



Article AudienceMR: Extending the Local Space for Large-Scale Audience with Mixed Reality for Enhanced Remote Lecturer Experience

Bin Han 💿 and Gerard Jounghyun Kim * 💿

Computer Science and Engineering, Korea University, Seoul 02841, Korea; hanbin97@korea.ac.kr * Correspondence: gjkim@korea.ac.kr

Abstract: AudienceMR is designed as a multi-user mixed reality space that seamlessly extends the local user space to become a large, shared classroom where some of the audience members are seen seated in a real space, and more members are seen through an extended portal. AudienceMR can provide a sense of the presence of a large-scale crowd/audience with the associated spatial context. In contrast to virtual reality (VR), however, with mixed reality (MR), a lecturer can deliver content or conduct a performance from a real, actual, comfortable, and familiar local space, while interacting directly with real nearby objects, such as a desk, podium, educational props, instruments, and office materials. Such a design will elicit a realistic user experience closer to an actual classroom, which is currently prohibitive owing to the COVID-19 pandemic. This paper validated our hypothesis by conducting a comparative experiment assessing the lecturer's experience with two independent variables: (1) an online classroom platform type, i.e., a 2D desktop video teleconference, a 2D video screen grid in VR, 3D VR, and AudienceMR, and (2) a student depiction, i.e., a 2D upper-body video screen and a 3D full-body avatar. Our experiment validated that AudienceMR exhibits a level of anxiety and fear of public speaking closer to that of a real classroom situation, and a higher social and spatial presence than 2D video grid-based solutions and even 3D VR. Compared to 3D VR, AudienceMR offers a more natural and easily usable real object-based interaction. Most subjects preferred AudienceMR over the alternatives despite the nuisance of having to wear a video seethrough headset. Such qualities will result in information conveyance and an educational efficacy comparable to those of a real classroom, and better than those achieved through popular 2D desktop teleconferencing or immersive 3D VR solutions.

Keywords: mixed reality; augmented reality; human-computer interaction; public speaking; social presence

1. Introduction

Since the arrival of COVID-19, classroom and conference activities via video teleconferencing (such as Zoom [1]) have become the unwanted norm (see Figure 1). Many complaints, levels of dissatisfaction, and concerns have arisen with respect to the diminished educational and communication efficacy of both students and lecturers [2] Aside from such difficulties possibly emanating from the difference between an actual classroom and a 2D online setting, another potentially problematic aspect may have to do with the audience size. A 2D oriented and video grid type of communication inherently limits the number of life-sized (or reasonably large) depictions of audiences that are visible at one time. For large audiences, one typically has to browse through pages of a grid of 25 to 30 people. Ironically, online lectures seem to allow for or even encourage the class size to become relatively large because of the absence of physical limitations (e.g., classroom size). Small intimate lectures offline or online will always be preferred in terms of conveying information and the educational effect. However, large scale lectures might be simply a fact



Citation: Han, B.; Kim, G.-J. AudienceMR: Extending the Local Space for Large-Scale Audience with Mixed Reality for Enhanced Remote Lecturer Experience. *Appl. Sci.* 2021, *11*, 9022. https://doi.org/10.3390/ app11199022

Academic Editor: Enrico Vezzetti

Received: 3 September 2021 Accepted: 23 September 2021 Published: 28 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/). of life, and sometimes a large-scale audience and crowd might be desirable, as in concerts or sporting events.

This study focuses on the perspective of a lecturer (or performer) whose information delivery (or performance) capability can be negatively affected through the use of 2D oriented video teleconferencing. Thus, our question is that how one can provide the best experience for the lecturer (or the performer)? For instance, our own pilot survey with lecturers on the use of Zoom [1] brought about major complaints such as an inability to capture the general reception of, and properly empathize with, the audience, and adjust one's pace or accordingly change the content. It may also affect one's general attitude, resulting in a certain detachment, indifference to classroom situations, and less zealousness. Such problems can worsen with a large audience. For stage performances, the performers will certainly prefer the presence of a real world, such as a large crowd versus a limited video screen grid (see Figure 2).



Figure 1. (a) Realspace (b) portal identification algorithm and portal scores (c) extended lecturer space (d) the use of the video see-through headset.



Figure 2. Video screen grid for remote crowd on stage [3].

What might be lacking in 2D oriented video teleconferencing for lecturers in making the information convey as effective and usable as possible? One viable and emerging alternative is the use of an immersive 3D virtual reality (VR) environment, which tries to mimic an actual classroom in a realistic manner (see Figure 3). With VR, it will be more feasible to present a large crowd with a high level of spatial immersion and co-presence. In fact, commercially available solutions exist, such as in [4–6]. This article proposes going one step further and using mixed reality (MR).



Figure 3. Various platforms for online lecture/performance: (**a**) 2D video teleconference on the (**b**) desktop, (**c**) 2D video grid in VR, (**d**) 3D VR with video screen avatars and (**e**) with 3D full-body avatars (**f**) the use of headset/controller, and AudienceMR with (**g**) video screen avatars and (**h**) 3D full-body avatars; and (**i**) the use of the video see-through headset.

Our system is AudienceMR, a multi-user mixed reality space that seamlessly extends the local user space (such as one's office) to become a large shared 3D classroom or lecture hall where some of the audience (real or virtual) members are seen seated in the real space, and more through the extended portal (see Figure 3, bottom row). Like VR, AudienceMR is distinguished from a typical teleconferencing solution in which the lecturer will look at a small and flat grid of facial videos. AudienceMR can provide a sense of presence of a large-scale crowd/audience with the associated spatial context. In contrast to VR, however, with MR, the learner can deliver content or a performance from a real, actual, comfortable, and familiar local space (versus a virtual space), while interacting directly with real nearby objects, such as a desk, podium, educational prop, instrument, and/or office material. Such a design will elicit a different and more realistic user experience from a similarly designed virtual environment, possibly closer to that of an actual classroom experience (which is currently prohibited owing to the COVID-19 pandemic). In particular, the interaction will be more natural with the AudienceMR in comparison to a VR counterpart in the sense that the user can make use of objects as achieved in the real world, for example, looking at paper printouts, interacting with devices/computers (e.g., managing presentations, playing videos, or making annotations), and making bodily gestures, which would mostly be accomplished indirectly using a controller in VR. This article expects these distinguishing qualities to bring about an improved lecturer/performer experience, leading to a more effective delivery of content, lecturer satisfaction, and ultimate benefits for the students/audience.

This article validated this hypothesis by conducting a comparative experiment on the lecturer experience using two independent variables: (1) an online classroom platform

type, i.e., 2D Zoom on a desktop, 2D Zoom in VR, a 3D VR classroom, and AudienceMR, and (2) student depiction, i.e., a 2D upper-body video screen and 3D full-body avatar. Note that this study only investigates the experience of the content deliverer (i.e., lecturer) rather than that of the students/audience.

2. Related Works

Video teleconferencing has emerged as the most popular and convenient meeting/ classroom/conference solution for coping with the current COVID-19 pandemic [7,8]. At the same time, significant concern has arisen with regard to the lack of human "contact" [9–13]. It is generally believed that video conferencing might suffice for one-way lecture-oriented education but falls short for areas or situations where more intimate teacher–student interaction and cooperation are needed [13,14]. In any case, most of these concerns are taken from the student's side, for example, whether students are able to effectively follow and absorb the content of the materials delivered through online teaching. To the best of our knowledge, virtually no studies have seriously looked at the content on the deliverer's side. Our pilot survey with 24 professors revealed that more than 90% of them preferred an offline class mainly because they were unable to closely observe the individual students or interact with the entire class concurrently and felt distant from the students. Lecturers may be unintentionally inclined or drawn to neglect struggling students because of such disconnection.

For this reason (and others), 3D immersive VR solutions have started to appear on the market and are gaining in popularity [15]. In addition, 3D environments such as AR and VR have already been shown to exhibit superior information for at least certain areas, such as those requiring 3D spatial reasoning [16], close interaction and collaboration [17], and information spatially registered to the target objects [18]. Although it is true that the need to wear a headset and other limitations (e.g., close and natural interactions) are still significant obstacles to the wider adoption of 3D immersive VR, the younger generations seem to have much less resistance to such usability issues, exemplified by the recent explosion in the usage of VR-based "metaverses" and improved content choice and quality [4,15]. The same applies to AR in terms of continuously improving form factors of new AR glasses [19]. Some recent studies have compared the advantages and disadvantages of 2D video teleconferencing and 3D immersive VR classrooms/meetings [20]. As the general findings, whereas the former offers a high usability and accessibility and is mostly suitable for one-way oriented lectures and content delivery, the latter is deemed better for closer interactive meetings.

Telepresence AR is another option for remote encounters [21–23], offering an extremely high level of realism and natural interaction, but generally unable to host large crowds unless the local venue is sufficiently large. The idea of a spatial extension in this context appeared as early as in the "Office of the Future", where a remote space was merged into and extended off the local space [24]. AudienceMR attempts to combine these advantages.

One area related to this issue is dealing with the fear of public speaking or stage fright for which VR is often applied with good results owing to its ability to recreate the pertinent situation in a convenient and economic fashion [25–28]. Whereas excessive fear might be undesirable, an appropriate or natural level of nervousness or tension might equate to the excitement of facing a large crowd; thus, such a quality can be used as a measure of the closeness of the virtual classroom to its offline counterpart [27]. In this study, two measurement were used. The State-Trait Anxiety Inventory (STAI) [29] and Personal Report of Communication Apprehension (PRCA) [30] scores, which are the major questionnaires used to measure the general level of anxiety and anxiety regarding public speaking, respectively.

Finally, several studies have investigated the effects of the form of the avatars or participants in multi-user telepresence systems in terms of the level of immersion, preference, and efficacy of the collaboration and information conveyance [31,32]. Comparisons are often made in terms of realism versus an animation-like depiction, whole body ver-

sus partial body, and inclusion/exclusion of certain body parts (and associated bodily expressions) [33,34].

3. AudienceMR Design and Implementation

In designing the AudienceMR, this system particularly considered the following factors (and attempted to reflect them as much), some of which were found through the aforementioned pilot survey with the lecturers at our university (see Table 1).

Table 1. Aspects considered in the design of AudienceMR including feedback from the pilot survey with educators on promoting a fluid, natural, and effective content delivery.

Category	Comments
Realism	Appearance of the participants/space, 3Dness, observable features, audience behavior (e.g., clapping , nodding), life-sized students
Presence	Co-presence with students, self-presence in the classroom, proper level of stage tension, Large scale crowd
Spatiality	Spatial immersion, 3D sound (reverb effect), seamless extension
Interaction	Natural interaction, interaction with students, interaction with objects/props
Familiarity	Deliver lecture from own space
Others	Convenience, light headset, wireless sensors

The overall system design is illustrated in Figure 4. AudienceMR is designed as a mixed reality system (see Section 3.2 for a discussion of other alternatives). The lecturer/performer is situated in any local space, which is 3D scanned offline. For implement this system, Zed mini depth camera and 3D reconstruction software [35] were used for this purpose. The local space scanned as a point cloud is then noise filtered and converted into a mesh and used as the underlying 3D geometric information in the MR scene. The lecturer designates a particular position from which to deliver one's lecture or provide a performance (e.g., at the lecturer's own desk in an office). A region-growth based algorithm (see Section 3.1) is applied to secure a user interactable area beyond which candidate portals are identified and placed for a special extension. In the case of portals that are placed directly on the floor or below the ceiling, the same textures are used to make the extended space look continuous and seamless (see Figure 3, bottom row). As the space allows, a few real students can sit in, or a remote telepresence user (seen as a video screen or 3D avatar) can be augmented into the space. A larger virtual space is created off the portal to situate the rest of the large audience (see Figure 3). This combined mixed-reality space is shared between the lecturer and audience.

As can be seen in Figure 3, students are either depicted using individual video screens (mainly showing their upper body) or as whole-body life-sized 3D avatars (although any size is possible). The audience was simulated in this study. For 2D video-based participants, pre-recorded videos of students listening to lectures were used. When capturing the videos, the actors were asked to exhibit various random but natural behaviors, such as turning or nodding their head, and making facial expressions or small bodily gestures. For the 3D avatars, their faces were textured with the images of the students and moved with pre-programmed behaviors similar to the instructions given in the video capture. During the experiment, the subject was unaware that the audience was pre-recorded or simulated.

Although not implemented at this time, the staff and audience in remote locations are envisioned to have a camera for video capture (as used in regular video teleconferencing) or as a means to extract bodily features (e.g., facial texture/expression, skeletal pose, and limb movements) for use in a simple avatar recreation. The student's system will be able to share the same space and see other students/lecturers and other shareable materials (such as lecture slides). The lecturer uses a video see-through headset (Samsung Odyssey+ [36]) with a Zed mini camera mounted on it to view and augment the shared lecture space with the underlying (pre-reconstructed) 3D geometry information. Although markers are currently used to place a few telepresence audience members within the MR space, in the future, a more sophisticated space analysis can be conducted to automate this process (i.e., identify sufficiently large and flat places to place life-sized telepresence avatars). A large projection screen mounted with a depth camera looking into the environment (in front of the lecturer) is another possible display that will eliminate the cumbersome wearing of a headset (assuming that the lecturer does not move around).

This mixed reality setup allows the lecturer to freely interact naturally with physical objects in an interactable space. That is, the activities of such lecturers can be captured as a whole and broadcast to the audience. For instance, the lecturer's whole body can be seen as is while speaking and holding props or objects, reaching for the computer, conducting slide presentations, or starting the video playback, among other instances. If needed, a controller can be used to interact with augmented objects, or an audience located at a far distance.

The overall content/software was implemented using Unity 2019.4.3 [37], running on a PC (with an Intel i9 Core CPU [38] and a NVidia RTX 2080 graphics card [39]), with a performance of greater than 40 fps. Most of the features listed in Table 1 are addressed.



Figure 4. The overall system configuration of AudienceMR.

3.1. Portal Placement and Evaluation

This section provide a more detailed description of the portal placement algorithm. Once the lecturer position is designated (manually), using the 3D geometric information obtained from the offline scanning, a region growth is applied in the planar dimension toward the walls or large vertical objects (against which the portals will be put up) to an extent that considers the sizes of the objects and obstacles in the way. Portals by default are rectangular and extend from the floor to the ceiling and are adjusted depending on the obstacles. For instance, a portal extending upward from the top of a low bookshelf or one extending downward from the bottom of a high picture frame is possible. Portals pasted onto the floor or ceiling are not considered because their visibility, as well as the fatigue from having to look up or down, is considered undesirable. Portal candidates identified in this way are ranked through a simple scoring scheme that considers several factors, such as the size, visibility, and distance, to help the lecturer select a few. Figure 5 illustrates the various possible portal placements ranked and chosen by the user.

$$S_{d_i} = \frac{\|p_{p_i} - p_l\|}{Max(\|p_{p_{0,i}} - p_l\|)}$$
(1)

$$S_{v_i} = \left\| \cos(\theta_{pl}) \right\| \tag{2}$$

$$S_{s_i} = \frac{W_i * H_i}{Max(W_{0..i} * H_{0..i})}$$
(3)

$$S_{p_i} = w_d S_{d_i} + w_v S_{v_i} + w_s S_{s_i}$$
(4)

where S_d , S_v , S_s , p_p , θ_{pl} , W, H, and w refer to the scores by distance, visibility, size, positions of the portal and lecturer, angle between the portal and lecturer, width and height of the portal and weights among the three sub-scores. The subscript i refers to the *i*th portal. The scoring scheme favors the portals that are (1) closer and oriented toward the user and (2) large with user-adjustable weights.



Figure 5. Four different portal selections in various local spaces.

3.2. Other Alternatives

Although the features in Table 1 have played the most important role in our design of AudienceMR, this study also considered other possibilities (excluding the 2D desktop

video conferencing, which is the conventional baseline case), some of which also serve as our comparison groups in our validation experiment. The first was a 3D immersive VR. As already discussed, VR can satisfy many of the attractive features listed in Table 1, except for realistic and natural interactions. However, it should be noted, as an example, a controller-based interaction (the most prevalent form of interaction method in VR) can still be sufficiently convenient for many interactive tasks, and special sensors can be employed to realize a more natural interaction.

Despite the feedback from our preliminary survey, it is not obvious if making all objects into 3D is necessary. In fact, remote crowd solutions exist for the performance stages based on setting up a grid of video screens [3] (see Figure 2). As such, we consider emulating such a scheme in VR (dubbed as "Zoom VR," as indicated in Figure 3). Although limited in the number of crowds it can accommodate based on its 2D nature, it can still provide the spatial presence of a classroom or stage with a realistic appearance of the audience. The interactions are still less natural for a VR-based solution, however.

One way to improve the spatial realism and fluidity of the interactions might be to consider the AR-based tele-presence environment. Most AR-based shared spaces are confined to the local space itself, which limits the number of remote participants present owing to spatial limitations. One possibility is to use miniature avatars [40], which we believe will not provide a sufficient co-presence or realism or apply a large empty hall as the local space (and augment the audience to all seats), which is too costly or impractical. A classroom or performance hall with a full audience televised to the remote lecturer would be incompatible with the COVID-19 pandemic. AudienceMR subsumes the AR tele-presence approach. Table 2 summarizes the relative characteristics of the major environmental types considered.

Env. Types	Realism	Presence/Audience Presentable at a Time	Spatiality	Interaction	Usability
Desktop 2D Zoom	High	Low/Small	Low	As is	Good
ZoomVR	High	Medium/Medium	Medium	Low	Low
3D VR Immersive	Medium	High/Large	High	Low	Low
Tele-presence AR	High	High/Small	High	Medium	Low
Audience MR	Medium	High/Large	High	Medium	Low

Table 2. Possible environment types to achieve presence of large audience and crowd and their characteristics. Environment types written in bold are compared in the validation experiment.

4. Validation Experiment (Teachers Experience in Large Crowd Online Lecture Delivery

4.1. Experimental Design

The validation experiment was conducted to compare and assess the related lecture delivery experience among four different major remote lecture platforms: (1) Desktop 2D Zoom like teleconferencing (DTZ), (2) Zoom-like grid in VR (ZVR), (3) 3D immersive VR (3DVR), and (4) audience MR (AMR) (see also Figure 3). Another independent variable we considered was the form of the audience, represented as (1) a video screen (realistic) showing the upper body (indicated as -V) and a (2) 3D full-body (less realistic) avatar (indicated as -A), whose face was decaled with that of the user's picture and exhibited some rudimentary behavioral features (e.g., random nodding, gestural and body movements, eye blinks, head turns, and nodding). Note that both avatar representations were approximately life-sized. Thus, the experiment was designed as a two factor (with each factor having four and two levels) within-subject repeated measure. There was a total of six conditions for comparison because the DTZ and ZVR could not operate with 3D avatars. Table 3 lists

the experimental conditions and their labels. Note that we also measured the dependent variables for the offline classroom situation as the baseline.

Env. Types Avatar	Desktop 2D Zoom	Zoom VR	3D VR	AudienceMR	_
Video screen Upper body	DTZ-V	ZVR-V	3DVR-V	AMR-V	Offline Classroom
3D Avatar Full body	N/A	N/A	3DVR-A	AMR-A	-

Table 3. Experimental conditions and labels.

The experimental task involved a delivery of a short lecture on topics familiar to one and at the same time also carried out small object-interactive tasks (see next subsection). With respect to the audience, the lecturing task was mostly one-way with the interaction restricted to the objects near-by the lecturer.

For the dependent variables, we measured the anxiety and level of fear of public speaking to compare the level of tension elicited in the different environment types and forms of audience. As previously mentioned, we used the STAI [29] and PRCA [30] questionnaires modified for our specific purpose. We also collected data on social presence [41,42], usability, and general preference. A detailed survey questionnaire is provided in the Appendix A. Note that this study concentrated only on the lecturer's experience. In addition, we emphasize that by the term, "lecturer", we do not mean strictly professional educators, but simply anyone providing a casual presentation or lecture. Because the audience was simulated a future study of its reception, experience, and educational effects has been planned.

As our general hypothesis, the users of AudienceMR will exhibit anxiety and fear of public speaking at a level similar to the offline classroom lecture and more so than in other environments, particularly compared to Desktop 2D Zoom or Zoom VR. We also expect Desktop 2D Zoom to significantly decrease the spatial immersion and co-presence elicited, which we think will be correlated to the level of tension in front of the audience. Between 3D VR and AudienceMR, the subjects are expected to prefer the latter in terms of interaction usability. Finally, between the video of the upper-body and 3D full-body avatars, we hypothesize that video representations will be more effective owing to their realism, despite only half of the body being shown. Because the current experimental task is mostly one-way lecturing, and because few intimate interactions are needed, the full-body representation will have a smaller effect, at least for the given experimental task.

4.2. Experimental Task

As indicated, the experimental task involved the delivery of short lectures on a familiar topic while conducting small object-interactive tasks. Several topics of interest generally familiar to Korean college level students (e.g., famous landmarks and movies in Korea) were prepared in advance on five or six slides (designed to take 5-6 min to explain). The subjects were to look over the slide materials and choose the one that they felt most comfortable explaining to the audience. They were also given sufficient time to study the slides (among the six shown in Table 3) before delivering the lecture under the given test condition. While giving the lecture, the slides were shown behind the lecturer, as would occur in a real classroom. The subjects were also asked to carry out small interactive subtasks: (1) controlling a slide presentation (e.g., flipping to the next page and making a simple annotation), (2) operating a video, and (3) holding a 3D model (real or virtual depending on the test condition) and explaining or pointing to it. Figure 3 shows the subject carrying out the task under six test conditions. Depending on the test conditions, the interaction was carried out in the real world using real objects placed in the subject's hands (DTZ, AMR) or virtual objects using a controller (ZVR, 3DVR). With respect to the audience, the lecture was mostly one-way, with the interaction restricted to the objects near the lecturer. The participants were unaware that the audience was simulated and actually made to believe that they were connected online.

4.3. Experimental Set-Up

Six test environments were prepared. For DTZ-V, a Zoom-like desktop application was developed that showed pages of a video grid and shared screen (e.g., to show a PowerPoint or video presentation). The screen can show up to 25 people at a time. To flip through the video grid pages (and look over the entire audience), the flip back-and-forth button had to be pressed using a mouse. The videos were pre-recorded, and during the recording, the recorded person was instructed to make random gestures such as nodding, shaking their head, various facial expressions, and small body movements as if listening to a live lecture.

ZoomVR and 3D VR test beds were developed using Unity as an immersive 3D environment [37]. With Zoom VR, a large 2D video grid (5×10 , with each video screen scaled to life size) was set up, spanning approximately 150 degrees of the visual field, with an appropriate distance from the virtual camera point. The 3D VR was similar except for the 2D grid, and a lecture hall like environment was used in which the individual rectangular video screens or 3D avatars were seated. In both environments, as the lecturer was wearing the headset, the slides were shown not only behind the lecturer for the audience to view, but also in front (similar to a tele-prompter), allowing the lecturer to easily explain the given material. The interface for slide control was also available, which could be enacted through a ray-casting selection with the controller.

The simulated life-sized (and with random variations) 3D avatars were modeled using a body with general characteristics, and the faces were textured with images of an audience and programmed to exhibit some rudimentary behavioral features (e.g., random nodding, gestural and body movements, eye blinks, head turns, and nodding). The implementation details of AudienceMR (AMR test condition) are provided in Section 3.

4.4. Experimental Procedure

The experiments were started by collecting the basic subject information and familiarizing them with the six test platforms, the main lecture task, and interaction subtasks. A total of 26 participants (mean age of 27.0, SD = 6.99) recruited from a university participated in the study. All had prior experience in giving presentations in front of a large crowd of various sizes. All but three subjects also had prior experience using VR and a headset. Before the main experiment session, the subjects were first surveyed for their anxiety and fear based on their prior experiences in making offline presentations in front of a large crowd, which served as the baseline measure. A survey-based measurement (instead of an actual experiment) was taken because making a presentation was considered a sufficiently common and frequent activity for the subject pool, particularly in light of the current COVID-19 situation.

In the main experimental session, the user experienced all six different platform treatments, each of which was presented in a balanced order. For each treatment, the user picked different topics for their short lecture and studied the pre-prepared slides for 10 min ahead of time. The whole session took approximately 1.5 h. After each session, the users filled out a user experience survey that assessed their anxiety and fear, social presence, and general usability. A detailed survey is presented in the Appendix A.

5. Results

The survey scores were category-based (STAI, PRCA, social presence, and interactional usability) and summed for a statistical analysis. A one-way ANOVA with Tukey's HSD pairwise comparison was applied to analyze the effects of the environment type on the various dependent variables. A detailed statistical analysis is provided in the Appendix B. For the anxiety and fear of public speaking, statistically significant effects were found (F = 7.581; p < 0.001), (F = 9.312; p < 0.001). DTZ showed the lowest levels of anxiety and fear compared to the baseline offline lecture (p < 0.001) (p < 0.001) and also to AMR (p = 0.001) (p < 0.001). ZVR also showed a lower level of anxiety by STAI and PRCA (p = 0.047) (p = 0.049). In fact, we observe that DTZ and ZVR exhibited a lower level

of anxiety and fear than the Offline and AMR. In summary, the general trend was that compared to the Offline baseline, all cases (which were online conditions) exhibited lower or nearly equal levels of anxiety and fear on the part of the lecturer. AMR and 3DVR achieved the levels near to the baseline without statistically significant differences, while DTZ and ZVR were lower. (see Figure 6a,b).



Figure 6. The effects of environment type on the (**a**) levels of anxiety, (**b**) fear of public speaking and (**c**) social presence and (**d**) usability. The * mark indicate statistically significant effect.

Figure 5c shows the effects of the environment type to the level of the social presence. Again similarly to the trend for the anxiety level, both DTZ (p < 0.001) and ZVR (p = 0.02) showed relatively lower social presence compared to AMR. No statistically significant difference was found between 3DVR and AMR however.

As for the general usability (see Figure 6d), the DTZ was considered the most usable, an obvious choice as it does not require the cumbersome usage of the headset. When asked for the interactional usability, AMR was on par with the Desktop Zoom with statistically significant differences with the ZVR (p < 0.001) or 3DVR (p < 0.001). Interactional usability refers to the usability confined to delivering the lectures and carrying out the small subtasks, and excluding the effects of the headset donning.

Two-way ANOVA with Tukey's HSD pairwise comparison was applied to analyze for the effects of the avatar type to the various dependent variables. For all dependent variables, no statistically significant effects were found among the comparison groups (see Figure 7). The preference for video and 3D avatars were also mostly split (Figure 8). Despite many researches having demonstrated different perceptual effects by the form of the avatar and its behavioral depiction [31–34] our experiment did not indicate anything in particular. We believe that this was because the audience was only simulated with rather random behaviors and the lecturer's task was mostly one-way, non-intimate and non-interactive. The look of the audience alone was not a significant factor in the lecturer's experience.



Figure 7. Effects of avatar type on the (**a**) levels of anxiety, (**b**) fear of public speaking, and (**c**) social presence.

General comments from the participants were useful in interpreting the results. One of the foremost comments was how the relatively higher "spatial" realism and natural interaction in AudienceMR contributed to the enhanced and direct lecturer experience, particularly in reducing the feeling of detachment that can occur through 2D teleconferencing or in less realistically modeled VR environments. We posit that this aspect influenced the arousal of anxiety level, which was similar to that in an actual classroom.



Figure 8. Preference based on the environment type (left) and avatar type (right).

6. Discussion

In this paper, we presented AudienceMR, a shared mixed reality environment that extends the local space to host a large audience and provides the proper level of tension and spatial presence to the lecturer. The results of our experimental lecturer UX, in comparison to other alternatives, such as popular 2D desktop teleconferencing and immersive 3D VR solutions, mostly validated our projected objectives. That is, AudienceMR exhibited a level of anxiety and fear of public speaking closer to that of real classroom situation, as well as higher social and spatial presence than 2D video-grid-based solutions and even 3D VR. Compared to 3D VR, AudienceMR offers a more natural and easy-to-use real object-based interaction. Most of the subjects participating in our validation experiment preferred AudienceMR over the alternatives despite the nuisance of having to wear the video see-through headset (see Figure 8). Although not verified in this paper, our ultimate expectations and hope are that such qualities would result in the conveyance of information and an educational efficacy comparable to those of a real classroom, better than other popular 2D desktop teleconferencing or immersive 3D VR solutions. However, it can be argued that 2D desktop video teleconferencing has an advantage of incurring less tension on the part of the lecturer.

Our study has several limitations. First, focus was solely on the experiences from the lecturer's point of view. The audience was simulated using typical but random classroom behaviors. As indicated above, the final effect on conveying information and educational efficacy could not be measured. Future implementation with an actual audience would allow such an investigation and a more intimate two-way lecture-audience interaction. We are also interested in applying AudienceMR to live performances (see Figure 9) and measure how it can mimic a real stage with audiences at the scale of tens of thousands. There is much room to integrate AI to make the MR-based classroom and education more intelligent and easy to use [43,44]. This remains a possible future extension to AudienceMR.



Figure 9. AudienceMR for stage performance.

Author Contributions: B.H. conducted the study and analysis. G.J.K. supervised all activities. Both authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the IITP/MSIT of Korea under the ITRC supportprogram (IITP-2021-2016-0-00312).

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to the fact that all participants were adults and participated by their own will.

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

Data Availability Statement: Not applicable.

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

Appendix A. Questionnaire

STAI (State-Trait Anxiety Inventory)

- **S1** I feel calm.
- S2 I feel tense.
- S3 I feel at ease.
- **S4** I feel uncomfortable.
- **S5** I feel self confident.
- **S6** I feel nervous.
- **S7** I feel indecisive.
- **S8** I am relaxed.
- **S9** I am worried.
- S10 I feel pleasant.

PRCA (Personal Report of Communication Apprehension)

PR1 I have no fear of giving a speech.

- PR2 Certain parts of my body feel very tense and rigid while giving a speech.
- **PR3** I feel relaxed while giving a speech.
- PR4 My thoughts become confused and jumbled when I am giving a speech.
- **PR5** I face the prospect of giving a speech with confidence.
- **PR6** While giving a speech, I get so nervous I forget facts I really know.

Social Presence

- SP1 The people's behavior influenced my style of presentation.
- **SP2** The people's behavior had an influence on my mood.
- SP3 I reacted to the people's behavior.
- **SP4** I was easily distracted by the people.
- **SP5** Sometimes the people were influenced by my mood.
- SP6 Sometimes the people were influenced by my style of presentation.
- **SP7** The people reacted to my actions.
- **SP8** I was able to interpret the people's reactions.
- **SP9** I had the feeling of interacting with other human beings.
- **SP10** I felt connected to the other people.
- SP11 I had the feeling that I was able to interact with people in the virtual room.
- SP12 I had the impression that the audience noticed me in the virtual room.
- SP13 I was aware that other people were with me in the virtual room.
- **SP14** I had the feeling that I perceived other people in the virtual room.
- SP15 I felt alone in the virtual environment.
- **SP16** I feel like I am in the same space as the students.
- SP17 I can conduct the class similar to the actual offline class.

Usability

- **U1** I can easily use this system.
- **U2** It is easy to learn how to use the system.
- **U3** I feel tired while using this system.

Preference

PF1 I prefer to use this system.

PF2 Overall, I am satisfied with this system.

Appendix B. Statistical Analysis Table

Table A1. ANO	VA test resu	lt.
---------------	--------------	-----

Measure	F	Sig.
STAI	7.581	< 0.001
PRCA	9.312	< 0.001
Social Presence	14.259	< 0.001

Table A2. STAI post hoc—Tukey HSD (Statistically significance levels: * p < 0.05; ** p < 0.01; *** p < 0.001).

	Offline	DTZ-V	ZVR-V	3DVR-V	AMR-V
Offline	N/A	< 0.001 ***	0.005 **	0.054	0.941
DTZ-V	< 0.001 ***	N/A	0.785	0.295	0.001 **
ZVR-V	0.005 **	0.785	N/A	0.926	0.047 *
3DVR-V	0.054	0.295	0.926	N/A	0.282
AMR-V	0.941	0.001 **	0.047 *	0.282	N/A

Table A3. PRCA post hoc—Tukey HSD (Statistically significance levels: * p < 0.05; ** p < 0.01; *** p < 0.001).

	Offline	DTZ-V	ZVR-V	3DVR-V	AMR-V
Offline	N/A	< 0.001 ***	0.002 **	0.059	0.859
DTZ-V	< 0.001 *	N/A	0.476	0.059	< 0.001 ***
ZVR-V	0.002 **	0.476	N/A	0.823	0.049 *
3DVR-V	0.059	0.059	0.926	0.823	0.429
AMR-V	0.859	< 0.001 ***	0.049 *	0.429	N/A

Table A4. Social Presence post hoc—Tukey HSD (Statiscally significance levels: * p < 0.05; ** p < 0.01; *** p < 0.001).

	DTZ-V	ZVR-V	3DVR-V	AMR-V
DTZ-V	N/A	0.23	0.002 **	< 0.001 ***
ZVR-V	0.230	N/A	0.290	< 0.001 ***
3DVR-V	0.059	0.059	N/A	0.823
AMR-V	0.859	< 0.001 ***	0.049 *	N/A

References

- 1. Zoom. Available online: https://zoom.us/ (accessed on 31 August 2021).
- 2. Nambiar, D. The impact of online learning during COVID-19: Students' and teachers' perspective. *Int. J. Indian Psychol.* 2020, *8*, 783–793.
- 3. Bizwire, T.K. Online Concerts. Available online: http://koreabizwire.com/k-pop-innovates-with-online-concerts-amidst-coronavirus-woes/158714 (accessed on 31 August 2021).
- 4. Fidelity, H. High Fidelity. Available online: https://www.highfidelity.com/ (accessed on 31 August 2021).
- 5. Research, L. Sansar. Available online: https://www.sansar.com/ (accessed on 31 August 2021).
- 6. Mozilla Hubs. Available online: https://hubs.mozilla.com/ (accessed on 31 August 2021).
- Dietrich, N.; Kentheswaran, K.; Ahmadi, A.; Teychené, J.; Bessière, Y.; Alfenore, S.; Laborie, S.; Bastoul, D.; Loubière, K.; Guigui, C.; et al. Attempts, successes, and failures of distance learning in the time of COVID-19. *J. Chem. Educ.* 2020, 97, 2448–2457. [CrossRef]
- Klein, P.; Ivanjek, L.; Dahlkemper, M.N.; Jeličić, K.; Geyer, M.A.; Küchemann, S.; Susac, A. Studying physics during the COVID-19 pandemic: Student assessments of learning achievement, perceived effectiveness of online recitations, and online laboratories. *Phys. Rev. Phys. Educ. Res.* 2021, 17, 010117. [CrossRef]
- 9. Alqurashi, E. Predicting student satisfaction and perceived learning within online learning environments. *Distance Educ.* 2019, 40, 133–148. [CrossRef]

- Arora, A.K.; Srinivasan, R. Impact of pandemic COVID-19 on the teaching–learning process: A study of higher education teachers. *Prabandhan Indian J. Manag.* 2020, 13, 43–56. [CrossRef] [PubMed]
- 11. Henriksen, D.; Creely, E.; Henderson, M. Folk pedagogies for teacher transitions: Approaches to synchronous online learning in the wake of COVID-19. *J. Technol. Teach. Educ.* **2020**, *28*, 201–209.
- 12. Kuo, Y.C.; Walker, A.E.; Schroder, K.E.; Belland, B.R. Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *Internet High. Educ.* **2014**, *20*, 35–50. [CrossRef]
- 13. The Washington Post. Zoom Classes Felt like Teaching into a Void—Until I Told My Students Why. *The Washington Post*, 11 March 2021.
- 14. Markova, T.; Glazkova, I.; Zaborova, E. Quality issues of online distance learning. *Procedia-Soc. Behav. Sci.* 2017, 237, 685–691. [CrossRef]
- 15. vTime: Facebook's XR based Social App. 2020. Available online: https://www.facebook.com/vTimeNet/ (accessed on 31 August 2021).
- 16. HTC Vive. Available online: https://www.vive.com/ (accessed on 31 August 2021).
- 17. Popovici, D.M.; Serbcanati, L.D. Using mixed VR/AR technologies in education. In Proceedings of the IEEE 2006 International Multi-Conference on Computing in the Global Information Technology-(ICCGI'06), Bucharest, Romania, 1–3 August 2006; p. 14.
- 18. IKEA. IKEAPlace. Available online: https://apps.apple.com/us/app/ikea-place/id1279244498 (accessed on 31 August 2021).
- 19. Microsoft. HoloLens 2. Available online: https://www.microsoft.com/en-us/hololens/ (accessed on 31 August 2021).
- Campbell, A.G.; Holz, T.; Cosgrove, J.; Harlick, M.; O'Sullivan, T. Uses of virtual reality for communication in financial services: A case study on comparing different telepresence interfaces: Virtual reality compared to video conferencing. In Proceedings of the Future of Information and Communication Conference, San Francisco, CA, USA, 14–15 March 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 463–481.
- 21. Leap, M. Magic Leap Social. Available online: https://www.magicleap.com/en-us/news/product-updates/connect-with-friends-with-avatar-chat (accessed on 31 August 2021).
- 22. Mixtive. AR Call. Available online: http://arcall.com/ (accessed on 31 August 2021).
- 23. Systems System Spatial. Available online: https://spatial.io/ (accessed on 31 August 2021).
- 24. Raskar, R.; Welch, G.; Cutts, M.; Lake, A.; Stesin, L.; Fuchs, H. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, USA, 19–24 July 1998; pp. 179–188.
- 25. Harris, S.R.; Kemmerling, R.L.; North, M.M. Brief virtual reality therapy for public speaking anxiety. *Cyberpsychol. Behav.* 2002, *5*, 543–550. [CrossRef] [PubMed]
- 26. Parsons, T.D.; Rizzo, A.A. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *J. Behav. Ther. Exp. Psychiatry* **2008**, *39*, 250–261. [CrossRef] [PubMed]
- Price, M.; Anderson, P.L. Outcome expectancy as a predictor of treatment response in cognitive behavioral therapy for public speaking fears within social anxiety disorder. *Psychotherapy* 2012, 49, 173. [CrossRef] [PubMed]
- 28. Wallach, H.S.; Safir, M.P.; Bar-Zvi, M. Virtual reality cognitive behavior therapy for public speaking anxiety: A randomized clinical trial. *Behav. Modif.* **2009**, *33*, 314–338. [CrossRef] [PubMed]
- 29. Julian, L. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). *Arthritis Care Res.* 2011, *63*, 467–472. [CrossRef] [PubMed]
- 30. McCroskey, J.C.; Beatty, M.J.; Kearney, P.; Plax, T.G. The content validity of the PRCA-24 as a measure of communication apprehension across communication contexts. *Commun. Q.* **1985**, *33*, 165–173. [CrossRef]
- Latoschik, M.E.; Roth, D.; Gall, D.; Achenbach, J.; Waltemate, T.; Botsch, M. The effect of avatar realism in immersive social virtual realities. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden, 8–10 November 2017; pp. 1–10.
- 32. Schreer, O.; Feldmann, I.; Atzpadin, N.; Eisert, P.; Kauff, P.; Belt, H. 3D presence-a system concept for multi-user and multi-party immersive 3D video conferencing. In Proceedings of the 5th European Conference on Visual Media Production (CVMP 2008), London, UK, 26–27 November 2008.
- Latoschik, M.E.; Kern, F.; Stauffert, J.P.; Bartl, A.; Botsch, M.; Lugrin, J.L. Not alone here?! Scalability and user experience of embodied ambient crowds in distributed social virtual reality. *IEEE Trans. Vis. Comput. Graph.* 2019, 25, 2134–2144. [CrossRef] [PubMed]
- Yoon, B.; Kim, H.I.; Lee, G.A.; Billinghurst, M.; Woo, W. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 547–556.
- 35. Stereolabs. ZED SDK. Available online: https://www.stereolabs.com/developers/ (accessed on 31 August 2021).
- Samsung. Hmd Odyssey+ (Mixed Reality). Available online: https://www.samsung.com/us/support/computing/hmd/hmdodyssey/hmd-odyssey-plus-mixed-reality/ (accessed on 31 August 2021).
- 37. Unity. Available online: https://unity.com/ (accessed on 31 August 2021).
- 38. Intel. Intel[®] Core[™] i9 Processors. Available online: https://www.intel.com/content/www/us/en/products/details/ processors/core/i9.html (accessed on 31 August 2021).
- 39. NVIDIA. NVIDIA Graphic Card. Available online: https://www.nvidia.com/en-gb/geforce/20-series/ (accessed on 31 August 2021).

- Piumsomboon, T.; Lee, G.A.; Hart, J.D.; Ens, B.; Lindeman, R.W.; Thomas, B.H.; Billinghurst, M. Mini-me: An adaptive avatar for mixed reality remote collaboration. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13.
- 41. Poeschl, S.; Doering, N. Measuring co-presence and social presence in virtual environments–psychometric construction of a german scale for a fear of public speaking scenario. *Annu. Rev. Cybertherapy Telemed.* **2015**, 58.
- 42. IJsselsteijn, W.A.; De Ridder, H.; Freeman, J.; Avons, S.E. Presence: Concept, determinants, and measurement. In *Human Vision and Electronic Imaging V*; International Society for Optics and Photonics: San Jose, CA, USA, 2000; Volume 3959, pp. 520–529.
- 43. Perrotta, C.; Selwyn, N. Deep learning goes to school: Toward a relational understanding of AI in education. *Learn. Media Technol.* **2020**, *45*, 251–269. [CrossRef]
- 44. Gandedkar, N.H.; Wong, M.T.; Darendeliler, M.A. Role of Virtual Reality (VR), Augmented Reality (AR) and Artificial Intelligence (AI) in tertiary education and research of orthodontics: An insight. In *Seminars in Orthodontics*; Elsevier: Amsterdam, The Netherlands, 2021.