Neuropsychologia 45 (2007) 587–597

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion

Dave Saint-Amour [a], Pierfilippo De Sanctis [a], Sophie Molholm [a,b], Walter Ritter [a,b], John J. Foxe [a,b,*]

[a] *Cognitive Neurophysiology Laboratory, Nathan S. Kline Institute for Psychiatric Research, Program in Cognitive Neuroscience and Schizophrenia, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA*

[b] *Program in Cognitive Neuroscience, Department of Psychology, The City College of the City University of New York, 138th Street & Convent Avenue, New York, NY 10031, USA*

Received 26 October 2005; received in revised form 23 March 2006; accepted 31 March 2006
Available online 6 June 2006

Abstract

Seeing a speaker's facial articulatory gestures powerfully affects speech perception, helping us overcome noisy acoustical environments. One particularly dramatic illustration of visual influences on speech perception is the "McGurk illusion", where dubbing an auditory phoneme onto video of an incongruent articulatory movement can often lead to illusory auditory percepts. This illusion is so strong that even in the absence of any real change in auditory stimulation, it activates the automatic auditory change-detection system, as indexed by the mismatch negativity (MMN) component of the auditory event-related potential (ERP). We investigated the putative left hemispheric dominance of McGurk-MMN using high-density ERPs in an oddball paradigm. Topographic mapping of the initial McGurk-MMN response showed a highly lateralized left hemisphere distribution, beginning at 175 ms. Subsequently, scalp activity was also observed over bilateral fronto-central scalp with a maximal amplitude at ∼290 ms, suggesting later recruitment of right temporal cortices. Strong left hemisphere dominance was again observed during the last phase of the McGurk-MMN waveform (350–400 ms). Source analysis indicated bilateral sources in the temporal lobe just posterior to primary auditory cortex. While a single source in the right superior temporal gyrus (STG) accounted for the right hemisphere activity, two separate sources were required, one in the left transverse gyrus and the other in STG, to account for left hemisphere activity. These findings support the notion that visually driven multisensory illusory phonetic percepts produce an auditory-MMN cortical response and that left hemisphere temporal cortex plays a crucial role in this process.

Keywords: Multisensory integration; McGurk illusion; Mismatch negativity; Topography; Preattentive; Audio–visual speech

1. Introduction

The mismatch negativity (MMN) is a well-known electrophysiological component reflecting preattentive detection of an infrequently presented auditory stimulus ('deviant') differing from a frequently occurring stimulus ('standard') (Näätänen & Alho, 1995; Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995). Generation of the MMN is believed to reflect the cortical processes involved in comparing current auditory input with a transient memory trace (lasting ∼10–20 s) of ongoing regularities in the auditory environment; when there is a perceptible change, there is an MMN response (Näätänen, 2001). As such, the MMN serves as an index of auditory sensory (echoic) memory and constitutes the only available electrophysiological signature of auditory discrimination abilities (Picton, Alain, Otten, Ritter, & Achim, 2000). Changes along several physical dimensions such as duration, intensity, or frequency of sounds can generate the MMN, including changes in spectrally complex stimuli like phonemes (Näätänen et al., 1997). Source analysis of magnetic (Hari et al., 1984; Sams et al., 1985) and electrical scalp-recordings (e.g. Giard, Perrin, Pernier, & Bouchet, 1990; Scherg & Berg, 1991) as well as intracranial recordings in animals (e.g. Csèpe, Karmos, & Molnar, 1987; Javitt, Steinschneider, Schroeder, Vaughan, & Arezzo, 1994) and humans (e.g. Rosburg et al., 2005) have shown that the principal neuronal generators of MMN are located on the supratemporal plane in auditory

* Corresponding author. Tel.: +1 845 398 6547; fax: +1 845 398 654.
E-mail address: foxe@nki.rfmh.org (J.J. Foxe).

0028-3932/$ – see front matter © 2006 Published by Elsevier Ltd.
doi:10.1016/j.neuropsychologia.2006.03.036

cortex, although additional regions including frontal and parietal cortices are also likely associated with MMN processing (Marco-Pallares, Grau, & Ruffini, 2005; Molholm, Martinez, Ritter, Javitt, & Foxe, 2005).

Although the MMN is typically elicited by physical changes in the regularity of acoustical signals and attention is not required for its generation, evidence suggests that MMN is associated with purely subjective changes in auditory percepts in the absence of any actual acoustical variation. This was first reported by Sams et al. (1991) using the so-called McGurk illusion, a remarkable multisensory illusion whereby dubbing a phoneme onto an incongruent visual articulatory speech movement can lead to profound illusory auditory perceptions (McGurk & MacDonald, 1976). Depending upon the particular combination of acoustic phoneme and visual speech articulation used to evoke the McGurk illusion, the resultant percept will tend to be a fusion of the mismatched auditory and visual speech inputs (e.g. auditory /ba/ dubbed onto visual /ga/ results in the percept of /da/) or will be dominated by the visual speech input (e.g. auditory /ba/ dubbed onto visual /va/ results in the percept of /va/). Even when initially naïve subjects are fully apprised of the nature of the illusion, the illusion is so dominant that subjects continue to report hearing the illusory speech percept rather than the presented auditory speech sound. Sams and colleagues showed that infrequent deviations in the visual articulation of audio–visual syllables (acoustic /pa/ and visual /ka/) interspersed in a sequence of congruent trials (acoustic /pa/ and visual /pa/) gave rise to magnetic mismatch fields in the supratemporal region (see also Möttönen, Krause, Tiippana, & Sams, 2002 for similar results). This MMN-like response associated with the McGurk effect was also found in scalp-recorded ERPs with maximal amplitude between 200 and 300 ms after auditory onset (Colin et al., 2002; Colin, Radeau, Soquet, & Deltenre, 2004).

The present study aimed to further characterize the McGurk-MMN, using high-density electrical mapping and source-analysis to investigate the underlying cortical sources of this activity. In previous studies, the McGurk-MMN has been investigated by comparing the magnetic or electrical responses evoked by the standard and the deviant audio–visual speech stimuli. Problematically, this comparison will yield not only any potential McGurk-MMN effect, but also any differential responses due to the physically different visual stimuli used to make up the standards and deviants. In the present study we ran additional conditions where only the visual stimuli were presented so that we could (1) subtract out the evoked visual responses from the auditory–visual responses from which the McGurk-MMN was derived; (2) characterize the visual activity to the visual-alone stimuli to assess the extent to which they do contribute to activity in the latency range of the auditory response. Controlling for effects due to physical differences in the visual stimuli, we find clear evidence for a McGurk-MMN. That is, perceived phonemic changes in the absence of actual acoustic changes elicited the MMN (Colin et al., 2002, 2004). In addition, topographic mapping and dipole modeling revealed a dominance of left hemispheric cortical generators during the early and late phases of the McGurk-MMN, consistent with the well-known left hemispheric dominance for the processing of speech.

2. Methods

2.1. Participants

Eleven adult volunteers (ages: 19–33 years; mean: 25.6; five males) participated in the experiment and were naive with regard to the intent of the study. After the study, all subjects were debriefed to ensure that they experienced strong McGurk illusions.[1] The participants reported that they had no hearing or neurological deficits. They possessed normal or corrected-to-normal vision and were right-handed (except for one) as assessed by the Edinburgh handedness inventory. All subjects provided written informed consent in accordance with the Declaration of Helsinki, and the Institutional Review Board of the Nathan Kline Research Institute approved all procedures.

2.2. Stimuli and procedure

Stimuli were generated by digitally recording video (frame rate: 25 images/s; audio sample rate: 44.1 KHz in 16 bits) of the natural articulations of a male English speaker's mouth saying the syllables /ba/ and /va/. Although the visual information for both syllables is different in terms of place of articulation, the duration of mouth movement was similar for both stimuli with respect to the onset of the acoustical signal. There was a small difference of approximately 40 ms between the onsets of the two visual articulatory movements relative to the onset of the respective acoustic signals (i.e. −320 ms for /ba/ versus −360 ms for /va/). Productions began and ended in a neutral closed mouth position. The illusory McGurk audio-visual pair was created by synchronously dubbing the spoken syllable /ba/ onto the video of /va/. This particular combination elicits a particularly strong McGurk illusion in which the auditory perception is dominated by the visual information—that is, observers usually report hearing /va/, which corresponds to the visual portion of the bisensory stimulus (e.g. Jones & Callan, 2003; Rosenblum & Saldana, 1992; Summerfield & McGrath, 1984). The auditory /ba/ stimulus (duration: 370 ms; intensity: 60 dB SPL) was presented binaurally over headphones (Sennheiser-HD600). The visual stimuli were presented to the center of a computer CRT monitor located 100 cm from the subject (Iiyama VisionMaster Pro 502, 1024 × 768 pixels, 75 Hz). Stimuli subtended $10^\circ \times 7.5^\circ$ of visual angle and subjects were instructed to look at the speaking face stimulus and pay attention to the mouth articulation.

McGurk-MMN was generated using an oddball paradigm in two conditions: visual alone (/ba/ as 'standard' and /va/ as 'deviant') and audio–visual (congruent audio /ba/ and visual /ba/ as 'standard' and incongruent audio /ba/ and visual /va/ as 'deviant'). Hence, in the critical audio–visual condition, the phoneme /ba/ was presented auditorily on every trial but was perceived as /va/ when presented with an incongruent visual articulation. The inter-stimulus interval was 1630 ms and the probability of a deviant trial was 20%. Because the task involved simple fixation on the speaker's mouth, very short stimulation blocks (approximately 1½ min each) were administrated to minimize fatigue effects. Visual alone and audio–visual blocks (35–40 stimuli/block) were presented in random order and separated by short breaks for a total of approximately 1420 trials per condition. The rationale for using the visual alone condition was twofold. First, we aimed to rule out the possibility that the McGurk-MMN might be attributable to visual mismatch processes (Pazo-Alvarez, Cadaveira, & Amenedo, 2003). Second, this condition controls for any sensory response differences due to differential mouth movements in the standard and the deviant stimuli. For the main analysis, the standard and deviant visual responses were subtracted from the corresponding auditory–visual responses. As such, MMN activity related to the illusory phonemic change was examined by comparing the resulting 'auditory' standard and deviant responses. This subtraction procedure, of course, does not mean that the resulting ERP responses are purely auditory as they will also include some integrative processing (e.g. Molholm, Ritter,

[1] A 12th subject was excluded from the analysis because he reported not experiencing strong McGurk illusions during post-experiment debriefing.