IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 October 2022, accepted 10 November 2022, date of publication 17 November 2022,
date of current version 6 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3222804

RESEARCH ARTICLE

Mouth-in-the-Door: The Effect of a Sound Image of an Avatar Intruding on Personal Space That Deviates in Position From the Visual Image

KEIICHI ZEMPO[1], **(Member, IEEE), AZUSA YAMAZAKI**[2], **NAOTO WAKATSUKI**[1], **KOICHI MIZUTANI**[1], **AND YUKIHIKO OKADA**[1,3]

[1]Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan
[2]Graduate School of Science and Technology, University of Tsukuba, Tsukuba 305-8573, Japan
[3]Center for Artificial Intelligence Research, University of Tsukuba, Tsukuba 305-8577, Japan

Corresponding author: Keiichi Zempo (zempo@iit.tsukuba.ac.jp)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Faculty of Engineering, Information and Systems, University of Tsukuba, under Application No. 2020R427, and performed in line with the Declaration of Helsinki.

ABSTRACT In this paper, we examine the audiovisual experience in virtual reality (VR) service context that enables a more effective interaction between a user immersed in a virtual environment (VE) and an avatar as a store staff. By utilizing the characteristics of VE experiences, we find the effects of this unrealistic relationship between the visual and auditory positions of the avatar presented to the user variable rather than uniformly presented in the same position. In this study, we conducted an experiment to investigate how the positional deviation between the sound and visual images can be tolerated in VE, the effect of positional deviation on the interpersonal distance to the avatar, and the possibility of manipulating the impression of the avatar by deviating the sound image from the visual image. For the experiment, we prepared a space resembling a VE store and conducted proximity experiments with 16 gender-balanced participants and six types of avatars. By utilizing the superiority of visual information over auditory information revealed in the experiments, we constructed an interpersonal situation with an avatar playing the role of store staff in which only the sound image intruded into the user's personal space, and we investigated users' impressions of the avatar. We also investigated users' impressions of the avatar. We found the following two phenomena in the experimental conditions where the positional difference was allowed: 1) Even when the positional difference was allowed, it caused an "uncanny valley"-like phenomenon that led to a decrease in rapport; and 2) In the conditions where the positional difference was allowed when the sound image was closer than the visual image to the participant, the rapport was greater with the avatar playing the role of the store staff. This phenomenon is similar to the "foot-in-the-door" phenomenon in which small unconscious consent (i.e., allowing a sound image to intrude on one's personal space) leads to an improvement in the evaluation of the other person (i.e., the rapport with another person). The techniques proposed in this paper, such as the positional difference between the sound and visual images, significantly improve the value of the service experiences obtained through interaction with others in VE.

INDEX TERMS Interaction, interpersonal service, personal space, service on metaverse, ventriloquism effect.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

I. INTRODUCTION

A. BACKGROUND

Services such as stores and offices in virtual environments (VEs) currently exist, and it is expected that the use of VE and communication based on the metaverse will become mainstream in the future [1]. The virtual reality (VR) market is expected to grow from $5 billion in 2021 to $12 billion by 2024.[1] Meta Platforms, Inc., a leader in the development of the metaverse, is investing more than $1 trillion per year in the development of platforms to realize the metaverse. Microsoft Corporation, the developer of the HoloLens mixed reality (MR) terminal, is developing Mesh for Microsoft Teams, a virtual meeting tool that combines MR and Teams functions [2]. The commoditization of these technologies is progressing, and HIKKY Corporation, which provides VR development engines and organizes VR events, has been holding events called the ''Virtual Market'' in VE since 2018.[2] Major electronics stores, apparel stores, department stores, entertainment stores, and other companies display their products in virtual space, which allows users to enjoy shopping from the comfort of their homes. In this way, there is an increasing number of opportunities for each person to become an avatar in a VE and to interact in meetings and customer service situations that involve physicality and movement [3]. The review of related fields also shows that many efforts are to streamline and customize interactions in the retail store and other service spaces, or workplaces for remote collaborative work, aiming at intelligent data analysis and intervention in real time [4], [5], [6]. Therefore, there is also a need for comfortable communication in VEs.

Personal space is one of the most important factors for a comfortable social life. Reference [7] described personal space as the area around a person where the intrusion of others causes discomfort. In this study, we define ''personal space'' as the area around a person where the intrusion of others causes discomfort, and ''interpersonal distance'' is the distance to the border centered on the person. Personal space is regulated dynamically and includes ''intimate'' (interpersonal distance: 0–0.45 m), ''personal'' (interpersonal distance: 0.45–1.20 m), ''social'' (interpersonal distance: 1.20–3.60 m) and ''public'' (interpersonal distance: >3.60 m) zones [8], which reflect the type of relationship that a person has with others. Since the magnitude of the interpersonal distance that forms personal space varies according to the situation [9], if the environment changes, then the appropriate length of interpersonal distance will also change accordingly [7], [8]. Proxemics, the study of interpersonal distance, was established in the 1950s and 1960s. Since then, research has been conducted on the effects of various factors on interpersonal distance. It has also been shown that personal space exists not only in the real environment (RE) but also in VE [10], [11].

B. OBJECTIVES AND CONTRIBUTIONS

This research proposes a more enriched shopping experience for users by having them interact with avatars in a way that parameterizes the relationship between sound and visual images, which is impossible with RE. In VE, unlike RE, it is possible to easily manipulate various avatar elements, such as appearance and voice. One's impression of others can be manipulated by changing the elements of the avatar in VE. Reference [12] showed that interpersonal distance changes with differences in avatars' facial expressions, and Bailenson et al.: [13] emphasized that interpersonal distance changes with differences in an avatar's eye movements. Furthermore, physical body movements in VE do not necessarily need to obey physical laws. For example, Ahuja et al.: [14] proposed a representation in which intentions are better conveyed by transforming them into emphasized anime-like movements depending on the context. In addition to the appearance of such avatars, it is possible to manipulate their positions. Unlike in RE, it is possible to realize effective person-to-person and person-to-agent interactions that ignore the laws of physics, such as teleporting to the immediate vicinity of the interlocutor when necessary, emphasizing the sound of the target of attention, and gradually becoming transparent.

In this paper, we examine the audiovisual experience in a VR service context that enables more effective interaction between the user immersed in the VE and the avatar as a store staff, as shown in Fig. 1. By utilizing the characteristics of VE experiments, we find the effects of this unrealistic relationship between the visual and auditory positions of the avatar presented to the user variable rather than uniformly presented in the same position.

The novelty of this study is twofold. First, it focuses on customer service in VR stores. The number of VR stores is currently increasing and is expected to expand in the future. Nevertheless, no studies have focused on the procedures of value co-creation with face-to-face customers in a VR space. Second, this study includes the parameterization of the positional relationship between the sound and visual images of the avatar in the VR environment. Although there have been studies that have changed avatars' appearance, voice tone, etc. as an approach unique to VR space that is not possible in physical space, this study is the first to manipulate the positional relationship of its sound image.

The main contributions of our work are as follows:

1) To clarify the differences in the shape of personal space formed in VEs under visual, auditory, or both conditions (in RQ#1);
2) To clarify the effective range of the ventriloquism effect when the position of the visual image and the sound image are different (in RQ#2);

[1] Statista. Forecast augmented (AR) and virtual reality (VR) market sizes worldwide from 2021 to 2024 (in billion U.S. dollars). Retrieved January 25, 2022, from www.statista.com/topics/2532/virtual-reality-vr/

[2] HIKKY Co., Ltd. Virtual Market Official Website. Retrieved January 25, 2022, from www.v-market.work/

3) To propose a higher quality of interpersonal interaction in VEs by utilizing the ventriloquism effect (in RQ#3); and
4) To provide experimental procedures unique to VR that separate the sound image from the visual image since the above efforts have never been made before (in §4).

II. RELATED WORKS

A. VENTRILOQUISM EFFECT

The ventriloquism effect is the illusion that a sound is perceived to originate from the location of a visual target when the two stimuli are presented at different locations [15], [16], [17], [18]. Although some theorize that this is due to the auditory signal being completely captured by the strong visual signal [19], [20], Alais and Burr: [15] showed that this effect can be explained by a simple model of the optimal combination of visual and auditory spatial cues (each mode is weighted by an inverse estimate of its variability). Because the ability to localize stimuli using visual cues is generally less variable than the ability to use only auditory cues, visual information will tend to bias responses to auditory stimuli when there is competition between these modalities. However, when visual stimuli are blurred and more difficult to localize, vision becomes worse than hearing, and conversely, the illusion that sound captures vision occurs.

There have been numerous studies on ventriloquism in azimuth [21], [22], [23], [24], [25], and they all conclude that the effect decreases as the angular difference between the position of the sound and visual stimuli increases. However, the range of thresholds reported in these studies is wide, and it ranges from 3° [21] to 20° [26]. These differences can be attributed to factors such as subject experience, audiovisual time differences, ‘‘persuasiveness’’ factors, and attention.

Ventriloquism effects also exist in the radial direction around the user, e.g., the ‘‘proximity imagery effect’’ [27], [28], [29], in which auditory stimuli are perceptually integrated with visual objects that are closer than the auditory target [30]. There is an asymmetry in the strength of this effect: if the visual target is farther away than the auditory target, then audiovisual unification fails more often [28].

Only a few studies have examined the effects of ventriloquism in VE. Reference [31] investigated the after-effects of ventriloquism in a VR audiovisual environment. The results showed that the apparent position of the auditory stimulus shifted in the direction of the visual object. This effect was greatest when the sound source was on the same side as the visual object. When it was on the opposite side, the localization shifted in the wrong direction, which confirms a phenomenon very similar to the real world in VE. However, this experiment was conducted with virtual loudspeakers and visual objects and did not reveal the effects of ventriloquism on avatars. Therefore, it is uncertain how the effect will behave in a situation where the user is supposed to interact with an avatar in VE, such as the experiments examined in this paper.

B. INTERPERSONAL DISTANCE RESEARCH

Research on interpersonal distance in RE has been conducted since approximately 1960. Ensuring adequate personal space allows people to interact with one another in a comfortable manner, which leads to feelings of safety [32], [33], [34], [35]. On the contrary, any intrusion into the personal space that an individual wants to secure is interpreted as a threat and triggers an anxious state [36], [37]. Personal space is generally oval in shape and is known to be larger in the front than in the back, or sides [10], [38]. The exact size and shape depend on a number of social and personal characteristics and environmental factors, for example, the movement of obstacles [39]. Other factors vary with appearance, attributes, and combinations [40], [41], [42], such as the tendency of women to prefer smaller interpersonal distances than men [43], [44], the possibility of variation with combinations of the two genders [7], [10], the influence of height [45], and the influence of facial expression and gaze [46], [47], [48]. For example, elements of the face of a face-to-face partner play an important role, and an angry facial expression and gaze from the other person can increase interpersonal distance [46], [49], [50], [51].

Reference [8] observed that the way that people feel about one another co-determines interpersonal distance. Similarly, Hayduk: [7] confirmed that liking someone leads to a tendency to reduce interpersonal distance, and Gifford: [52] used projection to confirm that showing that someone likes someone else leads to a closer distance. In addition, Little: [53] showed that when people’s relationship is closer, they stand closer. All of the above studies measured proximity by the stop-distance method [54], [55], [56], [57], [58].

The existence of personal space has been confirmed in VE, and the same trend as in many REs has been observed [12], [59], [60], [61], [62]. In addition, it is easier to control the conditions and to measure the participants in experiments in VE than in RE, which makes it easier to conduct research with an increased number of interpersonal people [63] and to measure the anxiety felt [43], [64]. Another feature of VE is that it is easy to change the attributes of the avatar itself. For example, Angelo et al.: [65] has shown that changing one’s own visibility can change one’s perception of the surroundings. However, to date, we have found no study that investigates the effect of manipulating the sound and visual images separately on personal space, as in the proposed method.

C. QUALITY OF INTERPERSONAL SERVICES

The quality of interpersonal services is priced by customers, and their perceptions and attitudes determine the value of these services [66]. Previous studies have shown that a store staff’s communication with customers influences their perceptions and attitudes toward the service [67], [68], [69], [70]. This quality of interpersonal service is called rapport, a concept that has been emphasized as a perception of a store staff in charge of customer service [71]. Reference [72]

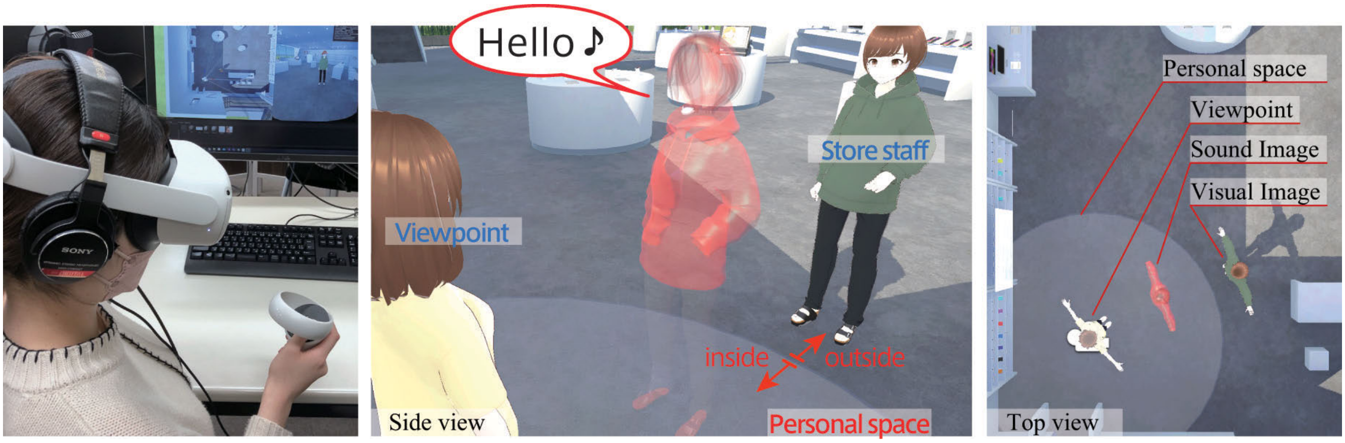

FIGURE 1. The "Mouth-in-the-door" technique enables more effective value co-creation, which utilize the characteristics of VE to make the visual and auditory positions of the avatar presented to the user variable.

defined rapport in service encounters and developed a scale. They found that rapport has an important influence on loyalty, word-of-mouth, and satisfaction.

The components of rapport include enjoyable interaction and personal connection. These are not the only two components of rapport, but they have been shown to be major components of service encounters. Enjoyable interaction is a pleasant interaction between a store staff and a customer and is positioned as an affective, cognitive evaluation of the interaction with the store staff. Personal interaction, in contrast, is based on the customer's perception of the psychological bond between the customer and the store staff. These characteristics have also been found to be enhanced by nonverbal communication, greetings, eye contact, and a pleasant tone of voice [73], [74], [75].

Rapport-building behaviors performed by skilled salespeople who are not involved in the dialog itself include backchannels, and interpersonal proximity [76], [77]. Although some experimental observations have shown the usefulness of these rapport-building behaviors in real stores that use sensor networks [78], [79], little remains clear in the context of customer service via VEs. In VR space interactions in general, regardless of the service context, the effects of changes in the avatar's appearance on the impression of the other person and on communication have been studied. Specifically, the impact of avatars' visual similarity and display method on the sense of their physical possession, presence, etc. References [80], [81], and [82] and the role of avatar nonverbal communication [83] have been identified. A few studies have also been conducted on the presence of store staff avatars in the service experience in VR stores. In VR stores, both the influence of the existence of avatars introducing products on the shopping experience [84], and the effect of interactions with store staff avatars on the shopping experience and brand evaluations [85], [86] are evident. However, the effects of the visual position of the interlocutor and the voice of the interlocutor in the service encounter through VE have not been clarified.

III. RESEARCH QUESTIONS

In this paper, we propose an interaction that intentionally generates a positional dissociation between the sound and visual images of the avatar that the user faces in VE, as shown in Fig. 1. This positional dissociation is expected to change the interpersonal distance in VE and the quality of service through interaction with the avatar.

Figure 2(d) shows an interpersonal situation that can occur only in VE. A situation in which only sound exists cannot occur in RE unless the illusionary phenomenon of acoustic AR devices such as [87], [88], and [89] is used. In RE, a sound source does not suddenly appear from a single point in an empty space; there is always an object that can be visually confirmed as a sound source. Humans use various information, including room acoustics, volume, and binaural differences, to localize a particular sound. This ability is acquired by repeatedly correcting the sound source while checking its correspondence with information obtained from vision. For example, the Head-Related Transfer Function (HRTF), one of the essential factors in the ability to localize sound images, can be corrected by habituation to changes in the shape of the auricle and adaptation of others' HRTFs [90]. Although there are individual differences in the mechanisms of auditory localization, there is a commonality in most of them. Therefore, spatial spaciousness and a sense of localization can be obtained even when stereophonic sounds are presented using an acoustic renderer containing a library of standard HRTFs.

The authors investigated the difference in the shape of personal space between RE and VE by comparing the distances (a) for avatars with both visual and auditory information (V&A condition), $D_{\mathrm{M}}(\theta)$, (b) for avatars with only visual information (not speaking; V/o condition), $D_{\mathrm{V}}(\theta)$, and (c) for avatars with only auditory information (invisible; A/o condition), $D_{\mathrm{A}}(\theta)$, which has been investigated by Yamazaki et al.: [11]. After sufficient calibration of the RE and VEs, the results of the measurements for eight males are shown in the left part of Fig. 2. Although the shapes of the objects generally matched in each condition, a difference

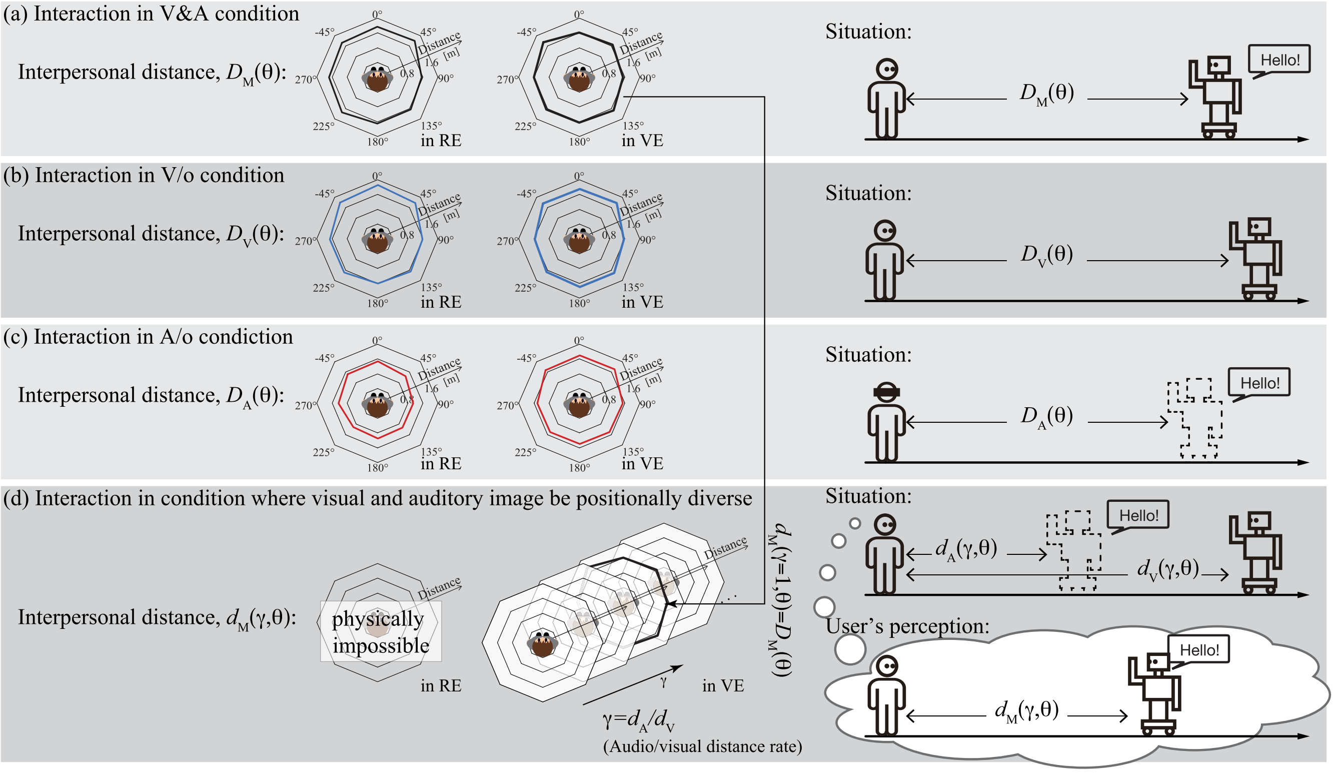
(a) Interaction in V&A condition
Interpersonal distance, D_M(θ):
in RE
in VE
Situation:
D_M(θ)
Hello!
(b) Interaction in V/o condition
Interpersonal distance, D_V(θ):
in RE
in VE
Situation:
D_V(θ)
(c) Interaction in A/o condiction
Interpersonal distance, D_A(θ):
in RE
in VE
Situation:
D_A(θ)
Hello!
(d) Interaction in condition where visual and auditory image be positionally diverse
Interpersonal distance, d_M(γ,θ):
physically impossible
in RE
in VE
γ=d_A/d_V
(Audio/visual distance rate)
d_M(γ=1,θ)=D_M(θ)
Situation:
d_A(γ,θ)
d_V(γ,θ)
Hello!
User's perception:
d_M(γ,θ)
Hello!
0°
45°
90°
135°
180°
225°
270°
-45°
Distance [m]
0.8
1.6

FIGURE 2. Conceptual diagram of the proposal: (a) interpersonal distance, $D_M(\theta)$, in the condition where the avatar is perceivable both visually and audibly; (b) interpersonal distance, $D_V(\theta)$, in the condition where the avatar is perceivable only visually; (c) interpersonal distance, $D_A(\theta)$, in the condition where the avatar is perceivable only audibly, which cannot exist in RE; and (d) Generalized interpersonal distance, $d_M(\gamma, \theta)$, which is affected by both the perceived position of the presented sound image, $d_A(\gamma, \theta)$, and the visual image, $d_V(\gamma, \theta)$, by using the audio/visual distance ratio, γ, as a variable. The condition $\gamma \neq 1$, which is shown in in Fig. 1, is impossible in RE. The left part shows the interpersonal distance for each condition as clarified by the authors in previous research [11]. (Note that, (c) in RE was achieved by having the participants wear an eye mask. See APPENDIX for the experimental conditions.)

in size was observed in the A/o condition, Fig. 2(c). This difference was not observed in condition V&A, Fig. 2(a), suggesting that the personal space based on auditory information is either very ambiguous or exists even though the HRTFs of the individual and the library are different.

Therefore, we expected that the ventriloquism effect would work within a certain range even if the sound and visual images deviated from one another in condition (a) and that they would be recognized as the same avatar. We proposed the following initial research question:

> RQ#1: To what extent is it acceptable for the sound and visual images of the same avatar to deviate in position in VE?

Next, the user's interpersonal distance to the avatar during positional difference, $d_M(\gamma, \theta)$, is expected to be affected by both the distance at which the avatar's image is presented and by the distance at which the sound image is present. Thus, the next research question is as follows:

> RQ#2: What is the effect of the positional discrepancy between sound and visual images on the magnitude of the interpersonal distance to the avatar in VE?

Finally, in proximity studies, there is a phenomenon in which interpersonal distance decreases in close relationships, and conversely, close interpersonal distance gives the illusion of a close relationship. Many interpersonal techniques utilize this phenomenon. Therefore, we consider the following research question:

> RQ#3: What is the effect of the intrusion of sound images into personal space on the impression of the avatar?

In the next chapter, we describe the investigation method and experimental results of the above research question using VE.

IV. EFFECTS OF THE POSITIONAL DEVIATION BETWEEN SOUND AND VISUAL IMAGES IN VR

A. EXPERIMENTAL DESIGN

The experimental participants included 16 people (8 males and 8 females) between the ages of 21 and 24 years who had no vision or hearing problems.[3] Eleven of the participants had previously experienced VR, and five had never experienced

[3] This study was approved by the Ethics Review Committee of the Faculty of Systems, Information and Engineering, University of Tsukuba and was conducted in accordance with the guidelines of the Declaration of Helsinki (2020R427, November 12, 2020).

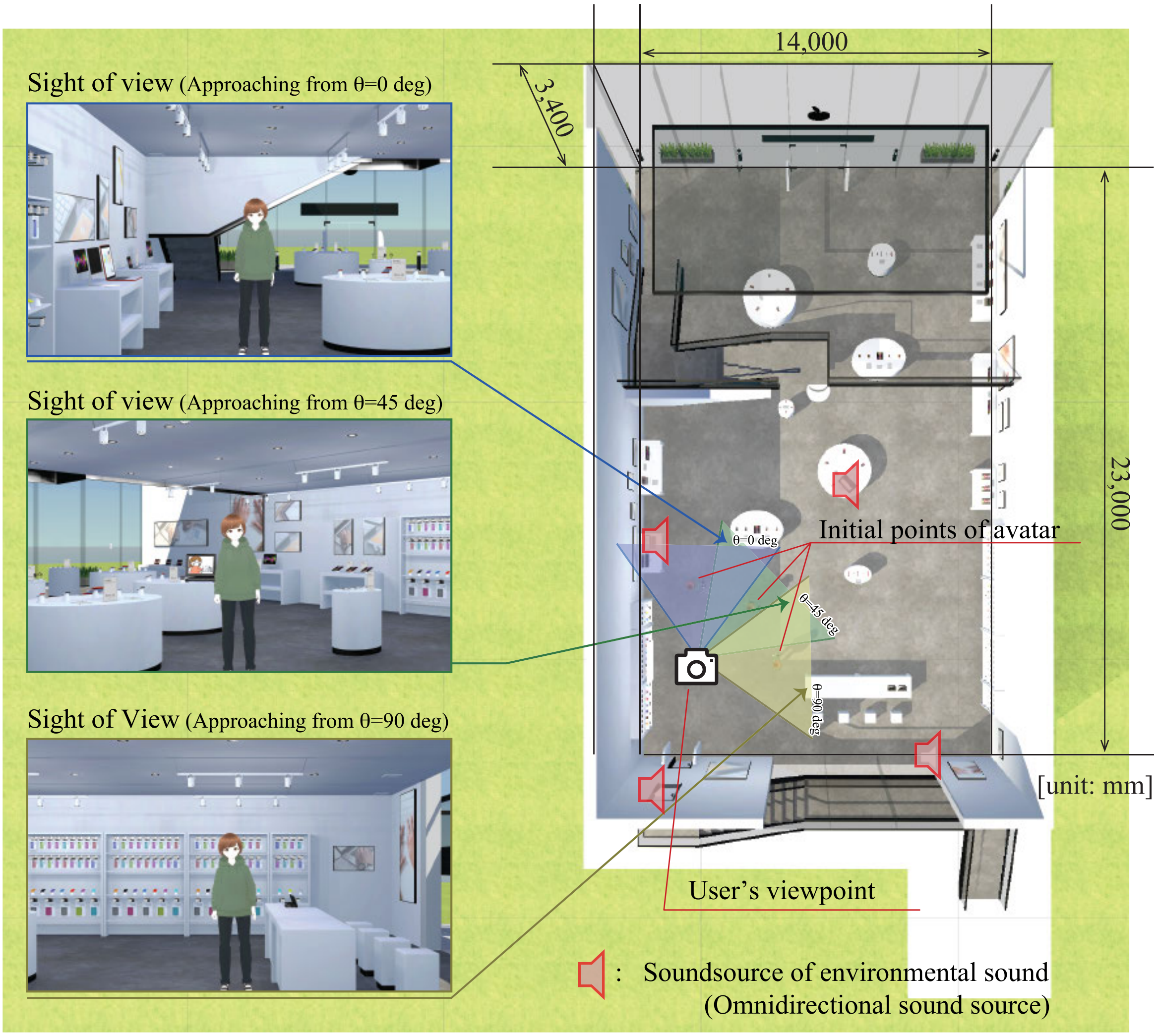
Sight of view (Approaching from θ=0 deg)
Sight of view (Approaching from θ=45 deg)
Sight of View (Approaching from θ=90 deg)
14,000
3,400
23,000
θ=0 deg
Initial points of avatar
θ=45 deg
θ=90 deg
[unit: mm]
User's viewpoint
: Soundsource of environmental sound
(Omnidirectional sound source)

FIGURE 3. Experimental setup diagram (capture of Unity space).

it. No one reported that they could not localize the sound image throughout the experiment.

In the experiment, we placed an avatar that played the role of store staff in a room that resembled an electronics retail store in VE (Fig. 3). The room used in the experiment was purchased from Unity's asset store,[4] and the avatar that played the role of the store staff member was created with VRoid Studio,[5] which is 3D character creation software. The questionnaire used in the experiment was presented on the left-hand side of the participant in VE. The room used for the experiment in Unity was 23 m × 14 m × 3.4 m, and the height of the avatar that played the store staff was 1.6 m.

We prepared six types of avatars to be used in the experiment, that is, one for each experimental condition (Fig. 4).

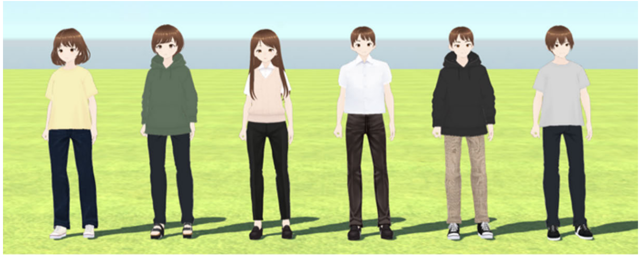

FIGURE 4. Store staff avatars used in the experiments.

There were three male avatars and three female avatars, and all of them were the same height. The avatar was a store staff member and said *"Irasshaimase"* (Welcome), *"Konnichiwa"* (Hello), *"Arigato gozaimasu"* (Thank you), and *"Yoroshiku onegai itashimasu"* (Hope you enjoy) in Japanese every second. As the environmental sounds that the participants could hear, the store's background music was output from

[4] Mixall, Electronics store - devices and furniture, assetstore.unity.com/packages/3d/props/interior/electronics-store-devices-and-furniture-184870

[5] pixiv Inc., VRoid Studio, vroid.com/en/studio, Last Access: February 14, 2022

the ceiling behind the left side of the participants, the air conditioning sound from the ceiling was output behind the right side, and the advertisement sound from the monitors on the product display tables was output in front of the right and left sides. For the sound output in VE, we used Steam audio, an audio environment developed by Valve to increase the sense of immersion along with the development of VR and MR so that the sound in the VE can be heard as 3D sound.

During the experiment, the participants were placed at the position of the camera icon in the figure with their basic posture in the upper direction of the overhead view. As in the general VR experience, the field of view changed according to the participant's own head movement. Similarly, the sound presented to both ears changed according to the head direction. For the approach direction, θ, in each experiment, we tested the approach from three directions: 0°, 45°, and 90°. This is because, in stores, store staff generally do not approach from behind to avoid startling customers. In addition, previous studies have shown that the shape of one's personal space is symmetrical when viewed from above. For this reason, the approaching situations were considered to be from the front, diagonally from the front, and from the side, and these three directions were selected. To eliminate the influence of habituation on the results, the order in which the experiments were conducted was changed for each participant in terms of the approach direction and avatar type. In this study, we assumed the scenario in which the approaching person was a store staff member whom one met for the first time, and we informed the experiment participants of this fact before conducting the experiment.

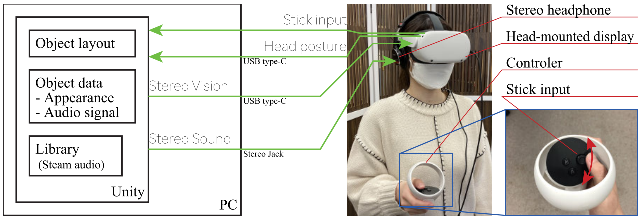

FIGURE 5. Experimental setup.

In the experiment in the VE, video and audio were presented using a head-mounted display (HMD) (Oculus Quest2, Meta, resolution: 1832 × 1920 each, 503 g), and headphones (SONY, MDR-CD 900ST) were used for presentation. The controller was supplied with the HMD used in the study (Fig. 5).

B. EXPERIMENTAL PROCEDURE

The experimental procedure for each participant is shown in Fig. 6. The experimental procedure consisted of the following four main steps: #1 time to get used to the VE; #2 measurement of the interpersonal distance in each auditory condition; #3 measurement of the tolerance for the positional deviation between the sound and visual images for the same avatar; and #4 measurement of the interpersonal distance and impression of the avatar with a positional deviation between the sound and visual images. The time required for each experiment was approximately 45 minutes, and a 5 minute break was taken at the end of the first half of the experiment.

1) STEP #1: GENERATING THE VE PERCEPTION

The first step was to adjust the eye line and get used to the VE. The participant wore the HMD and headphones and held the controller in their right hand. During the experiment, the participants stood and faced the top direction of the overhead view. The direction of the feet was fixed, but the participants could turn their heads to look left and right. First, the participants looked around the room in the VE where the experiment was conducted and answered whether they felt comfortable with their own eye level. If they felt uncomfortable, then they could change the camera position in Unity. Next, the participant experienced the VE, and the position of the avatar playing the role of the store staff was manipulated with the controller. In the VE, it was necessary to adapt to the stereophonic sound using the HRTF prepared as a library. The participants could freely manipulate the store staff avatar until they no longer felt uncomfortable with the audio and became accustomed to the VE. The analog stick of the controller was used to control the position of the avatar. Each time that the user moved the analog stick forward or backward or left or right, the avatar's position moved 0.1 m in this direction. By continuing to roll the analog stick, the position of the store staff avatar could be moved continuously.

2) STEP #2: MEASURING THE INTERPERSONAL DISTANCE IN EACH AUDIOVISUAL CONDITION

We measured the interpersonal distance for each avatar that played the role of a store staff member under different perceivable modal conditions (only visual image, only audio image and both). Half of the participants measured the interpersonal distance with visual information first ($D_{\mathrm{V}}(\theta)$), and the other half measured the interpersonal distance with auditory information first ($D_{\mathrm{A}}(\theta)$), which eliminated the effect of the experimental order on the results. In the visual-information-only condition (V/o condition), the avatar's appearance (visual image) was presented, but the voice (sound image) was not. In the auditory-information-only condition (A/o condition), the avatar's appearance (visual image) was not presented, but the voice (sound image) was. The stop distance method was used to measure interpersonal distance [54], [55], [56], [57], [58], [91]. The participants in the experiment moved the avatar's image or sound image closer by manipulating the analog stick of the controller to determine the position where further intrusion would cause discomfort. Since the measurement was made for each proximity angle, $\theta = 0°, 45°$, and 90°, the controller accepted manipulations only in the radial direction centered on the user. The visual image or sound image of the avatar playing the role of the store staff could be moved up to the position of the experimental participant but not behind the participant.

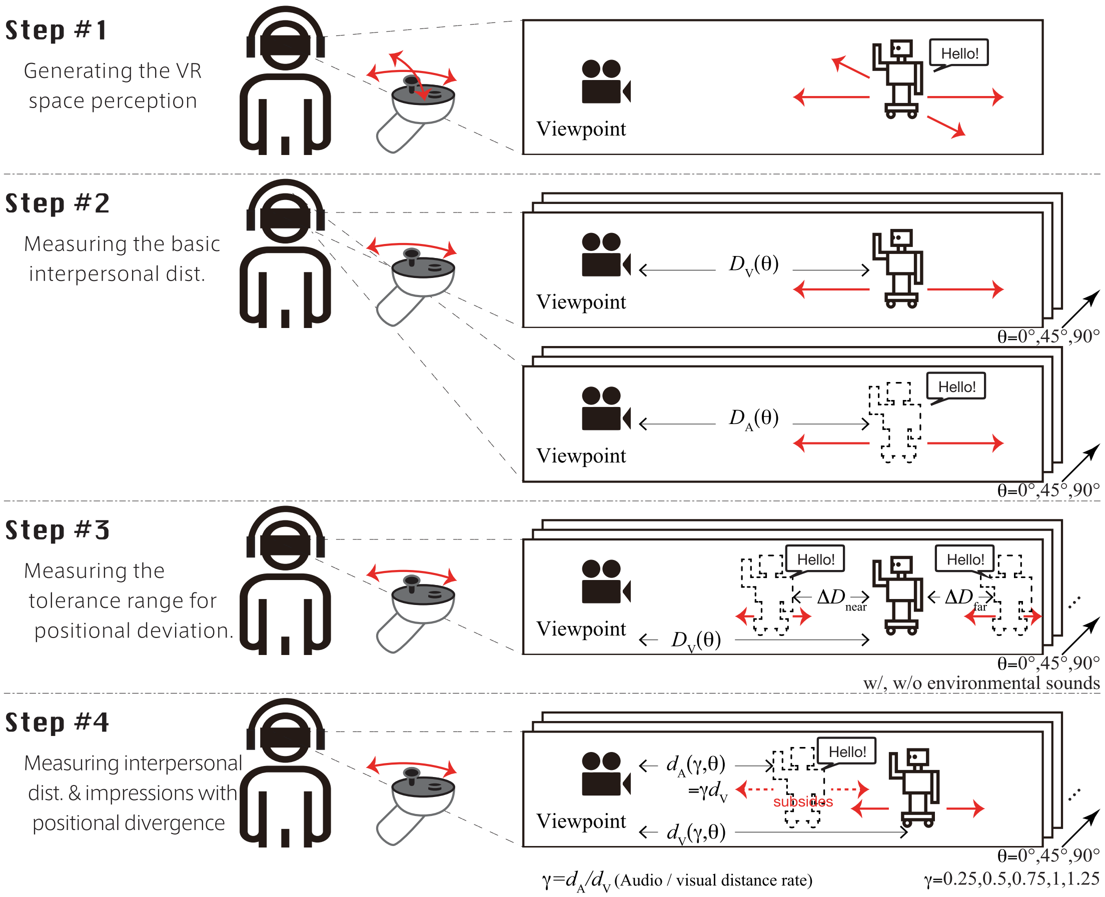

FIGURE 6. Experimental procedure.

3) STEP #3: MEASURING THE TOLERANCE RANGE FOR THE POSITIONAL DEVIATION BETWEEN THE SOUND AND VISUAL IMAGES OF THE SAME AVATAR

We measured the range in which the position of the sound image of the avatar that played the role of the store staff could be recognized as that of the same avatar (the range in which the ventriloquism effect worked) when it was separated from the visual image and manipulated in the radial direction. First, we placed the image of the avatar that played the role of the store staff at the position of the interpersonal distance, $D_V(\theta)$, which is determined by the V/o condition in Step #2. The initial position of the sound image was placed at the same position as the visual image, and the participant manipulated the sound image in the radial direction with the analog stick of the controller. The participant decided the position of the sound image in front of and behind the video image to where the participant felt that if the sound image was farther away from the visual image, then it would be unrecognizable as belonging to the same avatar. It was also possible for the participants to return the sound image to the initial position by pressing the reset button on the controller. The participants could press the reset button at any time during the experiment to confirm the position where the sound and visual images matched. This measurement was performed under the conditions of the presence and absence of environmental sounds (background music, air conditioning, and advertisements).

4) STEP #4: MEASURING THE INTERPERSONAL DISTANCES AND IMPRESSIONS FOR AVATARS WITH A POSITIONAL DIFFERENCE BETWEEN AUDIO AND VIDEO

We investigated the interpersonal distances, $d_V(\gamma, \theta)$ and $d_A(\gamma, \theta)$, and the impressions of the avatar playing the role of the store staff for each audio/video distance ratio, γ. The sound/image distance ratio was the distance from the participant to the sound image when the distance from the participant to the visual image was set to 1. For example, for $\gamma = 0.5$, when the participants saw the visual image of the avatar at a distance of 2 m, the sound image was at a distance of 1 m. Conversely, if $\gamma = 1.25$, when the participant saw the virtual avatar at a distance of 1 m, then the sound image was located at a distance of 1.25 m. When $\gamma = 1$, the relationship between the visual image and the sound image was the same as the relationship between the visual image and the sound image in RE, and the positions of the visual image of the avatar playing the store staff and the sound image were always the same. Since the purpose of this study was to bring the sound

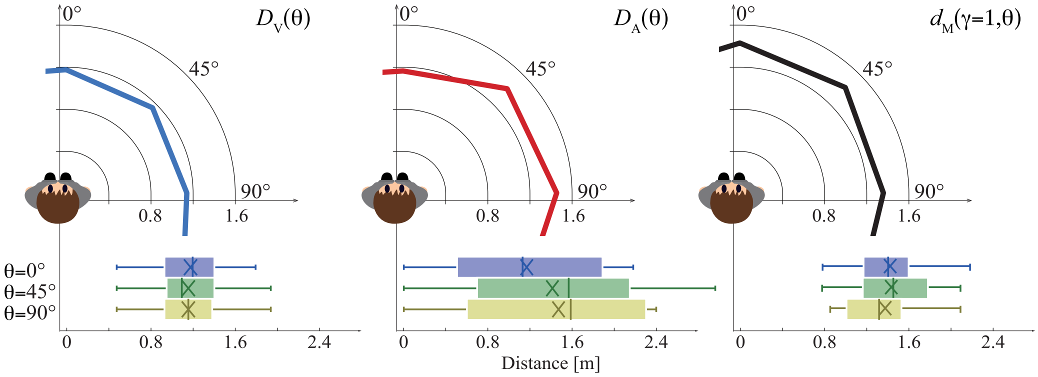

FIGURE 7. Interpersonal distances, $D_V(\theta)$, $D_A(\theta)$ and $d_M(\gamma = 1, \theta)$ $(= D_M(\theta))$, for each condition, namely, the V/o, A/o and V&A conditions, respectively.

images closer in proximity, we prepared many γ smaller than 1. To eliminate the influence of the avatar's appearance and voice type preference on the experimental results, the γ assigned to each avatar varied depending on the experimental participant. The interpersonal distance for each avatar was determined using the stop distance method as in Step #2 to measure the interpersonal distance in each audiovisual condition. We surveyed the participants' impressions of each avatar using a questionnaire each time. The questionnaire was written on the wall on the left-hand side of the participant in the VE where the experiment was conducted, and the participant answered it orally. The questionnaire items were developed based on the rapport questionnaire developed by Gremler and Gwinner: [72] and measured on a 7-point scale (from *1. not at all* to *7. very true*).

C. DISCOMFORT WITH THE POSITIONAL DISCREPANCY BETWEEN SOUND AND THE VISUAL IMAGES (RESULT OF RQ#1)

1) RESULTS

The shape of the interpersonal distance in this experimental condition is shown in Fig. 7. The shape of the interpersonal distance, $d_M(\gamma = 1, \theta)$, in the V&A condition is larger in the frontal direction than in the lateral direction. This tendency was consistent with previous studies and the results of previous studies by the authors. Compared to the interpersonal distance to the avatar's visual image, $D_V(\theta)$, the interpersonal distance to the sound image, $D_A(\theta)$, tended to have a larger variance.

The interpersonal distance for the avatars that had both the sound and visual images, $d_M(\gamma = 1, \theta)$, was larger than the interpersonal distance for the avatars that did not speak, $D_V(\theta)$, in all approach direction conditions. This difference was not observed when compared to the interpersonal distance for the avatars with invisible images, $D_A(\theta)$, even taking into account the effect of the unusual situation of ‘‘not being able to see the person with whom you are interacting'' in the VE experience.

In addition, we investigated the acceptable range at which the participants could recognize that the sound and visual images belonged to the same avatar despite a positional discrepancy between them. The results are shown in Fig. 8. These results show the range where the ventriloquism effect works in the radial direction around the user in VE. In visual localization, the absolute positional relationship is determined optically. However, auditory localization is relative because it can be localized to some extent using the HRTF of another person, and an effect of habituation is well known. Therefore, we compared the results of the two conditions, one with and the other without sound sources that could be localized in addition to the speech of the avatar that played the store staff.

Figure 8 shows the interpersonal distance, $d_M(\gamma = 1, \theta)$, and the range of the effect of ventriloquism on the sound image for each participant. The ratio, $\Delta D/d_M(\gamma = 1, \theta)$, where ΔD is the range where the ventriloquism effect works and is normalized by $d_M(\gamma = 1, \theta)$, to cancel out the differences in the size of the interpersonal distance between individuals, is shown as a shaded area, and the average value is given as a numerical value. In general, the range of the ventriloquism effect is narrower in the condition where environmental sound is present than in the condition where environmental sound is absent. In the presence of environmental sound, $D_M(\theta = 90°)$ was found to be significantly smaller than $D_M(\theta = 0°)$ and $D_M(\theta = 45°)$. It was also confirmed that $D_M(\theta = 90°)$ in the presence of environmental sound was predominantly smaller than that in the absence of environmental sound. These were tested at the 5% level by a t-test.

2) DISCUSSIONS

Since the interpersonal distance for the avatars that had both sound and visual images, namely, $d_M(\gamma = 1, \theta)$, was larger than the interpersonal distance for the avatars that did not speak, namely, $D_V(\theta)$, in all approach direction conditions, and it is presumed that the voice affects interpersonal

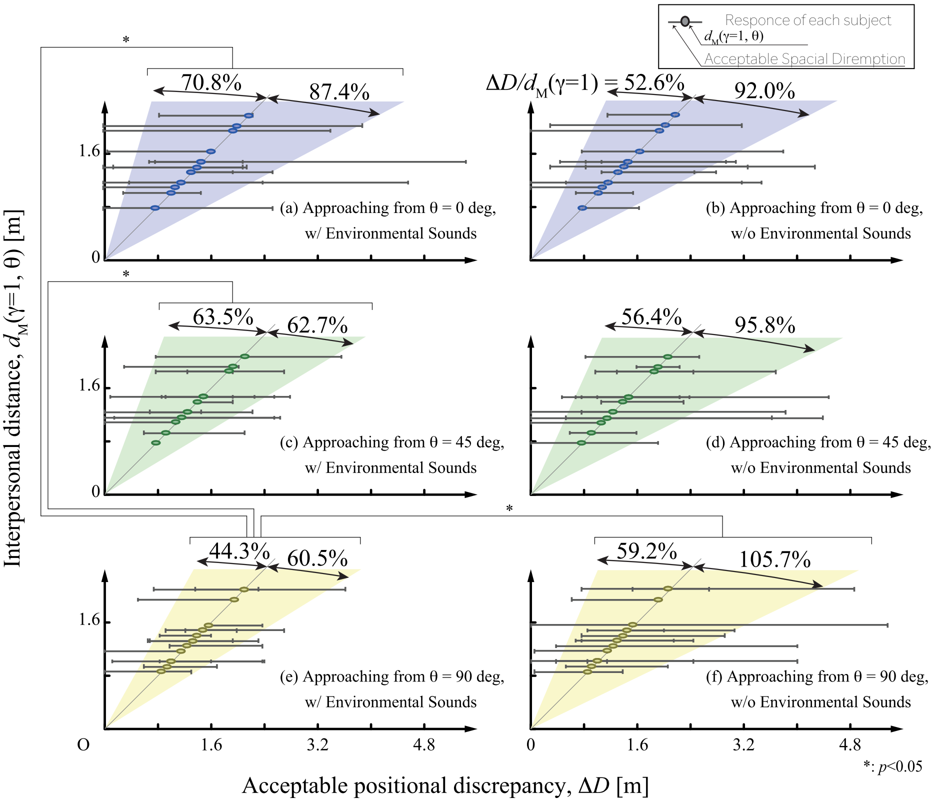

FIGURE 8. The range in which the position of the sound image can deviate from the interpersonal distance $D_V(\theta)$ to the avatar's visual image compared to $d_M(\gamma = 1, \theta)$, from $D_V(\theta) - \Delta D_{near}$ to $D_V(\theta) + \Delta D_{far}$. The approach from the $\theta = 0°$) direction is represented in (a) and (b), the $\theta = 45°$) direction is represented in (c) and (d), and the $\theta = 90°$) direction is represented in (e) and (f). The effects of the presence or absence of environmental sounds other than the avatar's voice are also compared, with (a), (c) and (e) indicating the presence of environmental sounds and (b), (d) and (f) indicating the absence of environmental sounds.

distance. We also consider the range in which the ventriloquism effect works.

The main factors in human source localization are the 1) level difference, 2) binaural time of arrival difference, and 3) spectral cues [92], [93], [94], [95], [96], [97], [98], [99]. However, since 3) spectral cues work only when the localization target is a familiar sound source and auricle, it is difficult to imagine that they are the dominant factors in the results of this experiment. In the presence of ambient sound, the 2) binaural arrival time difference is smaller in the $\theta = 0°$ direction than in the $\theta = 90°$ direction. Therefore, it is possible that ΔD is widened because the difference in distance cannot be made closer due to the decrease in the auditory resolution of the experimental participants. Comparing the conditions with and without environmental sound, we can see that the acceptable range is narrower in the condition with environmental sound than in the condition without environmental sound, except for the condition where the sound sources in the 0° direction are close together. It seems that the presence of environmental sound in VE as a comparison object contributes to the narrowing of the range of the ventriloquism effect because of the improvement of the localization performance due to the relative 1) level difference from other sound sources.

D. INTERPERSONAL DISTANCE FOR THE AVATARS WITH A POSITIONAL DIFFERENCE BETWEEN SOUND IMAGES AND VISUAL IMAGES (RESULT OF RQ#2)

1) RESULTS

We investigated the interpersonal distance, $d_M(\gamma, \theta)$, for an avatar that played the role of a store staff member with a positional discrepancy between the sound and visual images by changing the audio/visual distance ratio, γ. The interpersonal distance to the visual image of the avatar that played the role of the store staff, $d_V(\gamma, \theta)$, and the interpersonal distance to its sound image, $d_V(\gamma, \theta)$, are shown in Fig. 9 with the unimodal condition, $D_V(\theta)$ and $D_A(\theta)$. We also tested the differences for each condition for the angle of approaching, θ, and audio/video distance ratio, γ.

First, we present the results for $d_V(\gamma, \theta)$. As mentioned in the previous section, there is a difference between $d_V(\gamma = 1, \theta)$ and $D_V(\theta)$ for all directions. In addition to the condition $\gamma = 1$, all of $d_V(\gamma, \theta)$ has a larger mean value than $D_V(\theta)$ across almost all audio/video distance ratio and approach direction conditions, which indicates that there is a difference between $d_V(\gamma, \theta)$ that is not determined solely by visual conditions, contrary to intuition. However, even when γ varied, there was no difference on average between $d_V(\gamma, \theta)$ except for some proximity from the $\theta = 0°$ direction.

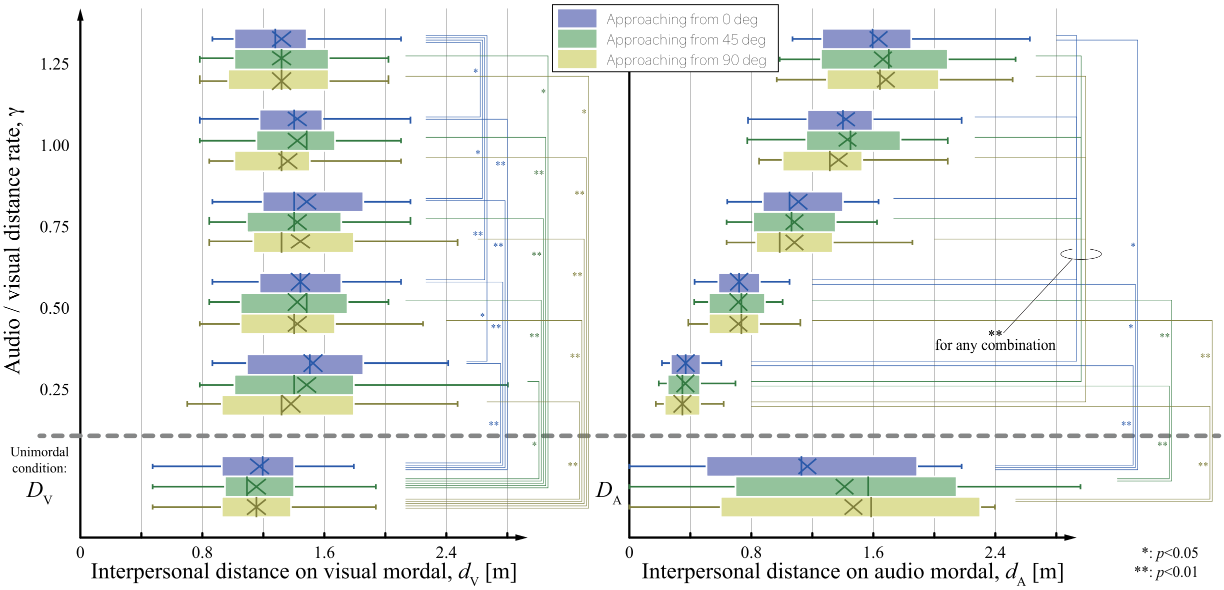

FIGURE 9. Box-and-whisker plot showing the difference in the interpersonal distance for each audio/visual distance ratio, γ. (Left) Distance between the experimental participant and the avatar playing the role of the store staff as measured by visual positioning, $d_V(\gamma, \theta)$, and a unimodal condition with only visual images, $D_V(\theta)$, (Right) Distance measured by auditory positioning, $d_A(\gamma, \theta)$, and the unimodal condition with only the sound images, $D_A(\theta)$. Comparing the left and right figures, we can see that in the left plots, changing γ does not cause much change, while in the right plots, changing γ causes a large change in the value of $d_A(\gamma, \theta)$.

Next, we discuss the results for $d_A(\gamma, \theta)$. Although there is a large difference in the variance between $d_A(\gamma = 1, \theta)$ and $D_A(\theta)$ due to individual differences, there is not a large difference in the mean. However, due to the variation in γ, there is a difference between $d_A(\gamma, \theta)$ and $D_A(\theta)$. It is also confirmed that as γ decreases, $d_A(\gamma, \theta)$ also decreases significantly compared to all combinations of $\gamma = \{0.25, 0.5, 0.75, 1, 1.25\}$. These were tested at the 5% level by a t-test.

2) DISCUSSIONS

This indicates that although the presence of sound has a significant effect on interpersonal distance, $d_M(\gamma, \theta)$, the effect of the location of the sound image on the visual interpersonal distance from the avatar is slight.

To compare the interpersonal distances of each individual to the positions of the visual and sound images in the respective audio/video distance ratio and approach direction conditions, we normalized the interpersonal distance to the visual image of the avatar playing the role of the store staff, $d_V(\gamma, \theta)$, and it to the sound image, $d_A(\gamma, \theta)$, in the unimodal conditions, $D_V(\theta)$ and $D_A(\theta)$, respectively, and made scatter plots for each of the audio/visual distance ratios, γ, and approach directions, θ (as shown in Fig. 10). In this graph, the horizontal axis represents $d_V(\gamma, \theta)/D_V(\theta)$, and the vertical axis represents $d_A(\gamma, \theta)/D_A(\theta)$. When each value is less than 1, this indicates that the visual image or sound image is intruding on the interpersonal space of the individual. In the condition where γ is less than 1, most $d_A(\gamma, \theta)/D_A(\theta)$ have values less than 1. However, in the condition where γ is greater than or equal to 1, there is no condition where $d_V(\gamma, \theta)/D_V(\theta)$ takes a value of less than 1.

From the above, it can be confirmed that when the face-to-face sales avatar exists in terms of both a sound image and a visual image, the information obtained visually is the dominant factor that determines the interpersonal distance. The avatar's visual image has difficulty intruding into the individual's personal space, but the sound image of the avatar can intrude into the individual's personal space.

E. EFFECT OF SOUND IMAGE INTRUSION INTO PERSONAL SPACE ON RAPPORT (RESULT OF RQ#3)

1) RESULTS

In the experimental procedure, the interpersonal distance $d_M(\gamma, \theta)$ to the avatar was measured for each of the audio/visual distance ratios, γ, and approaching directions, θ.

Based on the questionnaires in previous studies, six items of enjoyable interaction and five items of personal connection were utilized. The questionnaire used in this experiment was a 7-point Likert scale questionnaire with six questions that could be judged by first impressions such as greetings and classified into the three categories of impression, familiarity, and trust. The questions and results of the questionnaire are shown in Fig. 11. However, since the VR avatar could not serve customers in this experiment, we changed the questionnaire items to the future tense. The changed parts of the questions are written in capital letters.

From the results of the previous section, it was expected that the results of each questionnaire would be improved under the condition of $\gamma < 1$ because the sound image position intruded into the personal space. However, contrary

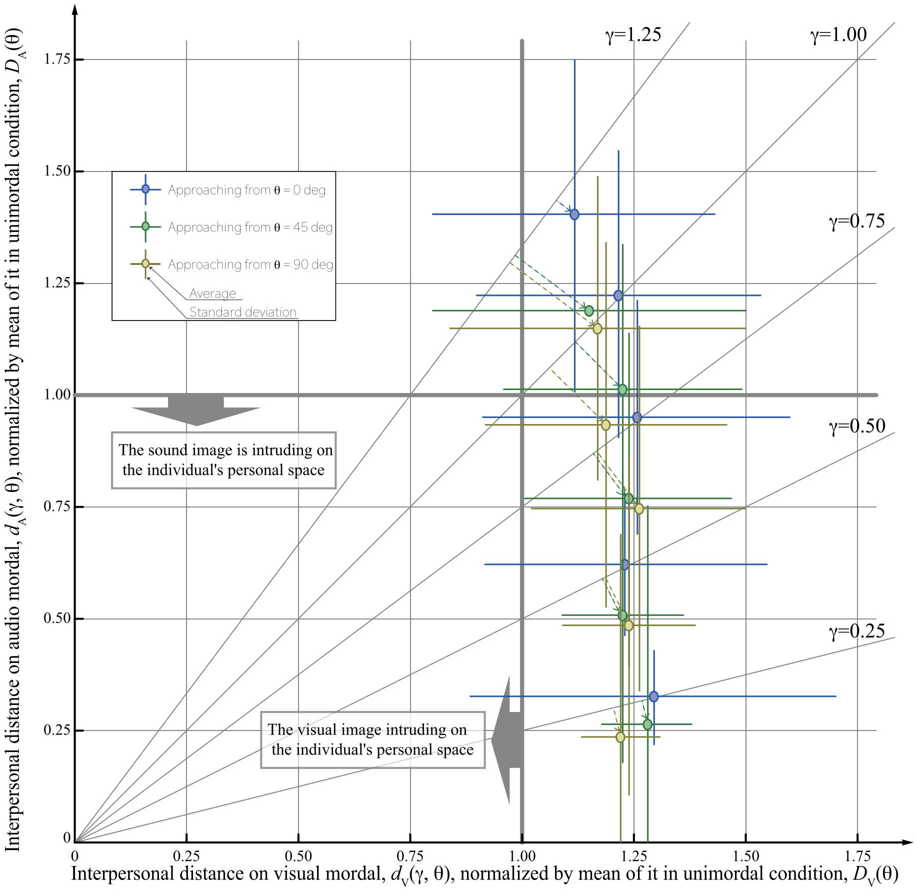

FIGURE 10. Scatter plots with error bars of $d_V(\gamma, \theta)$ and $d_A(\gamma, \theta)$ for each audio/visual distance ratio, γ, and approach angle, θ, as normalized by them in the unimodal condition, $D_V(\theta)$ and $D_A(\theta)$. The error bars represent the standard deviation. In the region $d_V(\gamma, \theta)/D_V(\theta)<1$, the avatar is visually intruding into personal space, and in the region $d_A(\gamma, \theta)/D_A(\theta)<1$, the avatar is audibly intruding into personal space. For any θ, changing γ does not cause $d_V(\gamma, \theta)/D_V(\theta)$ to be less than 1, but decreasing γ causes $d_A(\gamma, \theta)/D_A(\theta)$ to be conditional on taking a value of less than 1.

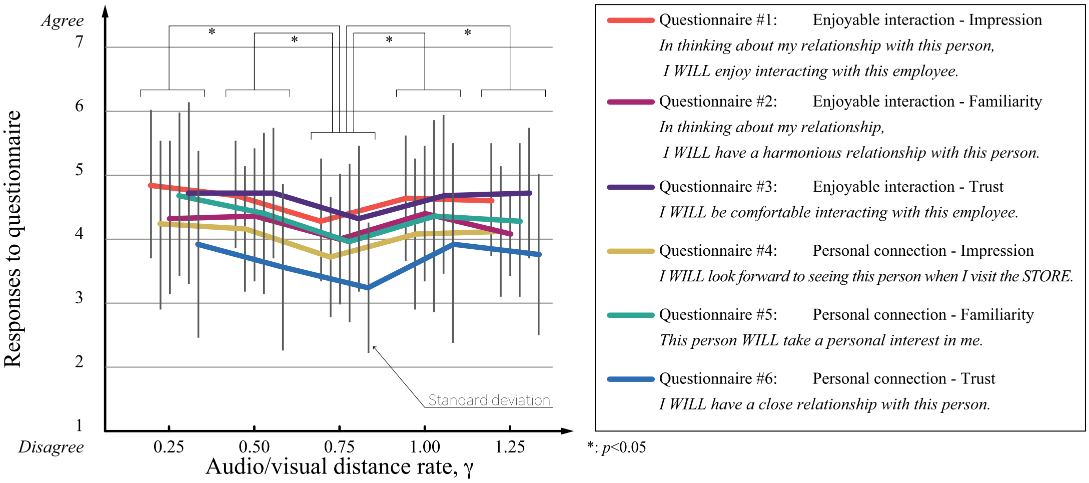

FIGURE 11. Questionnaire results for the rapport components for each audio/video distance ratio, γ.

to our expectations, the results of the questionnaire did not show this tendency. In addition, the responses to the questionnaires decreased significantly, where $\gamma = 0.75$. These were tested by performing the Mann–Whitney U test at the 5% level on the mean of the responses to the six questionnaires.

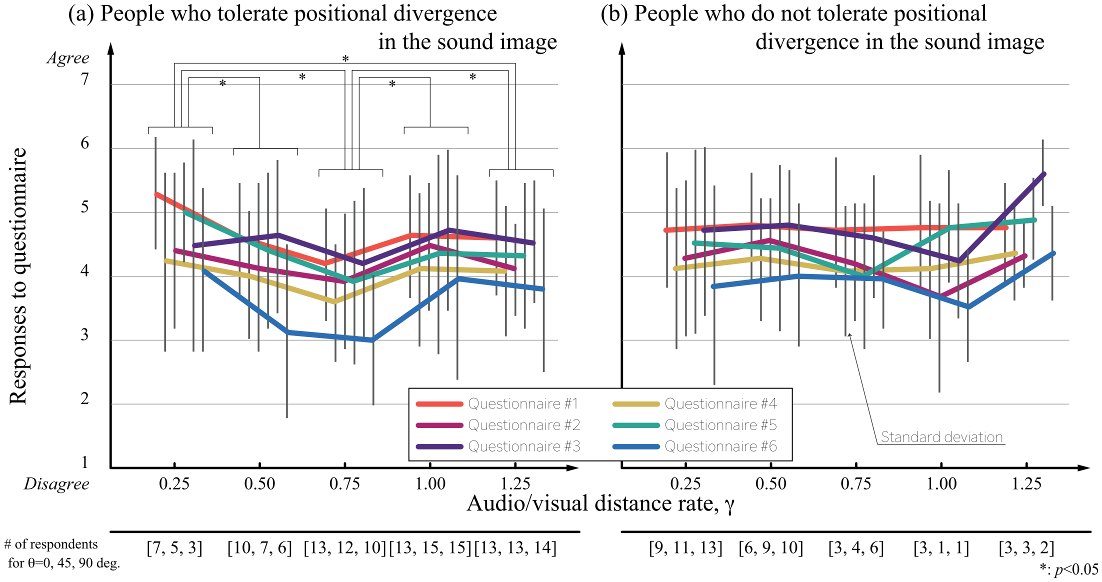

FIGURE 12. The results of the questionnaire for the rapport components for each audio/visual distance ratio, γ, (a) rapport of those whose sound image position deviated from the visual image within a tolerable range, and (b) rapport of the others. The bottom row shows the number of responses for each γ and θ condition. The sum of the number of responses for the same γ and θ on both sides is 16, which is the number of participants in this experiment.

2) DISCUSSIONS

Considering the effect of ΔD for each individual (as clarified in Section 4.3), we divided the questionnaire results into two groups, one in which $d_{\mathrm{A}}(\gamma, \theta)$ determined by each sound image/video distance ratio was in the range from $D_{\mathrm{V}}(\theta) - \Delta D_{\mathrm{near}}$ to $D_{\mathrm{V}}(\theta) + \Delta D_{\mathrm{far}}$ and the other in which it was not, as shown in Fig. 12. The figure on the left, which shows the rapport of the group in which the sound image was tolerant of positional deviation from the visual image, shows that the response to the questionnaire decreased significantly, where $\gamma = 0.75$. However, the response to the questionnaire tended to improve as γ decreased in the region of $\gamma = [0.25, 0.75]$. The questionnaire response of $\gamma = 0.25$ was found to be significantly different from the response of $\gamma = 0.50, 0.75, 1.25$. The right figure shows the response of the group in which the sound image was not tolerated to deviate positionally from the visual image. In this group, the region of $\gamma = [1.00, 1.25]$ had a small number of samples. Except for this region, there was no noticeable change in the rapport with the variation in γ, and no significant difference was observed. There was no significant difference between the left and right graphs for each γ.

To clarify the cause of the change in rapport in the group in which the sound image was tolerated to deviate from the visual image Fig. 12(a), we compared the rapport of each component and found that it increased at $\gamma = 0.25$ in Questions #1, #4 and #6. At $\gamma = 0.75$, we can see that Questions #1, #2, and #6 are decreasing. At $\gamma = 0.25$, personal connections, such as Questionnaire #4 and #6, are thought to unconsciously improve the response. This is a similar phenomenon to that observed in the hospitality/interpersonal technique, which is a technique to improve or create the illusion of intimacy with another person by intentionally sneaking into the other person's personal space.

F. SUMMARY

Through the above experiments, we investigated three RQs related to the ventriloquism effect, personal space, and rapport formation in VE. The conclusions for each RQ are summarized below.

1) RQ#1: TO WHAT EXTENT IS IT ACCEPTABLE FOR THE SOUND AND VISUAL IMAGES OF THE SAME AVATAR TO DEVIATE IN POSITION IN VE?

In the V&A condition in the VE, the boundary of the range where the sound and visual images of an avatar can deviate in position and the sound image and visual image can be recognized as belonging to the same avatar, $[D_{\mathrm{V}}(\theta) - \Delta D_{\mathrm{near}}, D_{\mathrm{V}}(\theta) + \Delta D_{\mathrm{far}}]$, which were manipulated by the participants in the experiment. The results show that the user's posture depended on the position of the avatar, with approximately 75% of the distance to the avatar in the frontal direction ($\theta = 0°$), 60% in the oblique direction ($\theta = 45°$), and 50% in the side direction ($\theta = 90°$) in average. In addition, the presence of environmental sounds can have an effect. The width of the range varied depending on the presence of environmental sounds; when there were no environmental sounds, the range in the horizontal direction became much larger.

2) RQ#2: WHAT IS THE EFFECT OF THE POSITIONAL DISCREPANCY BETWEEN SOUND AND VISUAL IMAGES ON THE MAGNITUDE OF THE INTERPERSONAL DISTANCE TO THE AVATAR IN VE?

We investigated the interpersonal distance to an avatar playing the role of store staff, $d_{\mathrm{M}}(\gamma, \theta)$, in a VE where there was a positional difference between the sound and visual images ($\gamma \neq 1$). We found that when the face-to-face avatar existed both acoustically and visually, the information obtained from the visual image of the avatar was more significant than the information obtained from the sound image. As a result, we found that when the face-to-face avatar existed both sonically and visually, the information obtained from the image was the dominant factor that determined the interpersonal distance. Although the presence of sound had a significant effect on interpersonal distance, the effect of the position of the sound image on visual interpersonal distance from the store staff avatar was slight. That is, the image of the avatar that played the role of the store staff had difficulty intruding on the individual's personal space, but the avatar's sound image was able to intrude on the individual's personal space.

3) RQ#3: HOW DOES THE INTRUSION OF THE SOUND IMAGE INTO THE PERSONAL SPACE AFFECT THE IMPRESSION OF THE AVATAR?

When the sound image was located in the personal space ($\gamma < 1$), the user's report of the avatar was investigated through a questionnaire. In all questionnaire results, rapport decreased when $\gamma = 0.75$, and no contribution to the improvement of rapport was observed when γ decreased. However, only in the group where the sound image was tolerated to deviate positionally from the image was confirmed that decreasing γ contributed to the increase in rapport in the $\gamma = [0.25, 0.75]$ region.

V. GENERAL DISCUSSION

A. "UNCANNY VALLEY"-LIKE PHENOMENA

We discuss the V-shape of rapport shown in Fig. 12(a). An alternate explanation for the phenomenon observed at $\gamma = 0.75$ is a phenomenon similar to the "uncanny valley".

In the $\gamma = 0$ condition, which is not provided in this experiment, the sound image is localized in the head even though the participants can see the avatar image. This situation is similar to the situation in which we see the face of the other person on display and hear the other person's voice through earphones in a video conference. Since the pandemic caused by COVID-19, we have been increasingly communicating with our interlocutors in this state, and many of us naturally accept the presentation of the $\gamma = 0$ condition in which the visual and sound images deviate in position without feeling uncomfortable even though the laws of physics in RE allow only the $\gamma = 1$ condition. Therefore, it is possible that an unconscious trough of discomfort arises between $\gamma = 0$ and $\gamma = 1$ for those who can tolerate the positional discrepancy between the sound and visual images in the $\gamma = 0.25, 0.50$ and 0.75 conditions. In this experiment, we could not measure this "valley of discomfort", but it may be one of the reasons why the shape of the rapport of the group in which the sound image is allowed to deviate from the image in position is this way.

B. APPLICATIONS

This technique may be helpful in a variety of service designs in VEs. The application of this technique, which distorts the Euclidean space where the sound and visual images geometrically correspond, is described for the reader's understanding according to the following two scenarios. One is the case where the service provider is the one authorized to perform this space-distorting operation proactively, and the other is the case where the customer performs it.

1) CASE A: WHEN THE SERVICE PROVIDER HAS THE AUTHORITY

When the service provider has authority when providing interpersonal services, it is possible to produce a high rapport for individual store staff since the liking for the store staff is directly related to the value of the service [66]. For example, assuming a VE environment where both parties can move around arbitrarily, the procedure is as follows.

> The store staff sets its sound source position perceived by the customer to be $\gamma < 1$ from when the target customer is far enough away, such as when the customer enters the store. When providing service through dialog, as with RE, both persons move closer to one another to communicate with the other person sufficiently. Since each has its own personal space, they converge to a certain distance, $d_{\mathrm{M}}(\gamma, \theta)$, through repeated interactions ($\because$ result of RQ #1). The interpersonal distance a visual avatar should take to a customer is nearly constant regardless of γ (Fig. 9). Therefore, if the visual avatar maintains the interpersonal distance, it is possible in many cases to place the sound image within the auditory personal space by setting $\gamma \leqq 0.75$ ($\because$ result of RQ #2; Fig. 10). If we use the mouth-in-the-door technique with a condition such as $\gamma = 0.25$ instead of a halfway condition such as $\gamma = 0.75$, then we can improve the rapport obtained from the customer by a certain percentage. Based on the result of RQ #3, the percentages are $43.8\% (= 7/16)$, $31.2\% (= 5/16)$, and $18.9\% (= 3/16)$. This is effective for customers who are unaware of the discrepancy between the sound and visual images due to the ventriloquism effect but has no negative effect on customers who are sensitive to the consistency between the sound and visual images location ($\because$ result of RQ #3; Fig. 12). Therefore, by setting $\gamma \leqq 0.75$, the service provider can obtain a certain percentage of customers' favorable impression of the store staff

without any disadvantage. As a result, the evaluation of the service and the resulting conversions are expected to improve.

However, note that this technique requires the customer's prior consent, such as when immersed in a VR space.

2) CASE B: WHEN THE CUSTOMER HAS THE AUTHORITY

The customer can use this technique to create an acoustically comfortable space and interact effectively with others, such as store staff. As many of us have experienced, having several small chats with people around us simultaneously in a physical gathering is not easy in a remote meeting. Even a brief conversation requires talking to everyone or setting up a breakout room, which is inconvenient. Therefore, there is still a demand for dialog that takes advantage of spatial positioning. Accordingly, it is expected that there will be a demand for a service space in a VE where various sound sources are spatially arranged to provide a sense of presence and reality. However, this is nothing more than a recreation of the real space, and the voice of the specific person with whom one wishes to interact will be buried by the surrounding noise and the representatives of the surrounding crowd. The characteristics of VEs are expected to lead to the development of acoustic interfaces that zoom in on the sound from a specific interaction partner or sound source.

To realize such an interface, we need to solve the dilemma between "spatial acoustic signals that are consistent with the arrangement of sound sources" and "acoustic signals that emphasize the sound of a specific target". Increasing the sound volume without consideration does not satisfy the former condition, and decreasing the distance to the object to better hear a specific sound source may increase one's discomfort since it may allow the sound source to intrude into one's own personal space. Therefore, it is possible to realize an acoustic space that satisfies these two conditions by calculating an acoustic space that assumes that the target sound source is close to oneself. In this way, γ is adjusted to the extent that the sound image position of the target does not intrude its own D_A, which is measured in advance (Fig. 9). This is expected to create an acoustically comfortable space for the customer.

C. LIMITATIONS

The following limitations can be considered in this study. First, the COVID-19 pandemic may have caused a change in the shape of people's personal space. In this experiment, there was no problem in the process of reaching this conclusion because the experiments were conducted at the same time under controlled conditions. However, there is a possibility that conventional proximity techniques that use personal space cannot be used in the future. In addition, the participants in this experiment were all Japanese undergraduate and graduate students. It is not possible to draw conclusions about the differences in personal space caused by generational or cultural differences.

Next, the calibration was not tailored to the individual's audiovisual characteristics. In this experiment, as with many VR experiences and services, most people do not perform a precise calibration but rather a simplified one. HRTFs, one of the essential factors in localizing the sound image, vary significantly among individuals. Accurate habituation to unfamiliar HRTFs requires a much more extended period than in this experiment [100]. Therefore, results may differ with a different audio device setup than the one used in this experiment. This may be an essential factor for the future development of the Metaverse and other VR experiences and services.

In this experiment, we evaluated only the first impression. It is also impossible to conclude the effect of the sound image that deviates from the image in a high context of person-to-person or person-to-agent interaction. The Prometheus effect, which is a phenomenon in which one's personality and behavior change depending on one's gender, height, and appearance, is a unique limiting factor in customer service in VE. In this experiment, the condition was controlled because the participants did not observe themselves in the VE, but if customer service in VEs becomes more common, then this effect cannot be ignored. In addition, we cannot ignore habituation to interactions in VE. If the user interacts with VE on a daily basis, then the user's detection of and tolerance for the positional discrepancy between the sound and visual images may decrease.

VI. CONCLUSION

The theme of this research crosses a wide range of fields and is budding challenge research. The major contribution of the paper is the findings of the effect of the unrealistic variable relationship between the visual and auditory positions of the avatar presented to the user by utilizing the characteristics of VE experiments rather than uniformly presented in the same position.

In this study, we experimented with investigating how the positional deviation between the sound and visual images can be tolerated in VE, the effect of positional deviation on the interpersonal distance to the avatar, and the possibility of manipulating the impression of the avatar by deviating the sound image from the visual image. For the experiment, we prepared a space that resembled a store in VE, and 16 gender-balanced participants conducted proximity experiments with six types of avatars. The abovementioned research purpose was fulfilled by changing the degree of positional discrepancy between the sound and the visual image. As a result, the magnitude of the interpersonal distance (i.e., the shape of the personal space), which varied depending on the direction around the user and the RE, was confirmed, and trends were observed for the visual information-only condition, the auditory information-only condition, and the combined condition. In particular, the effect of the ambiguity of the auditory information on the interpersonal distance when the individual made the decision was remarkably confirmed. It was also confirmed that most of the participants relied on visual information, not on auditory information, to determine the interpersonal distance to the avatar. The position of the sound image in the VE was tolerated even if it deviated from the position of the image, and approximately 75%, 60%, and 50% deviations from the forward, oblique, and side directions in average, respectively, were tolerated in the radial direction centered on the user. Using this phenomenon, we constructed an interpersonal situation with an avatar playing the role of store staff in which only the sound image intruded into the user's personal space, and we investigated the users' impression of the avatar. In this study, we used rapport, a measure of the connection between sales staff and customers. In conditions where the deviation of the sound image position from the visual position was not tolerable, the change in the degree of deviation did not affect the rapport. However, in the experimental conditions where the deviation was tolerated, the following two phenomena were observed: 1) Even when the positional difference was allowed, it caused an "uncanny valley"-like phenomenon, which led to a decrease in rapport; and 2) In the conditions where the positional difference was allowed when the sound image was closer than the visual image to the participant, the rapport was greater with the avatar. This phenomenon is similar to the "foot-in-the-door" phenomenon, in which a small unconscious consent (i.e., allowing a sound image to intrude one's personal space) leads to an improvement in the evaluation of the other person (i.e., the rapport that one has with the avatar as a store staff).

The phenomenon proposed in this paper, such as the positional discrepancy between the sound and visual images, is possible because this experiment involved an interpersonal service in a VE, not an RE. To the best of our knowledge, no research mentions this effect. The techniques discussed in this paper will significantly improve the value of service experiences obtained through interaction with others in VE. In the same way that online shopping has a function using machine learning tools that recommend personalized products based on customers' emotions and brand recognition [101], [102], it is conceivable that the personal space itself can be customized for each customer in the future. Specifically, by incorporating the presentation of personalized sound image locations proposed in this study into existing VR stores and on VR store platforms that are expected to increase in number in the future, the comfort level of the service experience will be improved for customers. As a result, these stores can expect to increase their brand value and sales.

APPENDIX. EXPERIMENTAL DESIGN OF PREVIOUS RESEARCH (DIFFERENCE IN THE SHAPE OF PERSONAL SPACE BETWEEN RE AND VE)

A. SETUP

This section describes the experimental methodology for the results shown in Fig. 2 [11]. To investigate the effects of RE and VE and visual and auditory information, on interpersonal distance, the same space and the same confronting mannequin/avatar were prepared in an RE and a VE. Additionally, the visual and auditory conditions were restricted. Three within-subject factors were manipulated, namely, the environment, approach direction, and perceived audiovisual condition. Two environments were prepared: an RE and a VE. For the RE, a 7.2 m × 7.5 m × 3.1 m room was used. A 3D scan of the room used in the RE was taken by a 12.9-inch iPad Pro with Light Detection and Ranging (LiDAR) and imported into Unity. This allowed us to recreate the real room in a VE, as shown in Fig. 13. LiDAR is an optical sensor technology that identifies the distance to an object and its properties. For the sound output in VE, the audio environment "Steam audio" is used, and the sound is heard in three dimensions. In the VE experiment, video and audio were presented using a head-mounted display (HMD; Dell Visor VR118, Dell Inc., resolution: 1440 × 1440 on each side, 590 g) and headphones (SONY, MDR-CD900 ST). To match the viewing angle, the viewing angle was standardized to 110°, the same as the HMD, by wearing goggles that limited the viewing angle during the experiment in the RE. Three audio-visual conditions were prepared as follows: a bimodal condition in which the subject could see the approaching person and hear his or her voice (V&A condition; Fig. 2(a)); a visual-only condition in which the subject could see the approaching person but not hear his or her voice (V/o condition; Fig. 2(b)); and an auditory-only condition in which the subject was blindfolded and could not see the approaching person but could hear his or her voice (A/o condition; Fig. 2(c)). As an approaching person, an avatar was used in VE, and a mannequin with an iPad Pro that showed an image of the avatar's face fixed on its face was used in the RE. The voice of the approaching person was a recording of a voice saying "Konnichiwa"(Hello).

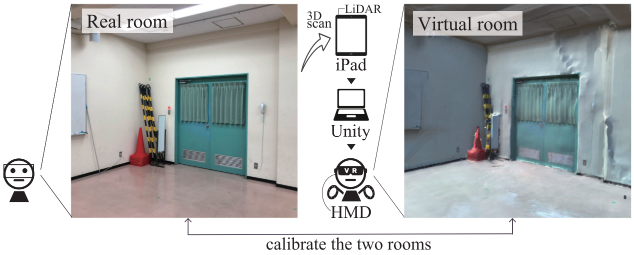

FIGURE 13. The real room (left) and the virtual room (right) used in the experiment. The virtual room was created by 3D scanning the real room. The size of each room was calibrated to match the perceived size of the room, and the interpersonal distance was then measured.

The approach directions were eight directions of 45° each, with the front direction being 0°. Each subject was tested in a random order concerning the experimental conditions. The experiment was conducted under the scenario that the approaching person was a store staff who had never met the subject before.

B. PROCEDURE

In the experiment, the RE and VE were calibrated to match in size, the interpersonal distance was determined by the stop-distance method, and the egocentric distance was mea-

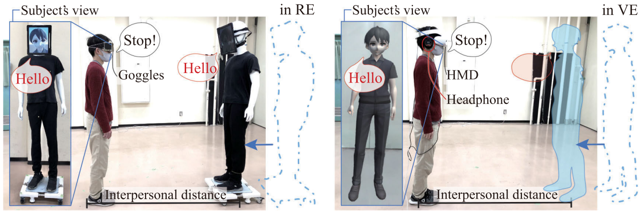

FIGURE 14. Scene of the experiment in the case of the approach from the front direction under V-A conditions in the RE (left) and VE (right). A mannequin was used as the approaching person in the RE, and an avatar was used in the VE. The interpersonal distance was determined with the stop-distance method, in which the approaching person was gradually brought closer to the subject, and the approaching person was signaled to stop at an appropriate distance.

sured by the blind walking task. Initially, calibration was performed to align the perceived size of the RE and VE. This is because it has been shown that distances are perceived as reduced in VEs compared to REs [103], [104], and the degree of reduction depends on the viewing angle and image quality of the device used [105]. The subject stood facing forward with his heels aligned with the markings in the center of the room in the RE and wore the HMD. The subjects alternately looked at or listened to the RE and the VE for the room size, eye level, and the appearance and loudness of the approaching person. If they differed, then the VEs were matched by adjusting their sizes. The next step was to determine the interpersonal distance using the stop-distance method. Figure 14 shows the stop-distance method, which is the most common method for measuring interpersonal distance [54], [55], [56], [57], [58], [91]. The approaching person gradually approached from a distance of 2.5 m from the subject, and when the subject felt uncomfortable if the person approached any closer, the subject signaled the approaching person to stop. Next, the interpersonal distance was measured with a blind walking task [106], [107], [108]. The subject understood the position of the approaching person based on the audiovisual information. They then closed their eyes and walked to where they thought the approaching person was. The distance walked was measured. This made it possible to measure the egocentric distance, which is the distance that is perceived.

The subjects were eight males between the ages of 22 and 25 years with no problems with walking, vision, or hearing. Since the length and characteristics of the interpersonal distance vary depending on gender [44], [109], the subjects in this study were standardized as males. No one reported that they could not localize the sound image throughout the experiment.

ACKNOWLEDGMENT

(*Keiichi Zempo and Azusa Yamazaki are co-first authors.*)

REFERENCES

[1] M. Meißner, J. Pfeiffer, C. Peukert, H. Dietrich, and T. Pfeiffer, "How virtual reality affects consumer choice," *J. Bus. Res.*, vol. 117, pp. 219–231, Sep. 2020.

[2] *Mesh for Microsoft Teams Aims to Make Collaboration in the 'Metaverse' Personal and Fun*, Microsoft Corporation, Redmond, WA, USA, Feb. 2022.

[3] M. Speicher, S. Cucerca, and A. Krüger, "VRShop: A mobile interactive virtual reality shopping environment combining the benefits of on- and offline shopping," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–31, Sep. 2017.

[4] T. Kliestik, A. Novak, and G. Lăzăroiu, "Live shopping in the metaverse: Visual and spatial analytics, cognitive artificial intelligence techniques and algorithms, and immersive digital simulations," *Linguistic Phil. Invest.*, vol. 2022, pp. 187–202, Jan. 2022.

[5] G. H. Popescu, C. F. Ciurlau, and C. I. Stan, "Virtual workplaces in the metaverse: Immersive remote collaboration tools, behavioral predictive analytics, and extended reality technologies," *Psychosociol. Issues Hum. Resour. Manag.*, vol. 10, no. 1, pp. 21–34, 2022.

[6] G. H. Popescu, K. Valaskova, and J. Horak, "Augmented reality shopping experiences, retail business analytics, and machine vision algorithms in the virtual economy of the metaverse," *J. Self-Governance Manag. Econ.*, vol. 10, no. 2, pp. 67–81, 2022.

[7] L. A. Hayduk, "Personal space: An evaluative and orienting overview," *Psychol. Bull.*, vol. 85, no. 1, p. 117, 1978.

[8] E. T. Hall, *The Hidden Dimension*, vol. 609. New York, NY, USA: Doubleday, 1966.

[9] R. Sommer, "Studies in personal space," *Sociometry*, vol. 22, no. 3, pp. 247–260, 1959.

[10] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Equilibrium theory revisited: Mutual gaze and personal space in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 10, no. 6, pp. 583–598, 2001.

[11] A. Yamazaki, N. Wakatsuki, K. Mizutani, Y. Okada, and K. Zempo, "Effects of audio-visual information on interpersonal distance in real and virtual environments," in *Proc. IFIP Conf. Hum.-Comput. Interact.* Berlin, Germany: Springer, 2021, pp. 405–410.

[12] A. Bönsch, S. Radke, J. Ehret, U. Habel, and T. W. Kuhlen, "The impact of a virtual agent's non-verbal emotional expression on a user's personal space preferences," in *Proc. 20th ACM Int. Conf. Intell. Virtual Agents*, Oct. 2020, pp. 1–8.

[13] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Interpersonal distance in immersive virtual environments," *Pers. Social Psychol. Bull.*, vol. 29, no. 7, pp. 819–833, Jul. 2003.

[14] K. Ahuja, E. Ofek, M. Gonzalez-Franco, C. Holz, and A. D. Wilson, "CoolMoves: User motion accentuation in virtual reality," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 2, pp. 1–23, Jun. 2021.

[15] D. Alais and D. Burr, "The ventriloquist effect results from near-optimal bimodal integration," *Current Biol.*, vol. 14, no. 3, pp. 257–262, Feb. 2004.

[16] P. W. Battaglia, R. A. Jacobs, and R. N. Aslin, "Bayesian integration of visual and auditory signals for spatial localization," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 20, no. 7, pp. 1391–1397, 2003.

[17] M. Brainard and E. Knudsen, "Experience-dependent plasticity in the inferior colliculus: A site for visual calibration of the neural representation of auditory space in the barn owl," *J. Neurosci.*, vol. 13, no. 11, pp. 4589–4608, Nov. 1993.

[18] G. H. Recanzone, "Rapidly induced auditory plasticity: The ventriloquism aftereffect," *Proc. Nat. Acad. Sci. USA*, vol. 95, no. 3, pp. 869–875, Feb. 1998.

[19] H. L. Pick, D. H. Warren, and J. C. Hay, "Sensory conflict in judgments of spatial direction," *Perception Psychophys.*, vol. 6, no. 4, pp. 203–205, Jul. 1969.

[20] P. Bertelson and M. Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," *Perception Psychophys.*, vol. 29, no. 6, pp. 578–584, Nov. 1981.

[21] J. Lewald, W. H. Ehrenstein, and R. Guski, "Spatio-temporal constraints for auditory–visual integration," *Behav. Brain Res.*, vol. 121, nos. 1–2, pp. 69–79, Jun. 2001.

[22] C. R. André, É. Corteel, J.-J. Embrechts, J. G. Verly, and B. F. G. Katz, "Subjective evaluation of the audiovisual spatial congruence in the case of stereoscopic-3D video and wave field synthesis," *Int. J. Hum.-Comput. Stud.*, vol. 72, no. 1, pp. 23–32, Jan. 2014.

[23] M. T. Wallace, G. E. Roberson, W. D. Hairston, B. E. Stein, J. W. Vaughan, and J. A. Schirillo, "Unifying multisensory signals across time and space," *Exp. Brain Res.*, vol. 158, no. 2, pp. 252–258, Sep. 2004.

[24] F. Melchior, S. Brix, T. Sporer, T. Roder, and B. Klehs, "Wave field syntheses in combination with 2D video projection," in *Proc. Audio Eng. Soc. Conf., 24th Int. Conf. Multichannel Audio, New Reality*. New York, NY, USA: Audio Engineering Society, 2003, Paper 47.

[25] P. Mannerheim, "Spatial sound and stereoscopic vision," in *Proc. Audio Eng. Soc. Conv.* New York, NY, USA: Audio Engineering Society, 2011, Paper 8424.

[26] S. Komiyama, "Subjective evaluation of angular displacement between picture and sound directions for hdtv sound systems," *J. Audio Eng. Soc.*, vol. 37, no. 4, pp. 210–214, 1989.

[27] D. H. Mershon, D. H. Desaulniers, T. L. Amerson, and S. A. Kiefer, "Visual capture in auditory distance perception: Proximity image effect reconsidered," *J. Auditory Res.*, vol. 20, no. 2, pp. 129–136, 1980.

[28] P. Zahorik, "Auditory and visual distance perception: The proximity-image effect revisited," *The J. Acoust. Soc. Amer.*, vol. 113, no. 4, p. 2270, 2003.

[29] L. Hládek, C. C. Le Dantec, N. Kopčo, and A. Seitz, "Ventriloquism effect and aftereffect in the distance dimension," in *Proc. Meetings Acoust.* Melville, NY, USA: Acoustical Society of America, 2013, Art. no. 050042.

[30] M. B. Gardner, "Proximity image effect in sound localization," *J. Acoust. Soc. Amer.*, vol. 43, no. 1, p. 163, 1968.

[31] F. Honbolygo, L. Veller, and V. Csepe, "Ventriloquism aftereffect in a virtual audio-visual environment," in *Proc. IEEE 3rd Int. Conf. Cogn. Infocommun. (CogInfoCom)*, Dec. 2012, pp. 475–478.

[32] R. Royston, "How humans relate: A new interpersonal theory by John Birtchnell. Published by Praeger, 1994," *Brit. J. Psychotherapy*, vol. 11, no. 4, pp. 637–638, 1995.

[33] J. Strayer and W. Roberts, "Children's personal distance and their empathy: Indices of interpersonal closeness," *Int. J. Behav. Develop.*, vol. 20, no. 3, pp. 385–403, Apr. 1997.

[34] J. A. Feeney, "Adult romantic attachment and couple relationships," in *Handbook of Attachment: Theory, Research and Clinical Applications*. New York, NY, USA: Guilford Press, 1999, pp. 355–377.

[35] M. Kaitz, Y. Bar-Haim, M. Lehrer, and E. Grossman, "Adult attachment style and interpersonal distance," *Attachment Hum. Develop.*, vol. 6, no. 3, pp. 285–304, Sep. 2004.

[36] D. M. Lloyd, "The space between us: A neurophilosophical framework for the investigation of human interpersonal space," *Neurosci. Biobehav. Rev.*, vol. 33, no. 3, pp. 297–304, Mar. 2009.

[37] A. Perry, O. Rubinsten, L. Peled, and S. G. Shamay-Tsoory, "Don't stand so close to me: A behavioral and ERP study of preferred interpersonal distance," *NeuroImage*, vol. 83, pp. 761–769, Dec. 2013.

[38] L. A. Hayduk, "The shape of personal space: An experimental investigation," *Can. J. Behav. Sci./Revue Canadienne Des Sci. du Comportement*, vol. 13, no. 1, p. 87, 1981.

[39] M. Gérin-Lajoie, C. L. Richards, and B. J. McFadyen, "The negotiation of stationary and moving obstructions during walking: Anticipatory locomotor adaptations and preservation of personal space *Motor Control*, vol. 9, no. 3, pp. 242–269, Jul. 2005.

[40] T. Amaoka, H. Laga, and M. Nakajima, "Modeling the personal space of virtual agents for behavior simulation," in *Proc. Int. Conf. CyberWorlds*, 2009, pp. 364–370.

[41] T. Iachini, Y. Coello, F. Frassinetti, V. P. Senese, F. Galante, and G. Ruggiero, "Peripersonal and interpersonal space in virtual and real environments: Effects of gender and age," *J. Environ. Psychol.*, vol. 45, pp. 154–164, Mar. 2016.

[42] R. Gifford, *Environmental Psychology: Principles and Practice*. Boston, MA, USA: Allyn & Bacon, 2007.

[43] C. Beaulieu, "Intercultural study of personal space: A case study," *J. Appl. Social Psychol.*, vol. 34, no. 4, pp. 794–805, Apr. 2004.

[44] M. Aliakbari, E. Faraji, and P. Pourshakibaee, "Investigation of the proxemic behavior of Iranian professors and university students: Effects of gender and status," *J. Pragmatics*, vol. 43, no. 5, pp. 1392–1402, Apr. 2011.

[45] J. J. Hartnett, K. G. Bailey, and C. S. Hartley, "Body height, position, and sex as determinants of personal space," *J. Psychol.*, vol. 87, no. 1, pp. 129–136, May 1974.

[46] L. Wagels, S. Radke, K. S. Goerlich, U. Habel, and M. Votinov, "Exogenous testosterone decreases men's personal distance in a social threat context," *Hormones Behav.*, vol. 90, pp. 75–83, Apr. 2017.

[47] M. L. Patterson, "A sequential functional model of nonverbal exchange," *Psychol. Rev.*, vol. 89, no. 3, p. 231, 1982.

[48] H. M. Rosenfeld, B. E. Breck, S. H. Smith, and S. Kehoe, "Intimacy-mediators of the proximity-gaze compensation effect: Movement, conversational role, acquaintance, and gender," *J. Nonverbal Behav.*, vol. 8, no. 4, pp. 235–249, 1984.

[49] C. Darwin, *The Expression of the Emotions in Man and Animals*. Chicago, IL, USA: Univ. of Chicago, 2015.

[50] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Social Psychol.*, vol. 17, no. 2, p. 124, 1971.

[51] R. W. Buck, V. J. Savin, R. E. Miller, and W. F. Caul, "Communication of affect through facial expressions in humans," *J. Pers. Social Psychol.*, vol. 23, no. 3, p. 362, 1972.

[52] R. Gifford, "Projected interpersonal distance and orientation choices: Personality, sex, and social situation," *Social Psychol. Quart.*, vol. 45, no. 3, pp. 145–152, 1982.

[53] K. B. Little, "Personal space," *J. Exp. Social Psychol.*, vol. 1, no. 3, pp. 237–247, 1965.

[54] D. P. Kennedy, J. Gläscher, J. M. Tyszka, and R. Adolphs, "Personal space regulation by the human amygdala," *Nature Neurosci.*, vol. 12, no. 10, pp. 1226–1227, Oct. 2009.

[55] G. Schoretsanitis, A. Kutynia, K. Stegmayer, W. Strik, and S. Walther, "Keep at bay!—Abnormal personal space regulation as marker of paranoia in schizophrenia," *Eur. Psychiatry*, vol. 31, pp. 1–7, Jan. 2016.

[56] R. Welsch, C. von Castell, and H. Hecht, "The anisotropy of personal space," *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0217587.

[57] H. Hecht, R. Welsch, J. Viehoff, and M. R. Longo, "The shape of personal space," *Acta Psychol.*, vol. 193, pp. 113–122, Feb. 2019.

[58] H. C. Miller, A.-S. Chabriac, and M. Molet, "The impact of facial emotional expressions and sex on interpersonal distancing as evaluated in a computerized stop-distance task," *Can. J. Exp. Psychol./Revue Canadienne Psychologie Expérimentale*, vol. 67, no. 3, p. 188, 2013.

[59] A. Bonsch, B. Weyers, J. Wendt, S. Freitag, and T. W. Kuhlen, "Collision avoidance in the presence of a virtual agent in small-scale virtual environments," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2016, pp. 145–148.

[60] S. Narang, A. Best, T. Randhavane, A. Shapiro, and D. Manocha, "PedVR: Simulating gaze-based interactions between a real user and virtual crowds," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol.*, Nov. 2016, pp. 91–100.

[61] F. A. Sanz, A.-H. Olivier, G. Bruder, J. Pettrè, and A. Lècuyer, "Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2015, pp. 75–80.

[62] T. Iachini, Y. Coello, F. Frassinetti, and G. Ruggiero, "Body space in social interactions: A comparison of reaching and comfort distance in immersive virtual reality," *PLoS ONE*, vol. 9, no. 11, Nov. 2014, Art. no. e111511.

[63] J. Llobera, B. Spanlang, G. Ruffini, and M. Slater, "Proxemics with multiple dynamic characters in an immersive virtual environment," *ACM Trans. Appl. Perception*, vol. 8, no. 1, pp. 1–12, Oct. 2010.

[64] N. E. Miller, "Liberalization of basic SR concepts: Extensions to conflict behavior, motivation and social learning," *Psychol., Study Sci., Study*, vol. 2, pp. 196–292, 1959.

[65] M. D'Angelo, G. di Pellegrino, and F. Frassinetti, "Invisible body illusion modulates interpersonal space," *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, Dec. 2017.

[66] A. Parasuraman, V. Zeithaml, and L. Berry, "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," *J. Retailing*, vol. 64, no. 1, pp. 12–40, Spring 1988.

[67] M. Söderlund, E.-L. Oikarinen, and T. M. Tan, "The hard-working virtual agent in the service encounter boosts customer satisfaction," *Int. Rev. Retail, Distrib. Consum. Res.*, vol. 32, no. 4, pp. 388–404, 2022.

[68] T. Otterbring, F. Wu, and P. Kristensson, "Too close for comfort? The impact of salesperson-customer proximity on consumers' purchase behavior," *Psychol. Marketing*, vol. 38, no. 9, pp. 1576–1590, May 2021.

[69] M. Söderlund and E.-L. Oikarinen, "Joking with customers in the service encounter has a negative impact on customer satisfaction: Replication and extension," *J. Retailing Consum. Services*, vol. 42, pp. 55–64, May 2018.

[70] M. Söderlund, "The proactive employee on the floor of the store and the impact on customer satisfaction," *J. Retailing Consum. Services*, vol. 43, pp. 46–53, Jul. 2018.

[71] J. D. Gfeller, S. J. Lynn, and W. E. Pribble, "Enhancing hypnotic susceptibility: Interpersonal and rapport factors," *J. Pers. Social Psychol.*, vol. 52, no. 3, p. 586, 1987.

[72] D. D. Gremler and K. P. Gwinner, "Customer-employee rapport in service relationships," *J. Service Res.*, vol. 3, no. 1, pp. 82–104, Aug. 2000.

[73] D. S. Sundaram and C. Webster, "The role of nonverbal communication in service encounters," *J. Services Marketing*, vol. 14, no. 5, pp. 378–391, Sep. 2000.

[74] W.-C. Tsai, "Determinants and consequences of employee displayed positive emotions," *J. Manage.*, vol. 27, no. 4, pp. 497–512, Aug. 2001.

[75] W.-C. Tsai and Y.-M. Huang, "Mechanisms linking employee affective delivery and customer behavioral intentions," *J. Appl. Psychol.*, vol. 87, no. 5, p. 1001, 2002.

[76] F. J. Bernieri, J. S. Gillis, J. M. Davis, and J. E. Grahe, "Dyad rapport and the accuracy of its judgment across situations: A lens model analysis," *J. Pers. Social Psychol.*, vol. 71, no. 1, p. 110, 1996.

[77] D. D. Gremler and K. P. Gwinner, "Rapport-building behaviors used by retail employees," *J. Retailing*, vol. 84, no. 3, pp. 308–324, Sep. 2008.

[78] T. Arai, Y. Chida, Y. Okada, and K. Zempo, "Sensor network to measure MAAI on value co-creation process: Feasibility study of MAAI optimization on customer service," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, Sep. 2019, pp. 1–4.

[79] K. Zempo, T. Arai, T. Aoki, and Y. Okada, "Sensing framework for the internet of actors in the value co-creation process with a beacon-attachable indoor positioning system," *Sensors*, vol. 21, no. 1, p. 83, Dec. 2020.

[80] D. Jo, K. Kim, G. F. Welch, W. Jeon, Y. Kim, K.-H. Kim, and G. J. Kim, "The impact of avatar-owner visual similarity on body ownership in immersive virtual reality," in *Proc. 23rd ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2017, pp. 1–2.

[81] P. Heidicker, E. Langbehn, and F. Steinicke, "Influence of avatar appearance on presence in social VR," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2017, pp. 233–234.

[82] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch, "The effect of avatar realism in immersive social virtual realities," in *Proc. 23rd ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2017, pp. 1–10.

[83] D. Maloney, G. Freeman, and D. Y. Wohn, "'Talking without a voice' understanding non-verbal communication in social virtual reality," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, pp. 1–25, Oct. 2020.

[84] S.-A.-A. Jin and J. Bolebruch, "Avatar-based advertising in second life: The role of presence and attractiveness of virtual spokespersons," *J. Interact. Advertising*, vol. 10, no. 1, pp. 51–60, Sep. 2009.

[85] J. H. Moon, E. Kim, S. M. Choi, and Y. Sung, "Keep the social in social media: The role of social interaction in avatar-based virtual shopping," *J. Interact. Advertising*, vol. 13, no. 1, pp. 14–26, Jan. 2013.

[86] Y. Zhao, N. Baghaei, A. Schnack, and L. Stemmet, "Assessing telepresence, social presence and stress response in a virtual reality store," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2021, pp. 52–56.

[87] N. Kuratomo, H. Miyakawa, S. Masuko, T. Yamanaka, and K. Zempo, "Effects of acoustic comfort and advertisement recallability on digital signage with on-demand pinpoint audio system," *Appl. Acoust.*, vol. 184, Dec. 2021, Art. no. 108359.

[88] Y. Ochiai, T. Hoshi, and I. Suzuki, "Holographic whisper: Rendering audible sound spots in three-dimensional space by focusing ultrasonic waves," in *Proc. CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, May 2017, pp. 4314–4325.

[89] Y. Mashiba, R. Iwaoka, H. E. B. Salih, M. Kawamoto, N. Wakatsuki, K. Mizutani, and K. Zempo, "Spot-presentation of stereophonic earcons to assist navigation for the visually impaired," *Multimodal Technol. Interact.*, vol. 4, no. 3, p. 42, Jul. 2020.

[90] P. M. Hofman, J. G. A. Van Riswick, and A. J. Van Opstal, "Relearning sound localization with new ears," *Nature Neurosci.*, vol. 1, pp. 417–421, Sep. 1998.

[91] D. Uzzell and N. Horne, "The influence of biological sex, sexuality and gender role on interpersonal distance," *Brit. J. Social Psychol.*, vol. 45, no. 3, pp. 579–597, Sep. 2006.

[92] D. J. Tollin, J. L. Ruhland, and T. C. T. Yin, "The role of spectral composition of sounds on the localization of sound sources by cats," *J. Neurophysiol.*, vol. 109, no. 6, pp. 1658–1668, Mar. 2013.

[93] M. E. Nilsson and B. N. Schenkman, "Blind people are more sensitive than sighted people to binaural sound-location cues, particularly interaural level differences," *Hearing Res.*, vol. 332, pp. 223–232, Feb. 2016.

[94] M. L. Hawley, R. Y. Litovsky, and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Amer.*, vol. 115, no. 2, pp. 833–843, Feb. 2004.

[95] M. S. Brainard, E. I. Knudsen, and S. D. Esterly, "Neural derivation of sound source location: Resolution of spatial ambiguities in binaural cues," *J. Acoust. Soc. Amer.*, vol. 91, no. 2, pp. 1015–1027, Feb. 1992.

[96] F. Keyrouz, "Advanced binaural sound localization in 3-D for humanoid robots," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 9, pp. 2098–2107, Sep. 2014.

[97] P. Voss, F. Lepore, F. Gougoux, and R. J. Zatorre, "Relevance of spectral cues for auditory spatial processing in the occipital cortex of the blind," *Frontiers Psychol.*, vol. 2, p. 48, Mar. 2011.

[98] R. Baumgartner, D. K. Reed, B. Tóth, V. Best, P. Majdak, H. S. Colburn, and B. Shinn-Cunningham, "Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 36, pp. 9743–9748, Sep. 2017.

[99] A. R. Palmer, T. M. Shackleton, and D. McAlpine, "Neural mechanisms of binaural hearing," *Acoust. Sci. Technol.*, vol. 23, no. 2, pp. 61–68, 2002.

[100] P. Stitt, L. Picinali, and B. F. G. Katz, "Auditory accommodation to poorly matched non-individual spectral localization cues through active learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, Dec. 2019.

[101] T. Kliestik, K. Zvarikova, and G. Lăzăroiu, "Data-driven machine learning and neural network algorithms in the retailing environment: Consumer engagement, experience, and purchase behaviors," *Econ., Manag. Financial Markets*, vol. 17, no. 1, pp. 57–69, 2022.

[102] E. Hopkins, "Machine learning tools, algorithms, and techniques," *J. Self-Governance Manag. Econ.*, vol. 10, no. 1, pp. 43–55, 2022.

[103] J. A. Jones, J. E. Swan, G. Singh, and S. R. Ellis, "Peripheral visual information and its effect on distance judgments in virtual and augmented environments," in *Proc. ACM SIGGRAPH Symp. Appl. Perception Graph. Vis. (APGV)*, 2011, pp. 29–36.

[104] I. V. Piryankova, S. de la Rosa, U. Kloos, H. H. Bülthoff, and B. J. Mohler, "Egocentric distance perception in large screen immersive displays," *Displays*, vol. 34, no. 2, pp. 153–164, Apr. 2013.

[105] L. E. Buck, M. K. Young, and B. Bodenheimer, "A comparison of distance estimation in HMD-based virtual environments with different HMD-based conditions," *ACM Trans. Appl. Perception*, vol. 15, no. 3, pp. 1–15, Aug. 2018.

[106] C. S. Sahm, S. H. Creem-Regehr, W. B. Thompson, and P. Willemsen, "Throwing versus walking as indicators of distance perception in similar real and virtual environments," *ACM Trans. Appl. Perception*, vol. 2, no. 1, pp. 35–45, Jan. 2005.

[107] J. Andre and S. Rogers, "Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis," *Perception Psychophys.*, vol. 68, no. 3, pp. 353–361, Apr. 2006.

[108] J. M. Loomis and J. W. Philbeck, "Measuring spatial perception with spatial updating and action," in *Proc. Carnegie Symp. Cogn.* Pittsburgh, PA, USA: Psychology Press, 2008, pp. 17–60.

[109] G. W. Evans and R. B. Howard, "Personal space," *Psychol. Bull.*, vol. 80, no. 4, p. 334, 1973.

KEIICHI ZEMPO (Member, IEEE) received the B.Sc. degree in physics from the College of Natural Science, University of Tsukuba, in 2008, the M.B.A. degree from the Department of Business Administration and Public Policy, University of Tsukuba, in 2010, and the Ph.D. degree in engineering from the Department of Intelligent Interaction Technologies, University of Tsukuba, in 2013.

He worked with the Center for Service Engineering, National Institute of Advanced Industrial Science and Technology (AIST), from 2013 to 2014. He is currently an Assistant Professor with the University of Tsukuba, a PRESTO Researcher with Japan Science and Technology Agency, and a CEO of Xtrans tech Inc. His research interests include human augmentation, sense substitution, service engineering, telepresence, and xR.

Dr. Zempo is a member of the Association for Computing Machinery (ACM), the Acoustics Society of Japan (ASJ), the Society for Serviceology (SfS), the Japanese Society for Artificial Intelligence (JSAI), and the Virtual Reality Society of Japan (VRSJ).

AZUSA YAMAZAKI received the B.Eng. degree from the University of Tsukuba, in 2021, where she is currently pursuing the master's degree with the Graduate School of Science and Technology. Her research interest includes spatial design.

NAOTO WAKATSUKI received the B.Eng., M.Eng., and D.Eng. degrees from the University of Tsukuba, in 1993, 1995, and 2004, respectively.

He was with Okayama University, from 1995 to 2001, and with Akita Prefectural University, from 2001 to 2006. He is currently an Associate Professor with the University of Tsukuba. His research interests include acoustic instrumentation, simulation-based visualization, vibration sensors and actuators, acoustical engineering, musical acoustics, and inverse problems.

Dr. Wakatsuki was affiliated with academic societies, including the Acoustical Society of Japan, the Acoustical Society of America, the Society of Agricultural Structures, and the Japan Society for Simulation Technology.

KOICHI MIZUTANI graduated from the National Defense Academy (NDA), Japan, in 1979. He received the Ph.D. degree in engineering from Kyoto University, in 1990.

He was a Researcher with the Department of Electrical Engineering, NDA, from 1984 to 1988, and the Department of Research, Communication and Intelligence School, Japanese Ground Self Defense Force (JGSDF), from 1988 to 1990. He was the Deputy Director of the Secretariat and the Director General of the National Defense Agency, from 1991 to 1992. He retired from JGSDF at the rank of Major. He joined the Institute of Applied Physics, University of Tsukuba, as a Faculty Member, as an Assistant Professor, in 1992, as an Associate Professor, in 1998, and as a Full Professor, in 2004, where he is currently a Researcher (full time) with the Faculty of Engineering, Information and Systems. He was given the title of a Professor Emeritus at the University of Tsukuba, in 2021. His research interests include ultrasonic electronics, medical electronics, welfare technologies, complementation of human sensory functions, robot sensing, communication systems in sensing grids, environment monitoring, applied optics, applied acoustics, musical acoustics, food and agricultural engineering, and the health monitoring engineering of livestock.

Dr. Mizutani is a member of the Acoustics Society of Japan (ASJ), the Marine Acoustics Society of Japan (MASJ), the Society of Agricultural Structures, Japan (SASJ), the Japan Society of Civil Engineering (JSCE), and the Japan Society of Applied Physics (JSAP).

YUKIHIKO OKADA received the Ph.D. degree in commerce and management from Hitotsubashi University, in 2006. Since 2006, he has been working with the University of Tsukuba, where he is currently an Associate Professor. Since 2010, he has been a Visiting Associate Professor with the Institute of Statistical Mathematics. Since 2017, he has also been a Chief Scientist with the Department of Service Engineering, Center for Artificial Intelligence Research, University of Tsukuba. In 2010, he was awarded the Japan Accounting Association Award by the Japan Accounting Association for his research of Service Target Costing. In 2021, he was awarded the Engineering Education Award by the Japanese Society for Engineering Education for his achievement in establishing and managing the ‘‘Master's Program in Service Engineering.’’