Acoustics'08
Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Relationship between a visual stimulus with a feeling of depth and its equivalent sound pressure level (ESPL)

Hiroshi Hasegawa, Hirotaka Ono, Takumi Ito, Ichiro Yuyama, Masao Kasuga and Miyoshi Ayama

Utsunomiya Univ., 7-1-2 Yoto, Tochigi-ken, 321-8585 Utsunomiya-shi, Japan
hasegawa@is.utsunomiya-u.ac.jp

This study investigated the equivalent perception between a visual stimulus and its associated sound. Experiments were performed of an auditory-visual stimulus presentation using an audio-video clip of a man beating a drum on a road, which had a feeling of depth with a perspective view of the road. There were four kinds of distance between the visual target (a man beating a drum) and the video camera to capture the target of 5, 10, 20, and 40 m. First, we evaluated the sound pressure level matching with the visual target in each presentation pattern (equivalent sound pressure level; ESPL). Next, we evaluated the point of subjective simultaneity (PSS) between the auditory and visual stimuli in each presentation pattern. Finally, based on the results of ESPL and PSS, we combined the visual and auditory stimuli in each distance with varying both the sound pressure level from −12 dB to 12 dB of the ESPL and the time delay between the auditory and visual stimuli from −8 F to 8 F (1 F = 1/30 s), where "+" indicates that the visual event preceded the sound, and carried out an experiment of the auditory-visual stimulus presentation. As a result, the ESPL intended to decrease when the delay time increased (the sound was delayed).

1 Introduction

Recent multimedia technologies have made it possible to construct various audio-video environments. Ii is, however, difficult to reproduce an auditory-visual space with feeling of being in the actual space. It is known that the feeling of correspondence between the auditory and visual information is one of the most important factors for reproducing an auditory-visual space with actual feeling. Many studies have been made on auditory-visual interactions from a psychological viewpoint [1]–[7]. There are, however, few studies on the interactions applied to actual auditory-visual environments [8]–[10].

In this paper, we focused on the equivalent perception between a visual stimulus and its associated sound. We employed a video clip of a man beating a drum on a road, which had a feeling of depth with a perspective view, as the visual stimulus and its drum sound as the auditory stimulus. First, we evaluated the equivalent sound pressure level (ESPL), that is the sound pressure level to provide a perceptual strength equivalent to that of the visual target, in each presentation pattern. Next, we estimated the point of subjective simultaneity (PSS) between the auditory and visual stimuli in each presentation pattern. Finally, based on the results of ESPL and PSS, we carried out an experiment to investigate the equivalent perception between the auditory and visual stimuli when varying both the sound pressure level and the time delay between the auditory and visual stimuli at various distances from the visual target.

2 Experimental environment

2.1 Experimental apparatus

Figure 1 shows a block diagram of the experimental apparatus. The visual stimulus was played using a digital video player (SONY HDR-HC1) and was projected onto a screen using a projector (EPSON EMP-TW600). The projected area on the screen was 2.09 m (W) × 1.17 m (H) as shown in Fig. 2. The pixel number of the projector display was 1440 (W) × 1080 (H). The subject was seated on a chair placed at a distance of 2.6 m from the center of the screen. The viewing angles from the subject to the projected area were 43.8 degrees in the horizontal direction and 25.4 degrees in the vertical direction. The auditory stimulus was presented via headphones (SENNHEISER HD-595).

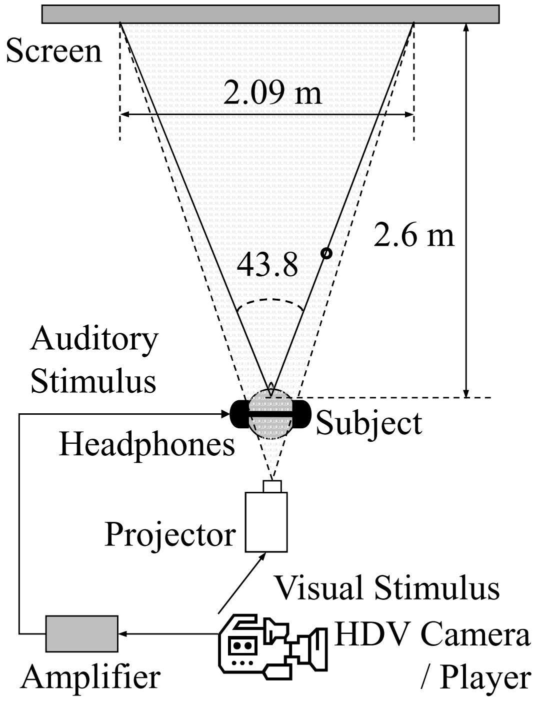


Figure 1: Experimental apparatus. The visual stimulus was played using a digital video player and was projected onto a screen using a projector. The sound stimulus was presented via headphones to the subject.

2.2 Auditory-visual stimulus

A video clip of a man beating a drum on a road (Fig. 2) and its drum sound were used as the visual and auditory stimuli, respectively. The visual stimulus, which had a feeling of depth with a perspective view of the road, was captured using a digital video camera (SONY HDR-HC1). The distance between the visual target (a man beating a drum) and the video camera was set at 5, 10, 20, or 40 m. Zooming level was set to give the same perspective as that of human visual system so that perceptual distance of the visual target in the video clip was approximately the same as the physical distance between the target and the video camera. We call the latter as "the presentation distance" in this study.

The auditory stimulus was recorded using a microphone (B&K 4190) near the drum, and the auditory stimulus at each presentation distance was produced

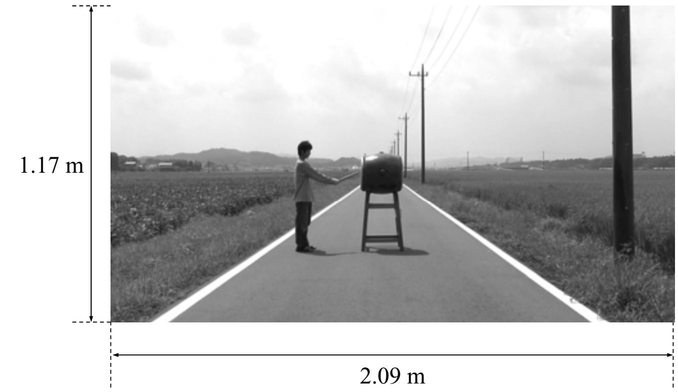


Figure 2: Projected area on the screen. A video clip of a man beating a drum on a road was presented to the subject. In this case the distance between the visual target and the video camera was 10 m.

by convoluting the recorded stimulus with the impulse response corresponding to each presentation distance. Thus, in this study, we took into account not only the SPL but also the impulse response of the auditory stimulus in each presentation distance.

2.3 Subjects

Eight male subjects in their early 20's were employed in the experiment. All subjects had normal or corrected-to-normal vision and normal hearing acuity.

3 Measurement of ESPL

In this section, we evaluated the equivalent sound pressure level (ESPL), i.e., the sound pressure level to provide a perceptual strength equivalent to that of the visual target.

3.1 Procedure

Four video clips of a man beating a drum at distances of 5, 10 (Fig. 2), 20, and 40 m from the video camera (presentation distance) and nine sound level differences of 0, ±3, ±6, ±9, and ±12 dB compared to the SPL (sound pressure level) by actual measurement of the beating sound at each distance were employed as the experimental stimuli, as shown in the center column of Table 1. We combined the video clip in each presentation distance and the corresponding auditory stimulus at each sound level difference, and then we produced 36 auditory-visual stimuli (4 presentation distances × 9 sound levels). Here, we took into account the time delay of sound at the distance from the drum to the video camera, i.e., $t_{\mathrm{d}} = 0$ (simultaneity of the visual event and its sound) was adjusted according to the calculated time delay (the center column of Table 2) corresponding to each presentation distance. We presented the auditory-visual stimuli to each subject in random order and repeated 7 times, i.e., we conducted 2016 trials (36 auditory-visual stimuli × 7 iterations × 8 subjects) in total. The duration of each presentation was about 5 s.

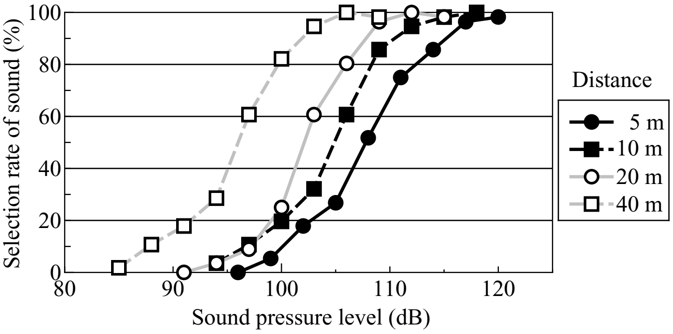


Figure 3: Selection rate of the answer that the sound stimulus was stronger than the visual stimulus. The vertical and horizontal axes denote the selection rate of the sound and the sound pressure level, respectively.

After each presentation, we asked the subject to answer the following question: "Which stimulus was stronger, the visual image or the sound?"

3.2 Result

Figure 3 shows the frequency of the answer that the sound was stronger than the visual image. The vertical axis denotes the selection rate of the sound, and the horizontal axis denotes the sound pressure level of the auditory stimulus. The symbols ●, ■, ○, and □ denote the cases the presentation distances were 5, 10, 20, and 40 m, respectively. In Fig. 3, the selection rates of the sound become large as the sound pressure level increases and as the presentation distance increases.

To determine the SPL of the sound stimulus matching with the visual target size, that is referred to as "ESPL (equivalent sound pressure level) [9]," we fitted the results in Fig. 3 using the following sigmoid logistic function:

$$f(x) = \frac{a}{1 + e^{-k(x - x_{\mathrm{c}})}}, \tag{1}$$

where x corresponds to the SPL, k is the slope coefficient related to the sharpness of the decision between "the sound was stronger" and "the visual event was stronger," and x_{c} is the value of x at $f(x) = a/2$, i.e., x_{c} shows the ESPL. $a = 100$ (%) corresponds to the maximum value of the answer rate that the sound stimulus was strong.

Figure 4 shows the ESPL depending on the presentation distance. In Fig. 4, the ESPL becomes lower as the presentation distance increases.

The right column of Table 1 shows values of the ESPL. In Table 1, the ESPL is almost equal to the standard SPL in each presentation distance.

4 Measurement of PSS

In this section, we estimated the point of subjective simultaneity (PSS) between the auditory and visual stimuli.

4.1 Procedure

We employed the same four video clips (5, 10, 20, and 40 m) in Sec. 3.1. The sound pressure level of the au-

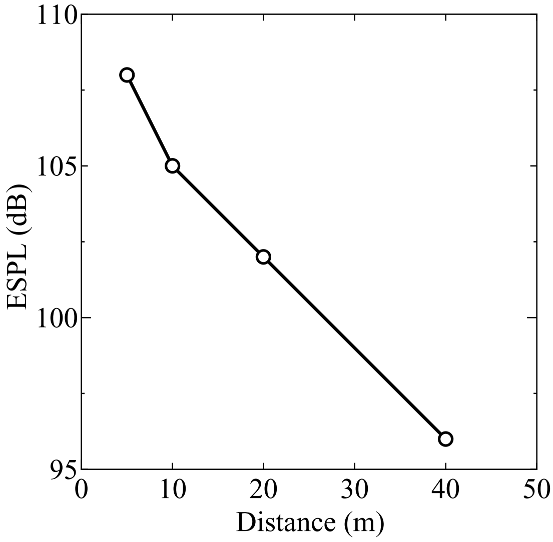


Figure 4: Equivalent sound pressure level (ESPL) depending on the presentation distance. The ESPL was obtained by curve fitting the results in Fig. 3 to Eq. (1).

Table 1: Standard SPLs and ESPLs of the drum sound. The standard SPLs were obtained from actual measurement.

Presentation distance (m)	Standard SPL (Peak) (dB)	ESPL (Peak) (dB)
5	108	108
10	106	105
20	103	102
40	97	96

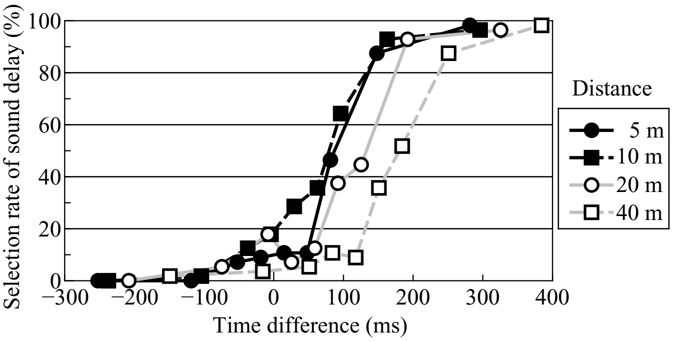


Figure 5: Selection rate of the answer that the sound stimulus was delayed relative to the visual stimulus. The vertical and horizontal axes denote the selection rate of the sound delay and the time difference between the visual event and its sound, respectively. "+" of the time difference denotes that the visual event preceded the sound.

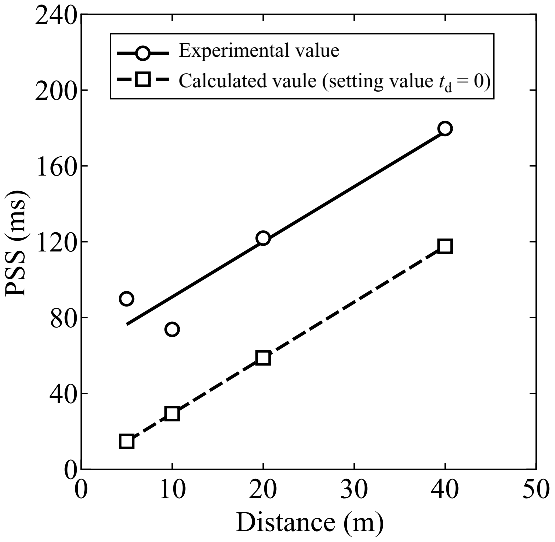


Figure 6: Point of subjective simultaneity (PSS) depending on the presentation distance. The PSSs (empty circles) were obtained by curve fitting the results in Fig. 5 to Eq. (1).

ditory stimulus corresponding to each presentation distance was set at the ESPL derived in Sec. 3.2 as shown in the right column of Table 1. Time delay between the auditory and visual stimuli was set at 0, ±1, ±2, ±4, or ±8 F (1 F = 1/30 s), where "+" and "−"indicate that the sound was delayed with respect to the visual event and vice versa, respectively, based on the calculated values in the center column of Table 2 as $t_\mathrm{d} = 0$. We combined the video clip and the corresponding auditory stimulus at each time delay, and then we produced 36 auditory-visual stimuli (4 presentation distances × 9 time delays). We presented the auditory-visual stimuli to each subject in random order and repeated 7 times as the same as Sec. 3.1, in total we conducted 2016 trials. The duration of each presentation was about 5 s.

After each presentation, we asked the subject to answer the following question: "Which stimulus preceded the other, the visual image or the sound?"

4.2 Result

Figure 5 shows the frequency of the answer that the sound was delayed with respect to the visual image. The vertical axis denotes the selection rate of the sound delay, and the horizontal axis denotes the time difference between the visual event and its sound. The symbols ●, ■, ○, and □ denote the cases the presentation distances were 5, 10, 20, and 40 m, respectively. To determine the point of subjective simultaneity (PSS) between the auditory and visual stimuli, we applied Eq. (1) to the results in Fig. 5. Here, in Eq. (1), x corresponds to the time difference, SPL, k is the slope coefficient related to the sharpness of the decision between "the sound was delayed" and "the visual event was delayed," and x_c shows the PSS.

Figure 6 shows the PSS depending on the presentation distance. In this figure, the solid and dashed lines show the straight-line approximation of the PSS and the calculated value of the time delay (setting value as $t_\mathrm{d} = 0$). In Fig. 6, the PSS is larger around 60 ms than the calculated value, i.e., the subjects felt the auditory-visual event more far away in the virtual space (experimental environment) than in the real space.

The right column of Table 1 shows values of the PSS. The PSSs were smaller about 60 ms than the calculated values.

Table 2: Point of subjective simultaneity (PSS) between the auditory and visual stimuli. The calculated values were obtained corresponding to each presentation distance.

Presentation distance (m)	Calculated value (ms)	PSS (ms)
5	14.7	76.4
10	29.4	90.9
20	58.8	120.0
40	117.6	178.1

5 Experiment

In this section, we investigated the equivalent perception between the auditory and visual stimuli when varying both the sound pressure level and the time delay between the auditory and visual stimuli based on the results of ESPL and PSS derived in Sec. 3.2 and 4.2, respectively.

We performed some training sessions for the subjects to accustom to the auditory-visual stimulus in each presentation pattern before this experiment. The training sessions were repeated until satisfying the following conditions in each subject:

(1) the error between the ESPL obtained from the training sessions and the standard ESPL (in Table tab:ESPL) in each presentation pattern is within 3 dB,

(2) the average of all the above errors of (1) is within 1.5 dB.

When each subject satisfied the above conditions, we judged that the subject was well-trained to the auditory-visual stimuli, and carried out the following experiment to the well-trained subjects.

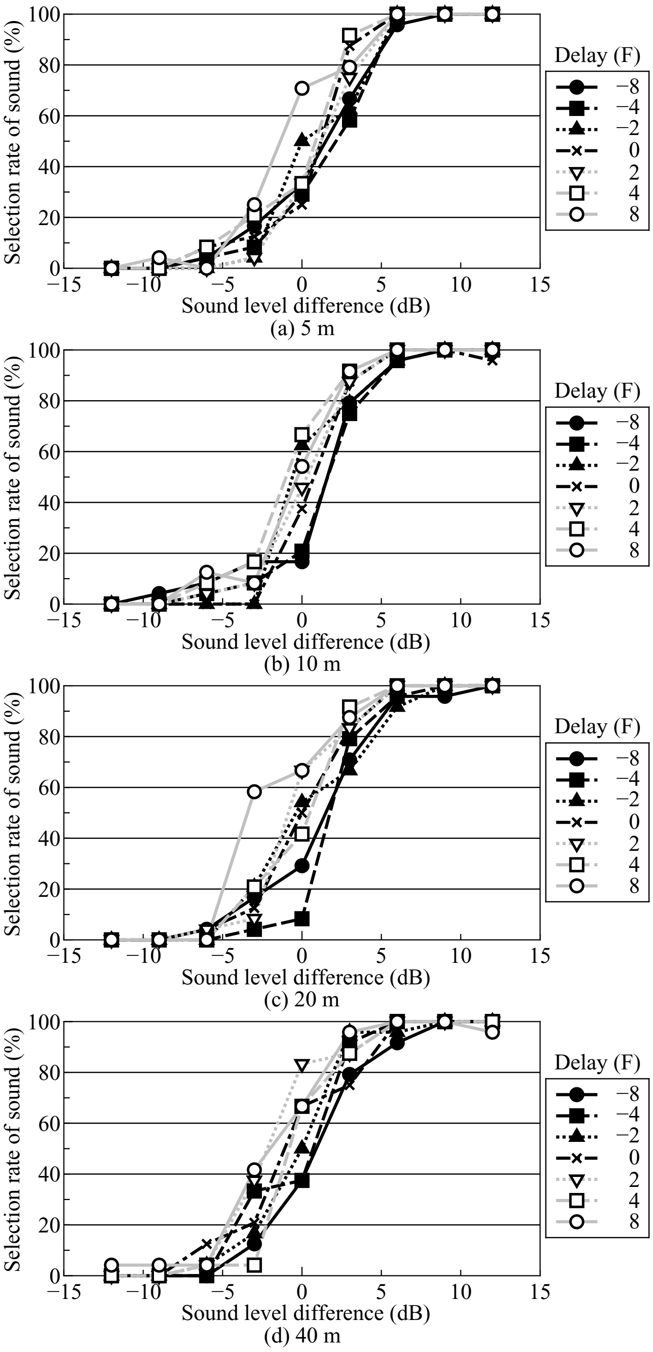
5.1 Procedure

We employed the same four video clips (5, 10, 20, and 40 m) in Sec. 3.1 and 4.1. The sound pressure level of the auditory stimulus corresponding to each presentation distance was set at 0, ±3, ±6, ±9, and ±12 dB based on the ESPL derived in Sec. 3.2. And, the time delay between the auditory and visual stimuli was set at 0, ±1, ±2, ±4, or ±8 F (1 F = 1/30 s) based on the PSS derived in Sec. 4.2. We combined the four video clips, the nine sound pressure levels, and the seven time delays, and then we produced 252 auditory-visual stimuli (4 presentation distances × 9 sound levels × 7 time delays). We presented the auditory-visual stimuli to each subject in random order and repeated 3 times, i.e., we conducted 6048 trials (252 auditory-visual stimuli × 3 iterations × 8 subjects) in total. The duration of each presentation was about 5 s.

After each presentation, we asked the subject to answer the following question: "Which stimulus was stronger, the visual image or the sound?"

5.2 Result

Figure 7 shows the frequency of the answer that the sound was stronger than the visual image. (a) – (e) correspond to the results at the presentation distances 5 – 40 m, respectively. The vertical axis denotes the selection rate of the sound, and the horizontal axis denotes the sound level difference from the ESPLs in Table 1. The symbols ●, ■, ▲, ×, ▽, □, and ○ denote the cases when the time delay t_d = −8, −4, −2, 0, 2, 4, and 8 F, respectively. In all cases, selection rates of the sound become large as the sound level difference increases.



(a) 5 m

(b) 10 m

(c) 20 m

(d) 40 m

Figure 7: Selection rate of the sound stimulus. (a), (b), (c), and (d) correspond to the results at the presentation distances of 5, 10, 20, and 40 m, respectively.

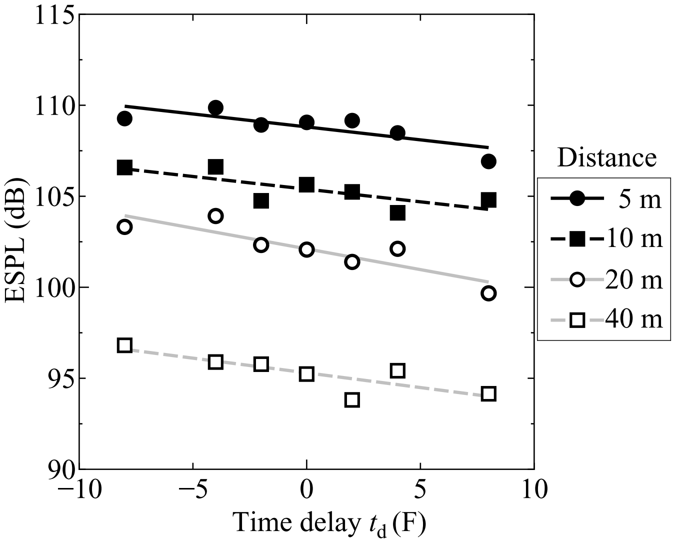


Figure 8: ESPL (equivalent sound pressure level) at each presentation distance. The vertical and horizontal axes denote the ESPL the time delay t_d, respectively. Each line denotes the straight-line approximation at each presentation distance.

To further analyze the above results, we applied Eq. (1) to each case shown in Fig. 7. Figure 8 shows the results of ESPL depending on the time delay t_d between the auditory and visual stimuli at each presentation distance. The vertical axis denotes the ESPL, and the horizontal axis denotes the time delay t_d. The symbols ●, ■, ○, and □ correspond to the cases when the presentation distances were 5, 10, 20, and 40 m, respectively. Each line denotes the straight-line approximation at each presentation distance.

In Fig. 8, ESPL becomes large as the presentation distance becomes near. It means that perceptual strengths of both the auditory and visual stimuli increase when the event occurred close to the subject, and the degree of the increase may depends on strength of the conceptual relationship between the auditory and visual stimuli. About the approximation lines, their slop coefficients have negative values in all cases of the presentation distance. This result shows an opposite tendency to our previous work [10]. It is thought that one of the reasons why they displayed opposite tendencies is that both the SPL and the impulse response corresponding to each presentation distance were considered in the auditory stimulus of this study while only the SPL of each presentation distance was considered in our previous work [10].

6 Conclusion

In this paper, we investigated the equivalent perception between a visual stimulus and its associated sound. We used a video clip of a man beating a drum on a road, which had a feeling of depth with a perspective view, as the visual stimulus and its drum sound as the auditory stimulus. We measured the ESPL (equivalent sound pressure level) and the PSS (point of subjective simultaneity) between the auditory and visual stimuli in each presentation pattern. From the results of ESPL and PSS, we carried out an experiment to investigate the equivalent perception between the auditory and visual stimuli when varying both the sound pressure level and the time delay between the auditory and visual stimuli at various presentation distances. As a result, we obtained that the ESPL intended to decrease when the delay time increased, i.e., the sound was delayed.

References

[1] T. C. Weerts and W. R. Thurlow, “The effects of eye position and expectation on sound localization,” *Perception & Psychophysics*, **9**, 1A, 35 – 39 (1971).

[2] W. R. Thurlow and C. E. Jack, “Certain determinants of the ‘ventriloquism effect’,” *Perceptual and Motor Skills*, **36**, 1171 – 1184 (1973).

[3] C. E. Jack and W. R. Thurlow, “Effects of degree of visual association and angle of displacement on the ‘ventriloquism’ effect,” *Perceptual and Motor Skills*, **37**, 967 – 979 (1973).

[4] C. V. Jackson, “Visual factors in auditory localization,” *Quarterly Journal of Experimental Psychology*, **5**, 52 – 65 (1953).

[5] G. J. Thomas, “Experimental study of the influence of vision on sound localization,” *Journal of Experimental Psychology*, **28**, 163 – 177 (1940).

[6] M. Radeau and P. Bertelson, “Auditory-visual interaction and the timing of inputs — Thomas (1941) revisited —,” *Psychological Research*, **49**, 17 – 22 (1987).

[7] E. A. Lavelace and D. M. Anderson, “The role of vision in sound localization,” *Perceptual & Motor Skills*, **77**, 843 – 850 (1993).

[8] H. Hasegawa, M. Ayama, S. Matsumoto, A. Koike, K. Takagi, M. Kasuga “Evaluation of the corresponding degree between a visual and its associated sound under dynamic conditions on a wide screen,” *IEICE Trans. Fundamentals*, **E87-A**, 1409 - 1416 (2004).

[9] H. Hasegawa, H. Nakane, M. Ono, M. Kasuga, and M. Ayama, “Equivalent perception between a visual image and its associated sound on a wide screen,” *Proc. of Forum Acusticum 2005*, 1625 – 1628 (2005).

[10] H. Hasegawa, H. Nakane, H. Ono, M. Ono, M. Kasuga, and M. Ayama, “A study on the equivalent sound pressure level (ESPL) of a visual stimulus with a feeling of depth,” *Proc. of the 9th Western Pacific Acoustics Conference (WESPAC IX 2006)*, hu-2-2-290 (2006).