NEUROSCIENCE
RESEARCH ARTICLE

A. M. Proverbio et al. / Neuroscience 385 (2018) 215–226

Electrophysiological Indexes of Incongruent Audiovisual Phonemic Processing: Unraveling the McGurk Effect

Alice Mado Proverbio,[a]* Giulia Raso[a] and Alberto Zani[b]

[a] *Neuro-Mi Center for Neuroscience, Dept. of Psychology, University of Milano-Bicocca, Italy*

[b] *IBFM-CNR, Milan Research Area 3, Italy*

Abstract—In this study the timing of electromagnetic signals recorded during incongruent and congruent audiovisual (AV) stimulation in 14 Italian healthy volunteers was examined. In a previous study (Proverbio et al., 2016) we investigated the McGurk effect in the Italian language and found out which visual and auditory inputs provided the most compelling illusory effects (e.g., bilabial phonemes presented acoustically and paired with non-labials, especially alveolar–nasal and velar–occlusive phonemes). In this study EEG was recorded from 128 scalp sites while participants observed a female and a male actor uttering 288 syllables selected on the basis of the previous investigation (lasting approximately 600 ms) and responded to rare targets (/re/, /ri/, /ro/, /ru/). In half of the cases the AV information was incongruent, except for targets that were always congruent. A pMMN (phonological Mismatch Negativity) to incongruent AV stimuli was identified 500 ms after voice onset time. This automatic response indexed the detection of an incongruity between the labial and phonetic information. SwLORETA (Low-Resolution Electromagnetic Tomography) analysis applied to the difference voltage incongruent–congruent in the same time window revealed that the strongest sources of this activity were the right superior temporal (STG) and superior frontal gyri, which supports their involvement in AV integration.

Key words: language, ERP, illusion, audiovisual, incongruence, MMN.

INTRODUCTION

This illusion, discovered by McGurk and MacDonald (McGurk and MacDonald, 1976) in English speakers, has also been observed in different languages (Sekiyama and Tohkura, 1991; Green and Norrix, 1997; Massaro and Palmer, 1998; Tiippana et al., 2004; Jiang and Bernstein, 2011), including Italian.

The McGurk effect occurs when a clearly audible syllable with one consonant is presented simultaneously with a visual presentation of a speaker articulating a syllable with a different consonant and the resulting percept is a syllable with a consonant other than the auditorily presented one. There are several studies about this effect, but, for the Italian language, it has only been examined at a behavioral level (for example Bovo et al., 2009; Proverbio et al., 2016).

In the literature there are several examples of perceptual illusions due to audiovisual (AV) integration. One is the *ventriloquist effect:* seen when two different stimuli (one visual and one acoustic) come from different spatial zones. The localization of acoustic stimulation is attributed to the spatial location of visual stimulation presentation (e.g. Shams et al., 2000). This effect demonstrates how visual stimuli can alter auditory perception. Another example is the *sound-induced flash illusion:* when a single flash is accompanied by two or more acoustic beeps. Subjects often report perception of multiple flashes too (Shams et al., 2002). The study shows how acoustic stimuli can alter visual perception. Time synchrony and/or spatial coincidence play a crucial role in modulating the effects of multisensory integration (Engel and Dougherty, 1971; Stein and Meredith, 1993; Zampini et al., 2005).

Multimodal audiomotor neurons located in the posterior superior temporal sulcus (pSTS) and in the medial temporal gyrus (MTG) respond both to sounds and to visual images of objects and animals (Wright et al., 2003), along with the right Heschl's gyrus and the inferior frontal cortex (Alpert et al., 2008). Speech stimuli

*Corresponding author. Address: Dept. of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy.
E-mail address: mado.proverbio@unimib.it (A. M. Proverbio).
Abbreviations: AV, audiovisual; BEM, boundary element model; ERP, evoked response potential; IFG, inferior frontal gyrus; LORETA, Low-Resolution Electromagnetic Tomography; MTG, medial temporal gyrus; pMMN, phonological Mismatch Negativity; PMv, ventral premotor; POp, pars opercularis; pSTS, posterior superior temporal sulcus; SMG, supramarginal gyrus; STG, superior temporal; STp, posterior superior temporal; STS, superior temporal sulcus; TVSA, temporal visual speech area.

https://doi.org/10.1016/j.neuroscience.2018.06.021

massively activate audio visuomotor neurons at temporal and premotor areas (Skipper et al., 2005, 2006, 2007; Pulvermüller and Shtyrov, 2006). A specific region located posteriorly and ventrally to STS, named the *temporal visual speech area* (TVSA), seems particularly responsive to it (Bernstein and Liebenthal, 2014; Calvert et al., 1999, 2000, Calvert and Campbell, 2003), along with the inferior frontal gyrus (IFG) (Hasson et al., 2007). STS is also stimulated by AV incongruence of musical gestures and related sounds in expert musicians, suggesting its role in the multimodal representation of motor actions (including phono-articulatory ones) and their visual and acoustic properties (Lam et al., 1999; Proverbio et al., 2014, 2017). Proverbio et al. (2011) have shown that visual processing alone may stimulate the auditory cortex (superior temporal gyrus, BA41, BA42, BA22) if objects or persons displayed (e.g., a man playing a saxophone, or using a jackhammer*)* convey strong auditory associations.

While listeners tend to base their speech comprehension more on the auditory than on the visual information (Chen and Massaro, 2004; Gentilucci and Cattaneo, 2005) visual processing of labial information is crucial to speech perception. If the auditory input is degraded, recognition may drop from 45% to 6% without labial information (Middelweerd and Plomp, 1987). This may explain why, when inconsistent AV information is presented, McGurk illusory effects may occur. According to fMRI data (e.g., Skipper et al., 2007) AV speech would elicit in the listener a motor plan for the production of the phoneme that the speaker might have been attempting to produce, and that feedback in the form of efference copy from the motor system ultimately influences the phonetic interpretation. The circuits would involve primary auditory cortex (A1), posterior superior temporal (STp) areas, supramarginal gyrus (SMG), somatosensory cortices (SI/SII), ventral premotor (PMv) cortex, and the pars opercularis (POp).

Electrophysiological studies have investigated the timing of AV integration. Kumar et al. (2016) compared processing of congruent or incongruent AV speech, in which auditory input might precede (−450 ms), follow (+450 ms) or be simultaneous with visual stimulation, finding that the McGurk illusions occurred more frequently for simultaneous AV stimulation. Van Wassenhove et al. (2007) demonstrated that the temporal window allowing fusion and distorted perception ranged between −30 ms and +170 ms of asynchrony.

The role of the two cerebral hemispheres in McGurk illusion generation is not clear.

On one side a large literature supports the role of the right superior temporal sulcus (STS) for processing of faces, voices, and face-voice integration, AV integration (e.g., Ethofer et al., 2013; Watson et al., 2014; Hagan et al., 2009), and the ventriloquism illusion (Macaluso et al., 2004). Interestingly, preserved McGurk effects, an illusion that requires integration of auditory and visual speech, were shown in a patient with left tempo-parietal area lesion (Baum et al., 2012), thus suggesting a role of the right homologous counterpart in AV integration. On the other hand, some studies hint at a specific role of the left STS in AV integration, due to the linguistic nature of stimuli (phonemes). For example, an fMRI study investigated the neural underpinnings of inter-individual variability in the perception of the McGurk illusion (Nath and Beauchamp, 2012) finding that the amplitude of the response in the left STS was significantly correlated with the likelihood of perceiving the McGurk effect. In addition, Pratt et al. (2015), investigating the spatio-temporal distribution of cortical activity during audio–visual congruent/incongruent (McGurk) stimulation, showed the crucial role of left hemispheric regions in the first 200 ms of processing within the auditory cortex (superior parietal cortex, middle temporal cortex Wernicke's area). A similar left hemispheric asymmetry was observed by Bernstein et al. (2008) in the activation of the supramarginal and angular gyrus (SMG/AG) specialized in phonetic processing in early (100–200) AV speech processing.

The problem with previous investigations on the McGurk illusion is that sometimes contrasts are limited to one or two AV pairing (only /ba/ vs. /ga/ or vs. /va/), which hardly can represent speech perception. Furthermore, it should be considered that AV incongruent stimulation does not necessarily lead to illusory percepts that can be instead generated by a different neural mechanism. For example (as hypothesized here) a left hemispheric STS involvement might reflect AV integration relying on early phonetic processing carried out by AG/SMG areas, while McGurk illusions might depend more on a right hemispheric processing of lip motion (bodily) information.

The aim of the research was to investigate the temporal course of brain activation during the perception of incongruent AV stimulation (leading to the McGurk effect to an estimated 1/4th of the cases, according to Proverbio et al., 2016) through the use of an implicit task. Participants had to respond to rare target phonemes (e.g. /ru/) which were presented randomly intermixed with congruent or incongruent non-targets. On the basis of the available literature we expected to find a pMMN (phonological Mismatch Negativity) elicited by incongruent AV stimulation (Colin et al., 2002; Stekelenburg and Vroomen, 2012; Eskelund et al., 2015). We therefore applied source reconstruction to the incongruent-congruent differential activity to unravel the neural bases of incongruent AV perception. pMMN is an electrophysiological linguistic component indicating the occurrence of the phonological stage of processing (Proverbio and Zani, 2010). It belongs to the wider class of Mismatch Negativity (MMN) responses, mainly generated in the auditory cortex, whose amplitude depends on the degree of variations/changes in the expected auditory percept, thus reflecting the cortical representation of auditory-based information (Näätänen, 1995). The MMN is elicited when the auditory input does not match the actual or predicted sensory information encoded and is generated by an automatic memory-based change-detection mechanism that operates independently of the listener's attention or behavioral goals (Näätänen et al., 2007; Czigler and Winkler, 2010). For this reason it is particularly suitable for implicit paradigms where subjects' attention is diverted elsewhere.