Audiovisual speech integration in the brain: Semantics and temporal synchrony

Efthymios Tsilionis¹, Argiro Vatakis¹,²,*

¹Department of Philosophy and History of Science, University of Athens, Panepistimioupoli, 15771, Athens, Greece
²Cognitive Systems Research Institute (CSRI), Makedonomachou Prantouna 7, 11525 Athens, Greece

Abstract

In everyday life we merge information from our different senses in order to perceive the external world. The binding of multisensory input, known as multisensory integration, results in a coherent and unified experience of external events, increases their detection, disambiguates their discrimination, leads to faster reaction times, and improves accuracy as opposed to unisensory input (e.g., Stein & Meredith, 1993). However, the accomplishment of integration depends on numerous low- and high-level factors. Low-level structural factors refer to temporal synchrony and spatial location of the stimuli, as well as any temporal correlation between them. Higher-level cognitive factors refer to semantic congruency, perceptual grouping, and phenomenal causality (Spence, 2007). One very popular theory in multisensory perception is the ‘unity assumption’ according to which, when signals from different modalities have many common amodal properties, it is more probable the perceptual system to treat them as originating from the same source (Vroomen & Keetels, 2010).

A typical event that is being influenced from all the above factors is audiovisual speech. Face-to-face communication requires the integration of the auditory (voice) and visual (lip and facial movements) information of a given speaker(s). One important aspect of speech is the tolerance of larger stimulus onset asynchronies (SOAs) as compared to nonspeech signals (e.g., bigger temporal window of integration for speech as compared to object actions; e.g., van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2007). Apart from the growing size of behavioral studies on the topic, the past decade a number of neuroimaging studies have focused on the neural basis of speech integration. A number of brain regions that have repeatedly been reported in audiovisual speech integration include high-level associative or integrative cortices such as the superior temporal sulcus (STS), intraparietal sulcus (IPS), inferior frontal gyrus (IFG), and insula, as well as subcortical regions such as the superior colliculus (SC) and primary sensory cortices (Calvert, Campbell, & Brammer, 2000; Macaluso et al., 2004; Stevenson et al., 2010). One common problem in studies investigating speech perception is the use of continuous speech which is a quite complex stimulus leading to larger and more variable temporal windows of integration as compared to brief speech stimuli (e.g., syllables, words;

*Corresponding author. Tel.: +30 211 012 4543; fax: +30 211 012 4544.
E-mail address: argiro.vatakis@gmail.com

1877-0428 © 2014 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of the Organizing Committee of the International Conference on Timing and Time Perception.
doi:10.1016/j.sbspro.2014.02.354
Vatakis & Spence, 2010). Another issue pertains to the possible recalibration effects due to the lengthy exposure to the asynchrony presented (e.g., Navarra et al., 2005).

In the present study by using functional magnetic resonance imaging (fMRI), we intend to examine the effect of semantics and temporal synchrony in unity as well as how this translates in brain activations. More specifically, we are going to use words and pseudo-words (two syllables long at most, thus controlling for complexity and recalibration effects), which will be presented in both the visual and auditory modality under different SOA’s. The sensory inputs will be presented in congruent (matching voice and lip/facial movements) and incongruent (mismatching voice and lip/facial movements) format (cf. Vatakis & Spence, 2008). Two tasks will be utilized in a block format. An explicit task, where participants will be required to detect the stimulus congruency, and an implicit task, where participants will be required to report the order of stimulus presentation (i.e., temporal order judgment task; TOJ). Both tasks will allow for the identification of activations related to the semantic relatedness of the stimuli, while the latter task will allow for the examination of visual and auditory speech influence in audiovisual speech integration. This experimental paradigm will allow us to identify the determinants of speech perception, given that our stimuli do not differ along physical (same temporal structure and complexity) and linguistic dimensions (same phonology) except for the semantics.

© 2014 The Authors. Published by Elsevier Ltd. Selection and peer-review under responsibility of the Organizing Committee of the International Conference on Timing and Time Perception.

Keywords: Speech integration; Unity effect; fMRI; Synchrony perception; auditory;visual

References


