Vision Research 50 (2010) 2093–2099

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Sound can prolong the visible persistence of moving visual objects

Souta Hidaka [a,*], Wataru Teramoto [a,b], Jiro Gyoba [a], Yôiti Suzuki [b]

[a] Department of Psychology, Graduate School of Arts and Letters, Tohoku University, Sendai, Miyagi 980-8576, Japan
[b] Research Institute of Electrical Communication, Tohoku University, Sendai, Miyagi 980-8577, Japan

ARTICLE INFO

Article history:
Received 19 March 2010
Received in revised form 23 July 2010

Keywords:
Visible persistence
Frequency changes of sounds
Audio–visual interaction
Object representation
Apparent motion

ABSTRACT

An abrupt change in a visual attribute (size) of apparently moving visual stimuli extends the time the changed stimuli is visible even after its physical termination (visible persistence). In this study, we show that elongation of visible persistence is enhanced by an abrupt change in an attribute (frequency) of the sounds presented along with the size-changed apparently moving visual stimuli. This auditory effect disappears when sounds are not associated with the visual stimuli. These results suggest that auditory attribute change can contribute to the establishment of a new object representation and that object-level audio–visual interactions can occur in motion perception.

1. Introduction

For our perceptual systems, which continuously receive large amounts of input, it is tremendously important to be able to distinguish a single object from the input and preserve single-object continuity across time and space. In visual perception, object-based processing, in which recently sampled information is integrated with an existing representation of the scene, is assumed to be a solution for the brain to realize the distinction and preservation of information (e.g., "object file" theory; Kahneman, Treisman, & Gibbs, 1992). Although the spatiotemporal proximity or continuity of inputs is a predominant factor (Scholl, 2001), the consistency of each object's features is also likely to be an important cue. Moore, Mordkoff, and Enns (2007) demonstrated that an abrupt change in the attribute (size) of an object in an apparent motion sequence extended the time for which the changed object is visible (visible persistence) even after its physical termination (Fig. 1). They also showed that this elongation of visible persistence did not occur when the visual stimulus moved behind an occluder containing a small hole so that there was a reasonable change in size of the visual stimuli. These results suggest that the object-level changes are likely to cause this phenomenon rather than the retinal-level changes. In a motion display, the visual system is considered to reduce the duration of visible persistence to establish the perception of a single moving object from continuous physical inputs (Burr, 1980). Moore et al. (2007) suggested that this suppression mechanism does not work in cases where the object representations between the two neighboring frames of a sequence of visual stimuli are different; the perceptual systems establish a new object representation so that the visible persistence of the size-changed stimuli increases in motion sequences (see also Moore & Enns, 2004).

Our internal representations are established by combining or integrating multisensory inputs, depending on the accuracy and/or reliability of each input (Welch & Warren, 1986), to make coherent and robust percepts (Ernst & Bülthoff, 2004). Thus far, from this perspective, many studies have investigated audio–visual cross-modal contributions to primates' perception. Although the superiority of vision over audition has been reported frequently (e.g., ventriloquism effect, Howard & Templeton, 1966), recent studies have shown that auditory stimuli can modulate the perception of visual stimuli. For instance, transient auditory stimuli can capture the temporal position of visual stimuli (temporal ventriloquism effect; Vroomen & de Gelder, 2004). It has also been shown that transient auditory stimuli can enhance the perception of visual stimuli in regard to perceived intensity (Stein, London, Wilkinson, & Price, 1996), perceived vividness (Sheth & Shimojo, 2004), and the saliency of visual changes (Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008). Furthermore, the presence of multiple auditory stimuli can cause a static visual flash to be perceived as multiple flashes (Shams, Kamitani, & Shimojo, 2000).

The manner in which object representations are maintained or updated has been investigated mainly with regard to visual modality. Given that our internal representations are established by means of information from a variety of sensory modalities, object representations can also have multimodal characteristics.

* Corresponding author. Present address: Department of Psychology, College of Contemporary Psychology, Rikkyo University, 1-2-26, Kitano, Niiza-shi, Saitama 352-8558, Japan. Fax: +81 48 471 7164.
E-mail address: hidaka@rikkyo.ac.jp (S. Hidaka).

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.07.021

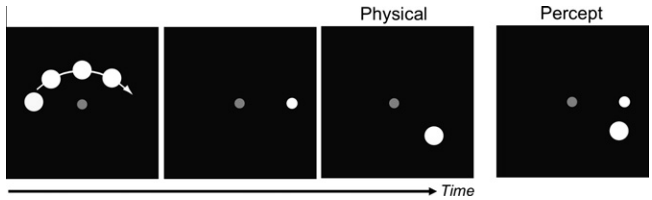

Fig. 1. Schematic illustration of the stimuli and the percept demonstrated by Moore et al. (2007). The size change in apparently moving stimuli induces the perception of multiple stimuli in the display where only one stimulus is physically presented.

Several studies with findings consistent with this idea have shown that object recognition reflects the semantic congruencies of inputs from different sensory modalities (Alpert, Hein, Tsai, Naumer, & Knight, 2008; Hein et al., 2007; Laurienti, Kraft, Maldjian, Burdett, & Wallace, 2004). This finding indicates that multimodal object representations are established in cognitive or semantic processing stages. However, it is not clear whether the object representations containing multimodal information can be formed in the perceptual processing stage, which would mediate the establishment of cognitive- or semantic-level object representations. In a motion display, object representations are being established and updated in response to incoming physical inputs at the perceptual level (Moore et al., 2007). In the current study, we measured the duration of visible persistence of an apparently moving object to investigate whether a radical change in an attribute (frequency) of sounds presented along with moving visual stimuli could contribute to the establishment of a new object representation (Moore et al., 2007).

In Experiment 1, we showed that frequency changes of auditory stimuli could alter the visible persistence of apparently moving visual stimuli. In Experiments 2 and 3, we confirmed that the abrupt onset or offset of auditory stimuli did not have any effect. In Experiment 4, we demonstrated that the effect of frequency changes in auditory stimuli on visible persistence disappeared when the auditory stimuli were difficult to be associated with apparently moving visual stimuli. These findings suggest that object representations can be formed by means of the multisensory integration of auditory and visual information in motion perception and that object-level audio–visual interactions can occur at a perceptual level.

2. Experiment 1

In Experiment 1, we investigated the effect of auditory attribute changes on the visible persistence of apparently moving visual stimuli. The frequency of auditory stimuli was abruptly changed in the penultimate frame, where the visual stimuli either did or did not change its size in apparent motion sequences.

2.1. Methods

2.1.1. Participants and apparatus

Written consent was obtained from each participant before all experiments. The experiments were approved by the local ethics committee of Tohoku University. In Experiment 1, there were 11 participants—two of the authors (S.H. and W.T.) and nine students of Tohoku University who were naive to the purpose of this experiment. All the participants had normal or corrected-to-normal vision and normal hearing. The visual stimuli were presented on a CRT display (Sony Trinitron GDM-FW900, 24-in.) with a resolution of 1600 × 1200 pixels and a refresh rate of 75 Hz. The auditory stimuli were presented via headphones (SENNHEISER HDA200). A customized PC (Dell-Dimension 8250) and MATLAB (The Mathworks, Inc.) with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) were used to control the experiment. The participants were instructed to place their heads on a chin rest.

2.1.2. Stimuli

A white disc (0.6° in diameter, 49.11 cd/m^2) was sequentially presented against a gray background (7.28 cd/m^2) as an apparent motion stimulus. The disc moved by 15° in every frame of 80 ms, following a circular trajectory whose center was located at a blue fixation point (5.31 cd/m^2, the CIE coordinates were 0.15 and 0.11; see Fig. 2). The radius of the imaginary circle of the motion trajectory was 3°. Two types of pure tone—600 Hz (L tone; SPL = 77.5 dB) and 3000 Hz (H tone; SPL = 71 dB)—were used as auditory stimuli. The duration of each tone was 80 ms with 8 ms of cosine ramp at the onset and offset and was consistent with the duration of each visual frame.

2.1.3. Procedure

In an apparent motion display, the disc was sequentially presented, in synchrony with the tone. In half of the trials, the size of the disc remained constant in all the frames (Without-Visual-Change condition); in the other half of the trials, the disc became smaller (0.45° × 0.45°) at the penultimate frame and returned to the previous size in the last frame (Visual-Change condition) (Fig. 2). As with previous research (Moore et al., 2007), our preliminary experiment confirmed that a change in size from smaller to larger did not lead to a sufficient effect. Moreover, in our experimental procedure, the effect of such a size change was found to be highly individual among the participants. We, therefore, introduced only larger-to-smaller changes in the Visual-Change condition. A sequence of either L or H tones—a tone for each frame—was constantly presented in the Without-Auditory-Change condition. In the Auditory-Change condition, a sequence of either L or H tones was presented up to the frame before the penultimate frame. Then, the tone was changed from L to H at the penultimate frame for the L tone sequence or vice versa for the H tone sequence and was returned to the previous tone at the last frame (i.e., LHL or HLH in the last three frames). The fundamental tone was the H tone for half of the trials and the L tone for the other half for the Without-Auditory-Change and Auditory-Change conditions. We also set a condition in which the auditory stimuli did not appear in all frames (Without-Sound condition). In addition, we manipulated the number of discs that were actually presented in the last frame. One disc was presented at the final location of the motion

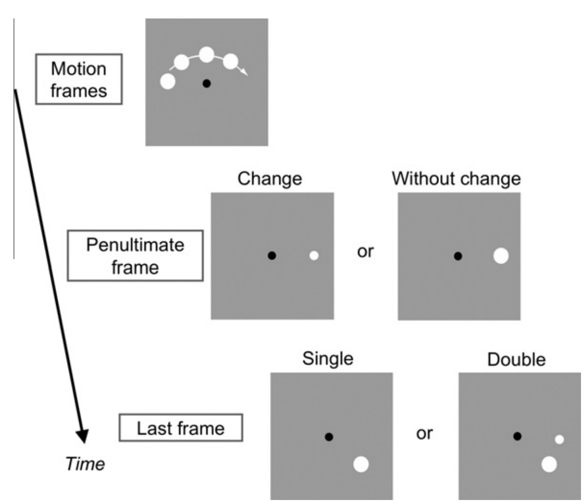

Fig. 2. Schematic illustration of time course and experimental design of visual stimuli used in Experiment 1. The fixation circle was blue, instead of black, in the experiment.