Human Factors in Virtual Environments and Game Design, Vol. 50, 2022, 278–284
https://doi.org/10.54941/ahfe1002089

Effect of Visibility of Auditory Stimulus Location on Ventriloquism Effect Using AR-Head-Mounted Display

Kaoru Kawai and Kenji Muto

Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548, Japan

ABSTRACT

Herein, we investigate the effect of the visibility of a loudspeaker on the ventriloquism effect. For this purpose, the discrimination threshold of angle at which the locations of a voice and an avatar image are being perceived to be displaced was measured. The discrimination threshold of angle was measured under two conditions: where the loudspeaker was visible and where the loudspeaker was not visible. From the results, the discrimination threshold of angle was more significant under the condition where the loudspeaker was visible than under the condition where the loudspeaker was not visible, demonstrating that the ventriloquism effect is considerably increased when the loudspeaker is visible.

Keywords: Augmented reality, Cross-modality, Audiovisual stimulus

INTRODUCTION

Virtual reality (VR) or augmented reality (AR) games using head-mounted displays (HMDs) are becoming increasingly popular in recent games. These games can present wider visual stimuli than TVs or handheld games. Moreover, in these games, the location of auditory stimuli is presented at the same location as visual stimuli. Currently, a method of presenting auditory stimuli at the same location as visual stimuli is stereophonic sound, which is realized using headphones or by controlling audio radiated from loudspeakers. The problem with this method is that it does not allow users to talk to each other, and the equipment is extremely large for daily life. Therefore, we propose presenting visual stimuli to auditory stimuli rather than presenting auditory stimuli to visual stimuli. When presenting visual stimuli to auditory stimuli, it is necessary to specify how far the locations of the sound source and visual stimuli can be shifted. Thus, we examined varying degrees of spatial disparity between auditory and visual stimuli to determine whether they are still perceived as originating from the same location.

A cross-modality between the locations of auditory and visual stimuli is the ventriloquism effect (Jackson, 1953), which refers to perceiving the location of auditory stimuli as coming from the location of visual stimuli when the two locations are displaced. Many researchers examined varying degrees of spatial disparity between auditory and visual stimuli and determined whether they are still perceived as originating from the same location. Hairston et al.

examined the spatial unity of auditory and visual stimuli (Hairston et al, 2003). Kimura et al. investigated the differential threshold of angle between auditory and visual stimuli location (Kimura et al, 1999). Mikko et al. examined the angle at which the location of audiovisual stimuli seemed to be separated (Kytö et al, 2015). The subjects used in these previous studies are visually unaware of loudspeakers radiating auditory stimuli because of the stereoscopic audio. However, when visual stimuli are presented through a loudspeaker via AR (i.e., AR glasses), the locations of loudspeakers are recognized. Moreover, there is no study on the effect of visibility of a loudspeaker playing a sound on the ventriloquism effect.

In this study, we aim to clarify the effect of visibility of a loudspeaker playing a sound on the ventriloquism effect. For this purpose, we conducted AR and VR condition experiments to determine whether auditory and visual stimuli are originating from the same location when there are varying degrees of spatial disparity between them. We defined the discrimination threshold angle as the boundary angle at which auditory and visual stimuli are perceived to originate from the same location when they are presented at different angles. Comparing these two experiments will clarify the effect of visual perception of the source of sound on the ventriloquism effect.

METHODS

This study aims to clarify the effect of the visibility of a loudspeaker playing a sound on the ventriloquism effect. For this purpose, two experiments, namely, AR and VR condition experiments, were conducted to investigate varying degrees of spatial disparity between auditory and visual stimuli and determine whether they are still perceived as originating from the same location. In the AR condition experiment, a voice was played through a single loudspeaker in front of the subjects. A 3D avatar image was shown to the subjects at a different angle using a see-through HMD. The subjects determined their perception of the shift in the locations of auditory and visual stimuli. In the VR condition experiment, multiple loudspeakers were lined up, and a voice was played from one of the loudspeakers. A 3D avatar image was presented at an angle different from that of the playing loudspeaker using an HMD. The subjects determined their perception of the shift in the locations of auditory and visual stimuli. The results of these measurements were analyzed using the constant method, a discrimination threshold of angle measurement method, and the discrimination threshold of angles was calculated. The differences in the discrimination threshold of angles between these two experiments were compared. Thus, the influence of a visual stimuli, a loudspeaker playing a sound, on the ventriloquism effect was clarified.

Measurement Method of Discrimination Threshold of Angle

The discrimination threshold of angle was examined on the basis of a subject's experience. The subjects were presented with a voice and 3D avatar image, and they determined a mismatch in the voice and image's locations.

The discrimination threshold of angle measurement method used in the experiment was the constant method in which the subjects were presented with audiovisual stimuli in multiple conditions with varying shifts in location and asked to make two-choice responses for each change range 20 times. The differential threshold is the stimulus value that becomes the 50% response point when the cumulative standard normal distribution is fitted to the response rate obtained by dividing results of two-choice responses by the total number of responses. In this study, when the visual stimulus presentation location was changed to the auditory stimulus presentation location, the subjects were asked to respond when they believed that the location of auditory and visual stimuli did not match. Then, as a response rate, the mismatch rate was calculated as the percentage of subjects who responded that the auditory and visual stimuli were not matched.

Audiovisual Stimuli

This section describes the audiovisual stimuli used in this experiment. The left side of Figure 1 shows audiovisual stimuli presented in the AR condition experiment, and the right side shows audiovisual stimuli presented in the VR condition experiment. Audio-visual stimuli included voice audio and video of a speaking 3D avatar. The auditory stimulus was the voice sound of self-introduction, "Konnichiwa Watashi Unity-chan" (The meaning of this phrase is "Hello, I am Unity-chan"), included in the Unity-chan package. The length of the voice data was 2.5 s. The voice level was 65 dB of A-weighted sound pressure level (LAeq) at the center of the subject's head. The visual stimulus was a 3D female avatar, "Unity-chan," provided by Unity Technology Japan. The avatar was presented in a standing location, and lip-synching was performed to the voice sound used for auditory stimuli. In the AR condition experiment, a loudspeaker on a tripod was presented as the background. In the VR condition experiment, black ground and a blue sky were presented as the background.

EXPERIMENTS

Discrimination Threshold of Angle Measurement Experiment Under AR Condition

The experiment was conducted at a large laboratory, Toyosu Campus, Shibaura Institute of Technology. Figure 2 shows the locations of the loudspeaker (Loudspeaker Unit: FE103En, Loudspeaker Enclosure: 180 × 200 × 160 [mm]) and HMD (Oculus Quest) used. The loudspeakerwas used as an auditory stimulus presentation device. The loudspeaker was placed 2 m in the frontal direction of the subject (0°). The height of the loudspeaker was 1.4 m, which corresponded to the mouth level of an average-height Japanese woman. Auditory stimuli were output from the HMD to the subject, radiated by the loudspeaker, and connected to an audio amplifier (YAMAHA, A-S01). The audio amplifier outputs auditory stimuli to the loudspeaker, and the HMD was used as a visual stimulus presentation device, and one of its functions, "Passthrough," was used as a see-through HMD to project outside

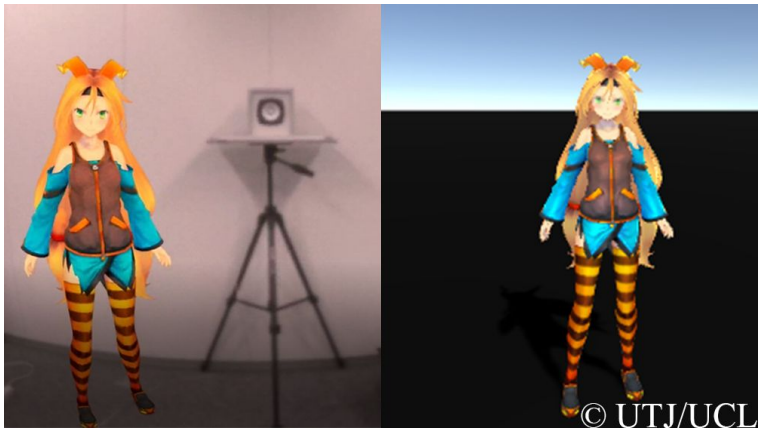

Figure 1: Posture of the female avatar used in the experiment. The left of the figure shows the view of the AR experiment condition, whereas the right side of the figure shows the view of the VR experimental condition. Lip-sync was used to make the mouth move correspondingly to the voice data.

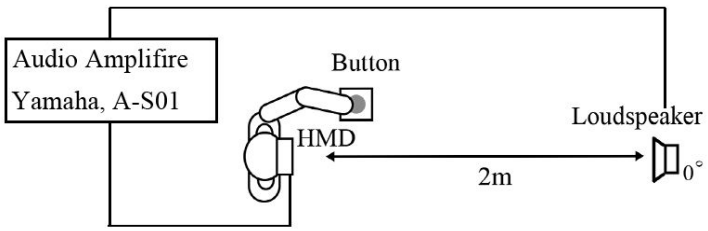

Figure 2: Experimental setup for AR condition.

images on the HMD. The presented stimuli were controlled using an Android application created using the 3D game engine Unity.

Each subject was brought in the experimental room and seated in a height-adjusted chair while wearing the HMD. First, as shown in Figure 2, we adjusted the subject's direction to set the loudspeaker's direction to 0°. Next, we adjusted images displayed on the HMD such that the 0° of the HMD corresponded to the 0° of the experimental system. The avatar direction was presented in the range of 0°–20° in intervals of 4°: 0°, 4°, 8°, 12°, 16°, and 20°. (A positive value shows the right-hand side of a subject.) There are six audiovisual stimulus patterns. The subjects were shown these six audio-visual stimulus patterns 20 times for each combination for a total of 120 times. These audiovisual stimuli were presented in random combinations. The sub-jects were in-structed to press the button of the HMD if they believed that the locations of the sound and image were shifted.

Discrimination Threshold of Angle Measurement Experiment Under VR Condition

The VR condition experiment was conducted in a large room, the same room as the AR condition experiment (Kaoru et al, 2021). The subject,

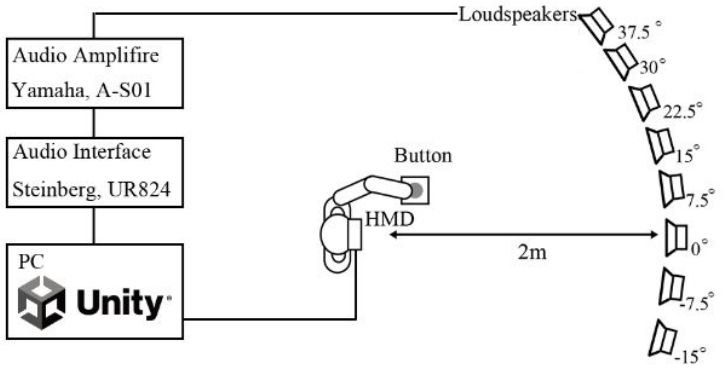

Figure 3: Experimental setup for VR conditions.

loudspeaker, and HMD are located, as shown in Figure 3. The auditory stimulus presentation device was a loudspeaker. When the visual presentation range is more expansive than the loudspeaker placement range, it becomes obvious that the sound is not presented from the visual location. Therefore, it is necessary to place the speakers in the same range as the visual presentation range. Eight loudspeakers were circularly set on a horizontal plane centered on the subject in intervals of 7.5°: −15°, −7.5°, 0°, +7.5°, +15°, +22.5°, +30°, and +37.5° (a positive value represents the right-hand side of a subject). The loudspeaker array had a radius of 2.0 m and height of 1.4 m. The visual stimulus presentation device is an HMD. The background in this VR condition experiment was a large floor and blue sky. HMD was connected to a PC, and audiovisual stimuli were presented using software created with Unity. Visual stimuli were displayed on the HMD. Auditory stimuli were presented by selecting the loudspeaker using Unity, audio interface (Steinberg, UR824), and audio amplifier. Each subject was seated in a height-adjustable chair, ensuring that the height of the loudspeaker array and the height of the center of the subject's head was similar.

Each subject was seated in a height-adjusted chair while wearing the HMD. First, as shown in Figure 3, we adjusted the HMD such that the 0° direction to the loudspeaker's direction was set to 0°. Next, we adjusted the subject's direction to set the loudspeaker's direction to 0°. Auditory stimuli were presented from a loudspeaker in the 0° direction. However, visual stimuli were presented from the 0° of the loudspeaker at 5° increments in the range of 0°–25°. As dummy stimuli to prevent subjects from perceiving the location of auditory stimuli, we presented auditory stimuli from the 15° and 30° loudspeakers. Visual stimuli were presented from the 15° loudspeaker at 5° increments in the range of −25°–25° and the 30° loudspeaker at 5° increments in the range of 0°–25°. The subjects were shown these 23 audiovisual stimulus patterns 20 times for each combination for a total of 460 times. These audiovisual stimuli were presented in random combinations, and the subjects were then instructed to press the button of the HMD when they felt a mismatch in location between the audio and image.

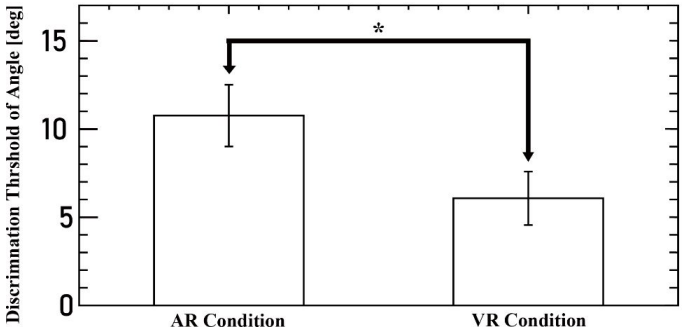

Figure 4: Measurement result of differential threshold of angle in AR and VR condition experiments (*: $p < 0.05$).

RESULT

Experimental Result

Nine subjects (seven males and two females in the age range of 19–21 years, and a mean age of 21.0) participated in the AR condition experiment. Moreover, seven subjects (five males and two females, within the age range of 19–21 years, and a mean age of 20.6) participated in the VR condition experiment. Seven subjects (five males and two females) participated in both condition experiments. The VR condition experiment was conducted first, followed by the AR condition experiment. All subjects were native Japanese speakers having a normal hearing acuity.

Figure 4 shows the experimental results of the AR and VR condition experiments. The discrimination threshold of angle is the vertical axis of the graph. Each bar is the mean of the discrimination threshold of angle, whereas the error bars are the standard division of the discrimination threshold of angle. The left is the discrimination threshold of angle in the AR condition experiment, and the right is the differential threshold in the VR condition experiment. The discrimination threshold of angle for the AR condition experiment was 10.6° ($SD = \pm 1.7°$, $n = 9$), whereas that for the VR condition was 6.1° ($SD = \pm 1.5°$, $n = 7$).

The mean of the discrimination threshold of angle in the AR condition experiment was greater than that of the discrimination threshold of angle in the VR condition experiment. The discrimination thresholds of angles in the AR and VR condition experiments were analyzed using Welch's t-test ($p = 0.05$). From the experimental results, there was a significant difference between the two condition experiments' mean values.

Discussion

In the previous study (Hairston et al, 2003; Kytö et al, 2015), the range of location shift between auditory and visual stimuli measured was under a condition where the sound was not visible. Here, we construct a system that allows everyone to experience extended reality by simply projecting images on loudspeakers. This study suggested that the ventriloquism effect is more significant when the location where the audio is being played is visible.

The discrimination threshold of angle under the VR condition, where the loudspeaker location was not visible, was approximately 6.2°. The mismatch rate of this experiment probably agreed with the mismatch rate of Hairston et al.'s experiment (Hairston et al, 2003) where they did not know which loudspeaker was playing, and the audio could not determine the place where auditory stimuli were being presented.

In the AR condition experiment, we considered two reasons for this increase in the discrimination threshold of angle. The first reason was that two visual stimuli, a loudspeaker and an avatar, caused the discrimination threshold of angle to become significant in the AR condition experiment. These visual stimuli forced the subjects to simultaneously focus on both the loudspeaker and 3D avatar image, which may have reduced their ability to concentrate. The discrimination threshold angle can be expected to become more significant when multiple visual stimuli are presented. The second reason is that images presented as background in the AR condition experiment were of low quality. In the AR condition experiment, we used "Passthrough," a feature of Oculus, to present images; however, the image quality was lower than that of the 3D avatar image. Therefore, certain subjects felt uncomfortable. We considered that this caused subjects not to concentrate on the location mismatch between auditory and visual stimuli, thus making the discrimination threshold angle to be significant.

CONCLUSION

In this study, we examined the differential threshold of angle of perceived location mismatch between auditory and visual stimuli using HMDs and loudspeakers through AR and VR condition experiments. From experimental results, the discrimination threshold of angle to the 0° loudspeaker was 10.6° under the AR condition and 6.2° under the VR condition. There was a significant difference in the differential threshold of angle between the AR and VR conditions. This result has implications for developing an audiovisual system based on previous studies on the ventriloquism effect. When the loudspeaker location is visible, the system can project a broader image to audio than when it is unclear.

REFERENCES

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. Journal of cognitive neuroscience, 15(1), 20–29.

Jackson, C. V. (1953). Visual factors in auditory localization. Quarterly Journal of Experimental Psychology, 5(2), 52–65.

Kaoru K, Kenji M. (2021). Difference between the discrimination angle of sound image to image and image to sound image. The 26th Annual Conference of the VRSJ, 2B3-6 (in Japanese).

Kimura, M., Kajii, M., Takahashi, M., & Yamamoto, K., (1999). The ventriloquism effect in peripheral vision, 4(1), 253–260.

Kytö, M., Kusumoto, K., & Oittinen, P. (2015, November). The ventriloquist effect in augmented reality. In 2015 IEEE International Symposium on Mixed and Augmented Reality (pp. 49–53). IEEE.