November 2020).
2. Obrist, M.; Gatti, E.; Maggioni, E.; Vi, C.T.; Velasco, C. Multisensory Experiences in HCI. *Ieee Multimed.* **2017**, *24*, 9–13. [CrossRef]
3. Davis, M.H. A multidimensional approach to individual differences in empathy. *Jsas Cat. Sel. Docu-Ments Psychol.* **1980**, *10*, 85.
4. Coates, C. Best Practice in making Museums more Accessible to Visually Impaired Visitors. 2019. Available online: https://www.museumnext.com/article/making-museums-accessible-to-visually-impaired-visitors/ (accessed on 30 November 2020).
5. Samantha Silverberg. A New Way to See: Looking at Museums through the Eyes of The Blind. 2019. Available online: https://www.pressreleasepoint.com/new-way-see-looking-museums-through-eyes-blind (accessed on 30 November 2020).
6. Vaz, R.; Freitas, D.; Coelho, A. Blind and Visually Impaired Visitors' Experiences in Museums: Increasing Accessibility through Assistive Technologies. *Int. J. Incl. Mus.* **2020**, *13*, 57–80.
7. Carrizosa, H.G.; Sheehy, K.; Rix, J.; Seale, J.; Hayhoe, S. Designing technologies for museums: Accessibility and participation issues. *J. Enabling Technol.* **2020**, *14*, 31–39. [CrossRef]
8. Hayhoe, S. *Blind Visitor Experiences st Art Museums*; Rowman & Littlefield: Lanham, MD, USA, 2017.
9. Jadyn, L. Multisensory Met: The Development of Multisensory Art Exhibits. Available online: http://www.fondazionemarch.org/multisensory-met-the-development-of-multisensory-art-exhibits.php (accessed on 30 November 2020).
10. Axel, E.S.; Levent, N.S. *Art Beyond Sight: A Resource Guide to Art, Creativity, and Visual Impairment*; American Foundation for the Blind: Arlington County, VA, USA, 2003.
11. Wilson, P.F.; Griffiths, S.; Williams, E.; Smith, M.P.; Williams, M.A. Designing 3-D Prints for Blind and Partially Sighted Audiences in Museums: Exploring the Needs of Those Living with Sight Loss. *Visit. Stud.* **2020**, *23*, 1–21. [CrossRef]
12. Klimt's 'Kiss', made with 3D Printer, to Touch and Feel. 2016. Available online: https://phys.org/news/2016-10-klimt-3d-printer.html (accessed on 13 February 2021).
13. Morelli, L. Designing an Inclusive Audio Guide Part 2: Tactile Reproductions. 2016. Available online: https://www.warhol.org/designing-an-inclusive-audio-guide-part-2-tactile-reproductions/ (accessed on 13 February 2021).
14. Wong, M.; Gnanakumaran, V.; Goldreich, D. Tactile spatial acuity enhancement in blindness: Evidence for experience-dependent mechanisms. *J. Neurosci.* **2011**, *31*, 7028–7037. [CrossRef]
15. Heller, M.A. Picture and Pattern Perception in the Sighted and the Blind: The Advantage of the Late Blind. *Perception* **1989**, *18*, 379–389. [CrossRef]
16. Taylor, B.; Dey, A.; Siewiorek, D.; Smailagic, A. Customizable 3D Printed Tactile Maps as Interactive Overlays. Proceedings of 18th International ACM SIGACCESS Conference on Computers and Accessibility, Reno, NV, USA, 24–26 October 2016; pp. 71–79.
17. Götzelmann, T. Visually Augmented Audio-Tactile Graphics for Visually Impaired People. *Acm Trans. Access. Comput.* **2018**, *11*, 1–31. [CrossRef]
18. Landau, S.; Gourgey, K. Development of a talking tactile tablet. *Inf. Technol. Disabil.* **2001**, *7*, 4.
19. Brule, E.; Bailly, G.; Brock, A.; Valentin, F.; Denis, G.; Jouffrais, C. MapSense: Multi-Sensory Interactive Maps for Children Living with Visual Impairments. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 445–457.
20. Shen, H.; Edwards, O.; Miele, J.; Coughlan, J.M. *CamIO: A 3D Computer Vision System Enabling Audio/Haptic Interaction with Physical Objects by Blind Users*; Association for Computing Machinery: New York, NY, USA, 2013.
21. Baker, C.M.; Milne, L.R.; Drapeau, R.; Scofield, J.; Bennett, C.L.; Ladner, R.E. Tactile Graphics with a Voice. *Acm Trans. Access. Comput.* **2016**, *8*, 1–22. [CrossRef]
22. Fusco, G.; Morash, V.S. The tactile graphics helper: Providing audio clarification for tactile graphics using machine vision. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, Lisbon, Portugal, 26–28 October 2015; pp. 97–106.
23. Holloway, L.; Marriott, K.; Butler, M. Accessible maps for the blind: Comparing 3D printed models with tactile graphics. In Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13.
24. Talking Tactile Exhibit Panels. Available online: http://touchgraphics.com/portfolio/exhibit-panels/ (accessed on 13 February 2021).
25. San Diego Museum of Art Talking Tactile Exhibit Panels. Available online: http://touchgraphics.com/portfolio/sdma-exhibit-panel/ (accessed on 13 February 2021).
26. Volpe, Y.; Furferi, R.; Governi, L.; Tennirelli, G. Computer-based methodologies for semi-automatic 3D model generation from paintings. *Int. J. Comput. Aided Eng. Technol.* **2014**, *6*, 88. [CrossRef]
27. Holloway, L.; Marriott, K.; Butler, M.; Borning, A. Making Sense of Art: Access for Gallery Visitors with Vision Impairments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–12.
28. Anagnostakis, G.; Antoniou, M.; Kardamitsi, E.; Sachinidis, T.; Koutsabasis, P.; Stavrakis, M.; Vosinakis, S.; Zissis, D. Ac-Cessible Museum Collections for The Visually Impaired: Combining Tactile Exploration, Audio Descriptions and Mobile Gestures. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Florence, Italy, 6–9 September 2016; pp. 1021–1025.
29. Reichinger, A.; Carrizosa, H.G.; Wood, J.; Schröder, S.; Löw, C.; Luidolt, L.R.; Schimkowitsch, M.; Fuhrmann, A.; Maierhofer, S.; Purgathofer, W. Pictures in Your Mind. *Acm Trans. Access. Comput.* **2018**, *11*, 1–39. [CrossRef]

30. Reichinger, A.; Fuhrmann, A.; Maierhofer, S.; Purgathofer, W. *A Concept for Reuseable Interactive Tactile Reliefs. Computers Helping People with Special Needs*; Miesenberger, K., Bühler, C., Penaz, P., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 108–115.
31. Vaz, R.; Fernandes, P.O.; Veiga, A.C.R. Designing an Interactive Exhibitor for Assisting Blind and Visually Impaired Visitors in Tactile Exploration of Original Museum Pieces. *Procedia Comput. Sci.* **2018**, *138*, 561–570. [CrossRef]
32. D'Agnano, F.; Balletti, C.; Guerra, F.; Vernier, P. Tooteko: A case study of augmented reality for an accessible cultural heritage. Digitization, 3D printing and sensors for an audio-tactile experience. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 207.
33. Quero, L.C.; Bartolomé, J.I.; Lee, S.; Han, E.; Kim, S.; Cho, J. An Interactive Multimodal Guide to Improve Art Accessibility for Blind People. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway, Ireland, 22–24 October 2018; pp. 346–348.
34. Bartolome, J.I.; Quero, L.C.; Kim, S.; Um, M.-Y.; Cho, J. Exploring Art with a Voice Controlled Multimodal Guide for Blind People. In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, Tempe, AZ, USA, 17–20 March 2019; pp. 383–390.
35. Cavazos Quero, L.; Bartolomé, L.C.; Cho, J.D. Accessible Visual Artworks for Blind and Visually Impaired People: Com-paring a Multimodal Approach with Tactile Graphics. *Electronics* **2021**, *10*, 297. [CrossRef]
36. Cho, J.D. Art Touch: Multi-sensory Visual Art Experience Exhibition for People with Visual Impairment. *J. Korean Soc. Exhib. Des. Stud.* **2020**, *34*. Available online: http://kiss.kstudy.com/thesis/thesis-view.asp?key=3845460 (accessed on 13 February 2021).
37. Furini, M.; Mirri, S.; Montangero, M. Gamification and Accessibility. In Proceedings of the 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 11–14 January 2019; pp. 1–5.
38. Ullmer, B.; Ishii, H. Emerging frameworks for tangible user interfaces. *Ibm Syst. J.* **2000**, *39*, 915–931. [CrossRef]
39. McGookin, D.; Robertson, E.; Brewster, S. Clutching at Straws: Using Tangible Interaction to Provide Non-Visual Access to Graphs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; pp. 1715–1724.
40. Manshad, M.S.; Pontelli, E.; Manshad, S.J. Trackable Interactive Multimodal Manipulatives: Towards a Tangible User Environment for the Blind. In Proceedings of the Constructive Side-Channel Analysis and Secure Design, Darmstadt, Germany, 3–4 May 2012; pp. 664–671.
41. Pielot, M.; Henze, N.; Heuten, W.; Boll, S. Tangible User Interface for the Exploration of Auditory City Maps. In *Constructive Side-Channel Analysis and Secure Design*; Springer International Publishing: New York, NY, USA, 2007; Volume 4813, pp. 86–97.
42. Petrelli, D.; O'Brien, S. Phone vs. Tangible in Museums: A Comparative Study. In Proceedings of the CHI18 ACM CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; p. 112.
43. Spence, C. Crossmodal correspondences: A tutorial review. *Atten. Percept. Psychophys.* **2011**, *73*, 971–995. [CrossRef] [PubMed]
44. Tanaka, A.; Parkinson, A. Haptic wave: A cross-modal interface for visually impaired audio producers. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 2150–2161.
45. Gardner, J.A.; Bulatov, V. *Scientific Diagrams Made Easy with IVEOTM. Computers Helping People with Special Needs*; Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1243–1250.
46. Goulding, C. *Grounded Theory: A Practical Guide for Management, Business and Market Researchers*; Sage: Newcastle upon Tyne, UK, 2002.
47. Spence, C. Scenting the Anosmic Cube: On the Use of Ambient Scent in the Context of the Art Gallery or Museum. *i-Perception* **2020**, *11*, 2041669520966628. [CrossRef] [PubMed]
48. Elgammal, I.; Ferretti, M.; Risitano, M.; Sorrentino, A. Does digital technology improve the visitor experience? A comparative study in the museum context. *Int. J. Tour. Policy* **2020**, *10*, 47–67. [CrossRef]
49. Drobnick, J. The museum as a smellscape. In *Multisensory Museum: Cross-Disciplinary Perspective on Touch, Sound, Smell, Memory and Space*; Levent, N., Pascual-Leone, A., Eds.; Rowman and Littlefield: Lanham, MD, USA, 2014; pp. 177–196.
50. Vega-Gómez, F.I.; Miranda-Gonzalez, F.J.; Mayo, J.P.; González-López Óscar, R.; Pascual-Nebreda, L. The Scent of Art. Perception, Evaluation, and Behaviour in a Museum in Response to Olfactory Marketing. *Sustainability* **2020**, *12*, 1384. [CrossRef]
51. Seubert, J.; Rea, A.F.; Loughead, J.; Habel, U. Mood induction with olfactory stimuli reveals differential affective responses in males and females. *Chem. Senses* **2009**, *34*, 77–84. [CrossRef]
52. Herz, R.S. The Role of Odor-Evoked Memory in Psychological and Physiological Health. *Brain Sci.* **2016**, *6*, 22. [CrossRef]
53. Dobbelstein, D.; Herrdum, S.; Rukzio, E. Inscent: A Wearable Olfactory Display as An Amplification for Mobile Notifica-Tions. In Proceedings of the 2017 ACM international Symposium on Wearable Computers, Maui, HI, USA, 11–15 September 2017.
54. Choi, Y.; Cheok, A.D.; Roman, X.; Sugimoto, K.; Halupka, V. Sound Perfume: Designing a Wearable Sound and Fragrance Media for Face-To-Face Interpersonal Interaction. In Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, Lisbon, Portugal, 8–11 November 2011.
55. Choi, Y.; Parsani, R.; Roman, X.; Pandey, A.V.; Cheok, A.D. Light Perfume: Designing a Wearable Lighting and Ol-Factory Accessory for Empathic Interactions. In *Advances in Computer Entertainment*; Springer: Berlin, Germany, 2012; pp. 182–197.
56. Whitehurst, G.J.; Falco, F.L.; Lonigan, C.J.; Fischel, J.E.; DeBaryshe, B.D.; Valdez-Menchaca, M.C.; Caulfield, M. Accelerating language development through picture book reading. *Dev. Psychol.* **1988**, *24*, 552. [CrossRef]

57. Bara, F. Exploratory Procedures Employed by Visually Impaired Children During Joint Book Reading. *J. Dev. Phys. Disabil.* **2013**, *26*, 151–170. [CrossRef]
58. DeBaryshe, B.D. Joint picture-book reading correlates of early oral language skill. *J. Child Lang.* **1993**, *20*, 455–461. [CrossRef]
59. Monache, S.D.; Rocchesso, D.; Qi, J.; Buechley, L.; De Götzen, A.; Cestaro, D. Paper Mechanisms for Sonic Interaction. In Proceedings of the Sixth International Conference on Information and Communication Technologies and Development, Cape Town, South Africa, 11–14 November 2012; p. 61.
60. Edirisinghe, C.; Podari, N.; Cheok, A.D. A multi-sensory interactive reading experience for visually impaired children; a user evaluation. *Pers. Ubiquitous Comput.* **2018**, 1–13. [CrossRef]
61. Palmer, A. Can Smell Be a Work of Art? Scent Artist Sissel Tolaas Uses Chemistry to Explore the Malodorous, Yet Beautiful, Scent of Decay in Central Park. February 2016. Available online: https://www.smithsonianmag.com/smithsonian-institution/can-you-smell-work-art-180958189/ (accessed on 30 November 2020).
62. Vlek, R.; van Acken, J.P.; Beursken, E.; Roijendijk, L.; Haselager, P. BCI and a User's Judgment of Agency. In *Brain-Computer-Interfaces in Their Ethical, Social and Cultural Contexts*; Springer: Dordrecht, The Netherlands, 2014; pp. 193–202.
63. Cornelio, P.; Maggioni, E.; Brianza, G.; Subramanian, S.; Obrist, M. SmellControl: The Study of Sense of Agency in Smell. In Proceedings of the 2020 International Conference on Multimodal Interaction, Utrecht, The Netherlands, 25–29 October 2020; pp. 470–480.
64. Haggard, P. Sense of agency in the human brain. *Nat. Rev. Neurosci.* **2017**, *18*, 196–207. [CrossRef]
65. Jacobs, L.F.; Arter, J.; Cook, A.; Sulloway, F.J. Olfactory Orientation and Navigation in Humans. *PLoS ONE* **2015**, *10*, e0129387. [CrossRef]
66. Hopkin, M. Tone Task Proves Blind Hear Better—Early vision loss leads to keener hearing. *Engl. Sci. Technol. Learn.* **2004**, 5. [CrossRef]
67. McElligott, J.; Van Leeuwen, L. Designing Sound Tools and Toys for Blind and Visually Impaired Children. In Proceedings of the Conference on Interaction Design and Children Building a Community—IDC '04, MD, USA, 1–3 June 2004; pp. 65–72. Available online: https://dl.acm.org/doi/abs/10.1145/1017833.1017842 (accessed on 30 November 2020).
68. Culbertson, H.; Schorr, S.B.; Okamura, A.M. Haptics: The Present and Future of Artificial Touch Sensation. *Annu. Rev. Control Robot. Auton. Syst.* **2018**, *1*, 385–409. [CrossRef]
69. Vi, C.T.; Ablart, D.; Gatti, E.; Velasco, C.; Obrist, M. Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition. *Int. J. Hum. -Comput. Stud.* **2017**, *108*, 1–14. [CrossRef]
70. Baumgartner, T.; Lutz, K.; Schmidt, C.F.; Jäncke, L. The emotional power of music: How music enhances the feeling of affective pictures. *Brain Res.* **2006**, *1075*, 151–164. [CrossRef]
71. Feeling Van Gogh, Van Gogh Museum, Amsterdam. Feel, Smell and Listen to the Sunflowers. Available online: https://www.vangoghmuseum.nl/en/visit/whats-on/feeling-van-gogh (accessed on 30 November 2020).
72. Carrières de Lumières (France) and Bunker de Lumières (Korea). Available online: https://carrieres-lumieres.com (accessed on 30 November 2020).
73. Cytowic, R.E. Touching tastes, seeing smells–and shaking up brain science. *Cerebrum* **2002**, *4*, 7–26.
74. Jeong, H.; Kim, Y.; Kim Cho, J.D. Subjective Conformity Assessment of Matching between Artwork and Classical Music using Deep Learning based Imaginary Soundscape. In Proceedings of the International Conference on Convergence Technology, Jeju, Korea, 8–10 July 2020.
75. Shapiro, L.; Stolz, S.A. Embodied cognition and its significance for education. *Theory Res. Educ.* **2018**, *17*, 19–39. [CrossRef]
76. Brang, D.; Ramachandran, V.S. How do Crossmodal Correspondences and Multisensory Processes Relate to Synesthesia? In *Multisensory Perception*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 259–281.
77. Marks, L.E. Weak synesthesia in perception and language. In *The Oxford Handbook of Synesthesia*; Simner, J., Hubbard, E.M., Eds.; Oxford University Press: Oxford, UK, 2013; pp. 761–789.
78. Taggart, E. Synesthesia Artists Who Paint Their Multi-Sensory Experience. Available online: https://mymodernmet.com/synesthesia-art/ (accessed on 13 February 2021).
79. Martino, G.; Marks, L.E. Synesthesia: Strong and Weak. *Curr. Dir. Psychol. Sci.* **2001**, *10*, 61–65. [CrossRef]
80. Lawrence, E. *Marks the Unity of the Senses/Interrelationships Among the Modalities, Series in Cognition and Perception*; Academic Press: New York, NY, USA, 1978.
81. Marks, U. *Laura, the Skin of the Film: Intercultural Cinema, Embodiment, and the Senses*; Duke University Press: Durham, NC, USA, 2000; p. 162.
82. Cottin, M.; Faria, R.; Amado, E. *The Black Book of Colors*; Groundwood Books Ltd.: Toronto, ON, Canada, 2008.
83. Sathian, K.; Stilla, R. Cross-modal plasticity of tactile perception in blindness. *Restor. Neurol. Neurosci.* **2010**, *28*, 271–281. [CrossRef]
84. Vinter, A.; Orlandi, O.; Morgan, P. Identification of Textured Tactile Pictures in Visually Impaired and Blindfolded Sighted Children. *Front. Psychol.* **2020**, *11*, 345. [CrossRef]
85. Taras, C.; Ertl, T. Interaction with Colored Graphical Representations on Braille Devices. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, San Diego, CA, USA, 19–24 July 2009; pp. 164–173.
86. Cho, J.D.; Quero, L.C.; Bartolomé, J.I.; Lee, D.W.; Oh, U.; Lee, I. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Res. Appl.* **2021**, *46*, 103–116. [CrossRef]

87. Cho, J.D.; Jeong, J.; Kim, J.H.; Lee, H. Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments. *Electronics* **2020**, *9*, 1981. [CrossRef]
88. Bartolome, J.D.I.; Quero, L.C.; Cho, J.; Jo, S. Exploring Thermal Interaction for Visual Art Color Appreciation for the Visually Impaired People. In Proceedings of the International Conference on Electronics, Information, and Communication, (ICEIC), Barcelona, Spain, 19–22 January 2020; pp. 1–5.
89. Lee, H.; Cho, J.D. A Research on Using of Color-Concept Directed Scent for Visually Impaired Individuals to Appreciate Paintings. *Sci. Emot. Sensib.* **2020**, *23*, 73–92. [CrossRef]
90. Cappelletti, L.; Ferri, M.; Nicoletti, G. Vibrotactile color rendering for the visually impaired within the VIDET project. *Photonics East (Isamvvdciemb)* **1998**, *3524*, 92–97. [CrossRef]
91. Baumgartner, E.; Wiebel, C.B.; Gegenfurtner, K.R. A comparison of haptic material perception in blind and sighted individuals. *Vis. Res.* **2015**, *115*, 238–245. [CrossRef] [PubMed]
92. Shin, J.; Cho, J.D.; Lee, S. Please Touch Color: Tactile-Color Texture Design for The Visually Impaired. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020.
93. Ramsamy-Iranah, S.; Rosunee, S.; Kistamah, N. Application of assistive tactile symbols in a 'Tactile book' on color and shapes for children with visual impairments. *Int. J. Arts Sci.* **2017**, *10*, 575–590.
94. Stonehouse, P. Tactile Colours. (n.d.). Retrieved August. 2020. Available online: http://cpwstonehouse.com/tactile-colours/ (accessed on 13 February 2021).
95. Shin, J.; Cho, J.; Lee, S. Tactile-Color System for Accessibility of Color Education: 2.5D UV Printed Supplementary Material for Visually Impaired Students. In Proceedings of the International Conference on Ubiquitous Information Management and Commu-nication, Seoul, Korea, 4–6 January 2021.
96. Deville, B.; Bologna, G.; Vinckenbosch, M.; Pun, T. See Color: Seeing Colours with an Orchestra. In *Computer Vision*; Springer International Publishing: Berlin, Germany, 2009; pp. 251–279.
97. Palmer, S.E.; Schloss, K.B. An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8877–8882. [CrossRef]
98. Banf, M.; Blanz, V. Sonification of Images for The Visually Impaired Using a MultiLevel Approach. In Proceedings of the 4th Augmented Human International Conference on—AH '13, Stuttgart, Germany, 7–8 March 2013; pp. 162–169.
99. Cavaco, S.; Henriques, J.T.; Mengucci, M.; Correia, N.; Medeiros, F. Color Sonification for the Visually Impaired. *Procedia Technol.* **2013**, *9*, 1048–1057. [CrossRef]
100. Datteri, D.L.; Howard, J.N. The sound of color. In Proceedings of the International Conference on Music Perception and Cognition (ICMPC), Sapporo, Japan, 25–29 August 2008; pp. 767–771.
101. Firth, I. On the linkage of musical keys to colors. *Specul. Sci. Technol.* **1981**, *4*, 501–508.
102. Bartolomé, J.I.; Cho, J.D.; Quero, L.C.; Jo, S.; Cho, G. Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People. *Electronics* **2020**, *9*, 1939. [CrossRef]
103. Akazue, M.; Halvey, M.; Baillie, L.; Brewster, S. The Effect of Thermal Stimuli on the Emotional Perception of Images. In Proceedings of the CHI Conference on Human Factors in Computing Systems; Association for Computing Machinery (ACM), San Jose, CA, USA, 8–13 May 2016; pp. 4401–4412.
104. Peiris, R.L.; Peng, W.; Chen, Z.; Chan, L.; Minamizawa, K. Thermovr: Exploring Integrated Thermal Haptic Feedback with Head Mounted Displays. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 5452–5456.
105. Lee, W.; Lim, Y.K. Explorative research on the heat as an expression medium: Focused on interpersonal com-munication. *Pers. Ubiquitous Comput.* **2012**, *16*, 1039–1049. [CrossRef]
106. Luo, R. *Encyclopedia of Color Science and Technology*; Springer: Berlin, Germany, 2016.
107. De Valk, J.M.; Wnuk, E.; Huisman, J.L.A.; Majid, A. Odor–color associations differ with verbal descriptors for odors: A comparison of three linguistically diverse groups. *Psychon. Bull. Rev.* **2016**, *24*, 1171–1179. [CrossRef] [PubMed]
108. Gilbert, A.N.; Martin, R.; Kemp, S.E. Cross-Modal Correspondence between Vision and Olfaction: The Color of Smells. *Am. J. Psychol.* **1996**, *109*, 335. [CrossRef] [PubMed]
109. Kemp, S.E.; Gilbert, A.N. Odor Intensity and Color Lightness Are Correlated Sensory Dimensions. *Am. J. Psychol.* **1997**, *110*, 35–46. [CrossRef]
110. Li, S.; Chen, J.; Li, M.; Lin, J.; Wang, G. Color Odor: Odor Broadens the Color Identification of The Blind. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 2746–2751.
111. Nehmé, L.; Barbar, R.; Maric, Y.; Jacquot, M. Influence of odor function and color symbolism in odor–color as-sociations: A French–Lebanese–Taiwanese cross-cultural study. *Food Qual. Prefer.* **2016**, *49*, 33–41. [CrossRef]
112. Stevenson, R.J.; Rich, A.; Russell, A. The Nature and Origin of Cross-Modal Associations to Odours. *Percept.* **2012**, *41*, 606–619. [CrossRef]
113. Maric, Y.; Jacquot, M. Contribution to understanding odour–colour associations. *Food Qual. Prefer.* **2013**, *27*, 191–195. [CrossRef]
114. Kim, Y.-J. Can eyes smell? cross-modal correspondences between color hue-tone and fragrance family. *Color Res. Appl.* **2011**, *38*, 139–156. [CrossRef]

115. Adams, C.; Doucé, L. What's in a scent? Meaning, shape, and sensorial concepts elicited by scents. *J. Sens. Stud.* **2017**, *32*, e12256. [CrossRef]
116. Runciman, J.F. Noises, Smells and Colours. *Music. Q.* **1915**, *1*, 149–161. [CrossRef]
117. Dobrzynski, M.K.; Mejri, S.; Wischmann, S.; Floreano, D. Quantifying Information Transfer Through a Head-Attached Vibrotactile Display: Principles for Design and Control. *Ieee Trans. Biomed. Eng.* **2012**, *59*, 2011–2018. [CrossRef] [PubMed]
118. Sherrick, C.E.; Cholewiak, R.W.; Collins, A.A. The localization of low-and high-frequency vibrotactile stimuli. *J Acoust. Soc. Am.* **1990**, *88*, 169–179. [CrossRef]
119. Brewster, S.A.; Brown, L.M.; Brewster, S.A.; Brown, L.M. Tactons: Structured Tactile Messages for Non-Visual Information Display. In Proceedings of the Australasian User Interface Conference, Dunedin, New Zealand, 18–22 January 2004; pp. 15–23.
120. MacLean, K.; Enriquez, M. Perceptual design of haptic icons. In Proceedings of the EuroHaptics, Dublin, Ireland, 6–9 July 2003; pp. 351–363.
121. Morioka, M.; Griffin, M.J. Perception Thresholds for Vertical Vibration at The Hand, Seat and Foot. In Proceedings of the European Acoustic Association forum Acusticum, Budapest, Hungary, 29 August–2 September 2005; pp. 1577–1582.
122. Saket, B.; Prasojo, C.; Huang, Y.; Zhao, S. Designing an effective vibration-based notification interface for mobile phones. In Proceedings of the Conference on Internet Measurement Conference; Association for Computing Machinery (ACM), Barcelona, Spain, 23–25 October 2013; p. 149.
123. Cho, D. A Study on Danwon Kim Hongdo's Painting of Shieuido. *Korean Soc. Sci. East. Art* **2017**, *37*, 208–240. [CrossRef]
124. Cho, I.H. A Study on the Poetry-based-Paintings of Lee Jing. *Dongak Art Hist.* **2011**, *12*, 103–124.
125. Calvert, G.; Spence, C.; Stein, B.E. *The Handbook of Multisensory Processes*; MIT Press: Cambridge, MA, USA, 2004.
126. Haverkamp, M. *Synesthetic Design: Handbook for a Multi-Sensory Approach*; Walter de Gruyter: Berlin, Germany, 2012.
127. Neves, J. Multi-sensory approaches to (audio) describing the visual arts. *Monti. Monogr. De Traducción E Interpret.* **2012**, *4*, 277–293. [CrossRef]
128. Wang, Q.; Janice Spence, C. The Role of Pitch and Tempo in Sound-Temperature Crossmodal Correspondences. *Multisens. Res.* **2017**, *30*, 307–320. [CrossRef]
129. Levitan, C.A.; Charney, S.; Schloss, K.B.; Palmer, S.E. The Smell of Jazz: Crossmodal Correspondences Between Music, Odor, and Emotion. In Proceedings of the CogSci, Pasadena, CA, USA, 22–25 July 2015; pp. 1326–1331.
130. DeRoy, O.; Crisinel, A.-S.; Spence, C. Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes. *Psychon. Bull. Rev.* **2013**, *20*, 878–896. [CrossRef]
131. Wang, Q.J.; Wang, S.; Spence, C. Turn up the taste: Assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch. *Chem. Senses* **2016**, *41*, 345–356. [CrossRef] [PubMed]
132. Van Der Zwaag, W.; Gentile, G.; Gruetter, R.; Spierer, L.; Clarke, S. Where sound position influences sound object representations: A 7-T fMRI study. *NeuroImage* **2011**, *54*, 1803–1811. [CrossRef] [PubMed]
133. Brunstrom, J.M.; Macrae, A.W.; Roberts, B. Mouth-State Dependent Changes in the Judged Pleasantness of Water at Different Temperatures. *Physiol. Behav.* **1997**, *61*, 667–669. [CrossRef]
134. Wnuk, E.; De Valk, J.M.; Huisman, J.L.A.; Majid, A. Hot and Cold Smells: Odor-Temperature Associations across Cultures. *Front. Psychol.* **2017**, *8*, 1373. [CrossRef]
135. Laska, M. Perception of trigeminal chemosensory qualities in the elderly. *Chem. Senses* **2001**, *26*, 681–689. [CrossRef]
136. Madzharov, A.V.; Block, L.G.; Morrin, M. The cool scent of power: Effects of ambient scent on consumer pref-erences and choice behavior. *J. Mark.* **2015**, *79*, 83–96. [CrossRef]
137. Mackenzie, C. How to Make Your Home Smell Like Cinnamon or Vanilla. Available online: https://www.hunker.com/12392677/how-to-make-your-home-smell-like-cinnamon-or-vanilla (accessed on 13 February 2021).
138. Belkin, K.; Martin, R.; Kemp, S.E.; Gilbert, A.N. Auditory Pitch as a Perceptual Analogue to Odor Quality. *Psychol. Sci.* **1997**, *8*, 340–342. [CrossRef]
139. Piesse, G.W.S. *The Art of Perfumery and the Methods of Obtaining the Odours of Plants*; Longmans, Green, and Company: Harlow, UK, 1879.
140. Crisinel, A.-S.; Spence, C. A Fruity Note: Crossmodal associations between odors and musical notes. *Chem. Senses* **2011**, *37*, 151–158. [CrossRef] [PubMed]
141. Crisinel, A.-S.; Jacquier, C.; DeRoy, O.; Spence, C. Composing with Cross-modal Correspondences: Music and Odors in Concert. *Chemosens. Percept.* **2013**, *6*, 45–52. [CrossRef]
142. Velasco, C.; Balboa, D.; Marmolejo-Ramos, F.; Spence, C. Crossmodal effect of music and odor pleasantness on olfactory quality perception. *Front. Psychol* **2014**, *5*, 1352. [CrossRef] [PubMed]
143. Jagiełło, K. Sound Cloud. Available online: https://soundcloud.com/kjagielo (accessed on 13 February 2021).
144. Perfumery Organ, Tasko Inc. Available online: https://www.ntticc.or.jp/en/archive/works/perfumery-organ/ (accessed on 13 February 2021).
145. Köhler, W. *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*; Liveright Pub. Corp.: New York, NY, USA, 1947.
146. Ramachandran, V.S.; Hubbard, E.M. Synaesthesia: A window into perception, thought and language. *J. Conscious. Stud.* **2001**, *8*, 3–34.
147. Adeli, M.; Rouat, J.; Molotchnikoff, S. Audiovisual correspondence between musical timbre and visual shapes. *Front. Hum. Neurosci.* **2014**, *8*, 352. [PubMed]

148. Parise, C.V.; Spence, C. Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Exp. Brain Res.* **2012**, *220*, 319–333. [CrossRef]
149. Hanson-Vaux, G.; Crisinel, A.-S.; Spence, C. Smelling Shapes: Crossmodal Correspondences Between Odors and Shapes. *Chem. Senses* **2013**, *38*, 161–166. [CrossRef]
150. Metatla, O.; Maggioni, E.; Cullen, C.; Obrist, M. Like Popcorn Crossmodal Correspondences between Scents, 3D Shapes and Emotions in Children. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, UK, 4–9 May 2019; pp. 1–13.
151. Schifferstein, H.N.J. Comparing Mental Imagery across the Sensory Modalities. *Imagin. Cogn. Pers.* **2009**, *28*, 371–388. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Electronics Editorial Office
E-mail: electronics@mdpi.com
www.mdpi.com/journal/electronics

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

ISBN 978-3-0365-3059-8