

Processing of vowels in supratemporal auditory cortex

David Poeppel^{a,*}, Colin Phillips^b, Elron Yellin^c, Howard A. Rowley^a,
Timothy P.L. Roberts^a, Alec Marantz^b

^a*Biomagnetic Imaging Laboratory, University of California San Francisco, 513 Parnassus Avenue, S-362, San Francisco, CA 94143-0628, USA*

^b*Department of Linguistics and Philosophy, MIT, Cambridge, MA 02139, USA*

^c*Mark of the Unicorn, Inc., Cambridge, MA 02138, USA*

Received 4 September 1996; revised version received 18 November 1996; accepted 8 December 1996

Abstract

The auditory evoked neuromagnetic fields elicited by synthesized vowels of two different fundamental frequencies F0 were recorded in six subjects over the left and right temporal cortices using a 37-channel biomagnetometer. Single equivalent current dipole modeling of the fields elicited by all vowel types localized activity to a well-circumscribed area in supratemporal auditory cortex in both hemispheres. There were hemisphere asymmetries in the amplitude and latency of the M100 response. We also observed changes in M100 latency related to vowel type, but not to F0. There was no clear effect of vowel type or F0 on dipole localization for the M100, but a possible vowel type by latency interaction. These M100 data provide further evidence that vowels are processed independently of their pitch. © 1997 Elsevier Science Ireland Ltd. All rights reserved

Keywords: Speech; Vowel; Auditory; M100; Functional neuroimaging; Magnetoencephalography; Magnetic source imaging; Language

In recent years several groups have investigated the neuromagnetic field evoked by vowel perception using magnetoencephalography (MEG) [3,5,7,8]. One objective has been to differentiate between speech and non-speech sounds. In particular, the source localization and temporal form of the evoked response may yield electrophysiological indices that characterize speech stimuli as phonemes independent of the typical acoustic variation that is ignored when speech is decoded, such as pitch or background noise [1,11].

The acoustic features that give vowels their unique perceptual identity (vowel quality) are the bands of energy peaks at particular frequencies, the formants. The distribution of formants in a vowel is largely a function of articulatory gestures (for example a lowered tongue height yields a higher frequency first formant). Because vowels are spoken at different characteristic pitches, ranging from very low bass to high-pitched children's voices, the speech perception system must be able to decode and label vowels across a range of fundamental frequencies (including atypical

occurrences, e.g. in whispering). At some level of speech perception, vowels and all other speech sounds must therefore be represented as pitch-independent phonemes. This identification is made on the basis of the formant values, which are relatively stable across speakers and articulatory contexts.

MEG data have been used to demonstrate phonetic invariance (the independence of evoked activity from phonetically irrelevant acoustic variation) of responses recorded from auditory cortex [1,11]. Furthermore, a recent experiment using pure tones suggests that, in addition to the probable spatial tonotopic organization of auditory cortex [4,10,14], there may also exist a temporal encoding mechanism for pitch [13].

To evaluate responses to vowels as a function of pitch (F0), phonetic category (formants), and hemisphere, we presented three synthetic vowels (/a/, /i/, and /u/) at two fundamental frequencies (100 and 200 Hz) to subjects while recording the evoked neuromagnetic field from the left or right temporal cortices. The latency, amplitude, and localization of the auditory evoked M100 neuromagnetic field were investigated as a function of pitch and of vowel type.

* Corresponding author. Tel.: +1 415 4766888; fax: +1 415 5024302; e-mail: poeppel@itsa.ucsf.edu

Three female and three male volunteers (two right handers and one left hander per group; mean age, 29.6 years) participated in the experiment. Subjects gave informed consent under a protocol approved by the UCSF and MIT committees on human studies. None of the subjects had any audiological abnormalities.

Vowels (300 ms) were created using a Klatt formant synthesizer (SenSynTM, Sensimetrics Inc., Cambridge, MA, USA). The source parameters were as follows: amplitude of voicing ramped up to 60 dB in 0–5 ms and remained constant until a rapid decline at 250 ms; amplitude of aspiration ramped up from 0 to 5 dB in the first 30 ms, remained constant till 290 ms, and sharply declined to 0 dB; the fundamental frequency varied from 120 to 90 Hz for the male vowels and from 220 to 190 Hz for the female vowels. Table 1 lists the formant values and formant bandwidths used for the six vowels. Stimuli were presented using PsyScope software [2]. Each vowel was presented 100 times in pseudorandom sequence to allow for signal averaging. The intertrial interval varied between 750 and 1250 ms.

The recordings were made in a magnetically shielded room using a 37-channel biomagnetometer (Magnes, Biomagnetic Technologies, Inc., San Diego, CA, USA). The sensor-array was placed over the left or right temporal lobe and adjusted to best capture the M100 field extrema for a 500 Hz pure tone. Six hundred millisecond epochs (100 ms pre-stimulus to 500 ms post-stimulus onset) were acquired around each stimulus. Neuromagnetic data were recorded from the contralateral hemisphere with a 1.0 Hz high-pass cutoff and a sampling rate of 2083.3 Hz. Following the simultaneous acquisition of the time-varying evoked neuromagnetic field in 37 channels for each epoch, data were sorted and averaged. The averaged waveforms were digitally filtered with a passband of 1.0 to 20 Hz; all further analyses were performed on the filtered data.

Localization analysis was restricted to the M100 evoked response component. The root mean square (RMS) of the field strength across the 37 channels was calculated for each sampled point in time. The M100 RMS peak and latency data were analyzed with standard analysis of variance methods using the UNIXSTAT software. To estimate the source parameters of the M100, a single

equivalent current dipole model was used. Representative single dipoles were chosen as follows: the latency of the M100 peak was determined and all dipoles in a 10 ms window around the peak were computed (i.e. peak latency \pm 5 ms). Of the dipoles in that interval, the one with the smallest model correlation confidence volume was chosen.

For each subject, high resolution volumetric magnetic resonance images (SPGR sequences, $128 \times 128 \times 124$ matrix; resolution, $\sim 1 \times 1 \times 1.5$ mm; TR, 36 ms, TE, 8 ms, flip, 70°) were acquired using a 1.5 T SIGNA magnetic resonance scanner (GE Medical Systems, Milwaukee, WI, USA). The left and right preauricular points and nasion were marked during the MEG recordings to allow coregistration of the MEG and magnetic resonance image (MRI) data sets. By superimposing these fiducial landmarks on the MRI of each subject, the position of the computed point sources can be visualized in anatomic context.

For three subjects (F1, M1, and M2), the M100 peak amplitudes measured from the left hemisphere were larger for each stimulus, in one subject (M3, left hander) the left response was larger for five of the six stimuli, and for one subject (F3, left hander) the response amplitudes from the right were always larger. Subject F2 showed no differential response effect of hemisphere. The main effect of hemisphere was not significant, $F_{(1,5)} = 0.72$, $P = 0.43$. There was no significant effect of fundamental frequency on amplitude of the M100 peak, $F_{(1,5)} = 1.76$, $P = 0.24$.

With the exception of the /u/ response, the left hemisphere response latencies exceeded the right hemisphere responses for all vowels, although the effect did not reach statistical significance, $F_{(1,5)} = 4.00$, $P = 0.10$. Statistically, the effect of phoneme type (/a/ versus /i/ versus /u/) on latency was highly significant, $F_{(2,10)} = 26.98$, $P = 0.001$. However, there were no effects of the fundamental frequency F0 on latency, $F_{(1,5)} = 0.87$, $P = 0.39$. For example, both types of /a/ had earlier M100 latencies than the other vowels in both hemispheres.

Fig. 1 plots, in a head-centered coordinate frame, the x (anterior-posterior), y (medio-lateral), and z (superior-inferior) coordinates of the single dipoles in the left and right hemisphere for one subject (M1). For all subjects, the localizations for vowels spatially overlapped with the localizations for independently presented tones and syllables [12]. In all three dimensions, the single dipoles cluster in the same area of auditory cortex. Fig. 2 shows one subject's (M1) coronal MRI with the dipoles for the (male and female) /a/ vowels coregistered (magnetic source images). Note that the vowel-elicited activity localizes to the supratemporal plane in both hemispheres.

The M100 dipole localizations were analyzed in the medio-lateral axial plane of each hemisphere (see Fig. 1) because (1) it was the localization plane with the most variance and (2) it is the plane of putative tonotopy [4,10,14]. Although there were no main effects of vowel type or fundamental frequency F0 in either hemisphere (vowel type, left hemisphere, $F_{(2,10)} = 0.51$, $P = 0.61$;

Table 1

Vowel formants: spectral composition (Hz)

Stimulus	F0	F1	Band-width	F2	Band-width	F3	Band-width
a	100	710	90	1100	100	2540	90
a (f)	200	850	90	1220	100	2810	90
i	100	280	90	2250	100	2890	90
i (f)	200	310	90	2790	100	3310	90
u	100	310	90	870	100	2250	90
u (f)	200	370	90	950	100	2670	90

F0, fundamental frequency; F1, 1st formant; F2, 2nd formant; F3, 3rd formant.

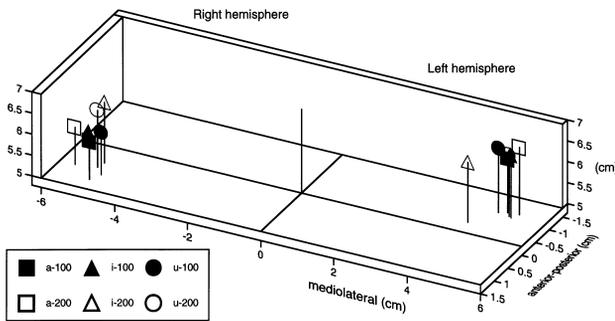


Fig. 1. Single dipoles for the three vowel types at both fundamental frequencies for one subject (M1). Each plotted point represents one M100 dipole (see text for choice of dipole). Filled symbols represent F₀ (100 Hz), open symbols represent F₀ (200 Hz). The localizations are to supratemporal auditory cortex. No effect of F₀ or vowel type is evident.

vowel type, right hemisphere, $F_{(2,10)} = 1.33$, $P = 0.31$; fundamental frequency, left hemisphere, $F_{(1,5)} = 0.015$, $P = 0.91$; fundamental frequency, right hemisphere, $F_{(1,5)} = 0.26$, $P = 0.64$, there was a notable vowel type \times fundamental interaction in the left temporal recordings ($F_{(2,10)} = 3.52$, $P = 0.074$) attributable to the male /u/ stimuli.

The data for one subject displayed in Fig. 1 illustrate that all the dipoles localized to a well circumscribed area. There is no hint of a vowel space, or any underlying ‘phonotopy’. It is important to bear in mind, also, that the dipoles for pure tone stimuli localized to the same coordinates [12], so the localizations are not distinguishing between speech and non-speech auditory stimulation.

With regard to the lateralization of responses for vowels, the findings to date suggest a moderate bias of processing towards the left temporal lobe [9,15,16]. Most analyses of the latencies and amplitudes of the major evoked response components have failed to detect any significant cortical asymmetries before the sustained field response (beginning approximately 200 ms after stimulus onset), at least with respect to vowels [5].

In MEG studies, Kuriki and Murase [8] localized M100 dipoles generated by vowels consistently more anterior to the M100 dipoles generated by tones in the left hemisphere. Eulitz et al. [5] recently reported a systematic difference between tone and vowel stimuli, but not as early as the M100. They recorded the evoked neuromagnetic field generated by pure tones and vowels from both hemispheres and showed that there was a clear (localizable) effect of the sustained field (SF) that distinguished tone and speech stimuli. Although these investigators found no hemispheric asymmetries between tones and vowels at the M100, they observed a larger amplitude in the SF to vowels as compared to the SF to tones. The SF difference was particularly strong in the left hemisphere. Moreover, the SF had a distribution that allowed them to localize its source to sites slightly anterior to the M100 source. They

attributed the larger left temporal SF response to speech-specific processing.

We observed in all subjects a hemispheric M100 latency difference that approached significance. Diesch et al. [3] also observed such a latency asymmetry for both their short and long vowels. If this effect is systematic, it is interesting in that it appears in the opposite direction to that which might be expected. A currently popular idea is that the left hemisphere is both specialized for speech and language and for fast temporal processing [6,17]. If that position is correct, it is relatively counterintuitive that the left auditory cortex uniformly generates longer latencies than the right for the major auditory evoked response component (M100) elicited by speech sounds.

The latency difference for vowel type that we find is quite similar to the data reported by Diesch et al. [3]. In particular, both studies found that /a/ produced significantly shorter M100 latencies than /u/ and /i/. We attribute this latency advantage of /a/ to the proximity of its first two formants to 1 kHz (see Table 1), which is close to the frequency that yields the fastest M100 latency for pure tones. However, the lack of coherent spatial grouping of vowels, and the vowel type \times fundamental interaction for localization in the medio-lateral plane question the existence of a spatial map for vowel phonemes. Although the interaction of vowel type and M100 latency was replicated, we find that varying F₀ is not associated with

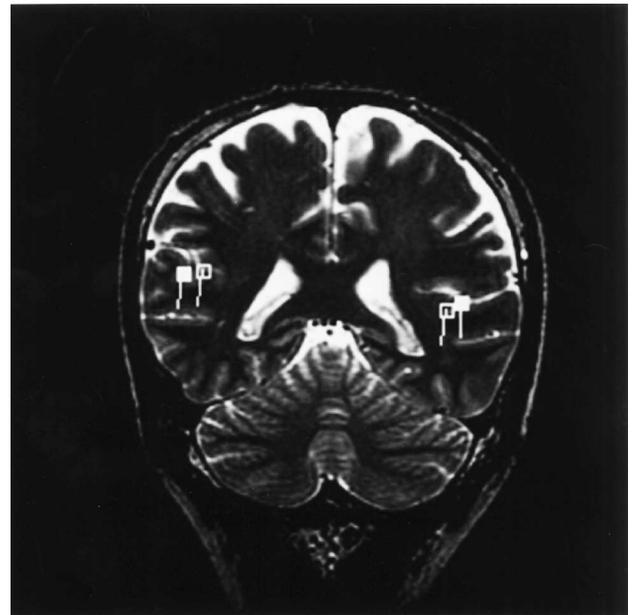


Fig. 2. Magnetic source image for subject M1. Overlay of the equivalent current source localizations (and orientation) derived from MEG recordings onto high resolution fast-spin echo T2-weighted MR image. Each localization corresponds to the dipole for the higher or lower-pitch /a/. The dipoles, projected onto one coronal slice, illustrate that the activity localizes to the supratemporal plane (superior bank of the superior temporal gyrus).

latency shifts despite the correlation between pure tone frequency and M100 latency [13].

For technical assistance we thank Susanne Honma, R.T. and Mary Mantle R.EEG.T. H.A.R. is partially supported by Biomagnetic Technologies, Inc. Financial support for this project came from the McDonnell-Pew Center for Cognitive Neuroscience at MIT, the NEC Corporation, NSF SGER grant #SBR-9423242, and the NSF-funded RTG program 'Language: Acquisition and Computation' (#DIR9113607).

- [1] Aulanko, R., Hari, R., Lounasmaa, O.V., Näätänen, R. and Sams, M., Phonetic invariance in the human auditory cortex, *NeuroReport*, 4 (1993) 1356–1358.
- [2] Cohen, J., MacWhinney, B., Flatt, M. and Provost, J., PsyScope: a new graphic interactive environment for designing psychology experiments, *Behav. Res. Methods*, 25 (1993) 257–271.
- [3] Diesch, E., Eulitz, C., Hampson, S. and Ross, B., The neurotopography of vowels as mirrored by evoked magnetic field measurements, *Brain Lang.*, 53 (1996) 143–168.
- [4] Elberling, C., Bak, C., Kofoed, B., Lebech, J. and Saermark, K., Auditory magnetic fields: source location and tonotopical organization in the right hemisphere of the human brain, *Scand. Audiol.*, 11 (1982) 61–65.
- [5] Eulitz, C., Diesch, E., Pantev, C., Hampson, S. and Elbert, T., Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli, *J. Neurosci.*, 15 (1995) 2748–2755.
- [6] Fiez, J., Tallal, P., Raichle, M.E., Miezin, F., Katz, W. and Petersen, S.E., PET studies of auditory and phonological processing: effects of stimulus characteristics and task demands, *J. Cognit. Neurosci.*, 7 (1995) 357–375.
- [7] Hayashi, M., Mashiko, T., Imada, T. and Odaka, K., Brain magnetic fields evoked by five Japanese vowels, *Inorg. Chim. Acta*, 14 (1992) 2–3.
- [8] Kuriki, S. and Murase, M., Neuromagnetic study of the auditory responses in right and left hemispheres of the human brain evoked by pure tones and speech sounds, *Exp. Brain Res.*, 77 (1989) 127–134.
- [9] Lund, E., Spliid, P., Andersen, E. and Bojsen-Moller, A., Vowel perception: a neuroradiological study of the perception of vowels in the human cortex, *Brain Lang.*, 25 (1986) 191–211.
- [10] Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G. and Wittkowski, W., Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields, *Electroenceph. clin. Neurophysiol.*, 69 (1988) 160–170.
- [11] Phillips, C., Marantz, A., McGinnis, M., Pesetsky, D., Wexler, K., Yellin, E., Poeppel, D., Roberts, T.P.L. and Rowley, H.A., Brain mechanisms of speech perception: a preliminary report. In C. Schütze, J. Ganger and K. Broihier (Eds.), *Papers on Language Processing and Acquisition*, Vol. 26, MITWPL, Cambridge, MA, 1995, pp. 153–191.
- [12] Poeppel, D., Yellin, E., Phillips, C., Roberts, T.P.L., Rowley, H.A., Wexler, K. and Marantz, A., Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds, *Cognit. Brain Res.*, (1997) in press.
- [13] Roberts, T.P.L. and Poeppel, D., Latency of auditory evoked M100 as a function of tone frequency, *NeuroReport*, 7 (1996) 1138–1140.
- [14] Romani, G., Williamson, S. and Kaufman, L., Tonotopic organization of the human auditory cortex, *Science*, 216 (1982) 1330–1340.
- [15] Rugg, M., The relationship between evoked potentials and lateral asymmetries of processing. In A. Gaillard and W. Ritter (Eds.), *Tutorials in ERP Research: Endogenous Components*, North-Holland, Amsterdam, 1983, pp. 369–431.
- [16] Studdert-Kennedy, M. and Shankweiler, D., Hemispheric specialization for speech perception, *J. Acoust. Soc. Am.*, 48 (1970) 579–594.
- [17] Tallal, P., Miller, S. and Fitch, R., Neurobiological basis of speech: a case for the preeminence of temporal processing, *Ann. N. Y. Acad. Sci.*, 682 (1993) 27–47.