

Publication III

Jääskeläinen, I. P., Kauramäki, J., Tujunen, J., and Sams, M. Formant transition specific adaptation by lipreading of left auditory cortex N1m. *NeuroReport*, 19(1): 93–97, January 2008.

© 2008 Lippincott Williams & Wilkins, Inc..
Reprinted with permission.

Formant transition-specific adaptation by lipreading of left auditory cortex N1m

Iiro P. Jääskeläinen^a, Jaakko Kauramäki^a, Juuso Tujunen^a and Mikko Sams^{a,b}

^aLaboratory of Computational Engineering and ^bBrain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, Espoo, Finland

Correspondence to Iiro P. Jääskeläinen, Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203, FIN-02015 TKK, Finland
Tel: + 358 50 560 9503; fax: + 358 9 451 4833; e-mail: iiro.jaaskelainen@tkk.fi

Work was carried out using the MEG device in the Brain Research Unit, analyses were carried out in the Laboratory of Computational Engineering

Received 28 September 2007; accepted 16 October 2007

To test for the feature specificity of adaptation of auditory-cortex magnetoencephalographic N1m responses to phonemes during lipreading, we presented eight healthy volunteers with a simplified sine-wave first-formant (F1) transition shared by /ba/, /ga/, and /da/, and a continuum of second-formant (F2) transitions contained in /ba/ (ascending), /da/ (level), and /ga/ (descending), during lipreading of /ba/ vs. /ga/ vs. a still-face baseline. N1m responses to the F1 transition were suppressed during lipreading, further, visual /ga/

(vs. /ba/) significantly suppressed left-hemisphere N1m responses to the F2 transition contained in /ga/. This suggests that visual speech activates and adapts auditory cortex neural populations tuned to formant transitions, the basic sound-sweep constituents of phonemes, potentially explaining enhanced speech perception during lipreading. *NeuroReport* 19:93–97 © 2008 Wolters Kluwer Health | Lippincott Williams & Wilkins.

Keywords: adaptation, auditory cortex, lipreading, magnetoencephalography, McGurk illusion, N1m response

Introduction

Perception is inherently multisensory. Seeing the speaker's articulatory gestures significantly enhances speech perception in noisy conditions [1]. Furthermore, mismatches between auditory and visual inputs may lead to audiovisual illusions [2], for instance, auditory /ba/ combined with lips forming /ga/ resulted in a percept of /da/ in 98% of experimental participants [2], and a mismatch between sound source location and visual location cues result in the ventriloquism effect.

Functional magnetic resonance imaging (fMRI) studies have shown that lipreading activates even the primary auditory cortex [3,4] (although see also Ref. [5]). Furthermore, audiovisual speech stimuli have been reported to cause both enhancement and suppression of auditory cortex and lateral temporal lobe hemodynamic responses as compared with the sum of responses to the respective unimodal stimuli [6,7]. Human intracranial-evoked responses recorded from posterior lateral superior temporal gyrus to auditory speech stimuli were also significantly more influenced by the addition of articulatory gestures than by meaningless gurning motion of the lips [8]. Recent fMRI studies have further shown that the site of audiovisual congruency effects depends on the type of stimuli [7].

In previous electroencephalography and magnetoencephalography (MEG) studies, amplitude suppression of the so-called N1 and N1m responses (recorded with electroencephalography and MEG, respectively) has been consistently reported when articulatory gestures are presented with auditory phonemes [9–11]. In macaque auditory cortex local field potential recordings, audiovisual species-specific

vocalizations caused both response enhancements and suppressions in auditory cortex neurons, with suppressive effects dominating when the delay from visual to auditory component was increased [12]. It is, however, presently not well known whether the modulatory effect of lipreading on auditory cortex responses is a general one, or whether the effects are specific to certain phonetic or acoustic features.

Phonemes are composed of formant transitions, which are sound sweeps occurring at specific frequency bands. Sound sweeps are the elementary acoustic features that auditory cortex (and subcortical auditory nuclei) neurons respond most robustly to, probably corresponding to the basic visual features processed by the visual cortex simple and complex cells. The phonemes /ba/, /ga/, and /da/ can be produced by simultaneously presenting sinusoidal sound glides that copy the first-formant and second-formant (F1 and F2) transitions. Although the F1 of these phonemes is identical, the F2 is a descending one in /ga/, level one in /da/, and ascending one in /ba/ (see Fig. 1). Here, we specifically hypothesized that lipreading will feature-specifically affect N1m response amplitudes to the F2 transitions (i.e. lipreading /ga/ vs. /ba/ differentially modulating the language dominant left-hemisphere N1m responses to descending vs. ascending F2 transitions).

Methods

Participants

Eleven healthy right-handed volunteers participated in the study, three of whom were discarded owing to technical reasons/poor data quality, thus yielding a total *N* of eight

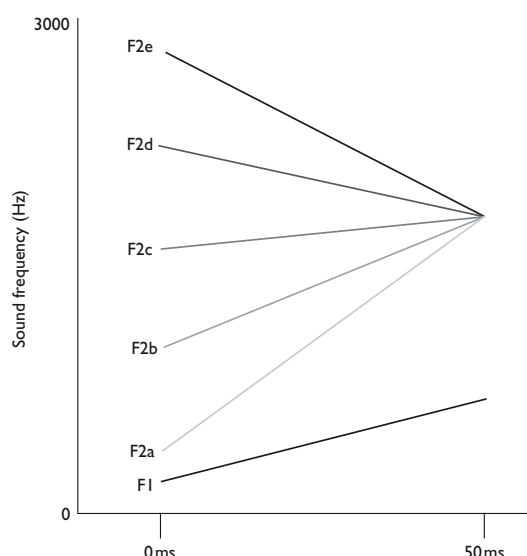


Fig. 1 The beginning and ending frequencies of the sine-wave sweep formant transition stimuli. The F1 is an ascending transition common to the /ba/, /ga/, and /da/ phonemes. F2a corresponds to /ba/ and F2e corresponds to /ga/, with the intermediate F2c corresponding to /da/.

(six men, two women, age 22–32 years, mean $\pm$ SD 26.4 $\pm$ 3.3). The participants were native Finnish speakers proficient in English. Each participant self-reported having normal hearing and vision, or corrected-to-normal vision. The participants signed a voluntary consent before participation, the experiments were run in accordance with the Helsinki Declaration, and an ethics approval was obtained from the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa, Finland.

Stimuli and task paradigm

Six different sine-wave sweep sounds, each lasting 50 ms, were used as auditory stimuli (see Fig. 1). The start and end frequencies of the sweeps were the following: (i) 200–700 Hz (F1), (ii) 400–1800 Hz (F2a), (iii) 1000–1800 Hz (F2b), (iv) 1600–1800 Hz (F2c), (v) 2200–1800 Hz (F2d), and (vi) 2800–1800 Hz (F2e). These were chosen to represent the first-formant transition common to /ba/, /ga/, and /da/ sounds, and a continuum of second-formant transitions ranging from the F2a in /ba/ to the F2e in /ga/. Notably, when the F1 sound is played simultaneously with each of the F2 sounds, a percept of phoneme ranging from /ba/ to /da/ to /ga/ is produced, however, in this study, the F1 and F2[a ... e] stimuli were played in isolation, thus not leading to phonetic perception, but allowing inspection of visual lipreading effects on the processing of the specific elementary features of the phonemes. Three different video clip stimuli were used, with a person articulating: (i) /ba/, (ii) /ga/, and (iii) a still picture of the face of the articulating person. Each video clip lasted for 1312 ms.

The sounds were presented binaurally to the participants via earplugs and the video was projected to a screen in front of the participants. Although the different sounds were

presented in random order (with a given sound never occurring twice in succession) with a random onset-to-onset ISI of 990–997 ms, the videos were presented in blocks with a given stimulus type repeated for 20–40 s (offset-to-onset ISI for video clips was 100–200 ms). The task of the participants was to pay close attention to the visual stimuli and press a button whenever the type of the visual stimulus changed (i.e. once every 20–40 s).

Magnetoencephalography recording

306-Channel whole head MEG (Vectorview, Elekta Neuro-mag, Finland) was recorded with a 601 Hz sampling rate (passband 0.01–172 Hz). Stimulus epochs time-locked to onset of the auditory stimuli were averaged, with epochs exceeding 3000 fT/cm (or $\pm 75 \mu$ V on electrooculogram) rejected as containing extracerebral artifacts. At least 100 artifact-free epochs were collected per condition.

Data analyses

The N1m response peak latencies and amplitudes were quantified from the channel pair with the highest response amplitude vector sum, separately for channels over the left and right auditory cortices (see e.g. [13]). Statistical analyses were conducted using analysis of variance with visual condition $\times$ formant transition type $\times$ hemisphere as the factors. Paired contrasts were calculated using the Scheffé test.

Results

Lipreading /ga/ significantly suppressed the amplitude of left-hemisphere N1m responses to the F1 transition when compared with the still-face condition ($P < 0.05$). Although the visual /ba/ failed to significantly suppress the left-hemisphere N1m responses to the F1 transition, there were no significant differences between lipreading effects caused by the visual /ga/ vs. /ba/ on response amplitudes to the F1 transition (see Fig. 2). We also failed to see any significant effects in the right-hemisphere response amplitudes or in

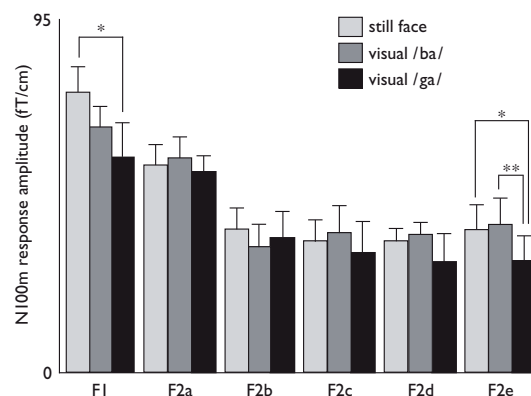


Fig. 2 The mean $\pm$ SEM left hemisphere N1m response amplitudes in the still-face baseline, visual /ba/, and visual /ga/ conditions. Although responses to sine-wave sweep stimuli copying the ascending first-formant transition (F1) that is common to the /ba/, /ga/, and /da/ phonemes were suppressed by /ga/, and tended to be suppressed during lipreading of /ba/, there was a specific suppression of responses to descending sine-wave sweep copying the second-formant (F2e) transition of /ga/ during lipreading of /ga/ (Scheffé test: * $P < 0.05$, ** $P < 0.01$).

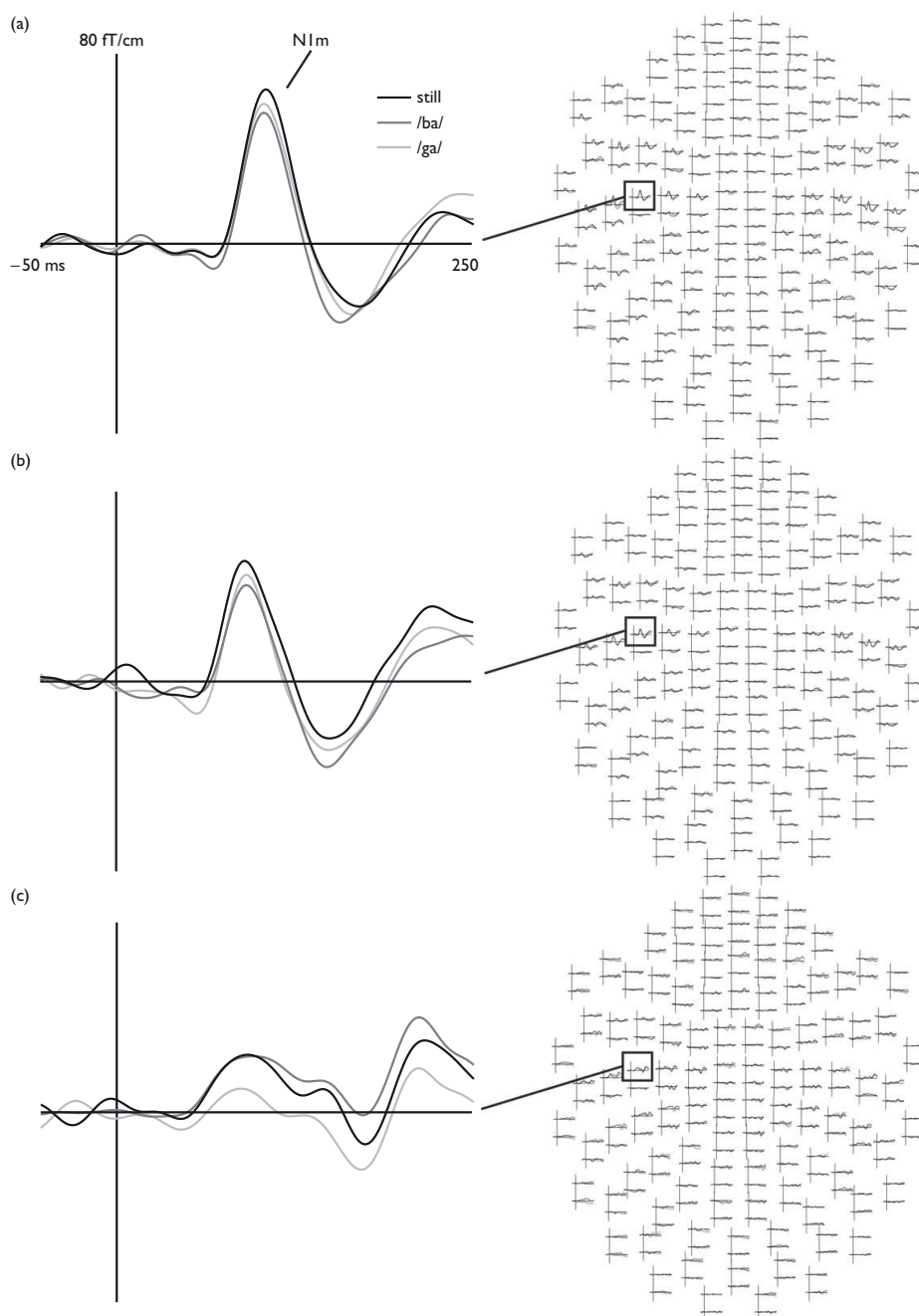


Fig. 3 Single participant-evoked magnetoencephalography (MEG) responses (shown are gradiometers on the right with a single channel over left auditory areas enlarged on the left). (a) Responses to F1 sounds during the still-face baseline, /ba/, and /ga/ lipreading conditions. Lipreading suppressed responses to F1 at the group level, with no significant differences between /ba/ and /ga/ lipreading conditions. (b) Responses to F2a sounds that are contained in /ba/ during the still-face baseline, /ba/, and /ga/ lipreading conditions. Lipreading failed to significantly affect responses to the F2a sounds at the group level. (c) Responses to F2e sounds the still-face baseline, /ba/, and /ga/ lipreading conditions. Lipreading /ga/ significantly suppressed NI00m responses to F2e sounds, as compared with both the still-face baseline and visual /ba/ conditions.

any of the response latencies. The averaged planar gradiometer response waveforms of a single participant are shown in Fig. 3.

During lipreading of /ga/, left-hemisphere N1m response amplitude to the descending F2e transition matching that contained in /ga/ (i.e. 2800–1800 Hz) was significantly suppressed, as compared with both the still-face ($P < 0.05$) and visual /ba/ ($P < 0.01$) conditions (see Fig. 2).

Discussion

In this study, lipreading of /ga/ specifically suppressed the amplitude of the left-hemisphere N1m responses to sine-wave sweeps copying the F2 transition contained in /ga/. This finding extends previous observations of suppressed N1 and N1m responses to phonetic stimuli during lipreading [9–11]. The present findings suggest that lipreading activates (and thus causes adaptation of) auditory cortex neural populations tuned to formant transitions, which are the basic sound-sweep constituents of phonemes.

Although lipreading /ga/ caused formant transition-specific adaptation, we failed to see corresponding effects during lipreading of /ba/. It is of course possible that this was due to insufficient signal-to-noise ratio in this study, however, it has been observed behaviorally that whereas auditory /ba/ combined with lips forming /ga/ resulted in a percept of /da/ (i.e. the McGurk effect) in 98% of experimental participants, the reverse combination of auditory /ga/ and lips forming /ba/ fails to produce the McGurk effect [2]. This suggests that /ga/ contains more salient visual articulatory cues that can be more accurately mapped to the relevant acoustic feature space by the brain.

Note that the lipreading effect in this study was not limited to the F2 frequency band, as responses to sine-wave sweeps at the F1 frequency were likewise suppressed. As the F1 transition, however, is common to /ga/ and /ba/, it is not surprising that lipreading /ba/ vs. /ga/ did not produce differential effects on the responses to the F1 transition stimulus.

How is the visual information conveyed to the auditory cortex? Currently there is not a definite answer to this question. Although there are direct anatomical connections to auditory cortex both from the visual cortex [14], and from heteromodal cortical areas [15], recent fMRI findings suggested that a feedback efference copy to the auditory cortex from prefrontal speech-production areas influences phonetic interpretation [16]. Specifically, visual /ka/ and auditory /pa/ that produced an illusory percept of /ta/ initially elicited auditory cortex activity patterns resembling those generated by congruent audiovisual /ka/ and /pa/, but at longer latencies the pattern began to resemble that elicited by audiovisual /ta/, which was also the dominant activation pattern in prefrontal areas [16]. These findings support studies that have suggested a key role for the 'mirror neuron' system (i.e. prefrontal neurons activated by both action and action observation, see Ref. [17]) in audiovisual speech perception [18–20]. Thus, whereas not directly tested here, it is possible that the present formant-transient-specific adaptation effects were mediated by feedback from the speech-production system. In contrast, nonspeech visual stimuli with learned auditory associations have also been observed to activate the posterior secondary auditory cortex [21]. Furthermore, it has been suggested that anterior auditory cortex areas are relevant in fine-grained

analysis of spectral information [22] and processing of speech-specific stimulus features [23,24], instead of the posterior auditory cortex areas that receive connections from the speech motor areas. Clearly, this important open question warrants investigation in future studies.

A recent fMRI study showed that activation of planum temporale during silent lipreading was significantly augmented when the participants selectively attended the articulatory gestures [25]. Given that the N1m may be (partially) generated within the planum temporale, it is important to also determine in future studies whether the suppression of N1m during lipreading is dependent on attentional factors. In this study, the task of pressing a button once every 20–40 s when the video changed between the still face, visual /ba/, and visual /ga/ conditions was a relatively easy and unengaging one for the participants, thus not allowing us to determine what role attention played in this study.

Conclusion

Our results indicate that lipreading of /ga/ specifically suppresses left-hemisphere N1m responses to the descending F2 transition contained in /ga/. This suggests that visual speech activates, and causes adaptation of, auditory cortex neural populations tuned to formant transitions, the basic sound-sweep constituents of phonemes. This might explain how seen articulatory gestures enhance speech perception in noise.

Acknowledgement

The study was supported by the Academy of Finland.

References

1. Sumbly WH, Pollack I. Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 1954; **26**:212–215.
2. McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature* 1976; **264**:746–748.
3. Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, et al. Activation of auditory cortex during silent lipreading. *Science* 1997; **276**:593–596.
4. Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, et al. Primary auditory cortex activation by visual speech: an fMRI study at 3T. *NeuroReport* 2005; **16**:125–128.
5. Bernstein LE, Auer ETJ, Moore JK, Ponton CW, Don M, Singh M. Visual speech perception without primary auditory cortex activation. *NeuroReport* 2002; **13**:311–315.
6. Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 2000; **10**:649–657.
7. Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ. The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 2007 (in press).
8. Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, et al. Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience* 2007; **145**:162–184.
9. Jaaskelainen IP, Ojanen V, Ahveninen J, Auranen T, Levanen S, Mottonen R, et al. Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *NeuroReport* 2004; **15**:2741–2744.
10. Besle J, Fort A, Delpeuch C, Giard MH. Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 2004; **20**:2225–2234.
11. Van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci USA* 2005; **102**:1181–1186.
12. Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 2005; **25**:5004–5012.
13. Ahveninen J, Kähkönen S, Tiitinen H, Pekkonen E, Huttunen J, Kaakkola S, et al. Suppression of transient 40-Hz auditory response by haloperidol

- suggests modulation of human selective attention by dopamine D2 receptors. *Neurosci Lett* 2000; **292**:29–32.
14. Budinger E, Heil P, Hess A, Scheich H. Multisensory processing via early cortical stages: connections of the primary auditory cortical field with other sensory systems. *Neuroscience* 2006; **143**:1065–1083.
 15. Schroeder CE, Foxe JJ. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cogn Brain Res* 2002; **14**:187–198.
 16. Skipper JL, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 2007; **17**:2387–2399.
 17. Rizzolatti G, Craighero L. The mirror-neuron system. *Annu Rev Neurosci* 2004; **27**:169–192.
 18. Nishitani N, Hari R. Viewing lip forms: cortical dynamics. *Neuron* 2002; **36**:1211–1220.
 19. Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikotis-Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport* 2003; **14**:2213–2218.
 20. Pulvermuller F, Huss M, Khefir F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y. Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci U S A* 2006; **103**:7865–7870.
 21. Bunzeck N, Wuestenberg T, Lutz K, Heinze HJ, Jancke L. Scanning silence: mental imagery of complex sounds. *NeuroImage* 2005; **26**:1119–1127.
 22. Jaaskelainen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levanen S, et al. Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci U S A* 2004; **101**:6809–6814.
 23. Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. *Cereb Cortex* 2005; **15**:1621–1631.
 24. Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, et al. Task-modulated 'what' and 'where' pathways in human auditory cortex. *Proc Natl Acad Sci U S A* 2006; **103**:14608–14613.
 25. Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Sams M. Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Hum Brain Mapp* 2006; **27**:471–477.