Visual Image Effects on Sound Localization in Peripheral Region under Dynamic Multimedia Conditions

Yoshinori KONO† , Hiroshi HASEGAWA† , Miyoshi AYAMA† , Masao KASUGA† , Shuichi MATSUMOTO‡ , Atsushi KOIKE‡ , and Koichi TAKAGI‡

† Department of Information Science, Faculty of Engineering, Utsunomiya University
7-1-2 Yoto, Utsunomiya 321-8585, Japan
Phone/Fax: +81-28-689-6287, E-mail: kono@vibra.is.utsunomiya-u.ac.jp
‡ KDDI R&D Laboratories Inc.
2-1-15 Ohara, Kamifukuoka 356-8502, Japan

Abstract— **This paper describes effects of visual information influencing sound localization in the peripheral visual field under dynamic conditions. Presentation experiments of an audio-visual stimulus were carried out using a movie of a moving patrol car and its siren sound. The following results were obtained: first, the sound image on the timing at the beginning of the presentation was more strongly captured by the visual image than that at the end, i.e., the "beginning effect" was occurred; second, in the peripheral regions, the "beginning effect" was strongly appeared in near the fixation point of eyes.**

I. Introduction

Reproduction of a high quality audio-visual space has become easy with the spread of multi-media apparatuses. However, it is not sufficient to reproduce an audio-visual space with a feeling of presence in the space. It is known that a feeling of identification of visual and sound information is one of the most important factors for reproducing the audio-visual space.

Numerous attempts have been made at audio-visual interactions from a psychological viewpoint[1]–[6]. Only few attempts, however, have so far been made at the interactions for the purpose of an application to actual audio-visual environments[7]–[9]. Furthermore, there are very few experiments on the interactions performed under dynamic situations using moving audio and visual stimuli[10]. And these experiments have been done only near the center of the visual field.

This paper is intended as an investigation of effects of visual information on the associated auditory information when they are presented simultaneously in the peripheral visual field under dynamic conditions. Presentation experiments of an audio-visual stimulus are carried out using a moving visual image and its corresponding sound image.

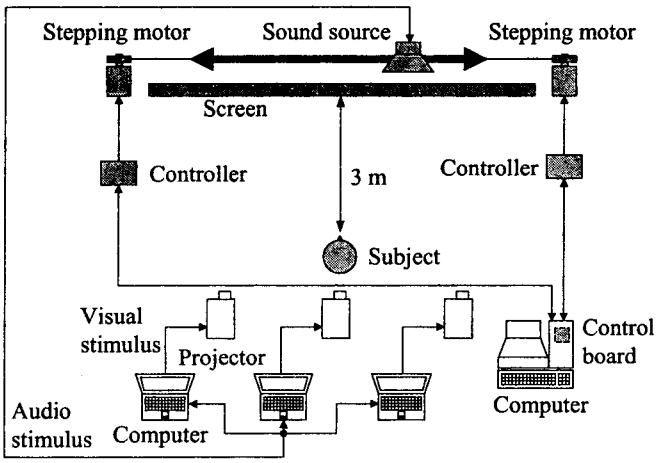


Fig. 1. Block diagram of the experimental apparatus.

II. Experiment

A. Apparatus and Stimulus

A block diagram of the experimental apparatus is shown in **Fig. 1**. A computer graphics movie of a moving patrol car and its siren sound were used as visual and sound stimuli, respectively. The visual stimulus was projected onto a screen of 0.85 m in height and 3.60 m in width using three projectors. The size of the region projected by each projector was 0.85 m (H) × 1.20 m (W). These three regions were defined as "the left peripheral visual region (L)", "the central visual region (C)", and "the right peripheral visual region (R)", respectively. In the region (C), there was a traffic signal which had a green light put on the center of the region used as the fixation point (**Fig. 2**). We instructed the subjects to gaze at the fixation point during the experiment. The patrol car, whose size was 8 cm (H) × 20 cm (W), moved horizontally at a speed of 2.5 m/s in any one of the three regions. A sample scene of the visual stimulus is shown in **Fig. 3**. In Fig. 3, the subscripts L, C, and R indi-

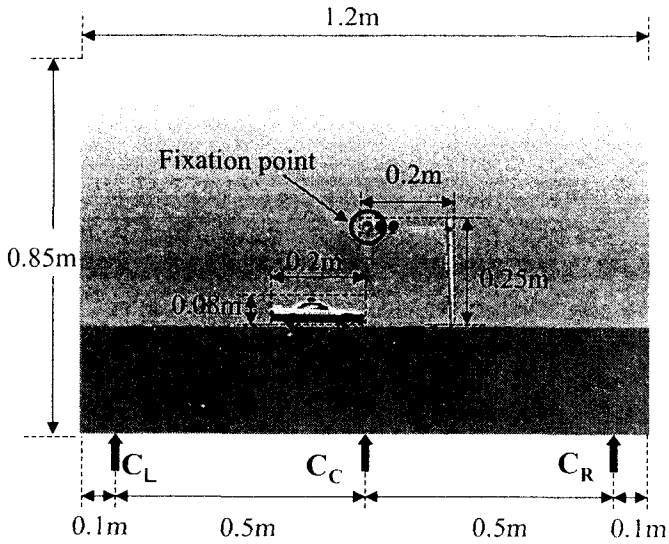


Fig. 2. Central visual region.

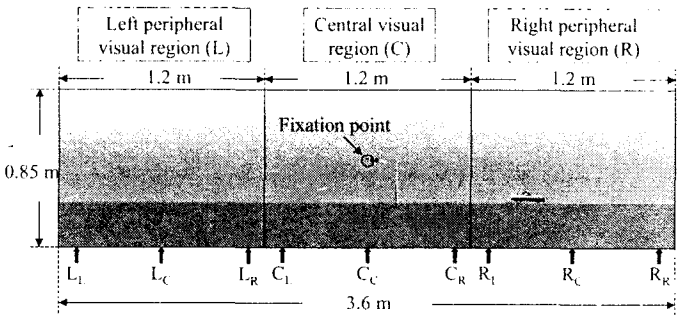


Fig. 3. Visual stimulus.

cate the positions of the left edge, the center, and the right edge of the moving range of a patrol car in each region, respectively. The associated sound stimulus was presented via a loudspeaker mounted on a truck, which could move along the railway laid on a horizontal plane behind the screen (**Fig. 4**). The truck was controlled by using stepping motors set up at the both ends of the railway. The loudspeaker was hidden behind a black grill cloth. A subject was seated on a chair placed three meters distant from the center of the screen. The head of the subject was fixed using a headrest installed in the chair.

B. Subject

Five males and five females, aged 21 and 22 years, served as subjects. All subjects had normal or corrected-to-normal vision and normal hearing acuity.

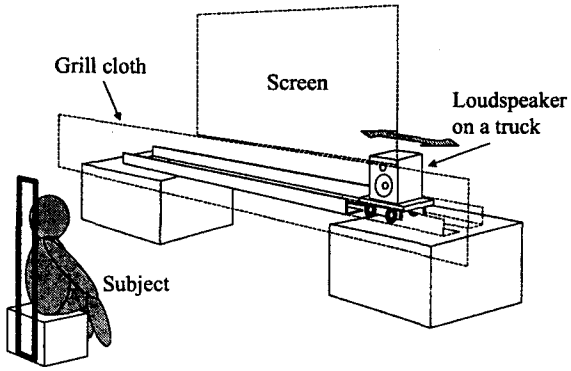


Fig. 4. Schematic diagram of the moving sound source.

C. Procedure

The visual stimulus was presented on any one of the three regions in a random order, and moved either from X_L to X_R or from X_R to X_L at a constant velocity of 2.5 m/s, where X represents L, C, or R denoted in Fig. 3. Then the sound stimulus was presented simultaneously in the same region under the following conditions:

(a) the sound source was fixed (i) at X_L, (ii) at X_C, or (iii) at X_R of the region,

(b) the sound source was moved in the same direction of the visual stimulus at velocities of (i) 2.0 m/s, (ii) 2.5 m/s, or (iii) 3.0 m/s,

(c) the sound source was moved in the opposite direction of the visual stimulus at velocities of (i) 2.0 m/s, (ii) 2.5 m/s, or (iii) 3.0 m/s.

Fig. 5 (a), (b), and **(c)** show the movement patterns of the visual image and the sound source under the conditions described in the above **(a)**, **(b)**, and **(c)**, respectively. In total, 54 movement patterns of the audio and visual stimuli were examined. Each pattern of the stimuli was presented to each subject for 3 times. After each presentation, the subjects were instructed to answer the perceived position where the sound image coincided with the visual image:

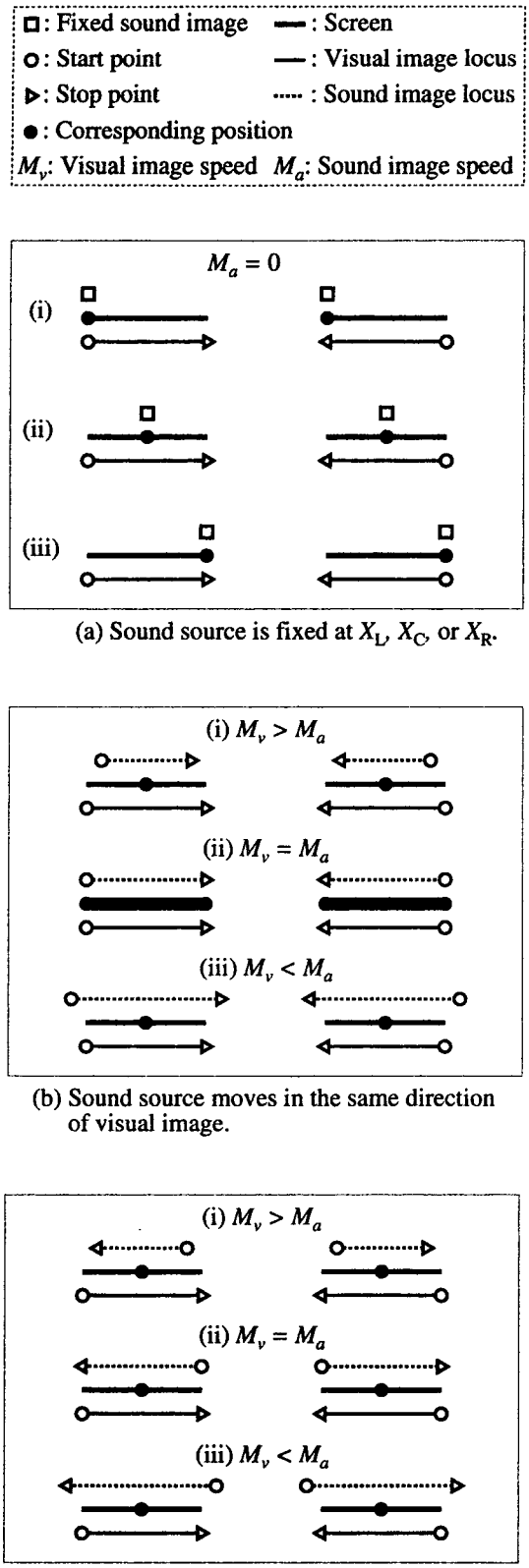
(1) entirely (at X_L, X_C, and X_R),

(2) partially (at X_L, X_C, X_R, X_L and X_C, X_R and X_C, or X_R and X_L),

(3) not at all.

III. Results

We evaluated a degree of the correspondence between the visual and sound images. The degree is defined by the ratio of the number of the times evaluated as a coincidence position to the total number of the presentation trials at each place (X_{Ln}, X_{Cn}, or X_{Rn}) of the screen in each presentation pattern, as shown in the following



(a) Sound source is fixed at X_L, X_C, or X_R.

(b) Sound source moves in the same direction of visual image.

(c) Sound source moves in the opposite direction of visual image.

Fig. 5. Movement patterns of the audio and visual stimuli.

equations:

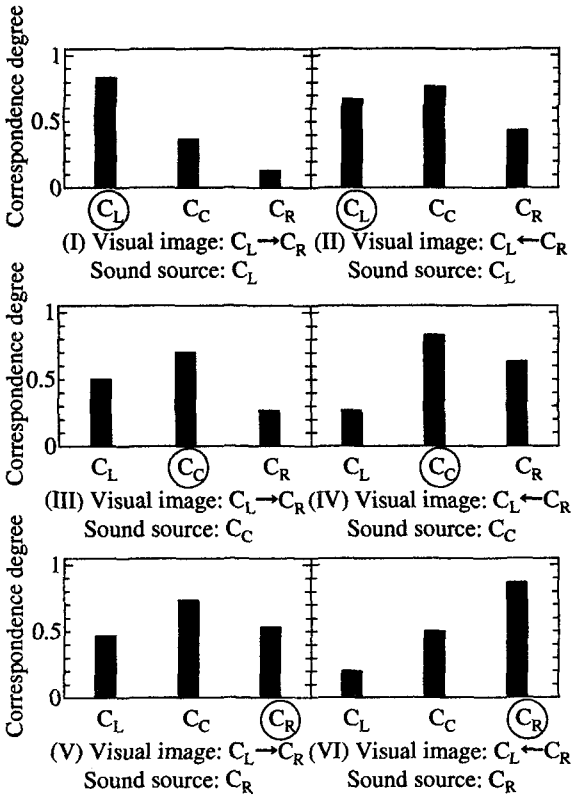
$$X_L = \frac{1}{N}\sum_{n=1}^{N} X_{Ln},$$
$$X_C = \frac{1}{N}\sum_{n=1}^{N} X_{Cn}, \quad (1)$$
$$X_R = \frac{1}{N}\sum_{n=1}^{N} X_{Rn},$$

where $N = 30$ (10 subjects × 3 trials) is the total number of the trials, and n is the index of each trial. In these equations, X_{Ln}, X_{Cn}, and X_{Rn} are set all zero as the default value, and when the answers of the coincidence at X_L, X_C, and X_R are obtained, then $X_{Ln} = 1$, $X_{Cn} = 1$, and $X_{Rn} = 1$ are set, respectively.

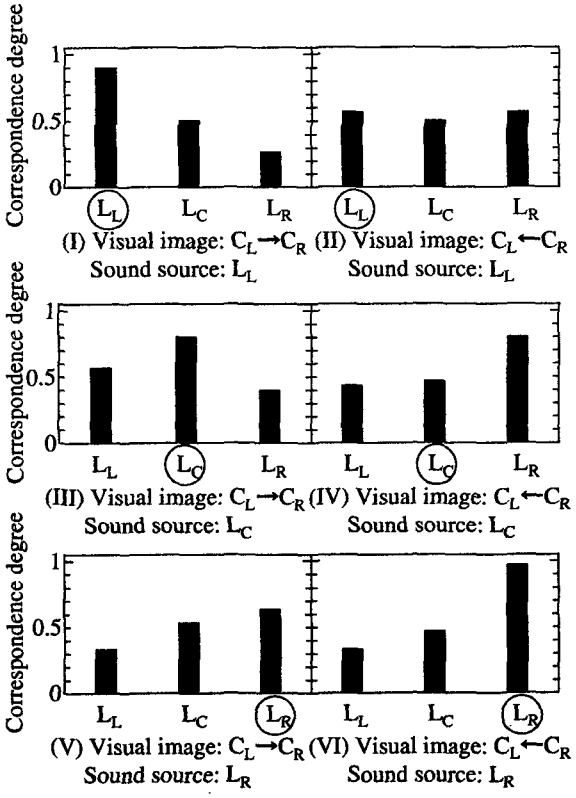
Fig. 6 (a) and **(b)** show the results of the central(C) and lift peripheral(L) visual regions, respectively, with fixed sound source conditions.

As shown in Fig. 6 (a), in most of the places where the visual image position coincides with the sound source position ((I)-C_L, (II)-C_L, (III)-C_C, (IV)-C_C, and (VI)-C_R), the perceptual coincidence degree tend to show higher values than those in other places, while one of such places ((III)-C_R) shows a moderate value. Among these physically coincided places, comparisons between (I)-C_L and (II)-C_L, and between (V)-C_R and (VI)-C_R, reveal that the degrees on the timing at the beginning of the presentation are higher than those at the end of the presentation ((I)-C_L > (II)-C_L, (VI)-C_R > (V)-C_L)). The same effect that a sound image is captured by the visual image at the beginning of the presentation is observed in the places where the visual image and sound source positions are not necessarily physically coincided. In (III) and (IV), for example, (III)-C_L and (IV)-C_R are higher than (III)-C_R and (IV)-C_L, respectively. In the following, we call this effect "beginning effect". In (II) and (V), the degrees at C_C are higher than those at (II)-C_L and (V)-C_R. In these conditions, the sound image may be captured more strongly in the places close to the beginning positions than in the places at the end of the presentation, instead of physical coincidence between visual image and sound source in the latter.

In Fig. 6 (b), in most of the places where the visual image position coincides with the sound source position ((I)-L_L, (III)-L_C, (V)-L_R, (VI)-L_R), the coincidence degree tend to show higher values than those in other places, while two of them ((II)-L_L, (IV)-L_C) show moderate values. The beginning effect in physically coincided places is found in the comparisons between (I)-L_L and (II)-L_L, and between (V)-L_R and (VI)-L_R, similar to Fig. 6 (a). Similar effect in physically non-coincided places is



(a) Sound source is fixed at C_L, C_C, or C_R, and visual stimulus is presented onto CV.



(b) Sound source is fixed at L_L, L_C, or L_R, and visual stimulus is presented onto LPV.

Fig. 6. Experimental results.

observed in (III) and (IV) ((III)-L_L > (III)-L_R, (IV)-L_R > (IV)-L_L), while the effect at the position near the fixation point ((IV)-L_R) is stronger than that at a distant position ((III)-L_L).

The velocity of the sound source does not affect the degree of coincidence.

The results of the visual image presented at the region (R) were almost symmetrical to the results at the region (L).

IV. Conclusion

In order to investigate effects of visual information on the corresponding auditory information in the peripheral visual field under dynamic conditions, presentation experiments were carried out using a computer graphics movie of a moving patrol car and its siren sound. As the results, the sound image at the beginning of an audio-visual presentation is more strongly captured with the visual image than that at the ending, i.e., the "beginning effect" was occurred. In the peripheral regions, the beginning effect was strongly appeared in near the fixation point of the eyes.

Acknowledgments

The authors gratefully acknowledge the partial supports of this work by a Grant-in-Aid from the Japanese Ministry of Education (No. 13650453), by a grant from HBF (Hōsō-Bunka Foundation), and by a grant from TAO (Telecommunications Advancement Organization of Japan).

References

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

[2] W. R. Thurlow and C. E. Jack, "Certain determinants of the "ventriloquism effect"," Perceptual and Motor Skills **36**, pp. 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**, pp. 967 – 979 (1973).

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

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

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

[7] K. Nakabayashi "Interaction between stereophonic sound image and video image," The Journal of the IIITE **37**, pp. 984 – 991 (in Japanese) (1983).

[8] K. Komiyama "Visual factors in sound localization for HDTV," The Journal of the ASJ **43**, pp. 664 – 669 (in Japanese) (1987).

[9] H. Hasegawa, K. Takahashi, M. Ayama, and M. Kasuga "Effects of visual information on sound image localization," The Journal of the IIITE **55**, pp. 455 – 462 (in Japanese) (2001).

[10] S. Mateeff, J. Hohnsbein, and T. Noack, "Dynamic visual capture: apparent auditory motion induced by a moving visual target, " Perception **14**, pp. 721 – 727 (1985).