Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans

Iiro P. Jääskeläinen,1,2,3CA Ville Ojanen,1 Jyrki Ahveninen,2 Toni Auranen,1 Sari Levänen,2,4 Riikka Möttönen,1 Iina Tarnanen and Mikko Sams1

1Laboratory of Computational Engineering, Helsinki University of Technology, Espoo, P.O. Box 9203, Innopoli 2, Teknikant. 14, FIN-02015 HUT Finland; 2Massachusetts General Hospital—Massachusetts Institute of Technology—Harvard Medical School Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA; 3Biomag Laboratory, Helsinki University Central Hospital, Helsinki, Finland; 4Low Temperature Laboratory, Helsinki University of Technology, Espoo, Finland

CA Corresponding Author and Address: iiro.jaaskelainen@hut.fi
Received 10 September 2004; accepted 14 September 2004

The technique of 306-channel magnetoencephalography (MEG) was used in eight healthy volunteers to test whether silent lip-reading modulates auditory-cortex processing of phonetic sounds. Auditory test stimuli (either Finnish vowel /æ/ or /o/) were preceded by a 500 ms lag by either another auditory stimulus (/æ/, /o/ or the second-formant midpoint between /æ/ and /o/), or silent movie of a person articulating /æ/ or /o/. Compared with N1 responses to auditory /æ/ and /o/ when presented without a preceding stimulus, the amplitudes of left-hemisphere N1 responses to the test stimuli were significantly suppressed both when preceded by auditory and visual stimuli, this effect being significantly stronger with preceding auditory stimuli. This suggests that seeing articulatory gestures of a speaker influences auditory speech perception by modulating the responsiveness of auditory-cortex neurons. NeuroReport 15:2741–2744 © 2004 Lippincott Williams & Wilkins.

Keywords: Audio-visual integration; Cortical adaptation; Human auditory cortex; Magnetoencephalography; N1m response; Phonetic; Vowel

INTRODUCTION

Seeing the articulations of a speaker significantly improves speech perception, especially in the presence of acoustic noise [1]. Further, incongruent visual information distorts the auditory percept, for instance, the so-called McGurk illusion refers to simultaneously presented visual /gæ/ and auditory /ba/ resulting in perception of /da/ [2]. This illusion further seems to reflect the complementary nature of auditory and visual inputs: visual speech effectively conveys the place of articulation and seems to bias auditory perception accordingly [3]. While these phenomena have been well described behaviorally, the underlying neural mechanisms have remained vague.

Using magnetoencephalography (MEG), Sams et al. [4] were the first to describe the influence of seen speech on auditory cortex processing of phonetic stimuli. Specifically, neuromagnetic mismatch negativity response (MMN), generated within the auditory cortex ~150 ms from stimulus onset, was elicited by infrequent McGurk stimuli (auditory /pa/ and visual /ka/) as compared with the responses elicited by audio-visually congruent ‘standard’ stimuli (auditory and visual /pa/) [4] (see also [5]). Subsequent fMRI studies have suggested that the posterior aspect of superior temporal sulcus (STS) is the locus of audio-visual speech integration [6]. Indeed, these fMRI findings agree well with the notion that there is a speech processing pathway traversing laterally and ventrally from the human primary auditory cortex [7].

Recent observations in rats suggested that heteromodal cortex, residing in between primary auditory and visual cortical areas and roughly corresponding to areas such as the STS in humans, contains neurons responding to both auditory and visual stimulation [8]. When near-threshold stimulation is used, the responses of these neurons are multiple times larger during audio-visual than during unimodal stimulation [8]. With higher stimulus intensities, that are typically used in human neuroimaging studies, the hemodynamic and electromagnetic responses seem to summate roughly linearly [6,9].

Both animal and human data have suggested that neurons respond to acoustic features of hierarchically increasing complexity as one progresses from the primary auditory cortex towards secondary auditory areas such as the STS [7,10], eventually responding to phonetic category [11]. Thus, it can be hypothesized that auditory and visual inputs converge upon neurons tuned on the acoustic features of phonetic stimulation within the auditory cortex or STS. Such a mechanism could well explain the McGurk illusion; visual information conveying the place of articulation would facilitate the responsiveness of neurons that are tuned on associated acoustic features, thus significantly biasing the auditory percept and resulting in the McGurk illusion. It has also been recently suggested that differential adaptation of distinct neural populations underlying the so-called N1 response give rise to the MMN [12] (for corroborating findings in animals, see [13,14]). Thus, the previous MMN results [4,5] could be explained by...

In the present study, it was hypothesized that seeing a visual articulation causes adaptation of auditory cortex MEG responses to a subsequently presented phonetic ‘test’ sound, reflecting initial facilitation and subsequent adaptation of feature-specific neurons within the auditory cortex and/or the posterior STS. Furthermore, it was hypothesized that this adaptation is significantly larger when the test stimulus is preceded by an auditory rather than a visual stimulus, reflecting additional adaptation of neurons tuned on simple acoustic features such as sound intensity. It was also hypothesized that adaptation caused by both preceding auditory and visual stimuli would be specific to phonetic category.

MATERIALS AND METHODS

Eight healthy volunteers (right-handed, students and research staff members, seven males, one female, ages 20–30 years) participated in the study after a voluntary consent was obtained, in accordance with the Helsinki declaration, and the study protocol was approved by the local ethics committee. Three separate stimulus conditions were presented to the subjects. In the first condition, auditory test stimulus (either the Finnish vowel /æ/ or /o/) was preceded by a 500 ms lag by another auditory stimulus (either /æ/, /o/, or the second-formant midpoint between the /æ/ and /o/). In the second condition the auditory /æ/ or /o/ test stimulus was preceded, again by a 500 ms lag, by 450 ms movie clips of a person articulating the /æ/ and /o/ (for snapshots of the articulations, see Fig. 1). In the third condition, the auditory /æ/ and /o/ test stimuli were presented without any preceding auditory or visual stimuli. The auditory stimuli were presented at 60 dB over the individually determined hearing threshold to the right ear of the subjects, and the inter-stimulus interval between two consecutive test stimuli was 3.5 s in each of the conditions. Monaural right-ear stimulation was used in order to elicit robust responses in the contralateral left hemisphere [12] that has been specifically implicated in the processing of simple acoustic features such as sound intensity. It was also hypothesized that adaptation caused by both preceding auditory and visual stimuli would be specific to phonetic category.

RESULTS

The amplitude of the left-hemisphere N1 response to test stimuli was significantly suppressed, both when the test stimuli were preceded by auditory (p < 0.001) and visual (p < 0.05) stimuli, as compared with the responses to the test stimuli when they were presented alone. Further, this effect was significantly greater when auditory stimuli preceded the test stimuli than when the preceding stimuli were visual (p < 0.01; Fig. 2, Fig. 3). Even though the peak latencies of the responses tended to be increased by preceding auditory and visual stimulation, these effects failed to reach statistical significance (Table 1). Likewise, trends towards category-specific adaptation of the responses (e.g. stronger adaptation when /æ/ preceded /æ/ than when /o/ preceded /æ/) failed to reach statistical significance.

DISCUSSION

The results of the present study tentatively support our hypothesis according to which seeing the articulatory lip movements of a speaker causes adaptation of a subset of feature-specific neurons within the human auditory cortex (possibly after initial facilitation of these neurons’ responsiveness). The latency of these adaptation effects (Table 1) corroborates well with previous studies that have described MEG responses associated with phonetic perception [11,16–18]. As expected, the response amplitude suppression caused by the preceding auditory stimuli was significantly larger than that caused by the preceding visual stimuli, possibly due to additional adaptation of auditory-cortex neurons to simple acoustic features such as stimulus intensity.

Tentatively, this modulating effect could explain the so-called McGurk illusion [2] wherein simultaneously presented visual and auditory phonemes result in perception of
a third phoneme. It has been recently proposed that phonological categorization emerges from specific spatial distribution of frequency-tuned neurons that give rise to the N1 response [16]. Thus, modulation by visual speech of a subset of such auditory-cortex neurons could easily bias auditory phonetic perception, giving rise to the McGurk illusion.

While we failed to observe statistically significant category-specific adaptation effects, it is possible that this was caused by the relative similarity of the /æ/ or /ə/ phonetic categories. Thus, the question of whether the adaptation effect is category specific and, further, whether the adaptation of responses due to seeing visual articulations is explained by the information on the place of articulation biasing categorical perception [3], remains to be determined in future studies (see also [19]).

While our results unequivocally demonstrate adaptation by temporally preceding visual articulations of neuromagnetic activity generated within the human secondary (i.e., the likely location of neurons receiving convergent audio-visual input [6]), or even primary, auditory cortex, non-invasive methods with higher spatiotemporal accuracy, such as combined fMRI and MEG [12,20], need to be utilized in future studies in order to more precisely localize these effects.

CONCLUSIONS
Seeing the visual articulations of a speaker adapts neuromagnetic N1 responses to subsequent auditorily presented phonetic sounds. This may reflect modulation by visual speech of the responsiveness of auditory cortex neurons, possibly underlying the McGurk illusion.

REFERENCES

Acknowledgements: Financially supported by the Academy of Finland and NIH-ROI HD040712.