Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Effect of change in sound pressure level on the equivalent perception between a visual stimulus and its associated sound

Shun YAMAMOTO(1), Hiroshi HASEGAWA(1), Tomoharu ISHIKAWA(1), Ichiro YUYAMA(1), Masao KASUGA(1) and Miyoshi AYAMA(1)

(1) Graduate School of Engineering, Utsunomiya University, 7-1-2 Yoto, Utsunomiya-shi, 321-8585, Japan

PACS: 43.66.Lj

ABSTRACT

This study investigated the equivalent perception between a visual event and its associated sound when the sound pressure level (SPL) was varied. We performed an experiment of an auditory-visual stimulus presentation using an audio-video clip of a man beating a drum on a road. The visual stimulus had a feeling of depth with a perspective view of the road. We produced auditory-visual stimuli at presentation distances of 5, 10, 20, and 40 m under various conditions, where we varied the SPL of the auditory stimulus (drum sound) from -12 to 12 dB based on the measured SPL and the rate of the presentation distance from -40 to 40%. The visual stimulus was projected onto a screen that had the viewing angles of 30.8 degrees (W) $\times$ 16.1 degrees (H), and the auditory stimulus was reproduced via headphones. We presented the auditory-visual stimuli to the experimental subjects and asked to subjectively evaluate whether the size of the visual event was larger or smaller compared with that imagined from the strength of its associated sound. Then we estimated the subjective feeling of depth of the visual event, which is the visual distance matching with the SPL of the sound, in each presentation distance. As a result, we obtained that the subjective feeling of depth intended to decrease when the SPL increased, that is, the subjects perceived the visual event being nearer when the associated sound level became higher.

INTRODUCTION

Recent multimedia technology has made it possible to construct various audio-video environments. It is, however, difficult to reproduce an auditory-visual space with a feeling of being in the actual space. It is known that the association between auditory and visual information is one of the most important factors for reproducing an auditory-visual space. There have been done many studies on auditory-visual interactions from a psychological perspective (Weerts and Thurlow 1971; Thurlow and Jack 1973; Jack and Thurlow 1973; Jackson 1953; Thomas 1940; Radeau and Bertelson 1987; Radeau and Bertelson 1993), but there is little known about the interactions to apply to multimedia environments.

We have been investigating interactions between auditory and visual information whose conceptual relationship was strong (Nakane et al. 2005; Hasegawa et al. 2008; Hasegawa et al. 2009). In this paper, we focused on the relationship between the feeling of depth of a visual event and the sound pressure level (SPL) of its associated sound. We carried out an experiment of auditory-visual stimuli presentation using video clips of a visual event, which had a feeling of depth with a perspective view, and its sound.

EXPERIMENTAL ENVIRONMENT

Apparatus

Figure 1 shows a block diagram of the experimental apparatus. Visual stimuli were played using a digital video player (SONY HDR-HC1) and were projected onto a screen using a projector (EPSON EMP-TW600). The projected area on the screen was 2.09 m in width and 1.17 m in height. The pixel number of the projector display was 1440 (W) $\times$ 1080 (H). An experimental 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. Auditory stimuli were presented via headphones (SENNHEISER HD-595).

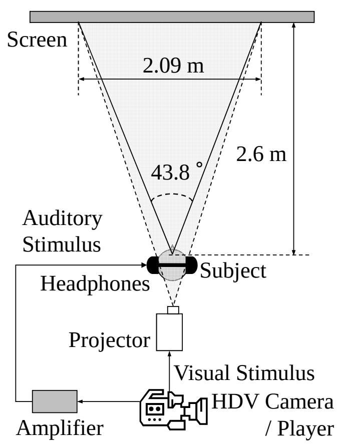
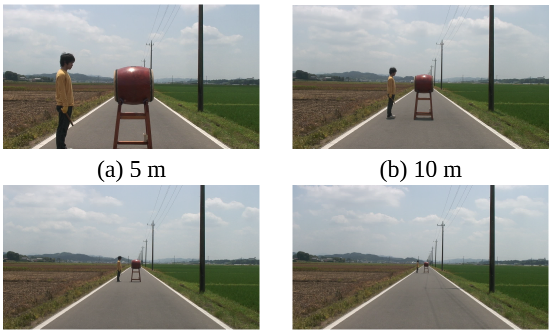


Figure 1: Experimental apparatus. Visual and auditory stimuli were presented by a digital video player and headphones, respectively. The size of the screen was 2.09 m (W) $\times$ 1.17 m (H).

Visual and auditory stimuli

A video clip of a man beating a drum on a road and its drum sound were used as the visual and auditory stimuli, respectively. The visual event (a man beating a drum), 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 and the video camera was set at 5, 10, 20, or 40 m as shown in Fig. 2. The 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 drum sound was recorded using a microphone (B&K 4190) near the drum, and the auditory stimulus at each presentation distance was produced by convoluting the recorded drum sound with the spatial impulse response corresponding to each presentation distance (Hasegawa et al. 2009). Thus, in this study, we took into account not only the SPL of the auditory stimulus but also the spatial impulse response in each presentation distance.



(a) 5 m (b) 10 m

(c) 20 m (d) 40 m

Figure 2: Visual stimuli on the screen. Video clips of a man beating a drum on a road were presented to the experimental subjects. (a), (b), (c), and (d) correspond to the presentation distances 5, 10, 20, and 40 m, respectively.

Subject

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

ESPL

First, we had to set the point of subjective equality (PSE) of the sound pressure level (SPL) of the auditory stimulus that provides a perceptual strength equivalent to that of the visual stimulus. We refer to this sound level as "the equivalent sound pressure level (ESPL)." The ESPL changes depending on the presentation distance and has a characteristic that is decreasing as the presentation distance increases. In this study, we employed the ESPLs obtained from our previous experiment (Hasegawa et al. 2010), since the ESPLs were almost equal to the standard SPLs obtained from actual measurements. The right column of Table 1 shows values of p_{PSE}. In Table 1, p_{PSE} is almost equal to the standard SPL in each presentation distance.

MEASUREMENT OF PSS

Here, we estimated the point of subjective simultaneity (PSS) between the auditory and visual stimulus.

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

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

Procedure

We employed four video clips of a man beating a drum at presentation distances of 5, 10, 20, and 40 m (Fig. 2) as the visual stimuli and its drum sounds corresponding to the presentation distance as the auditory stimuli. The SPL of the auditory stimulus corresponding to each presentation distance was set at the ESPL in the right column of Table 1. The time difference between the auditory and visual stimulus was set at ± 1, ± 2, ± 4, or ± 8 F (1 F $= 1/30$ s), where a positive value indicates the sound was delayed with respect to the visual event, respectively, based on the calculated values in the center column of Table 2 as $t_{\mathrm{d}} = 0$. We combined each visual stimulus and its corresponding auditory stimulus at each time delay, and then produced 36 auditory-visual stimuli (4 presentation distances $\times$ 9 time delays). We presented the auditory-visual stimuli to each subject in random order and repeated seven times, and we conducted 2016 trials 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 preceded the other, the visual event or the sound?"

Result

Figure 3 shows the frequency of the answer that the sound was delayed with respect to the visual event. 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.

To determine the PSS between the visual event and its sound, 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 time difference, 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} is the value of x at $f(x) = a/2$, i.e., x_{c} shows the PSS t_{PSS}. $a = 100$ (%) corresponds to the maximum value of the answer rate that the sound was delayed.

Figure 4 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 $t_{\mathrm{d}} = 0$), respectively. In Fig. 4, the experimental values are generally larger than the calculated values, i.e., the subjects felt the auditory-visual event more far away in the virtual space (experimental environment) than in the real space.

LEARNING TEST

As a preparation for the next experiment, we performed a learning test of the experimental stimuli to the subjects. We presented the standard audio-video clips to the subjects repeatedly of all presentation distances (5, 10, 20, and 40 m) applying the ESPLs and PSSs obtained in the previous sections.

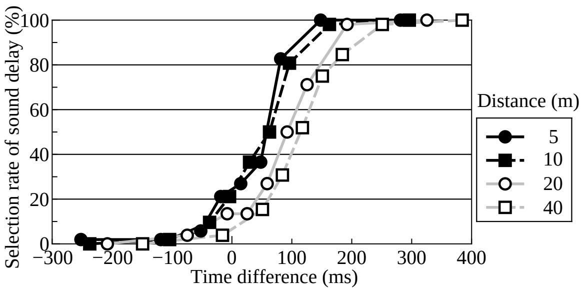


Figure 3: Selection rate of the answer that the sound stimulus was delayed compared to the visual stimulus. The vertical and horizontal axes denote the selection rate of the sound delay and the time difference between the visual event and its sound, respectively. A positive value of the time difference shows that the visual event preceded the sound.

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

Presentation distance (m)	Calculated time delay (ms)	PSS (ms)
5	14.7	49.5
10	29.4	59.1
20	58.8	79.0
40	117.6	118.4

To confirm the degree of learning to the auditory-visual stimuli, we carried out a verification test. Firstly, we prepared the visual stimuli at distance differences (feeling of depth of the visual stimulus) of 0, ±10, ±20, and ±40% with respect to the standard video clip at each presentation distance (Fig. 5). We then combined each visual stimulus with its corresponding auditory stimulus using the ESPL (Table 1) and PSS (Table 2) at each presentation distance. In total, 28 auditory-visual stimuli were produced (4 presentation distances × 7 distance differences). We presented the auditory-visual stimuli to each subject in random order and repeated three times. After each presentation, we asked the subject whether the size of the drum perceived from the visual image was larger or smaller compared with that imagined from strength of the sound.

We then calculated the PSSs at all presentation distances in each subject and judged whether the PSSs satisfied the following requirements or not.

> **Req. 1:** Each difference between the presentation distance and the equivalent distance calculated from the PSS was ≤ 20%.
> **Req. 2:** The average of the differences in **Req. 1** was ≤ 10%.

When the both requirements were satisfied, we judged that the subject got accustomed to the auditory-visual stimuli. After several repetitions of the learning test, all experimental subjects satisfied the both requirements, and they advanced to the next experiment.

EXPERIMENT

In this experiment, we investigated an effect of change in the SPL on the equivalent perception between the auditory and visual stimulus.

Procedure

We employed the same 28 video clips at presentation distances of 5, 10, 20, and 40 m with distance differences of 0, ±10, ±20, and ±40% in the previous section. The SPL of the auditory stimulus corresponding to each presentation distance was set at 0, ±3, ±6, ±9, or ±12 dB based on the ESPL in Table 1. Time delay between the auditory and visual stimulus corresponding to each presentation distance was set at the PSS in Table 2. We combined each visual stimulus with its corresponding auditory stimulus at each SPL, and then we produced 252 auditory-visual stimuli (4 presentation distances × 7 distance differences × 9 SPLs). We presented the auditory-visual stimuli to each subject in random order and repeated three times, in total we conducted 6048 trials (252 auditory-visual stimuli × 8 subjects × 3 times repetition). The duration of each presentation was about 5 s.

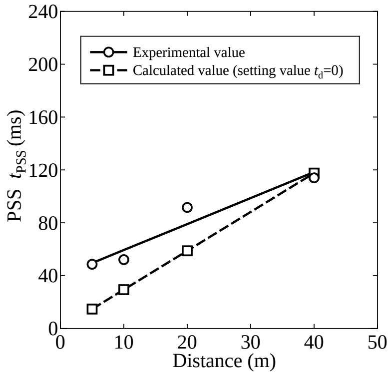


Figure 4: Point of subjective simultaneity (PSS) depending on the presentation distance. The PSSs (empty circles) were obtained by curve fitting the results in Fig. 3 to Eq. (1).

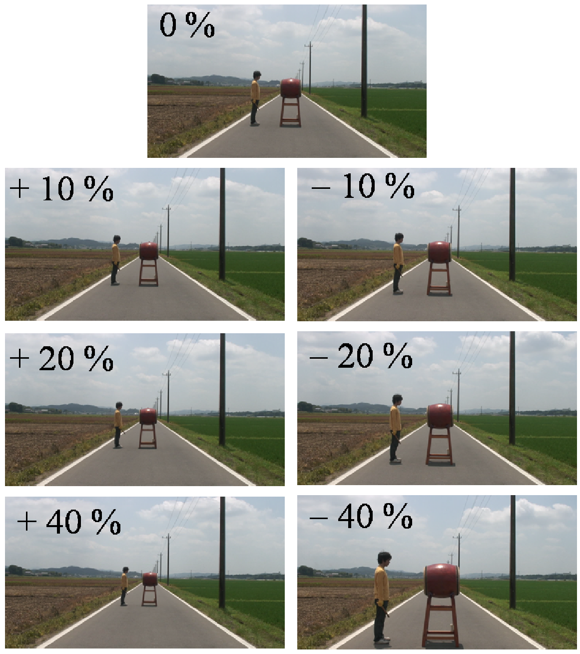
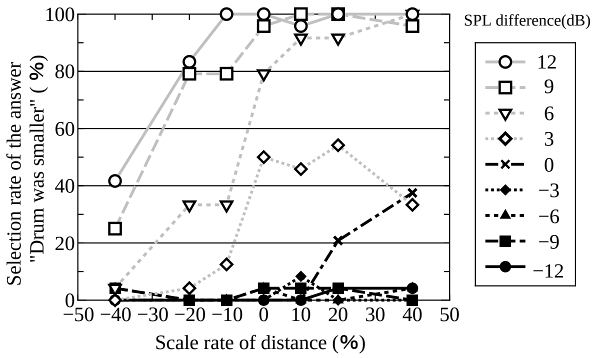


Figure 5: Visual stimuli at distance differences of 0, ±10, ±20, and ±40% with respect to the standard video clip at a presentation distance of 10 m

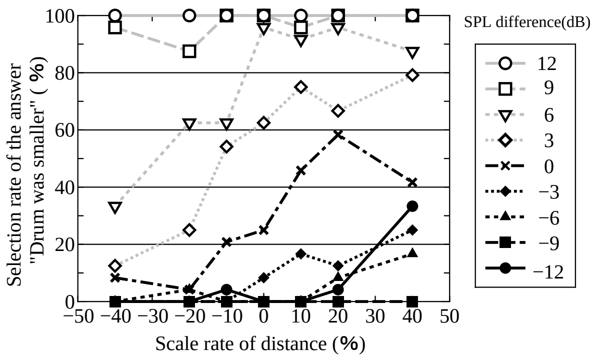
After each presentation, we asked the subject whether the size of the drum was larger or smaller compared with that imagined from loudness of the sound.

Result

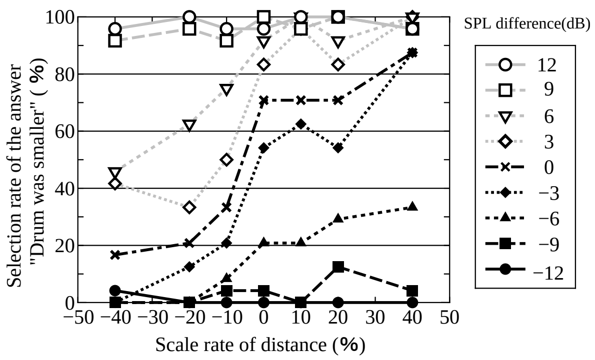
Figure 6 shows the frequency of the answer that the size of the drum was smaller than that imagined from loudness of the corresponding sound. (a), (b), (c), and (e) correspond to the results at the presentation distances 5, 10, 20, and 40 m, respec-



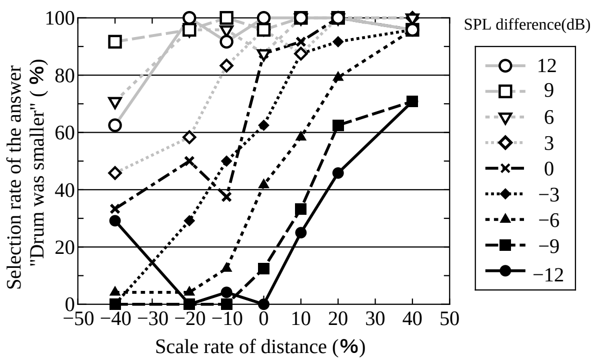
(a) 5 m



(b) 10 m



(c) 20 m



(d) 40 m

Figure 6: Selection rate of the answer that the visual size of the drum was smaller than that imagined from strength of the sound. (a), (b), (c), and (d) correspond to the results at the presentation distance of 5, 10, 20, and 40 m, respectively.

tively. The vertical axis denotes the selection rate of the answer that the visual size of the drum was smaller, and the horizontal axis denotes the scale ratio of the distance difference. The symbols filled circle, filled square, filled triangle, filled diamond, cross, empty diamond, empty triangle, empty square, and empty circle denote the cases when the SPL differences were −12, −9, −6, −3, 0, 3, 6, 9, and 12 dB, respectively. In most cases, the selection rates become large as the scale of the distance difference increases.

ANALYSIS

To quantitatively evaluate the perceived feeling of depth, we applied Eq. (1) to each case shown in Fig. 6. In Eq. (1), x corresponds to the scale rate of distance, k is the slope coefficient related to the sharpness of the decision between "the visual size of the drum was larger" and "the visual size of the drum was smaller," and x_c shows the point of subjective equality (PSE) of distance d_{PSE}. Here, this curve fitting was limited to the cases that changes of the selection rate were large, i.e., the cases of the SPL differences of 3, 6, and 9 dB at a presentation distance of 5 m, the cases of 0, 3, and 6 dB at 10 m, the cases of −6, −3, 0, 3, and 6 dB at 20 m, and the cases of −12, −9, −6, −3, and 0 dB at 40 m were adapted to the fitting. .

Figure 7 shows the results of the PSE of distance d_{PSE}, i.e., the subjective feeling of depth, depending on the SPL difference at each presentation distance. The vertical axis denotes the PSE of distance d_{PSE}, and the horizontal axis denotes the SPL difference. The symbols filled circle, filled square, empty circle, and empty square 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. 7, each approximation line has negative slope with respect to the SPL difference in all cases. This means that change in the SPL influenced on the subjective feeling of depth. The subjective feeling of depth was nearer when the SPL increased, and vice versa, i.e., the subjective feeling of depth shifted to the same direction as the SPL. This result shows that the subjective feeling of depth could be captured by the change in the SPL.

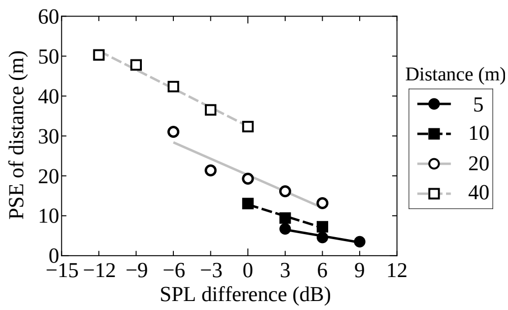


Figure 7: PSE (point of subjective equality) of distance at each presentation distance.

CONCLUSION

In this paper, we investigated the equivalent perception between auditory and visual information whose conceptual relation was strong. We carried out an experiment of auditory-visual stimuli presentation using video clips of a man beating a drum on a road, which had a feeling of depth with a perspective view of the road, and its drum sound. Then we analyzed the relationship between the subjective feeling of depth of the visual stimulus (a man beating a drum) and the sound pressure level (SPL) of auditory stimulus (drum sound). As a result, it is found that the subjective feeling of depth became nearer when the SPL increased, on the opposite, it became far away as the

SPL decreased. This result shows that the subjective feeling of depth could be captured by the change in the SPL.

REFERENCES

Hasegawa, H. et al. (2008). “Relationship between a visual stimulus with a feeling of depth and its equivalent sound pressure level (ESPL)”. *Proc. of Acoustics’08 Paris* ACOUSTICS2008, p. 1641.

Hasegawa, H. et al. (2009). “Effect of sound transfer function on the equivalent perception between a visual image with a feeling of depth and its associated sound”. *Proc. of the 16th International Congress on Sound and Vibration* T37-1, p. 174.

Hasegawa, H. et al. (2010). “Effect of the time delay between a visual stimulus and its associated sound on their Equivalent perception”. *to be presented in the 17th International Congress on Sound and Vibration*.

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

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

Nakane, H. et al. (2005). “Detection of onset asynchrony between auditory and visual information on a wide screen”. *Acoust. Sci*, pp. 469 –472.

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

— (1993). “The role of vision in sound localization”. *Perceptual and Motor Skills* 77, pp. 843 –850.

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

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

Weerts, T. C. and W. R. Thurlow (1971). “The effects of eye position and expectation on sound localization”. *Perception & Psychophysics* 9.1A, pp. 35 –39.