Relationship between gaze direction and sound localization in ventriloquism effect

Junsei Sato1,*, Kazutomo Fukue2,†, Yuichiro Kinoshita2,‡ and Kenji Ozawa2,§

1Advanced Image & Telecommunication Labs., SHARP Corporation, 1–9–2 Nakase, Mihama-ku, Chiba, 261–8520 Japan
2Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 4–3–11 Takeda, Kofu, 400–8511 Japan

(Received 19 May 2010, Accepted for publication 5 August 2010)

Keywords: Ventriloquism effect, Sound localization, Gaze direction, Loudspeaker, Amplitude panning

PACS number: 43.66.Lj, 43.66.Qp [doi:10.1250/ast.32.40]

1. Introduction

Sound localization for consistency in visual and sound images is important to the product quality of large televisions. Our previous study [1] revealed that the ventriloquism effect [2,3] is not robust enough to localize sound images at any location on the screen with a pair of loudspeakers placed at each side below the screen. In that study, however, the experimental results varied between subjects. To investigate the reason behind such variance, the relationship between gaze direction and sound localization in the ventriloquism effect is studied in this letter.

2. Experimental testing

2.1. Stimuli

A 65-inch television set and a pair of single cone loudspeakers were arranged as shown in Fig. 1. The distance between the subject and the surface of the television screen was 1.5 m. The loudspeakers were covered with acoustically transparent lace curtains which prevented the subjects from seeing them. The subjects’ gaze directions were recorded with an eye-tracking apparatus (DITECT, VIEW-TRACKER). A video clip of a metronome and its sound were used as visual and auditory stimuli (Fig. 2). A location of the metronome’s pendulum pivot on the clip represented the coordinates of the visual stimulus. A location of the stimulus was one of four cases: (1) nowhere, (2) at the center on the right half of the screen, (3) at the center on the left half of the screen, and (4) at the center of the screen (Fig. 3). In all cases, the vertical angle of the location was 7 degrees above the listener’s eye level. With regard to the auditory stimuli, the sound pressure level differences of the right and left channels were −12, −8, −4, −2, 0, 2, 4, 8 and 12 dB. The sound was presented three times at 60 beats per minute. The sound pressure level at the position of the subject’s head was 70 dB. All experiments were conducted in a soundproof room.

2.2. Procedure

As mentioned above, the visual stimulus was at one of four positions and there were nine sets of auditory conditions, so that the number of pair combinations was 36. The stimuli were presented randomly to eliminate order effects, and each combination was tested five times. Prior to each trial, a cross marker X with a one-degree angle was presented at a random location on the screen for two seconds to eliminate initial gaze direction effects. Then the visual and auditory stimuli were synchronously presented for three seconds after the cross marker disappeared. Each click of the metronome mimicked a real one, i.e., it was emitted at the same time as when the metronome in the video clip actually made the sound. Subjects were asked to judge the location of the sound image by clicking a mouse with the cursor on the screen. Clicking must be completed within five seconds after the end of video clip playback. No instruction concerning the gaze direction during the trials was given to the subjects. There were nine subjects aged from 20 to 29 years.

3. Results

3.1. Sound localization classified by visual stimuli

Figure 4 shows, as an example, horizontal angles of sound image directions as a function of level differences between the right and left signals. The plots with square markers correspond to case (1) in Section 2.1, and circular markers correspond to case (4). The range of the vertical axis corresponds to the width of the screen. The error bars represent standard deviations, and the dashed line describes the location of the visual stimulus.

The slope of the plots for case (4) were more gradual than that of case (1). Two-factor repeated measures ANOVA confirmed the interaction between the visual stimulus location and the sound pressure level differences between the right and left channels. Thus, the ventriloquism effect that the sound image shifts to the corresponding visual stimulus was obtained in this and all cases.

3.2. Gaze direction

In the trials, the subjects’ eye motions were continuously tracked throughout the period when the visual stimulus was displayed on the screen. The obtained data of the motions appeared to be categorized into three types: (i) staying at the visual stimulus, (ii) moving back and forth between the visual stimulus and another location at least once, and (iii) staying at other locations. In order to examine the relationship between the gaze motion and the ventriloquism effect, all data were subjectively classified into one of the above three types. Additionally a new type, (iv) others, was defined for data

*Email: sato.junsei@sharp.co.jp
†Email: g09mk030@yamanashi.ac.jp
‡Email: ykinoshita@yamanashi.ac.jp
§Email: ozawa@yamanashi.ac.jp
belonging to none of the first three types. Type (iv), for instance, represents the following cases: gaze directions wandered among more than two locations, or the eye-tracking apparatus failed to track the motions.

Figure 5 shows an example of the classification results where the visual stimulus condition was case (4) defined in Section 2.1. The vertical axis shows the ratio of the number of trials for each type to the total number of trials. The total number of trials was 45, i.e., five judgments were made by each of the nine subjects. The graph indicates that the ratio of gaze directions leaving the visual stimuli becomes high when the sound image is far from the visual stimulus location. The relationship between the gaze direction and sound localization is examined in the following subsection.

3.3. Relationship between sound image location and eye motion type

Figure 6 shows a comparison of sound image locations between two motion groups: one consists of motion type (i) defined in the previous subsection, and the other consists of types (ii) and (iii). The three graphs correspond to cases (2), (3) and (4) defined in Section 2.1. The horizontal axes of these graphs denote sound image locations for case (1), which were converted from the right and left level differences using the square plots in Fig. 4. The vertical axes are sound image locations for the other cases.

Each horizontal dashed line describes the location of each visual stimulus. As Fig. 5 shows, it can be observed that gaze directions rarely stayed at the visual stimuli when the sound image locations were very far from the stimuli. Therefore, to ensure reliability, the plots representing fewer trials than three
are omitted. The graphs in Fig. 6 show that the plots for the case that gaze direction stays at the visual stimuli were significantly distributed around the locations of the visual stimuli regardless of the sound condition; this means the ventriloquism effects were obtained. In contrast, the plots for the cases that gaze directions leave the visual stimuli lay around the diagonal dashed lines, indicating that the ventriloquism effects were not obtained. For verification, ANOVA was conducted and the interaction between gaze direction types and sound pressure level differences was statistically significant. An interpretation of this result is that sound image locations are modulated by gaze direction motions.

Overall, these results lead to two conclusions. 1. If a gaze direction stays at the visual stimulus, the ventriloquism effect is expected to be obtained even when the angle between the sound image and the visual stimulus reaches 10 or 20 degrees, 2. Otherwise, it is not obtained at all.

4. Conclusion

The results demonstrate a clear relationship between the ventriloquism effect and gaze direction. Unless subjects are staring at the visual stimuli, the ventriloquism effect may not be obtained at all. Further study is required to quantitatively verify this.

References