

## Article

# Design and Implementation of Impromptu Mobile Social Karaoke for Digital Cultural Spaces in the New Normal Era

Choonsung Shin <sup>1</sup> and Yangmi Lim <sup>2,\*</sup><sup>1</sup> Graduate School of Culture, Chonnam National University, Gwangju 61186, Republic of Korea; cshin@jnu.ac.kr<sup>2</sup> IT Media Department, Duksung Woman's University, Seoul 01369, Republic of Korea

\* Correspondence: yosimi@duksung.ac.kr

**Abstract:** Although music singing and sharing is a most common cultural activity, users are not allowed to instantly share their own songs due to fixed streaming and listening services on the music service platforms. In order to deal with instant singing and listening together with other users, this research was conducted to propose a mobile social karaoke system that supports group creation and music sharing via smartphones. The proposed karaoke system consists of a social music cloud that provides impromptu mobile singing and sharing services on the user's devices. The social music cloud manages a group of users and supports data streaming and message sharing among the users' devices. The users' devices enable users to utilize karaoke services based on touch and voice-based natural interfaces in consideration of mobile specifications. After testing the system's usability and stability, the results illustrated that the voice-based interface was effective in controlling and using the service according to the devices' mobility and availability. In addition, when karaoke services are utilized in small groups, music transmission/reception is possible without being affected by the number of users. This study has three implications: first, it presented an important possibility for the creation of more active digital cultural spaces and changing mobile life by providing users with a recreation function; second, it provides convenient touch and voice UI for mobile users; and, third, it improves performance and management through its distributed processing.

**Keywords:** mobile music sharing; impromptu mobile network; user voice interaction



**Citation:** Shin, C.; Lim, Y. Design and Implementation of Impromptu Mobile Social Karaoke for Digital Cultural Spaces in the New Normal Era. *Appl. Sci.* **2023**, *13*, 12319. <https://doi.org/10.3390/app132212319>

Academic Editor: Luis Javier Garcia Villalba

Received: 6 October 2023

Revised: 2 November 2023

Accepted: 7 November 2023

Published: 14 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Due to the development of mobile computing technologies and the COVID-19 pandemic, there has been an accelerated change in the way people enjoy content on the digital platform. Post-COVID-19, audiences drastically decreased in theaters and performance halls, which were the central traditional cultural spaces. However, the fields of digital entertainment industries like broadcasting, performances, and movie-enabled group viewing through online platforms have tremendously increased. In addition, music sharing social media services are creating a new digital cultural space that stimulates a particular sensibility while sharing music with friends, colleagues, and acquaintances [1,2].

In particular, in the digital age, music sharing has formed new social networks. Through the recommendations of close friends and the utilization of 'Like', social ties and groups began to be formed [3]; furthermore, in the current situation where people cannot see each other face to face due to COVID-19, social network users have gradually begun to demand strong social space bonds through music streaming sharing services [4,5]. In addition, Spotify, one of the largest music streaming service providers, provides a music sharing service based on a wide range of tastes and knowledge including a sense of belonging by listening to music, construction of social relationships, and associations between pieces of personal information [6,7]. Spotify users also follow their Facebook friends and receive music feedback from them. In addition, the expansion of smart cars is accelerating the change in mobile music services and introducing connected service environments [8,9].

However, the current music sharing services still have difficulties in the re-creation of listener-centered music and the improvement of bonds in each individual's independent digital cultural space [10]. Users want to enjoy the traditional lifestyle of listening to music together in the same time slot, sharing sensibilities together, and singing together in digital cultural spaces [11]. For this to happen, a service was needed that would enable users to create and share social groups, regardless of place and time, and to share spontaneous reactions of listeners or audiences. Although mobility as a service (MaaS) combines vehicles and mobility to present diverse technologies for entertainment content services and music sharing, few studies have been conducted to design a real-time music sharing service centered on listeners or audiences.

To meet the lifestyle changes according to the development of digital cultural spaces, this research was conducted to propose a mobile social karaoke system for users to listen to, to sing along to, and to share songs based on the connectivity of users in mobile environments, thereby satisfying individuals' needs and improving mutual bonds among users. In this paper, the proposed mobile social karaoke system supports the creation of extemporaneous music sharing groups and network-based music streaming so that users are able to create social music groups anytime, anywhere, and sing along with the users participating in the groups. In addition, the proposed karaoke system was configured into both a personal mode in which individuals listen and sing alone and a social mode in which users sing with and listen to other users so that users can satisfy their personal needs or improve mutual bonds with other users. Furthermore, considering users' different types of movement, the system provided a convenient interface based on voice recognition and text-to-speech (TTS) modality for the user's safety while driving. This study experimented with a group of Android smartphone users to test the system's usability and the results showed that users can sing and share songs alone or together with multiple users in a mobile environment. For the new normal era, wherein the user's lifestyle connectivity is regarded as important in the mobile environment, now and in the future, this study will help to respond to the changes that digital cultural spaces face.

The instant sharing of the users' own singing on music service platforms is restricted due to the following reasons: (i) copyright and licensing agreements must be followed by music service platforms. If users' own singing data are allowed to be uploaded, this can constitute an infringement of copyrighted material, especially relating to songs that are under copyright protection. To prevent legal complications related to copyright laws, shared music is reviewed by music platforms; (ii) high-quality music files (audio and video) must be uploaded to the service platforms, otherwise, low-grade, poorly recorded, or unsatisfactory content will overload the system which will affect the overall experience of listeners; and (iii) community guidelines and standards must be followed by music platforms to suppress problems due to inappropriate or offensive content. In addition, spam and overload algorithm recommendations and bandwidth cost factors are significant reasons to restrict the direct uploading of users' content.

In this work, a mobile-based social karaoke system is proposed to support group creation and music sharing activities using smartphones. The social music cloud existing in this model provides impromptu mobile singing and sharing services directly on the user's device panel. Instant data streaming and messaging facilities for karaoke users are allowed through touch and voice-based natural interfaces. The usability and stability of this model have been tested. Here, a three-step procedure is followed to implement the effective karaoke system, primarily the introduction of significant potential to foster cultural environments. Secondly, it offers user-friendly touch and voice interfaces and, finally, enhanced performance and management through the distribution process. The proposed work is beneficial for group activities without violating community guidelines and standards. Moreover, it reduces the overload of spam and irrelevant content on the music service platforms.

The composition of this paper is as follows. In Section 2, both connectivity in the changing mobile environment and connectivity from music sharing social services are

examined. In Section 3, the concept and structure of the mobile social karaoke system in a mobile environment are explained, and, in Section 4, the implementation and experiment are explained. In Section 5, conclusions are presented and future work is described.

## 2. Related Work

### 2.1. MaaS Entertainment Technology

Smart cars, represented by autonomous vehicles, are being studied as spaces where not only music is listened to but also music is shared and produced by users. Samsung Electronics' digital cockpit integrates a total of eight displays and eight cameras to provide information that is necessary for driving, safety support, and entertainment streaming services [12]. The vehicle audio system that combines Continental and Sennheiser's Ac2ated sound and AMBEO 3D audio technology aims at the spatialization of mobility with speakerless technology. It also aims to provide the experience of immersive 3D sound regardless of the position of passengers because sounds are produced on diverse surfaces in the vehicle [6]. Hyundai Motor Company's separated sound zone (SSZ) control system is a technology to form sound zones, which are spaces where sound exists so that different independent sounds can be heard in a space within a vehicle [13]. Therefore, a mobile entertainment service in MaaS is becoming important and extends to include more diverse services for future needs.

### 2.2. MaaS Music Sharing Service

MaaS is a field in which attention from all businesses and schools is concentrated. Zhiyuan studied IoV-based mobile media scenarios to explain a vehicle mobile living space that is created through a connection with the intelligent Internet of Vehicles (IoV) [14]. The vehicle mobile living space technology is recognized as having made a major change to the MaaS ecosystem [15]. It includes major technologies including multi-transmission mode, mobile payment, and media combination, and, in particular, he regards in-vehicle infotainment (IVI) media as a major industry and explains that IVI will bring about changes in movie, music, and game services [13]. The combination of MaaS and smart mobile in the IVI industry is further accelerating the music listening experience of people around the world. Falkowski-Gilski studied technologies that enable users to share their music, tastes, and moods with other users by just accessing the web by mobile [16]. Prey argues that the playlist sharing algorithm of Spotify is a new product recreated by users [17–19]. Alev explains that although the development of music services has been great to the extent that, as of March 2020, Apple Music has 49.5 million monthly users and Spotify has 47.7 million monthly users, the consumption patterns of users or listeners are expected to change hereafter [1,7]. He says that the economic value of technologies that support users' immersive or active participation while firsthand experiencing music will increase. This also includes sharing technology for enjoying music together, and virtual reality technology for network-based singing together [20]. Furthermore, a recent study utilized music videos composed of visual and audio information as a teaching medium for mobile users since music video clips are abundant on various SNS platforms [21]. However, mobile music sharing services are mainly controlled for business purposes and hardly support end-users' needs and participation according to changes in life patterns.

The user's music listening life pattern based on immersive and active participating behavior is combined with the vehicle's mobile living space technology. In particular, in 2020, due to COVID-19, the practice of listening to music has changed, as shown in Figure 1a,b. Figure 1c expresses the hypothesized appearance of future utilization of the mobile social karaoke that is proposed as a vehicle mobile living space.

In the case of Figure 1c, streaming sharing technology is necessary to form cultural spaces where individuals listen to the same song in their respective spaces. Currently, YouTube and Spotify utilize streaming technologies most representatively. Sok-Ian Sou developed the MuLink mobile social service system utilizing the professional services provided by YouTube. Using YouTube's hashtags, playlists, and profile archive data, it

creates playlists for sharing groups among users based on their shared music [22]. This study uses beacons and Bluetooth to build a social and geographic social group music sharing system. Currently, due to streaming technology and sound source protection in mobile environments, the system is changing into a central server type. The above-mentioned application was built with WiFi-based cloud computing technology and has the important advantage of preventing wasting server resources and related costs [23–26]. The application allows users or listeners to recreate and reshare music based on their experiences and skills.



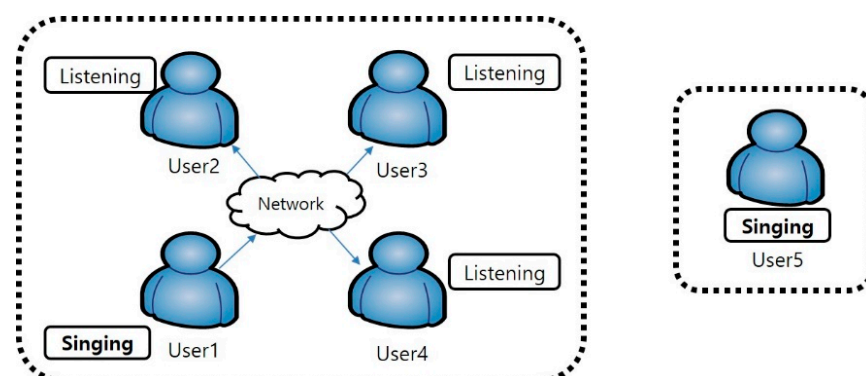
**Figure 1.** Changes in popular culture listening environments.

### 3. Design of the Impromptu Mobile Social Karaoke System

#### 3.1. Concept and System of Mobile Social Karaoke

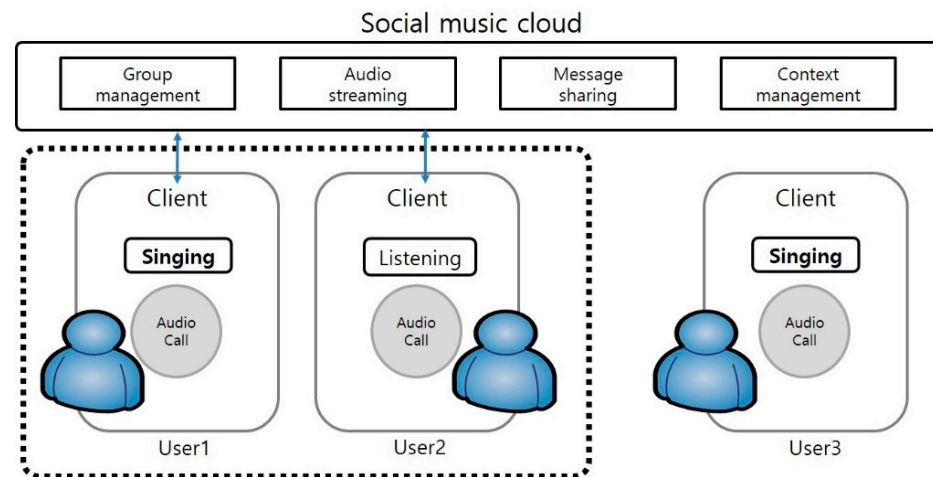
In this paper, the proposed mobile social karaoke system is a network-based social music sharing system that is extemporaneously created so that individuals or multiple users can sing and share songs together in a mobile environment. The service was designed to reflect the changing digital culture environment and individual-centered music creation and sharing so that, in the service, individuals can share songs while singing and create and share social music groups regardless of place and time. The social karaoke service consists of both a personal mode to personally listen and sing along and a social mode so a group can sing and listen to songs in turn. In the personal mode, the user can listen to, sing along to, or save and listen to the song again. In detail, the user can listen to and sing along to songs while moving or in a stationary state. The user can acquire songs online based on voice and touch, and sing along. The user can also record his/her own songs and listen to them again.

Unlike the personal mode of listening and singing alone, the social mode enables interested users to share music while singing or listening to songs. The social group is intended to enable users in the group to sing and listen together. Anyone can create the group at any time, and, when it has been created, anyone can participate in it. In this group, one user participates as a singer and the rest as listeners, and the order of the singers is changed randomly or sequentially. After each song is finished, users evaluate the song and assign a score to the best user. Figure 2 shows a block diagram of the personal mode and social mode in the mobile social karaoke service.



**Figure 2.** Social mode (left) and single mode (right).

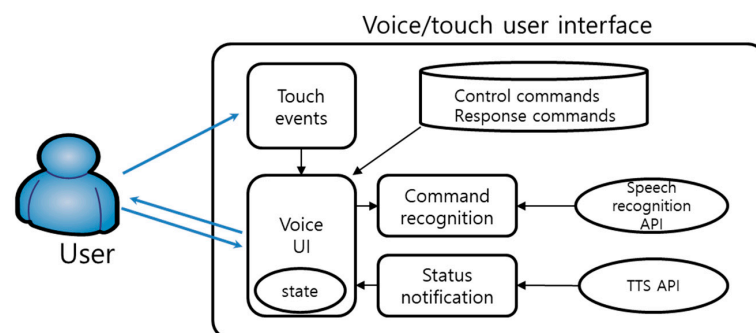
The system for the mobile social karaoke service is composed of a social music cloud server and a mobile device. As shown in Figure 3, the social music cloud server is shared by users and it conducts user management, group management, networking and streaming, and context management. The mobile device supports user interface, sound source acquisition and playback, sound source generation, storage, and streaming.



**Figure 3.** A system for the Mobile Social Karaoke service.

### 3.2. Voice/Touch-Based User Interface

Depending on how the user moves, the user interface is composed of voice/touch-based interactions as shown in Figure 4. Music services are controlled based on touch and voice recognition and TTS. The voice recognition is configured by interlocking the command set of user commands and answers with the TTS API. The voice commands consist of “alone mode”, “find song mode”, and “listen to song mode” in the personal mode, and “sing together mode”, “find song mode”, “listen to song mode”, and “evaluate mode” in the social mode. The answer commands consist of “right”, “yes”, “no”, and “nope”. In addition, in order to reduce errors and ensure quick responses, the interface supports voice recognition and responses while managing the state of the service. The service states consist of “start”, “single mode”, and “sing together mode”, and only the corresponding mode is selected from among the voice commands. In addition, the touch interface is configured on the screen to support interaction by hand when needed.



**Figure 4.** Voice/touch-based user interaction interface.

### 3.3. Social Music Information Sharing and Audio Streaming

The system includes a user datagram protocol (UDP) which supports connectionless data transmission for data streaming between mobile devices and the social music cloud. UDP packets capable of fast data transmission were configured for the extemporaneous creation of social music groups and music sharing. The packets consist of a control packet (CP) in charge of information and an opinion packet (OP) in charge of data. The CP contains



commands for group control, and the OP contains music streaming data. The OP is 7-byte data consisting of Course, Type, Error\_status, Group, and Command. The Course is the direction in which the packet is transmitted and expresses 0 in the case of the direction from the smartphone (HP) to the server and 1 in the opposite case. The Length represents the length of the CP, and the Type indicates transmission with 0 and responses with 1. The Error\_status indicates the status when there is an error during transmission/reception. The Group and Command are utilized for the states of users in the group and for control. The Group is a message type in the form of information that is sent from the device and the server and it represents notification delivery from the device (Group\_Info), notification delivery from the server (Group\_Server), and group session monitoring information (Group\_Monitor). The Group\_Info includes the delivery of group access, song start, song ends, and scores, and the Group\_Server includes the delivery of notices for the group, song start, request, score request, evaluation score notification, and the notification of song end in group. The Group\_Monitor is conducted periodically and it is a notification of the user device's entry into and exit from the session. Tables 1 and 2 describe this CP packet structure in detail. Based on these CP and OP packets, the user devices continuously communicate with the server to control the music sharing state and stream music.

**Table 1.** Structure of UDP CP.

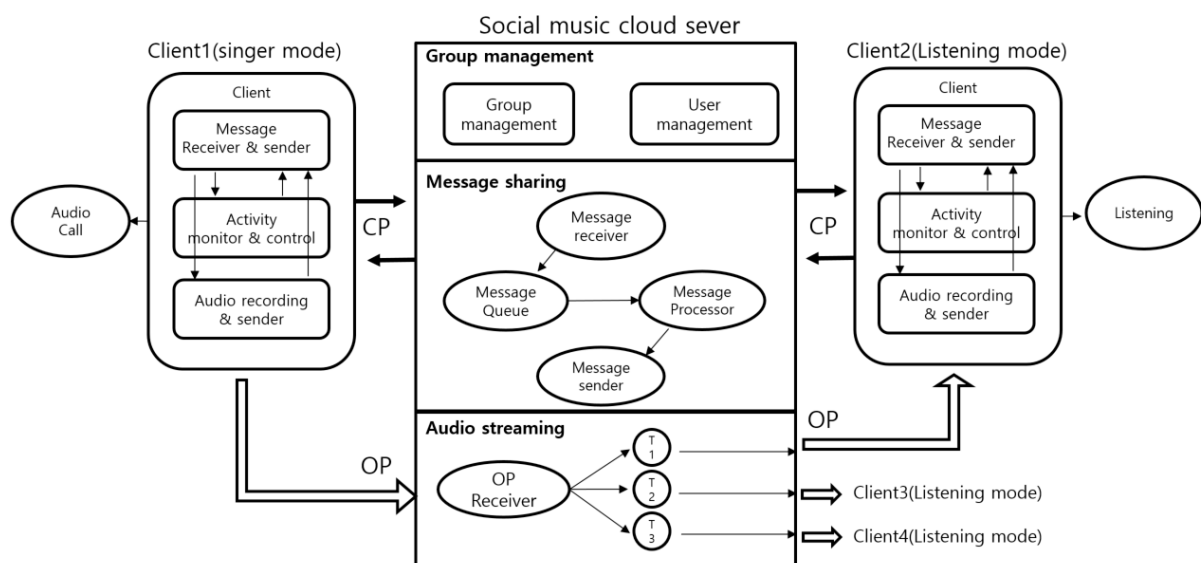
Item	Length	Remark
Course	1	0: Smart Phone(HP) -> server, 1: server -> HP
Length	2	Length of control message in bytes (CP length) Byte order
Type	1	0: send, 1: reply
Error status	1	When sending, the setting value is 0, and when responding, If the control message cannot be executed, the specified value is set.
Group	1	Group message type
Command	1	Command according to group message

**Table 2.** Group and command information in the CP packet.

Group	Command	Value	Remark
Group_Info	INFO_HP	1	Send HP's access ID and Nickname
	INFO_START	2	HP Song Start
	INFO_STOP	3	HP Song Stop
	INFO_SCORE	4	HP Song Score
Group_Server	SERVER_CONNLIST	1	Notify all HP users who have accessed the server
	SERVER_HPSTART	2	Notifies the affected HP accessor to sing
	SERVER_RSCORE	3	Request score
	SERVER_SCORE	4	The average score from each HP is averaged and sent to all visitors.
	SERVER_FINISH	5	All the songs of all members of the group are over.
Group_Monitor	MONITOR_SESSION	1	Send to server every 3 s
			If the last SESSION Time in the server is more than 10 s apart from the present, it is deleted from the list.
	MONITOR_MD	2	Delete from the sending list when singing
	MONITOR_GD	3	When a member leaves When the group is disbanded

As shown in Figure 5, the server exchanges CP with user devices to create a group, and, when a song starts, audio streaming occurs among users in the group. At this moment, the OP is utilized because OP is a packet to store voice data. The voice data are sampled from the singing device and are delivered to other users through the server. One session of a social music group consists of one person who is singing and the rest who are listening to the song. For this, UDP streaming for communication with the server is supported based

on an asynchronous socket. The UDP enables data transmission even before the connecting client agrees to connect so that low-delay delivery and 1:N information sharing is possible. When a mobile device generates audio data, they are delivered to the server, and the server obtains them from the mobile device and streams them to multiple members of the social group. In particular, the server uses a multithreading technique for effective audio sharing on multiple devices. This is a technique that creates multiple threads and delivers them at the same time in cases where a song sung by one user in a group is streamed to multiple devices. As shown in Figure 5, the social music group sharing and management of the mobile social karaoke system are conducted based on a user device (client singer mode and client 2 listening mode) and a cloud server (audio sharing server).

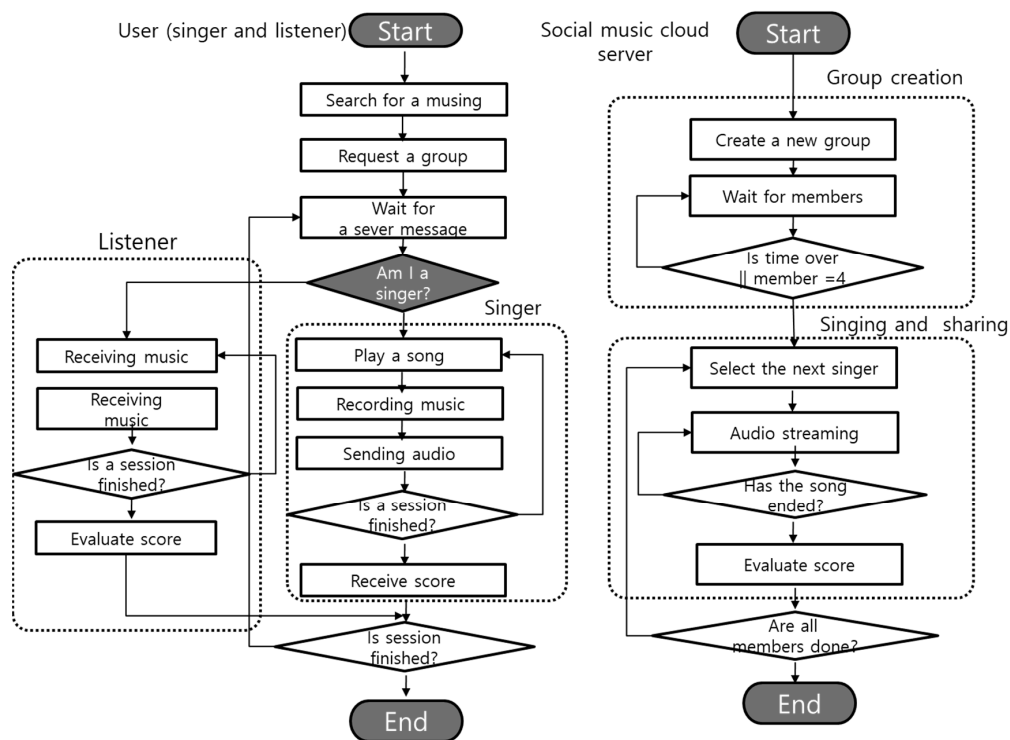


**Figure 5.** The structure for mobile social karaoke system.

### 3.4. Social Music Sharing Group Management and User Devices

The process of sharing and streaming of group information by the server and devices is shown in Figure 6. The server plays the role of managing users and groups, creating a music group according to the client's request, and sequentially selecting the next person to sing. The user device can be set to the personal mode for listening and singing alone, but in the social mode, the user participates in a karaoke group and uses a karaoke service based on linkage with the server.

In particular, in the social group mode, users and servers operate group control and music sharing services through communication. First, the user selects the social group mode, designates a singer and a song, and thereafter requests the group to participate in the session. In the singer mode, the search for songs, receiving the number of members in the user's group, singing, and requesting scores are conducted, and in the listening user mode, standby, receiving the number of members in the user's group, listening to songs, and evaluation are conducted. The server adds clients based on the group name and delivers this information to the members of the group. When group creation is completed, a person who will sing a song is selected and a message is delivered to the relevant user to be notified. When a song begins on the user's device, the server starts a voice streaming channel and streams the user's voice to other users. Thereafter, when the user has finished singing, the server requests an evaluation of the group users, calculates the average of the obtained scores, and delivers it to the person who sang the song. Then, the server checks whether all users in the group have sung, and if there is any user who has not sung, the server selects the next person who will sing in the group to sing, and, if not, the server terminates the session as all users in the group have sung.



**Figure 6.** Process of group information sharing and music streaming by the server and devices.

The karaoke system comprises several components, such as the user interface (UI), application logic, communication components, data storage cloud, external interfaces, security elements, monitoring system, and managing system. The role of the user interface (UI) is to facilitate easy interaction between the users and the karaoke system. It provides a panel for users to input commands and collect information. The communication components facilitate an easy exchange procedure between the different parts of the system. The storage cloud stores the data and manages their integrity, retransmission, and tenacity. The external interface property allows the karaoke system to connect with external services that extend the functionality of the system through external resources. Security elements protect against unauthorized access, data breaches, and vulnerabilities.

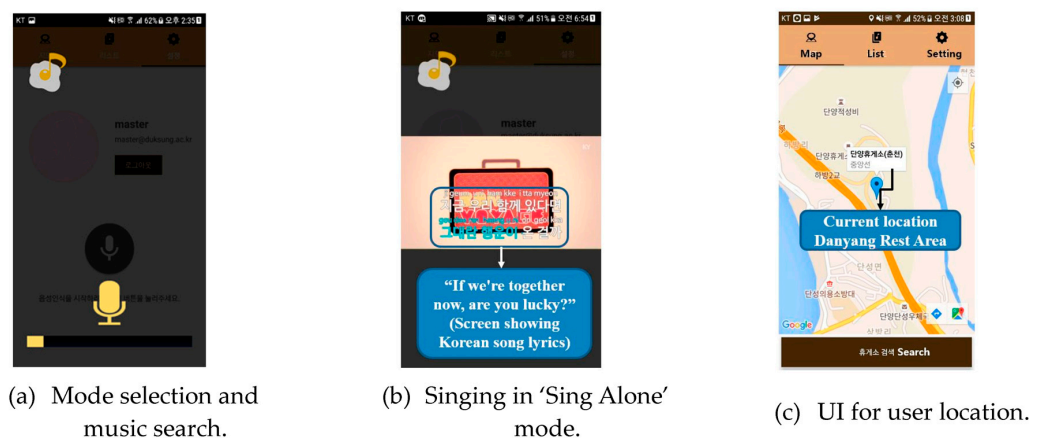
#### 4. Implementation and Evaluation of the Mobile Social Music Karaoke System

##### 4.1. Implementation

The proposed mobile social karaoke system was implemented based on Android smartphones and Apache web servers. We implemented the client app with Android smartphones powered by TTS, speech recognition, and intuitive interactions utilizing an API. In addition, music search and playback functions were added using a YouTube API to secure sound sources for karaoke. In addition, a ChatHead, which is an Android library that allows an app to be ready all the time, was configured to enable service through voice commands at any time so that providing full-time services could be possible in mobile situations. As for the web server, Apache and PHP were installed for web services such as login in the i9900K CPU and 32G RAM environment, and MariaDB was installed for data management. In addition, the server also supported real-time data exchange among smartphones by offering a UDP demon service. In addition, a 5G Wi-Fi network device was included to support the wireless network of the smartphones. User and group tables were configured so that user profiles and group profiles would be managed through MariaDB. In this environment, users' smartphones exchange information about the group through port 7000 of the server. The server would place port 50003 into standby mode for UDP data streaming when the group's voice sharing has started, and user devices would place port 50030 into standby mode in order to make it possible for audio streaming between users.

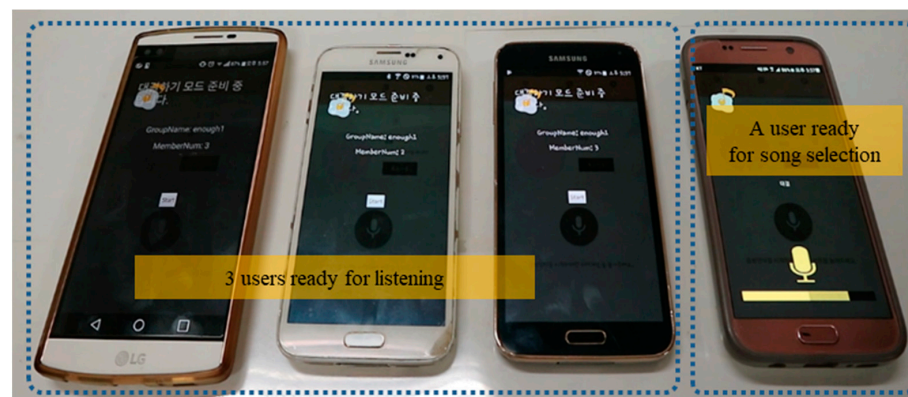


As shown in Figure 7, the proposed system works in two steps: starting the app and selecting singing mode. In the first step, a user can start the karaoke system by logging into the smartphone's karaoke app or clicking the chat head with a yellow icon overlaid on the screen. In the second step, each user needs to select a signing mode via touch or voice interface. Then, the user starts the karaoke system, the karaoke app requests voice to select the 'Sing Alone' or 'Sing Together' mode through the TTS function. When the user gives a voice command for 'Sing Alone', the personal mode runs. The karaoke app notifies that the 'Sing Alone' mode is running and notifies a request to 'tell the names of the song and singer'. Thereafter, when the names of the song and singer selected by the user have been supplied, the karaoke app retrieves the sound source through YouTube searches and executes it immediately. The user can sing along and save his/her song on his/her smartphone. In addition, the user is allowed to see his/her location to find other instant neighborhoods nearby. This function aims at knowing each other based on location, however, location sharing has not been implemented yet.



**Figure 7.** Social Karaoke Service UI.

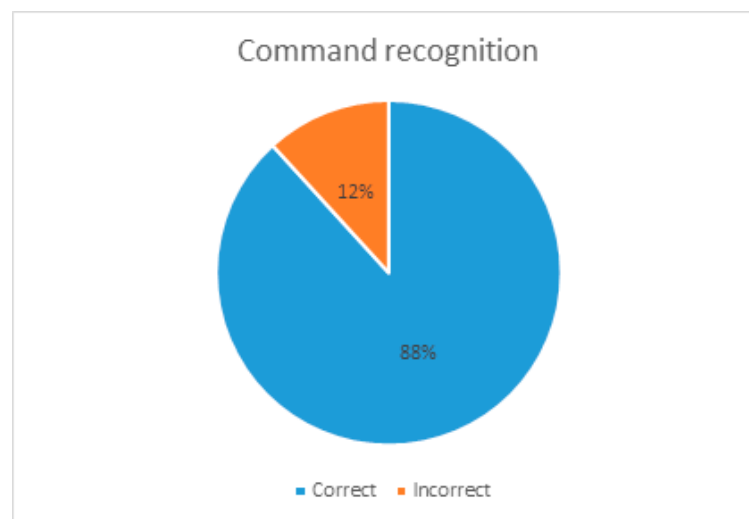
On the other hand, if the 'Sing Together' mode is selected in the first stage to select a mode after starting the karaoke app, the social mode in which singing in turns/control is possible will also be started. In this social mode, after first fixing the names of a song and the singer as in the 'Sing Alone' mode, the user waits until enough group participants are gathered. As shown in Figure 8, there were four smartphones that had joined the 'Sing Together' mode. When the number of participants reaches at least 3, if one of the users notifies the start or clicks start, the singing and listening will start. The remaining members who are not singing listen to the song and score the song, and notify the scores when the singing ends. The next person sings, and the karaoke service continues as such. In Figure 8, one person on the right is singing and three devices on the left are listening in a group of 4 people.



**Figure 8.** A group of 4 users in 'Sing Together' mode.

#### 4.2. Experiment and Analysis

In order to validate the proposed karaoke system, the performance of the user interface and mobile network stability with multiple devices were tested. The first experiment was to measure the accuracy of voice recognition while controlling the karaoke system. We utilized 4 smartphones and the voice commands consisted of mode selection, song and singer selection, and confirmation of starting and finalization. We collected voice recognition results while user groups were created and operated. During group singing mode, each user needed to select singing mode, specify the song, and confirm the selection. We obtained a total of 34 commands from 10 user sessions. As shown in Figure 9, the accuracy of the voice commands was 88% while the error of the voice commands was 12%. In other words, 30 of the 34 commands were recognized, while the remaining 4 commands failed. The 4 errors include system and timing errors in mode selection. In spite of such errors, the total recognition rate was high because the command was composed of very simple words to control the system.

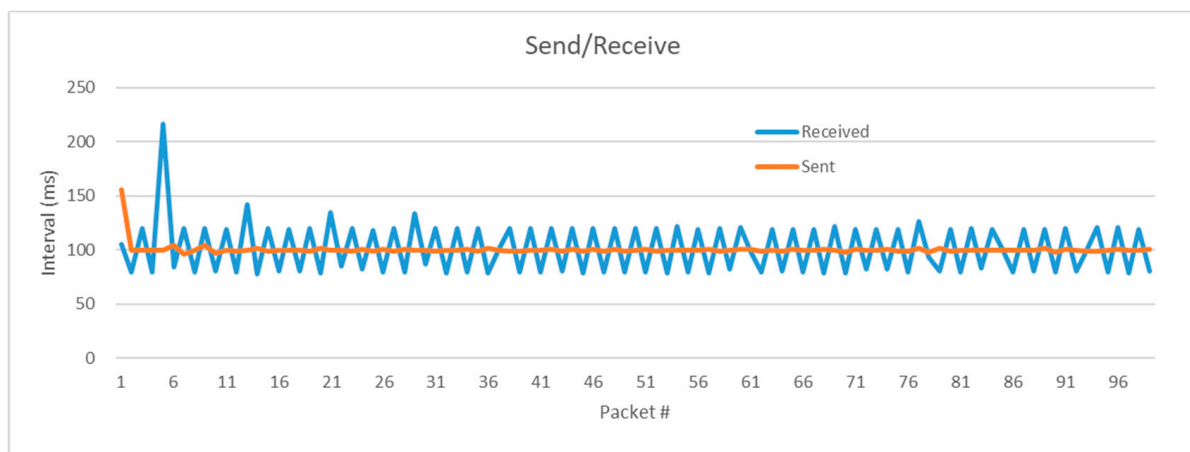


**Figure 9.** Evaluation of the accuracy of the voice interface.

In addition, the network stability of smartphones was tested during music sharing. In order to support the social karaoke system, it is important to support real-time music streaming from a singing user to other listening users. For this purpose, two experiments were conducted to evaluate the singing voice packet transmission/reception performance between users' smartphones, which is the most important function in karaoke services. The first experiment was about transmission/reception performance in the minimum group environment consisting of two users and the second experiment was about the voice packet transmission/reception performance of larger groups. In the first experiment to evaluate the network streaming performance, the smallest group of two people was formed and each performance was tested in singing and listening, respectively. Then, packet transmission between the members, UDP relay, and reception characteristics were measured. As shown in the experiment result in Figure 10, the sending smartphone transmitted 1.6k voice recordings and data packets every 0.1 s, and the UDP server received the data packets from the singer and retransmitted them to the registered group members. It could be seen that the receiving user's smartphone received and played the voice packets at 0.1 s (+/− 20 ms). That is, the transmission/reception is carried out at the rate of 10 packets/s. Although there were slight delays in reception, since karaoke users are usually far away from each other, the slight delays had almost no effect on listening to the music.

In the second experiment, the performance of singing data streaming was evaluated by increasing the number of group members. For this purpose, 3 small groups were tested by increasing the number of group users from 2 to 4. Four different types of smartphones,

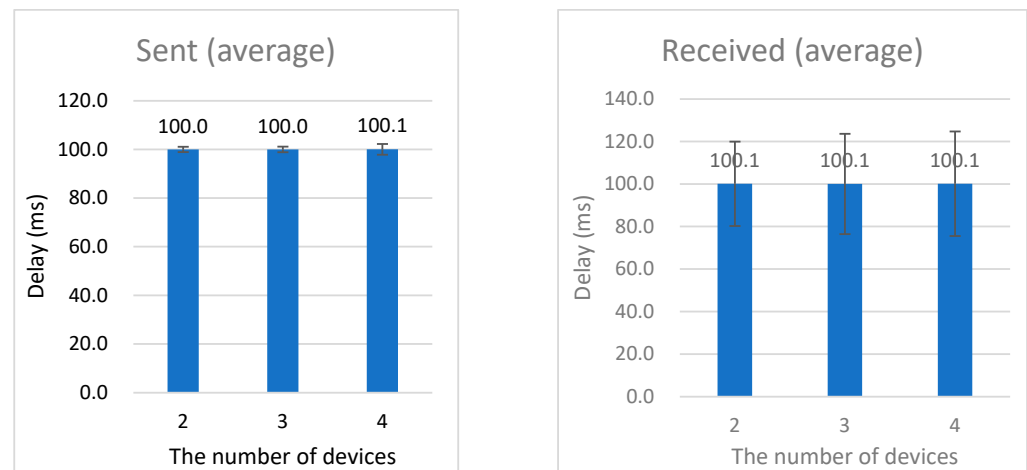
including recently released smartphones and low-performance smartphones, were utilized in order to test the possibility of smartphone performance. During each test, the ‘Sing Together’ mode was initiated, and small groups consisting of 2, 3, or 4 persons were formed. Thus, one user was singing and other users were listening. The smartphones were in the same location, where 5G Wi-Fi was covered, to validate the network performance between the social music cloud and smartphones. To evaluate the streaming transmission/reception performance, the same song was repeatedly sung and listened to by individual smartphones in turns, and the voice packet transmission/reception performance was measured. The sending smartphone of a singer transmitted 1.6k voice recordings and data packets every 0.1 s, and the UDP server received the data packets from the singer’s smartphone and retransmitted them to the registered group members.



**Figure 10.** Measurement of packet transmission/reception between group members.

As with the experiment result shown in Figure 11, while voice packets were transmitted at regular intervals, large deviations occurred in reception. While transmission was conducted regularly with deviations not larger than 3 ms per 100 ms on average, deviations up to  $\pm 25$  ms per 100 ms occurred in reception. Although this fluctuation in reception is mainly due to the characteristics of UDP, data transmission is very fast and could support more devices without preparing a socket connection. The transmission/reception deviations did not increase very much even when the number of users in the group increased. In particular, even in the group of 4 members, the time did not change significantly. This is because packet transmission delay time was minimized by including the server and then the delay was no longer than 1 ms. Although the deviation of packet reception was higher than the transmission, the frustration was almost similar to the 2 groups of users described in the first experiment. On the other hand, we found that there were differences in performance according to the smartphone specifications. Deviations occurred a little more in the smartphones with low performance and not much deviation occurred in the recently released smartphones. Therefore, it could be seen that large performance degradation does not occur because one mobile device transmits packets, and the server transmits packets to individual smartphones of group members separately.

The results of the experiments showed that the voice interface of voice recognition and TTS was effective in performing the desired functions while increasing accuracy based on the specified command set. Given the mobile and mobile environment, the role of a natural interface is important, and utilization to that end was possible. As for network-based music sharing, music packets could be quickly transmitted, relayed, received, and played so that music could be shared in real time and performance degradation did not occur even when the number of users increased.



**Figure 11.** Network Streaming Performance Evaluation 2.

#### 4.3. Mobility and Availability

In terms of user mobility and availability, the voice-based interface of the karaoke system is effective in several ways. The significant points are as follows: (i) Enhanced portability and accessibility: the voice-based interfaces, such as the speech recognition technique, liberate users from carrying physical input devices so that users can access the karaoke system at a wider distance due to the hands-free service when they are in motion, at a social gathering, or driving a vehicle. The mobility feature of the setup grants a dynamic and pleasant experience to users; (ii) User convenience: using the voice-based technique of the karaoke system, users can immediately find their favorite songs, control playback, and adjust settings just by giving voice commands. This service is more useful when they are in social places or groups to enjoy music with their smartphones; (iii) Real-time interaction and availability: real-time interaction with the karaoke system is possible due to the voice-based interface. Users can tune to their favorite songs, adjust the volume, and provide feedback without interfering with their activities.

The quick action of voice-based commands greatly enhances the effectiveness of the presented karaoke system by improving mobility and availability.

## 5. Discussion

### 5.1. Implication

(1) Provide an impromptu mobile music culture space. Music sharing is now changing from an environment where providers are determined to an environment where users can create and consume freely. In this paper, the proposed karaoke system enables users to recreate their own songs regardless of when and where they are and share them with other users extemporaneously. While most of the various music services surveyed in previous studies forced passive listening, the proposed karaoke system presents an important possibility for the creation of more active digital cultural spaces and changing mobile life by giving users a re-creation function. The model encourages healthy interaction and engagement for users without the need for extensive planning. It is user-friendly, it responds quickly, and it creates music for users together.

(2) Provide convenient touch and voice UI for mobile users. The voice and touch UI is useful even in situations where users are engaged in tasks, including driving or other activities. Most modern song services focus on listening to streaming services, but a more convenient interface in addition to touch is important to listen to and to sing along to in mobile environments. In the proposed karaoke service, voice recognition and TTS functions are provided based on a predetermined voice set and state to enable system control based on voice. Therefore, users are able not only to see and touch the screen but also to utilize voice commands to control music sharing while moving or driving. The design of the touch interface with its simplicity and intuitiveness provides recognizable icons, buttons, and

navigation panels for users. Users can easily initiate activities with voice and touch UI commands and can tune into music according to their interests.

(3) Improve performance and management through distributed processing. The configuration of services through collaboration between the cloud server and the user's smartphone is effective. When sharing music between users, the method of sharing music with nearby users has the disadvantage of increasing the network load of smartphone users. In the karaoke system, the cloud server manages the group and performs voice streaming to minimize the impact on the performance of music transmission and reception even when the number of users increases. The cloud server receives UPD data packets from a smartphone user and then copies and delivers them to other smartphones continuously, thus being independent of the number and performance of mobile devices. Parallel processing breaks down complex tasks into smaller segments and processes multiple nodes simultaneously. Because of that, overall execution time is reduced, and the tasks are completed in a short time.

(4) Impact of digital cultural spaces and users' mobile lives. There are several ways by which the proposed karaoke system has an impact on digital cultural spaces and users' mobile lives in a positive way. It acts as a significant platform for cultural expression and interaction. With this system, users can sing a song in their native culture and language that forecasts the diversity of cultural spaces. In this technology, people are connected to each other in the group, so, online community centers can be formed to share similar interests among music enthusiasts. These communities can serve as core aspects of users' mobile lives that offer a sense of membership and affiliation. It is also an entertaining and stress-relieving tool to have fun in daily life using smartphones. Moreover, the karaoke system provides a digital platform to share songs that promote user-generated content and culture. To improve their singing skills, users can rehearse using this system and gain valuable knowledge. In simple words, this system is a form of informal education integrated into users' mobile lives. It can also be utilized in the form of entrepreneurial opportunities. One of the useful properties is the data storage of user preferences, selective songs, and singing habits that can be utilized for content recommendation algorithms. Ultimately, all the features enhance digital cultural spaces and users' mobile lives.

## 5.2. Limitation

To create and share diverse songs, multiple groups in which many users participate should be supported. Currently, creation, standby, start, and service functions are tailored to one group. However, as the number of users increases, more than one group will be joined and, accordingly, group search and recommendations become important. For this, group profiles including group preferences, genres, and tags should be defined and managed. In addition, to further ensure the mobility of users, group creation/recommendation based on location/time is necessary.

In addition, the interface for users needs to be improved. In the case of the voice interface of the proposed system, the command set should be improved for natural recognition of commands other than the determined commands and for accuracy and faster movement so that the menu can be easily moved. In addition, information on groups and the user's profile should be provided on the user screen and linked to the voice interface for an intuitive interface.

Finally, the usability of music sharing services needs to be evaluated based on diverse users. In this paper, a basic performance evaluation was conducted for voice and network functions. However, a more comprehensive evaluation of usability, such as the functional and quality characteristics of music sharing and services, is needed. If the services are provided according to each user's different way of singing along, group creation, and song evaluation, it is expected that the service will be easier to use.

## 6. Conclusions

This research was conducted to propose a social karaoke system on the mobile phone where users can extemporaneously sing and share songs. The proposed karaoke system consists of a social music cloud and user devices to provide impromptu mobile singing and sharing services. The social music cloud managed a group of users and supported data streaming and message sharing among user devices. The user devices enabled users to utilize karaoke services based on a touch- and voice-based natural interface in consideration of mobile specifications. After testing its usability and stability, the results illustrated that the voice-based interface was effective in controlling and using the service. In addition, music transmission/reception was possible without being affected by the number of users when karaoke services are utilized in small groups.

Impromptu music sharing in a digital cultural environment is now in its incipient stage, so further studies will be needed. First, for practical music creation and sharing services, studies on group creation, management, recommendation, and profiles will be conducted for multiple groups. Currently, the proposed system only supports one group and more dynamic group management should be improved for multiple groups simultaneously. This extension also needs to include location-based music sharing and recommendations. In addition, voice- and visualization-based user interfaces need to be provided. In this work, we proposed the interfaces required/needed based on a cost–benefit analysis presented for static voice recognition but aims to target dynamic voice support groups in moving or driving situations. To set up a dynamic interface driven by cost benefits, elements such as a dynamic voice recognition interface, user profiles and settings, hands-free operations, voice command for navigation, scalability, etc., are required. While the proposed system basically works on voice recognition, a further study should include more dynamic voice commands to support groups while moving or driving. Finally, users' opinions and experiences need to be evaluated while various users will participate and create, share, and utilize music. The quality and method of singing and social music sharing could be improved by the more diverse participation of users in different locations.

**Author Contributions:** Conceptualization, C.S. and Y.L.; Methodology, C.S. and Y.L.; Formal analysis, Y.L.; Writing—original draft, C.S. and Y.L. All authors contributed toward data analysis, drafting and revising the paper, and agreed to be responsible for all aspects of this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (no. 2021R1F1A1047682) and by the MSIT (Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization Support Program (no. IITP-2022-RS-2022-00156287), supervised by the IITP (Institute for Information & Communications Technology Planning Evaluation).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

1. Alan, A.K.; Kabadayi, E.T.; Bakis, S. Creating Experience in Digital Music Services: The Role of Personal Innovativeness and Mobility—Creating Experience in Digital Music Services. In *Structural Equation Modeling Approaches to E-Service Adoption*; IGI Global: Hershey, PA, USA, 2019; pp. 1–23. [\[CrossRef\]](#)
2. Jones, S. Music and the Internet. In *The Handbook of Internet Studies*; Wiley-Blackwell: Oxford, UK, 2011; pp. 440–451.
3. Vriens, E.; Ingen, E. Does the rise of the Internet bring erosion of strong ties? Analyses of social media use and changes in core discussion networks. *SAGE J. New Media Soc.* **2018**, *20*, 2432–2449. [\[CrossRef\]](#) [\[PubMed\]](#)



4. Brandtzaeg, P.B. Social networking sites: Their users and social implications-A longitudinal study. *J. Comput.-Mediat. Commun.* **2012**, *17*, 467–488. [CrossRef]
5. Liu, Q.; Luo, M. Cross-Cultural Examination of Music Sharing Intentions on Social Media: A Comparative Studies in China and the United States. *Int. J. Hum.-Comput. Interact.* **2023**, *39*, 1303–1313. [CrossRef]
6. Cohen, M.; Martens, W.L. Spatial Soundscape Superposition and Multimodal Interaction. In *The Technology of Binaural Understanding; Part of Modern Acoustics and Signal Processing Book*, 13 August 2020; Springer: Cham, Switzerland, 2020; pp. 351–390. Available online: [https://link.springer.com/chapter/10.1007/978-3-030-00386-9\\_13](https://link.springer.com/chapter/10.1007/978-3-030-00386-9_13) (accessed on 7 November 2022).
7. Wendy's Taps Spotify Branded Moments to Catch You When You're Hungry. Available online: <https://adexchanger.com/advertiser/wendys-taps-spotify-branded-moments-catch-youre-hungry/> (accessed on 7 November 2022).
8. Coppola, R.; Morisio, M. Connected Car: Technologies, issues, future trends. *ACM Comput. Surv.* **2016**, *49*, 46. [CrossRef]
9. Steffen, G.; Christian, W.; Michael, R. How Automakers Can Survive the Self-Driving Era. Available online: <https://www.es.kearney.com/automotive/article?/a/how-automakers-can-survive-the-self-driving-era> (accessed on 7 November 2022).
10. Goffman, E. *The Presentation of Self in Everyday Life*; Penguin Press: London, UK, 1990.
11. Liu, K.T.; Reimer, R.A. Social playlist: Enabling touch points and enriching ongoing relationships through collaborative mobile music listening. In *Processors Mobile HCI '08*; ACM Press: New York, NY, USA, 2008; pp. 403–406. [CrossRef]
12. Choi, J.-K.; Klm, K.; Kim, D.; Choi, H.; Jang, B. Driver-adaptive vehicle interaction system for the advanced digital cockpit. In Proceedings of the 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon, Republic of Korea, 11–14 February 2018; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]
13. Kim, H.J. Apparatus and Method for Controlling Sound in Vehicle. U.S. Patent US20170133001A1, 18 December 2018. Current Assignee: Hyundai Motor Co, Kia Motors Corp. Available online: <https://patents.google.com/patent/US20170133001> (accessed on 5 October 2023).
14. Yu, Z.; Jin, D.; Song, X.; Zhai, C.; Wang, D. Internet of Vehicle Empowered Mobile Media Scenarios: In-Vehicle Infotainment Solutions for the Mobility as a Service (MaaS). *Sustainability* **2020**, *12*, 7448. [CrossRef]
15. Kamargianni, M.; Matyas, M. The business ecosystem of mobility-as-a-service. In Proceedings of the 96th, Transportation Research Board, Washington, DC, USA, 8–12 January 2017; pp. 1–14.
16. Falkowski-Gilski, P. On the Consumption of Multimedia Content Using Mobile Devices: A Year to Year User Case Study. *Arch. Acoust.* **2020**, *45*, 321–328. [CrossRef]
17. Prey, R. Locating Power in Platformization: Music Streaming Playlists and Curatorial Power, Social Media + Society. *Sage J.* **2020**, *6*, 2056305120933291. [CrossRef]
18. Secret Lives of Playlists. Available online: <https://watt.cashmusic.org/writing/theseecretlivesofplaylists> (accessed on 7 November 2022).
19. Storme, T.; De Vos, J.; De Paepe, L.; Witlox, F. Limitations to the car-substitution effect of MaaS. Findings from a Belgian pilot study. *Transp. Res. Part A Policy Pract.* **2020**, *131*, 196–205. [CrossRef]
20. Sullivan, J.L. The Platforms of Podcasting: Past and Present. *SAGE J. Soc. Media + Soc.* **2019**, *5*, 2056305119880002. [CrossRef]
21. Zhuo, K.; Guofeng, L. Popular music singing video teaching based on android mobile network and embedded system. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *13*, 91. [CrossRef]
22. Sou, S.; Chen, S.; Li, H.; Yang, P.; Wang, Y. Nearby x Music: A Mobile Social Service in Proximity based on YouTube Hashtag and Playlist. Institute of Computer and Communication Engineering. In Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 15–18 October 2019; pp. 517–518. [CrossRef]
23. Kang, P.; Bang, H.; Ham, S. Method and Apparatus for Playing Multimedia Content of Server via Streaming at Client Using Cache. Korea Patent KR20140105298A, 1 September 2014.
24. Chen, X.; Jiang, Z.; Xiong KShen, J. Practical Routing Protocol for Impromptu Mobile Social Networks. In Proceedings of the 2014 12th IEEE International Conference on Embedded and Ubiquitous Computing, Milano, Italy, 26–28 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 258–264.
25. Xu, Y.; Chen, X. Social-Similarity-Based Multicast Algorithm in Impromptu Mobile Social Networks. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 346–351.
26. Lim, V.; Khan, S.; Picinali, L. Towards a More Accessible Cultural Heritage: Challenges and Opportunities in Contextualisation Using 3D Sound Narratives. *Appl. Sci.* **2021**, *11*, 3336. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.