The role of auditory distance information in visual size perception

Daiki Yamasaki* and Hiroshi Ashida†

Graduate School of Letters, Kyoto University,
Yoshida Honmachi, Sakyo-ku, Kyoto, 606–8501 Japan

(Received 30 September 2018, Accepted for publication 8 May 2019)

Keywords: Audiovisual integration, Auditory distance perception, Visual size perception

PACS number: 43.66.Lj [doi:10.1250/ast.41.331]

1. Introduction

An object’s actual size can be visually inferred from its retinal size and distance from the observer, and distance information is indispensable for the visual system to uniquely determine the actual size. In daily life, distance information is mainly conveyed by the visual and auditory modalities. Auditory estimation of distance is generally poorer [1]. In ambiguous visual conditions, however, auditory cues become very meaningful in perceiving external events. To clarify how our brain achieves size perception, it is important to understand the role of auditory distance cues in size perception of ambiguous visual objects.

There are a few studies suggesting that visual size estimation achieves using not only visual but also multisensory cues. In the audiovisual domain, it has been demonstrated that the temporal asynchronies between pairs of audiovisual stimuli induced the perceived size changes [2]. Our recent study has shown that human participants utilize auditory information of changing distance to compensate for visual size perception of ambiguous objects [3]. Based on the evidence, it is suggested that the brain can integrate auditory and visual cues to accomplish size perception when visual distance information is available.

Here, the next question arises that whether and how auditory distance information improves visual size estimation when visual distance cues are ambiguous or even unavailable. The aim of the current study was to investigate the question in a direct way. We constructed the experimental situation in which participants could use the monocular image size of an object and auditory distance information. Participants performed a size adjustment task of visual targets paired with auditory distance information in the visual system, the adjusted diameters of visual targets should decrease when auditory distance information is available.

2. Experimental methods

2.1. Participants

Six adults with normal hearing and normal or corrected vision (mean age = 23.3, SD = 1.89, three females) took part in the experiment.

2.2. Apparatus and stimuli

The auditory stimuli consisted of binaural recordings of a sequence of five white-noise bursts, which was generated with Audacity (version 2.0.6.0). Each burst had 50 ms duration and each inter stimulus interval was 450 ms. The sounds were presented through the loudspeakers (SONY PCVA-SP2) located at 50, 100, and 200 cm in front of each participant’s head, and individually recorded using binaural microphones (Roland CS-10EM) and a PCM recorder (TASCAM DR-05). The intensity of the sound was about 75 dB SPL at 50 cm, measured roughly at the location of the participants’ ears using a noise meter (SMART SENSOR AR814). In the experiment, the auditory stimuli were presented through the earphones. The background noise level was about 31.0 dB SPL.

The visual stimuli were sequences of white-colored circles presented on a black OLED monitor screen (SONY PVM-A170, 60 Hz refresh rate) placed 100 cm in front of participants. Throughout the experiment, participants viewed the stimuli monocularly with the dominant eye.

The audiovisual stimuli were simultaneously presented using Psychopy software and its Pyo back-end for the sound presentation (version 1.92, [5]) on a computer (Apple Mac Pro). The relative timing of the audiovisual stimuli was measured and calibrated using an Arduino microcomputer attached to a photocell and a condenser microphone.

2.3. Procedure

Participants performed a size adjustment task, where they were exposed to a synchronized pair of audiovisual stimuli in each trial. Participants held a plastic ball with 4.4 cm diameter in their left hands as the reference stimulus throughout the experiment. We employed the real object reference in order to avoid possible effects of a visual reference stimulus, for example size aftereffect [6], on the estimated target size. Participants were asked to familiarize themselves with the reference ball before the experiment. They were instructed to

*e-mail: yamasaki.daiki.58n@st.kyoto-u.ac.jp
†e-mail: ashida@psy.bun.kyoto-u.ac.jp

VOLUME 41, NUMBER 1
adjust the actual diameter of the target to it of the reference ball. They were allowed to increase and decrease the target diameter by moving a computer mouse forward and backward, respectively. The initial target diameter was randomly set from either 1.0’ or 5.0’. When they felt that they adjust the size correctly, they pressed the mouse button and start the next trial. The reported diameter was recorded in visual angle (deg) in each trial.

Each participant performed the two Room illumination conditions, Dark room and lit room, in separate blocks. Participants were taken into the test room blindfolded for the purpose of preventing them from knowing the monitor distance. Each of the blocks consisted of 20 repetitions for each of the three auditory distances. Trial order was randomized for each block and participant. The first block was conducted in a dark room with normal room reverberation, and the second in the same room with the light on. Thus, there were in total six experimental conditions (three Auditory distances and two Room illuminations). After the experiment, we asked participants to verbally report the perceived distance of each auditory stimulus.

3. Results
Data analysis
We analyzed all the individual data with linear mixed models, using the lme4 and lmerTest packages on the R software (version 3.3.1, http://www.r-project.org/). We first made a statistical inference using a maximum model with all possible fixed and random effects, and then conducted an automated model selection with backward elimination by using the step function of the lmerTest package. In the present case of the balanced factorial design, a full-factorial ANOVA is known to provide test statistics for all fixed effects [7]. Thus, we conducted backward elimination only for random effects, and, based on the selected model, we implemented an ANOVA F test for fixed effects using the anova function of the lmerTest package. Further, we used the diffusmeans function of the lmerTest package to conduct multiple comparisons across the conditions. To correct p values for multiple comparisons, the Bonferroni method was applied.

3.1. Size adjustment task
We analyzed all the individual data using the linear mixed model that was selected with backward elimination. The model included all possible fixed effects and the random by-participant intercepts and slopes of Room illumination. Figure 1 shows the least-squares mean of the reported target diameters in each condition. The object with 4.4 cm diameter should subtend a visual angle of 2.52 deg. According to the formula: \( \theta = 360/\pi \tan^{-1}(4.4/D/2) \), giving ideal performances given by the auditory distances (depicted in a gray line). Thus, for example, the ideal performance for 100 cm auditory distance condition should subtend a visual angle of 2.52 deg.

To examine the hypothesis that the auditory influence on visual size estimation changes depending on Room illumination, we conducted a two-factorial ANOVA F test for fixed effects and revealed significant two-way interaction \((F(2,704) = 7.56, p < 0.001)\). There was significant main effect of Auditory distance \((t(2) = 67.91, p < 0.001)\), but not of Room illumination \((t(1) = 0.75, p = 0.43)\). These results show that the adjusted target size changed depending on auditory distance and, moreover, auditory distance information differently modulated visual size perception based on the room illumination.

Given the significant interaction, we conducted multiple comparisons across the conditions to investigate how the auditory and illumination factors affected the visual task in more detail. There was no significant difference between the illumination conditions at the auditory distance of 50 cm \((t(1,066) = -1.44, p = 0.302)\). On the other hand, the mean target diameter was significantly smaller for the Dark room condition than the Lit room condition at the auditory distance condition of 200 cm \((t(1,066) = -4.98, p < 0.001)\). This pattern of results indicates that auditory distance information more drastically altered the adjusted target size in the Dark room condition, compared to the other condition.

3.2. Perceived distances of the binaural recordings
Figure 2 shows the reported distances to the auditory stimuli. The perceived distance successfully increased with increasing the actual distance to the sound source. Contrary to the previous findings of a central tendency in human auditory distance perception (for review, [1,8]), participants generally overestimated the distances to the sound sources.

![Fig. 1](image1.png)  
**Fig. 1** The target diameter of each conditions estimated using a linear mixed model. Error bars refer to SE.

![Fig. 2](image2.png)  
**Fig. 2** Perceived distances of the binaural recordings. The dot refers to the individual report by each participant.
4. Discussion

The results showed that the size perception could be achieved by combining monocular image size and auditory distance information. As depicted in Fig. 1, the adjusted target size monotonically changed as a function of auditory distance. This result indicates that auditory distance information contributes to the scaling of visual size perception, both under darkened and well-lit circumstances. However, the adjusted diameters showed a central tendency and did not fit the ideal performances given by auditory distances, suggesting that the visual size scaling by using auditory distance information was suboptimal.

As hypothesized, the result showed that auditory distance information had a greater effect on visual size scaling when visual distance cues were unavailable (Dark room) compared to when available (Lit room). A possibly confounding factor is the effect of visual distance cues that were available in the Lit room condition. If the visual distance to the monitor (100 cm in the study) was used in the Lit room condition, the adjusted diameters should get near to the ideal performance at 100 cm. Given that the adjusted diameters tended to be rather larger in the Lit room condition, it is less likely to merely reflect the effect of such visual distance cue.

One might argue that the adjusted size could have been affected by the procedure, where participants needed to convert the tactiley perceived size to visual size. Previous research showed that the tactile size of an object is underestimated than actual [9]. Considering the constant overshooting of the adjusted diameters, however, the visuo-tactile conversion is unlikely to be a crucial factor. Besides, because the reference size was constant throughout the experiment and participants were satisfactorily familiarized with the size beforehand, possible influences of tactile sense, if any, hardly challenge that auditory distance information contributes to visual size perception.

To summarize, the current study demonstrated that humans utilize auditory distance information to rescale visual size perception. Future studies need to clarify the mechanism of the audiovisual integration of retinal size and auditory distance.

References