

Seeing speech: visual information from lip movements modifies activity in the human auditory cortex

Mikko Sams¹, Reijo Aulanko², Matti Hämäläinen¹, Riitta Hari¹, Olli V. Lounasmaa¹, Sing-Teh Lu¹ and Juha Simola¹

¹Low Temperature Laboratory, Helsinki University of Technology, Espoo (Finland) and ²Department of Phonetics, University of Helsinki, Helsinki (Finland)

(Received 14 January 1991; Revised version received 13 March 1991; Accepted 15 March 1991)

Key words: Audition; Auditory cortex; Audio-visual interaction; Intersensory convergence; Evoked responses; MEG; Speech perception; Neuro-magnetism; Man

Neuromagnetic responses were recorded over the left hemisphere to find out in which cortical area the heard and seen speech are integrated. Auditory stimuli were Finnish /pa/ syllables presented together with a videotaped face articulating either the concordant syllable /pa/ (84% of stimuli, V = A) or the discordant syllable /ka/ (16%, V ≠ A). In some subjects the probabilities were reversed. The subjects heard V ≠ A stimuli as /ta/ or /ka/. The magnetic responses to infrequent perceptions elicited a specific waveform which could be explained by activity in the supratemporal auditory cortex. The results show that visual information from articulatory movements has an entry into the auditory cortex.

Seeing the articulatory movements of a speaker's face provides complementary information for speech comprehension. The benefits from 'speech seeing' or lip reading become particularly advantageous in a noisy environment or when hearing is defective [3, 5, 17]. Visual information is helpful, e.g., in discriminating between bilabial /p/ and velar /k/ consonants, in which the acoustic cues are transient and their discrimination demands precise spectral and temporal resolution.

A convincing example of the integration of auditory and visual speech is the illusion called 'McGurk effect' [12, 13]: when an auditorily presented syllable /ba/ was associated with a videotaped face articulating /ga/, the auditory perception was altered by the visual input and the subject heard either /da/ or /ga/. This illusion is present in most adults but is less profound in children. Interestingly, the influence of visual information on auditory perception remains stable even when the subject knows the nature of this illusion. Up to now there has been no information about the actual neuroanatomical area where the visual information affects the auditory perception.

Magnetoencephalographic (MEG) recordings provide a powerful, completely noninvasive tool to investigate

cortical activity in human subjects. In this method, the weak magnetic signals associated with neural currents are recorded outside the head by means of SQUID (Superconducting Quantum Interference Device) magnetometers [9]. The field is measured at tens of locations and its cerebral source is often modelled with an equivalent current dipole (ECD), which is found by least-squares fitting to the experimental data. The parameters of the model are the location, orientation, and strength of the source.

We used neuromagnetic measurements to find out in which cortical area the heard and seen speech are integrated. Ten healthy adults (4 females and 6 males, 9 native speakers of Finnish and one of Swedish) were studied. The stimuli were edited from a video recording of a Finnish female articulating /pa/ and /ka/. The auditory stimulus was always the natural consonant-vowel syllable /pa/ with a duration of 215 ms and an intensity of about 70 dB sound pressure level (SPL). The auditory /pa/ syllable was dubbed to the visual /ka/ articulation, and combinations where the visual and auditory stimuli were in concordance (V = A, 84% of the stimuli) and where they were discordant (V ≠ A, 16% of the stimuli) were joined to a continuous film of a speaker articulating one or the other of the syllables 800 times. In 7 subjects, the probabilities of the audio-visual stimuli were also reversed (V ≠ A 84%, V = A 16%). The auditory-visual stimuli were presented approximately once per second.

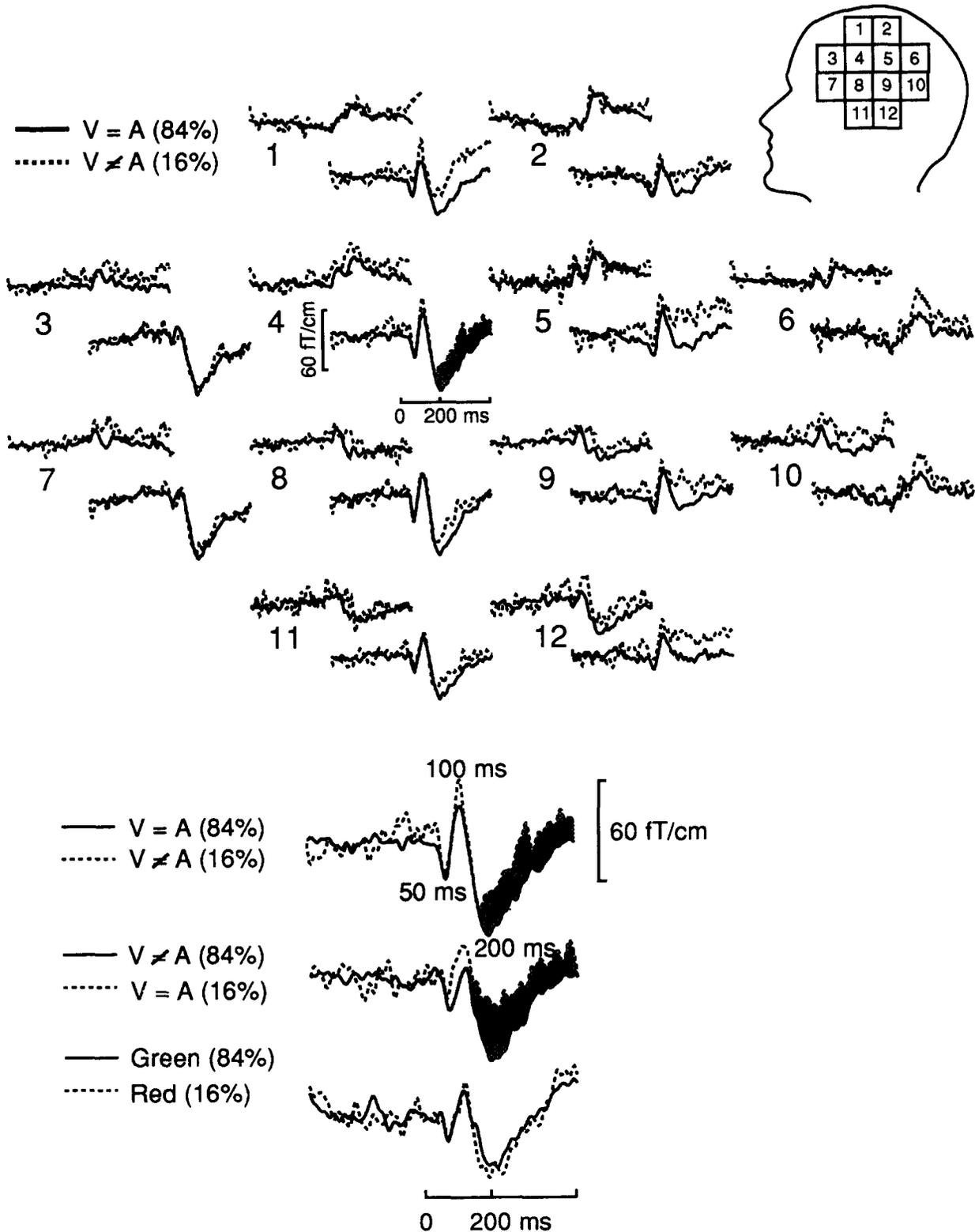


Fig. 1. Magnetic responses of one subject, measured with a 24-SQUID gradiometer over the left hemisphere, are shown in the top part of the figure. The upper traces of each pair show the field gradient in the vertical (y) and the lower one in the horizontal (x) direction. The exact locations and orientations of the gradiometers with respect to the head were determined by passing a current through three small coils, fixed on the scalp, and by analyzing the magnetic field thus produced. The number of averages is 500 for $V=A$ and 80 for $V \neq A$. The recording passband was 0.05–100 Hz; the responses have been digitally low-pass filtered at 40 Hz. The visually produced difference between the responses to the $V \neq A$ and $V=A$ stimuli was largest at locations 1, 4, and 5. The x -responses at location 4 during the 3 measurement conditions are shown enlarged in the bottom part of the figure. The three pairs of traces were recorded over the same area in consecutive measurements.

In a control condition, run with all subjects, the face was replaced by a green (84%) or red (16%) light stimulus (produced with an LED), which preceded the auditory syllable by 350 ms.

The subject was lying on a bed in a magnetically shielded room with his head firmly supported, and the auditory stimuli were led to his right ear while he was watching the video monitor through a 12-cm diameter hole in the wall. In the control condition, the LED was attached to the wall beside the hole. The task of the subject was to listen carefully to what the speaker was saying and to count silently the number of all auditory stimuli, and to report the count after the session.

Magnetic field maps were constructed on the basis of recordings over the left hemisphere with our newly commissioned 24-channel SQUID-gradiometer which samples two derivatives of the radial component of the magnetic field, $\partial B_r/\partial y$, and $\partial B_r/\partial x$ at 12 locations simultaneously. The instrument detects the largest signal just above a dipolar current source [11].

The subjects heard the $V \neq A$ stimuli as /ta/ or /ka/. The magnetic responses to the frequent $V=A$ stimuli typically consisted of three consecutive deflections, peaking at 50, 100, and 200 ms (Fig. 1). Similar deflections

are elicited by any kind of abrupt sounds and can be explained by equivalent current dipoles in the supratemporal auditory cortex [6, 8].

The magnetic responses to infrequent $V \neq A$ stimuli had similar 50-ms and 100-ms deflections as those to the $V=A$ stimuli. However, starting at about 180 ms, the two stimuli were different. A rather similar difference waveform (responses to the frequent stimuli subtracted from those to the infrequent ones) was elicited by infrequent $V=A$ stimuli among frequent $V \neq A$ stimuli. However, the signals to the syllables preceded by frequent green and infrequent red light stimuli were identical (Fig. 1).

The infrequent $V \neq A$ stimuli elicited a distinct difference waveform in 7 out of 10 subjects. Infrequent $V=A$ stimuli elicited such a waveform in 6 out of 7 subjects studied, including those three who did not show it to infrequent $V \neq A$ stimuli. Visual articulation presented alone, without the auditory input, elicited no response over the left temporal area in the two subjects studied.

The distributions of difference waveforms could be explained by ECDs at the supratemporal cortex, showing that visual information from the articulatory movements may have an entry into the human auditory cor-

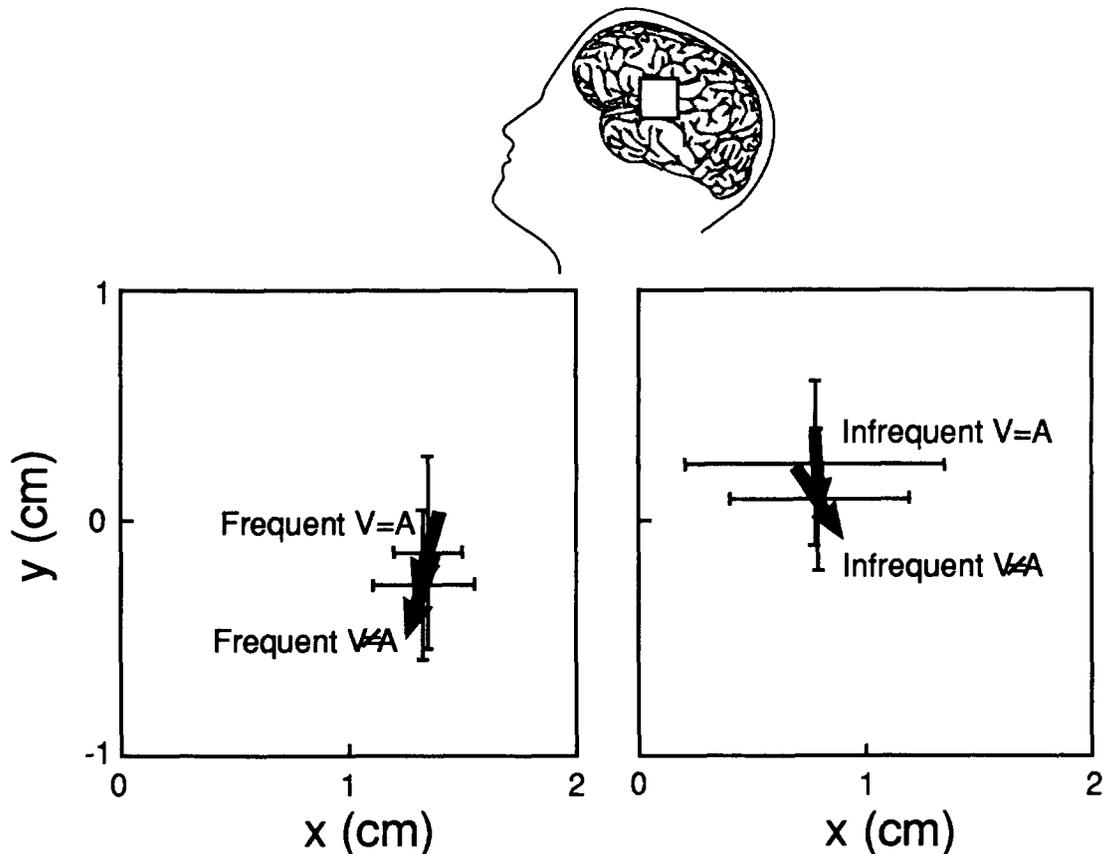


Fig. 2. The mean (\pm S.E.M.) xy -locations and orientations of ECDs for the 100-ms deflections (on the left, $n=6$) and for the difference waveforms (on the right, $n=6$ for $V=A$ and 5 for $V \neq A$). The approximate location of the plotting square is shown on the schematic head.

tex. This is consistent with the very vivid nature of the auditory illusion.

Fig. 2 on the left shows, on the xy -plane, the mean locations and orientations of the ECDs for the 100-ms deflections to the frequent $V=A$ and $V\neq A$ stimuli in those 6 subjects in whom it was calculated for both stimuli. The sources were practically identical, as were ECDs for the difference waveforms for the infrequent $V\neq A$ and $V=A$. All sources fit to the activation of the supratemporal auditory cortices. The ECDs explained $96\pm 2\%$ (mean \pm S.E.M.) and $95\pm 2\%$ of the field variance during the 100-ms deflection to the frequent $V=A$ and $V\neq A$ stimuli, respectively. The corresponding values for the difference waveforms elicited by the infrequent $V\neq A$ and $V=A$ stimuli were $91\pm 2\%$ and $88\pm 3\%$. The sources for the 100-ms responses to frequent stimuli were slightly more posterior and inferior than those for the difference waveforms. All ECDs were about 3 cm below the scalp, with no statistically significant differences in the depths.

It is important to note that the ECDs for both the difference waveforms and 100-ms responses are oriented downwards. In electric evoked potential recordings, such sources would produce a negative potential at the centro-parietal scalp. Therefore, the present difference waveform cannot be the magnetic counterpart of the well-known P300 deflection, elicited by various infrequent changes in a stimulus sequence.

Single- and multi-unit recordings in animals have shown convergence of visual and auditory inputs in many brain structures: superior colliculus, primary visual cortex, frontal cortex and superior temporal sulcus, temporo-parietal association areas, intraparietal sulcus, caudal inferior parietal lobule, and caudal superior temporal gyrus [1, 2, 10, 14, 15, 18]. However, we are not aware of any previous data showing an effect of visual stimuli on the reactivity of the auditory cortex.

The site of the equivalent current dipoles, calculated from our experimental data, probably reflects the resultant of activity in the primary auditory cortex and in the surrounding belt areas excited in synchrony. Therefore, the judgment of the exact cytoarchitectonic area in the auditory cortex where the visually produced activity change occurs is unwarranted at present.

According to classical accounts, the visual language, after its preliminary analysis in the occipital cortex, reaches the angular gyrus where it is reorganized into auditory form [7]. It has also been proposed, on the basis of brain damages, that the ability to lip read is a function of the left occipito-temporal cortex [4]. However, we did not see coherent activity in these two areas.

In face-to-face communication the speech can be 'seen' before it is heard; visual cues, e.g., from lip movements, exist in some cases hundreds of milliseconds be-

fore the corresponding auditory stimulus. Visual /ka/ information might prime such auditory neurons which are tuned to any non-labial consonant followed by an open vowel. Due to priming, the auditory /pa/ might activate the /ta/ and /ka/ 'detectors' more vigorously than the /pa/ detectors, giving rise to a biased perception. The present difference waveform might be analogous to a 'mismatch response' elicited by deviant auditory stimulus [16], but now associated with a deviant auditory perception of identical sounds. Our control condition with light stimuli suggests that the observed difference waveform cannot be explained by different degrees of attention allocated to the frequent and infrequent stimuli. Therefore, we interpret our results of neuromagnetic responses to acoustically identical but perceptually different auditory stimuli to indicate that it is possible to affect the processing of speech sounds in the human auditory cortex by visual input.

This study was supported by the Academy of Finland and by the Körber Foundation (Hamburg). We thank Mr. Olli Aaltonen and Dr. Lea Leinonen for comments on the manuscript, the personnel of the video studio at the Helsinki University of Technology for preparing the stimulus material, and our film star Pia Kempainen.

- 1 Benevento, L.A., Fallon, J., Davis, B.J. and Rezak, M., Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and orbital cortex of the macaque monkey, *Exp. Neurol.*, 57 (1977) 849-872.
- 2 Bental, E., Dafny, N. and Feldman, S., Convergence of auditory and visual stimuli on single cells in the primary visual cortex of unanesthetized unrestrained cats, *Exp. Neurol.*, 20 (1968) 341-351.
- 3 Binnie, C.A., Montgomery, A.A. and Jackson, P.L., Auditory and visual contributions to the perception of consonants, *J. Speech Hear. Res.*, 17 (1974) 608-618.
- 4 Cambell, R., The cerebral lateralization of speech. In B. Dodd and R. Cambell (Eds.), *Hearing by Eye: The Psychology of Lip-reading*, Lawrence Erlbaum, London, 1987, pp. 215-225.
- 5 Dodd, B., The role of vision in the perception of speech, *Perception*, 6 (1977) 31-40.
- 6 Elberling, C., Bak, B., Kofoed, B., Lebech, J. and Saermark, K., Magnetic auditory responses from the human brain, *Scand. Audiol.*, 9 (1980) 185-190.
- 7 Geschwind, N., Disconnection syndromes in animals and man. Part I-II, *Brain*, 88 (1965) 237-294.
- 8 Hari, R., Aittoniemi, K., Järvinen, M.-L., Katila, T. and Varpula, T., Auditory evoked transient and sustained magnetic fields of the human brain. Localization of neural generators, *Exp. Brain Res.*, 40 (1980) 237-240.
- 9 Hari, R. and Lounasmaa, O.V., Recording and interpretation of cerebral magnetic fields, *Science*, 244 (1989) 432-436.
- 10 Hikosaka, K., Iwai, E., Saito, H.-A. and Tanaka, K., Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey, *J. Neurophysiol.*, 60 (1988) 1615-1637.
- 11 Kajola, M., Ahlfors, S., Ehnholm, G.J., Hällström, J., Hämäläinen, M.S., Ilmoniemi, R.J., Kiviranta, M., Knuutila, J., Lounas-

- maa, O.V., Tesche, C.D. and Vilkmán, V., A 24-channel magnetometer for brain research. In S. J. Williamson, M. Hoke, G. Stroink, M. Kotani (Eds.), *Advances in Biomagnetism*, Plenum, New York, 1989, pp. 673–676.
- 12 MacDonald, J. and McGurk, H., Visual influences on speech perception processes, *Percept. Psychophysiol.*, 24 (1978) 253–257.
 - 13 McGurk, H. and MacDonald, J., Hearing lips and seeing voices, *Nature*, 264 (1976) 746–748.
 - 14 Meredith, M.A. and Stein, B.E., Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration, *J. Neurophysiol.*, 56 (1986) 640–662.
 - 15 Pandya, D.P. and Yeterian, E.H., Architecture and connections of cortical association areas. In A. Peters, E.G. Jones (Eds.), *Cerebral Cortex*, Vol. 4, Association and Auditory Cortices, Plenum, New York, 1985, pp. 3–61.
 - 16 Sams, M., Kaukoranta, E., Hämäläinen, M. and Näätänen, R., Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses, *Psychophysiology*, in press.
 - 17 Summerfield, Q., Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd and R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-reading*, Lawrence Erlbaum, London, 1987, pp. 3–51.
 - 18 Thompson, R.F., Johnson, R.H. and Hoopes, J.J., Organization of auditory, somatic sensory, and visual projection to association fields of cerebral cortex in the cat, *J. Neurophysiol.*, 26 (1926) 343–364