INTEGRATION OF AUDITORY AND VISUAL COMPONENTS OF ARTICULATORY INFORMATION IN THE HUMAN BRAIN

Reijo Aulanko and Mikko Sams*

Department of Phonetics, University of Helsinki, Finland
***Low Temperature Laboratory, Helsinki University of Technology, Finland**

ABSTRACT

In normal face-to-face conversation, both auditory and visual cues are used in speech perception. When the cues are contradictory, a perceptual "fusion" may arise, as in the "McGurk effect". Using magnetoencephalography (MEG), we measured the neural responses elicited by concordant and discordant audio-visual articulatory cues in the human brain. The auditory syllable [pa] was repeatedly presented to 10 subjects, together with a videotaped face articulating either [pa] or [ka]. The same auditory stimulus, presented with different visual face stimuli, elicited different magnetic responses in the auditory cortex. This indicates that visual articulatory information has an effect on the processing of auditory phonetic information in the auditory cortex.

1. INTRODUCTION

Speech perception is audio-visual in normal face-to-face conversation. Seeing the articulatory movements of a speaker's face provides complementary information for speech comprehension. The visual cues are especially needed in a noisy environment and by listeners with hearing defects [1, 3, 15].

Visual information is obviously helpful, e.g., in discriminating between labial and non-labial consonant articulations or between rounded and unrounded vowels, but other distinctions are also reflected in the muscular movements of the face [7]. Even the difference between falling and rising intonation can perhaps be conveyed by visual cues alone [4].

Visual articulatory information affects the perception of an auditory speech stimulus although people with normal hearing are not usually aware of this. A convincing example of the importance of visual cues is the illusion sometimes called "McGurk effect". It refers to the phenomenon where a subject is presented with conflicting articulatory information through the auditory and visual modalities causing him/her to perceive speech sounds which are combinations or fusions of the visual and auditory cues [8—11]. The most frequently cited classical example of this audio-visual illusion is the case of an auditory syllable [ba] presented with a videotaped face articulating [ga] eliciting an auditory perception of [da] [8, 10]. This illusion usually remains stable even after the subject is told about its nature.

There is no exact information about the actual neural basis of audio-visual speech perception. It has been stated that, after its preliminary analysis in the occipital cortex, the visual language reaches the angular gyrus where it is reorganized into auditory form [5]. 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 [2].

In this experiment [13] we made neuromagnetic measurements to locate the neuroanatomical area in which the integration of auditory and visual components takes place. As a first step towards this goal, we wanted to see if visual articulatory stimuli have an effect on the processing of an auditory phonetic stimulus in the human auditory cortex.

2. EXPERIMENT

2.1. Subjects

Ten healthy adults (4 females, 6 males; 9 native speakers of Finnish, one of Swedish) were studied individually.

2.2. Stimuli

The stimuli were edited from a video recording of a Finnish female speaker articulating the CV syllables [pa] and [ka]. 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 with an inter-stimulus interval of about one second. In seven subjects, the probabilities of the audio-visual stimuli were also reversed (V≠A 84%, V=A 16%). The auditory stimulus always remained the same syllable [pa] with a duration of 215 ms and an intensity of about 70 dB SPL. In a control condition, the face was replaced by a short green (84%) or red (16%) light (LED) stimulus, which preceded the auditory syllable by 350 ms.

2.3. Magnetoencephalography

The neuromagnetic responses elicited by the stimulation were measured using magnetoencephalographic (MEG) recordings. MEG provides 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 [6]. The field is measured at several locations and its cerebral source is often modelled with an equivalent current dipole (ECD). The parameters of the model are the location, orientation, and strength of the source.

2.4. Procedure

During the experiment, 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. Thus, the subject was not asked to react differently to the two stimuli. The only difference in reactions was supposed to be the different "silent identification". We could not ask the actual perceptual identity of each of the 800 stimuli from the subject during the experiment, but before the experiment we checked that the subject really heard the identical acoustic stimulus as two different syllables.

Magnetic field maps were constructed on the basis of recording over the left hemisphere with a 24-channel SQUID-gradiometer which samples two derivatives of the radial component of the magnetic field at 12 locations simultaneously. The instrument detects the largest signal just above a dipolar current source. 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 experiment consisted of presenting a frequent "standard" stimulus and an infrequent "deviant" stimulus in a pseudo-random order. In such conditions, an automatic neural difference detection process has been observed, the so-called mismatch response, which indicates that the nervous system has detected a change or difference in the repeated stimulation [12, 14].

3. RESULTS

The subjects perceived a strong audio-visual illusion: they heard the V≠A stimuli either as [ta] or [ka] or something in between.

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.

The magnetic responses to infrequent V≠A stimuli had 50-ms and 100-ms deflections similar to those elicited by the V=A stimuli. However, starting at approximately 180 ms, the two responses 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≠A stimuli. However, the signals to the auditory syllables preceded by frequent green and infrequent red light stimuli ~~were identical~~ (Fig. 1).

The infrequent V≠A stimuli elicited a distinct difference waveform in 7 out of the 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≠A stimuli. Visual articulation presented alone, without the auditory input, elicited no response over the left temporal area in the two subjects studied.

4. DISCUSSION

The results of this experiment indicate that visual articulatory information has an effect on the processing of the auditory phonetic information in the human brain. Identical auditory syllables, presented with two different visual face stimuli, were heard as two different syllables. The neuromagnetic responses to acoustically identical but perceptually different auditory stimuli suggest that the processing of speech sounds in the human auditory cortex can be affected by visual input. The neural activity originating from the auditory cortex was not correlated with acoustical energy but with auditory, especially phonetic, perception.

The response distributions in this experiment could be explained by ECDs at the supratemporal auditory cortex, showing that visual information from the articulatory movements may have an entry into the human auditory cortex. This is consistent with the very vivid nature of the auditory illusion. We did not see coherent activity in the two areas suggested by Geschwind [5] and Campbell [2], i.e. angular gyrus and occipito-temporal cortex.

In face-to-face communication speech can be "seen" before it is heard; visual cues from lip movements may exist in some cases hundreds of milliseconds before 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 biased perception. Our control condition with light stimuli shows that the found difference waveform clearly cannot be explained by different degrees of attention allocated to the frequent and infrequent stimuli.

5. REFERENCES

[1] BINNIE, C.A., MONTGOMERY, A. A. & JACKSON, P.L. (1974), "Auditory and visual contributions to the perception of consonants", *Journal of Speech and Hearing Research*, 17, 619-630.

[2] CAMPBELL, R. (1987), "The cerebral lateralization of lip-reading". In *Hearing by eye: The psychology of lip-reading* (B. Dodd & R. Campbell, eds), 215-226. London: Lawrence Erlbaum.

[3] DODD, B. (1977), "The role of vision in the perception of speech", *Perception*, 6, 31-40.

[4] FISHER, C.G. (1969), "The visibility of terminal pitch contour", *Journal of Speech and Hearing Research*, 12, 379-382.

[5] GESCHWIND, N. (1965), "Disconnexion syndromes in animals and man", *Brain*, 88, 237-294 & 585-644.

[6] HARI, R. & LOUNASMAA, O.V. (1989), "Recording and interpretation of cerebral magnetic fields", *Science*, 244, 432-436.

[7] JACKSON, P.L. (1988), "The theoretical minimal unit for visual speech perception: Visemes and coarticulation", *The Volta Review*, 90, 99-115.

[8] MACDONALD, J. & MCGURK, H. (1978), "Visual influences on speech perception processes", *Perception & Psychophysics*, 24, 253-257.

[9] MASSARO, D.W. & COHEN, M.M. (1983), "Evaluation and integration of visual and auditory information in speech perception", *Journal of Experimental Psychology: Human Perception and Performance*, 9, 753-771.

[10] MCGURK, H. & MACDONALD, J. (1976), "Hearing lips and seeing voices", *Nature*, 264, 746-748.

[11] MILLS, A.E. & THIEM, R. (1980), "Auditory-visual fusions and illusions in speech perception", *Linguistische Berichte*, 68/80, 85-108.

[12] NÄÄTÄNEN, R., GAILLARD, A. W.K. & MÄNTYSALO, S. (1978), "Early selective attention effect reinterpreted", *Acta Psychologica*, 42, 313-329.

[13] SAMS, M., AULANKO, R., HÄMÄLÄINEN, M., HARI, R., LOUNASMAA, O.V., LU, S.-T. & SIMOLA, J. (in press), "Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex", *Neuroscience Letters*.

[14] SAMS, M., HÄMÄLÄINEN, M., ANTERVO, A., KAUKORANTA, E., REINIKAINEN, K. & HARI, R. (1985), "Cerebral neuromagnetic responses evoked by short auditory stimuli", *Electroencephalography and Clinical Neurophysiology*, 61, 254-266.

[15] SUMMERFIELD, Q. (1987), "Some preliminaries to a comprehensive account of audio-visual speech perception". In *Hearing by eye: The psychology of lip-reading* (B. Dodd & R. Campbell, eds), 3-51. London: Lawrence Erlbaum.

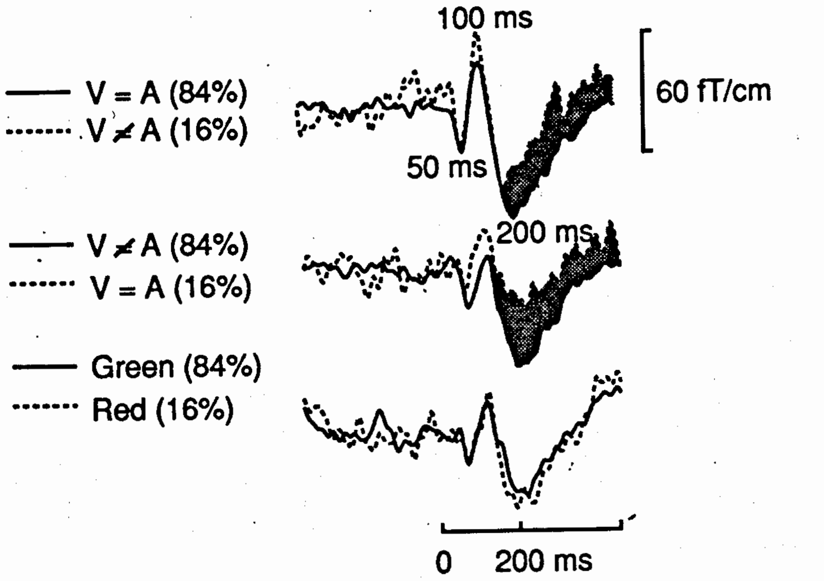

FIGURE 1. Magnetic responses of one subject, measured with a 24-SQUID gradiometer over the left hemisphere in three measurement conditions. Only one of the channels with the largest responses is shown. The three pairs of traces were recorded over the same area in consecutive measurements. The number of averages is 500 for the frequent stimuli (84%) and 80 for the infrequent stimuli (16%). The recording passband was 0.05–100 Hz, and the responses have been digitally low-pass filtered at 40 Hz. The visually produced difference between the responses to the identical auditory stimulus can be clearly seen in the two uppermost pairs of traces. The responses to the auditory syllables preceded by frequent green and infrequent red light stimuli were identical (lowermost pair of traces).