Effects of visual information on auditory presence

Kenji Ozawa1, Satoshi Ohtake1, Yōiti Suzuki2 and Toshio Sone3

1Faculty of Engineering, Yamanashi University,
4–3–11 Takeda, Kofu, 400–8511 Japan
2Research Institute of Electrical Communication/Graduate School of Information Sciences, Tohoku University,
2–1–1 Katahira, Aoba-ku, Sendai, 980–8577 Japan
3Faculty of Systems Science and Technology, Akita Prefectural University,
84–4 Ebinokuchi, Tsuchiya, Honjo, 015–0055 Japan

(Received 9 September 2002, Accepted for publication 2 October 2002)

Keywords: Auditory presence, Binaural reproduction, Visual information, Scheffe’s paired comparison

PACS number: 43.66.Lj, 43.66.Pn

1. Introduction
The authors have studied auditory presence as the first step to understanding the multimodal characteristics of the impression of presence [1,2]. The word “presence” here corresponds to the Japanese word “rinjōkan” [3], with the meaning that one has the illusion of being in some place or environment, even when one is physically situated in another. In this study two experiments were conducted to investigate the effects of visual information on auditory presence.

2. Psychoacoustical experiment
2.1. Stimuli
The seventeen sounds shown in Table 1 were binaurally recorded via a dummy head (Koken, Samurai), 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). Prior to the presentation, the frequency-response characteristics of the outer ears of the dummy head were rectified through digital signal processing in order to avoid the duplication of the characteristics in recording and reproduction [4]. The moving pictures were presented using a 50-inch display (Sony, KL-X9200J) located 2 m away from the subject.

2.2. Experiment 1: Presenting only sounds
The experiments were carried out using Scheffé’s method of paired comparison modified by Ura [5]. Six male and two female subjects, between 21 and 24 years of age, with normal hearing acuity participated in the experiments.

In Experiment 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 (= 17×19), because every possible permutation was presented to cancel out time order errors.

2.3. Experiment 2: Presenting sounds and moving pictures
In this experiment, both sounds and moving pictures were presented to the subjects. The experimental procedure was the same as that of Experiment 1 except for the number of stimuli. The sound stimuli which showed the highest and the lowest presence in Experiment 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 Experiment 1. The subjects were asked to evaluate the auditory presence irrespective of the quality of the moving pictures.

3. Results and discussion
A scaling procedure was applied separately to the data obtained in the two experiments. The results of scaling are shown in Fig. 1 classified by the three stimulus groups. In each panel, the upper and the lower scales indicate the results of Experiments 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 Experiment 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.

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 Fig. 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. Moreover, the tendency for auditory presence to increase was clear when visual objects in the picture moved.

Acknowledgements
This study was partially supported by the Ministry of Education, Culture, Sports, Science and Technology (No. 11832008), and carried out under the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University (No. H12/A05).

References
