Cross-modal effects of visual information on auditory presence

Kenji Ozawa[1], Yôiti Suzuki[2]

[1] *University of Yamanashi, 400-8511 Kofu, Japan, Email: ozawa@yamanashi.ac.jp*

[2] *Tohoku University, 980-8577 Sendai, Japan, Email: yoh@riec.tohoku.ac.jp*

Introduction

Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another. The authors have studied auditory presence as the first step to understanding the multimodal characteristics of presence [1, 2]. In this study two experiments were conducted to investigate the effects of visual information on auditory presence.

Psychoacoustical experiments

Stimuli

The seventeen sounds shown in Table 1 were binaurally recorded via a dummy head (Koken, Samrai), and the corresponding moving pictures were recorded using a digital video camera (Panasonic, DJ-100) in live situations. They are classified into three groups as shown in Table 1 with respect to the location of sound sources relative to the listener as follows:

Group I: Sound sources were moving.

Group II: Sound sources were moving with the listener.

Group III: Sound sources were stationary.

A segment with the duration of 15 s was used as a stimulus for presentation to a subject for each sound and its corresponding moving picture. The sounds were presented to a subject via headphones (Stax, SR-Lambda) with the binaural calibration [3]. The moving pictures were presented using a 50-inch display (Sony, KL-X9200J) located 2-m away from the subject.

Exp. 1: Audio experiment

The experiments were carried out using Scheffé's method of paired comparison modified by Ura [4]. Six male and two female subjects, between 21 and 24 years of age, with normal hearing acuity participated in the experiments.

In Exp. 1, only sounds were presented to the subjects. Two sounds, randomly selected from the seventeen sounds, were presented sequentially with an interval of 3 s. The subjects compared the two sounds (namely, A and B, in the presentation order) in terms of their auditory presence. They were instructed to give rating judgment on a seven-point category scale (from "−3" to "+3"). One end (−3) of the scale corresponded to the case when the sound A had much higher presence than the sound B; the other end (+3) to the opposite case; and the midpoint (0) to the case when both A and B had equivalent presence. The number of comparisons for each subject was 272 ($={}_{17}P_2$), because every possible permutation was presented to cancel out time order errors.

Table 1: Sounds classified into three groups I, II, and III. The two sounds with * were used in both experiments. The words in *italics* appear in Figure 1. (See the text for details.)

I	sound of a *train passing* (*) *another person's footsteps* on fallen leaves sound of *vehicles passing in a tunnel* sound of *vehicles passing on a road* *caws of a flying crow*
II	*the listener's footsteps on fallen leaves* *the listener's footsteps in a corridor* sound of a *playground slide with rolling bars* sound of the engine heard in *a moving car* sound heard on *a pedaled bicycle*
III	*murmuring of a stream* *roaring of a waterfall* (*) *classical music* performed in a concert hall *clapping of hands* in a concert hall sound of a *door* opening and *shutting* sound of a *warning signal at a railroad crossing* *birds singing*

Exp. 2: Audio and visual experiment

In this experiment, both sounds and moving pictures were presented to the subjects. The experimental procedure was the same as that of Exp. 1 except for the number of stimuli. The sound stimuli which showed the highest and the lowest presence in Exp. 1 were included in this experiment both with and without pictures. These stimuli were "sound of a train passing" and "roaring of a waterfall," respectively. Thus the number of stimuli in this experiment was nineteen instead of seventeen.

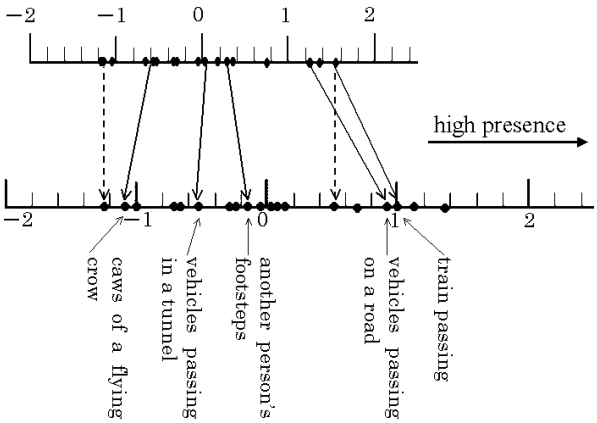
Experiment 2 was conducted about one month after Exp. 1. The subjects were asked to evaluate the auditory presence irrespective of the quality of the moving pictures.

Results and discussion

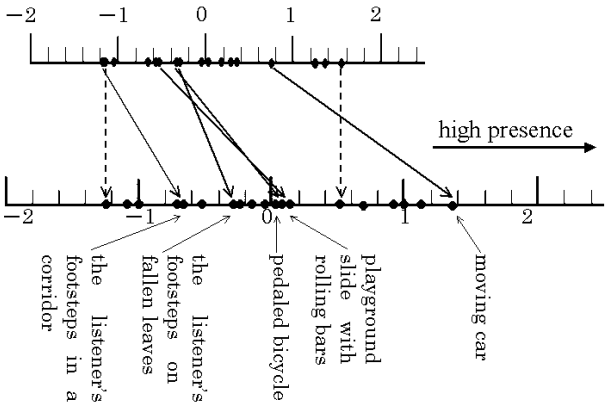
A scaling procedure was applied separately to the data obtained in the two experiments. The results of scaling are shown in Figure 1 classified by the three stimulus groups. In each panel, the upper and the lower scales indicate the results of Exps. 1 and 2, respectively. The two scales were adjusted so that the two stimuli included in both experiments meet in vertical positions as shown in the figure by broken arrows.

For the results of Exp. 1, in which only sounds were presented, the stimuli with distinctly moving sound sources or large sound images were rated higher in presence in general. This shows good agreement with the results in the authors' previous report [1].

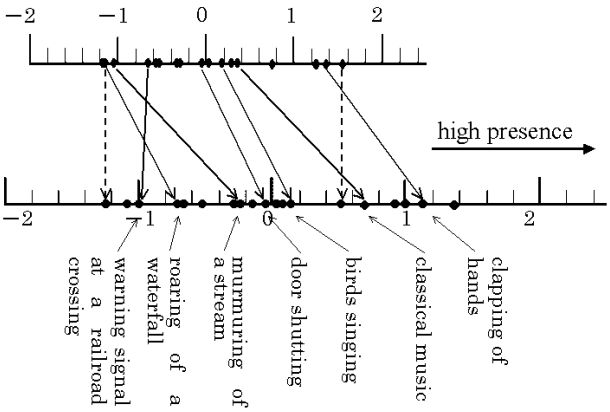
The effects of visual information will be discussed hereafter. Figure 1(a) shows the results for stimulus group I, i.e., the group of stimuli with moving sound sources. Attaching the visual stimulus to the auditory stimulus induced a clear increase in the auditory presence for the objects moving at high speed, such as "sound of a train passing" and "sound of vehicles on a road." However, auditory presence for the stimulus of "sound of vehicles passing in a tunnel" varied only slightly. This seems to be because the moving picture was so dark that there was little effect of visual information. As for the stimulus of "caws of a flying crow," the presence became lower. This is considered to be due to the fact that the flying crow did not appear in the corresponding picture.



(a) Results for stimulus group I.



(b) Results for stimulus group II.



(c) Results for stimulus group III.

Figure 1: Changes in evaluated presence of the stimuli by addition of visual information.

For stimulus group II, i.e., the group of stimuli with sound sources moving with the listener, the general tendency of auditory presence to become higher was significant as shown in Figure 1(b). As for these stimuli, the sound sources themselves were stationary relative to the listener when only sounds were presented. In contrast, the listeners could perceive their movement when the corresponding moving pictures were presented with the sounds. This might have contributed to the changes. As for "the listener's footsteps," the stimulus "in a corridor" showed a larger change than "on fallen leaves." Although the two walking-speeds were almost the same, the change in moving pictures was relatively large for "in a corridor" because the walls and ceiling were near the listener. This might have resulted in the larger change in auditory presence.

Figure 1(c) exhibits the results for stimulus group III, i.e., the group with stationary sound sources in their positions. There tended to be a higher presence for the stimuli with moving visual objects in the pictures. For the stimulus of "warning signal at a railroad crossing," the change in auditory presence was slight because the signal did not appear in the moving picture.

Conclusion

Auditory presence showed a marked change if the corresponding moving picture gave information on the original sound field correctly such as when the sound source appeared in the picture. Moreover, the tendency for auditory presence to increase was clear when visual objects in the picture moved.

Acknowledgements: The authors wish to thank Mr. Satoshi Ohtake of University of Yamanashi for his cooperation in executing the experiments. This study was partially supported by MEXT (No. 11832008) and the Cooperative Research Project Program of the Res. Inst. Elect. Comm., Tohoku University (No. H15/A08).

References

[1] K. Ozawa, Y. Chujo, Y. Suzuki and T. Sone, "Contents which yield high auditory-presence in sound reproduction," *Kansei Engineering International* **3**, 25-30 (2002).

[2] K. Ozawa, Y. Chujo, Y. Suzuki and T. Sone, "Psychological factors involved in auditory presence," *Acoust. Sci. and Tech.* **24**, 42-44 (2003).

[3] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics* **36**, 171-218 (1992).

[4] S. Ura, "An analysis of a paired comparison experiment," *Quality Control - J. of JUSE* **16**, 78-80 (1959).