Chapter 35
Audiovisual Integration of Natural Auditory and Visual Stimuli in the Real-World Situation

Xiaoyu Tang
Okayama University, Japan & Northeast Normal University, China

Yulin Gao
Okayama University, Japan

Weiping Yang
Okayama University, Japan

Ming Zhang
Northeast Normal University, China

Jinglong Wu
Okayama University, Japan

ABSTRACT

Bimodal audiovisual (AV) stimuli are detected or discriminated faster and more accurately than either visual or auditory unimodal stimuli. This effect is called audiovisual integration. Recently, researchers have been increasingly focused on the audiovisual integration of natural, auditory, and visual stimuli in real-world situations. There are some differences between audiovisual integration of naturalistic stimuli and non-naturalistic stimuli, such as the time of occurrence of audiovisual integration, and the neural mechanism. Factors affecting audiovisual integration in real-world situations are summarized here, with particular focus on temporal asynchrony and semantic matching. Stimuli of audiovisual integration in the real-world situation should be controlled strictly, especially emotional factors, familiarity factors, semantic matching, and the match of the naturalistic stimuli and non-naturalistic stimuli. In the future, researchers should study the influence of attention on audiovisual integration and the mechanism of audiovisual integration with naturalistic stimuli in the real-world situation.

DOI: 10.4018/978-1-4666-2113-8.ch035

BACKGROUND

Imagine the following scene: a ferocious man is pointing a gun towards you. You can see him pulling the trigger and hear the gunfire. If you run as fast as possible, then you might be able to escape from his shot at your brain or heart. Now suppose that you cannot hear the gunfire, and you have to run off only based on the visual information. As we know, in the former situation, the chances of survival are greater than in the latter situation. Multi-sensory integration is very important for humans, not only in crisis situations, but also in everyday life. Therefore, it is crucial to find out how the human brain processes multi-sensory stimuli, especially the main auditory and visual stimuli. Multisensory integration refers to the brain's ability to synthesize the information that it derives from two or more senses (McGurk & MacDonald, 1976; Stein & Meredith, 1990) and bimodal audiovisual (AV) stimuli are detected or discriminated faster and more accurately than either visual or auditory unimodal stimuli (Li, Wu, & Touge, 2010). Giard and Peronnet (1999), adopting the ERPs technique and comparison approach [AV-(A+V)], found that audiovisual integration occurred at an early stage (about 40ms) after stimuli onset, which showed in different distributions of ERPs (Giard & Peronnet, 1999).

INTRODUCTION

Traditionally, researchers have used simple non-natural stimuli, such as a colored square and pure tone, to investigate the mechanisms of audiovisual integration. Audiovisual integration could occur at an early stage (40ms after stimuli onset) and/or a relatively late stage (160ms) (Giard & Peronnet, 1999; Talsma & Woldorff, 2005). Early multisensory interactions occur primarily when stimuli in at least one modality are low in intensity (Daniel Senkowski, Saint-Amour, Höfle, & Foxe, 2011). However, those studies cannot draw conclusions about possible multisensory interactions in early evoked brain responses to stimuli with different intensity levels across modalities, such as a low intensity auditory stimulus paired with a high intensity visual stimulus. Selectively exogenous attention (bottom-up) could modulate multisensory integration, or the divided attention to auditory stimuli could be a prerequisite for early integration (Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2008; Theeuwes, 1991; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). The role of endogenous attention (top-down) mechanisms in audiovisual integration processing has been discussed (Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). Generally speaking however, there are many unsolved problems, such as the occurring stage or influence of selective and divided attention to multisensory integration; researches are still controversial now (Koelewijn, Bronkhorst, & Theeuwes, 2010; Mozolic, et al., 2008; Schroeder & Foxe, 2005; Stein & Stanford, 2008).

As we know, naturalistic multisensory stimuli in the real-world are more complex. For more ecological validity, researchers began to study multisensory integration with natural stimuli in real-world situations. Murray et al. (2006) used event related potentials (ERPs), comparing responses to man-made subjects versus the sounds of living subjects, in an oddball paradigm. Their results showed that the processing of the sounds of living and man-made objects were different in the human brain, in speed and spatiotemporal respects (Murray, Camen, Andino, Bovet, & Clarke, 2006). Additionally, Beauchamp et al. took advantage of the high spatial resolution of functional magnetic resonance imaging (fMRI) and used animals and man-made objects as stimuli materials. They found that the posterior superior temporal sulcus (pSTS) and the middle temporal gyrus (MTG)—which is involved in the processing of environmental sounds—were more active when auditory and visual objects' features were presented together, and showed signal increases for meaningful stimuli (Beauchamp, 2005; Beauchamp, Lee, Argall, &