

Short communication

MEG correlates of categorical perception of a voice onset time continuum in humans

Panagiotis G. Simos^{a,*}, Randy L. Diehl^b, Joshua I. Breier^a, Michelle R. Molis^b,
George Zouridakis^a, Andrew C. Papanicolaou^a

^a Department of Neurosurgery, The University of Texas—Houston Medical School, 6431 Fannin Suite 7.148 Houston, 77030, TX, USA

^b Department of Psychology, The University of Texas at Austin, Austin, TX, USA

Accepted 28 July 1998

Abstract

Event-related magnetic fields (ERFs) were recorded from the left hemisphere in nine normal volunteers in response to four consonant-vowel (CV) syllables varying in voice-onset time (VOT). CVs with VOT values of 0 and +20 ms were perceived as /ga/ and those with VOT values of +40 and +60 ms as /ka/. Results showed: (1) a displacement of the N1m peak equivalent current dipole toward more medial locations; and (2) an abrupt reduction in peak magnetic flux strength, as VOT values increased from +20 to +40 ms. No systematic differences were noted between the 0 and +20 ms stimuli or between the +40 and +60 ms CVs. The findings are in agreement with the results of multiunit invasive recordings in non-human primates regarding the spatial and temporal pattern of neuronal population responses in the human auditory cortex which could serve as neural cues for the perception of voicing contrasts. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Magnetoencephalography; Human; Speech; Voicing; VOT; Auditory cortex; N1m

The ability to perceive syllable-initial stop consonants as voiced (e.g., /g/) or voiceless (e.g., /k/) depends to a large extent on differences in voice onset time (VOT), which is acoustically defined as the interval between the noise burst produced at consonant release and the onset of waveform periodicity associated with vocal cord vibration, or voicing [9]. During this interval (in cases where voicing onset follows the release), the first formant is strongly attenuated, and the higher formants are excited aperiodically. Voicing distinctions are categorical in the sense that, given a number of stimuli that vary in VOT on a continuum, listeners tend to classify tokens with VOT values shorter than 20 ms as voiced and tokens with VOT values longer than 40 ms as voiceless. The exact location of the category boundary along the VOT dimension depends on various acoustic properties of the stimuli such as the onset frequency of the first formant [10]. This also appears to be true for young infants [1,2,8] and for members of certain

animal species [5]. These observations are consistent with the view that categorical perception of certain speech contrasts is based on neurophysiological mechanisms characteristic of a variety of auditory systems that are not unique to humans. In search of clues as to the nature of these mechanisms, researchers have examined population responses in the auditory cortex of at least two mammalian species using invasive recording techniques [12,25]. For instance, a potential neuronal cue that could be involved in the categorical-like encoding of VOT differences has been discovered using multiunit recordings from the primary auditory area in rhesus monkeys [25]. Consonants with long VOT values elicit two successive neuronal response peaks separated by a delay approximately equal to the VOT value of each stimulus. The second peak is absent in the response to stimuli with short VOT values. The first peak appears to be driven by the onset of the higher formants (F2/F3) and the second peak by the onset of voicing, which contains energy at considerably lower frequencies and usually coincides with the onset of the first formant (F1). Considering that the primary auditory cortex is tonotopically organized, one would expect that the two

* Corresponding author. Fax: +1-713-500-7787; E-mail: asimos@heart.med.uth.tmc.edu

peaks might show maximum amplitude within partially nonoverlapping populations. Specifically, cells with spectral preferences near the F2/F3 frequency range might be primarily responsive during the time window of the first peak, whereas cells with lower characteristic frequencies might be more responsive to the onset of voicing which, in turn, drives the second peak. Both populations are expected to contribute to the single peak elicited by stimuli with short VOT values. Reliable measures of population neuronal activity in the human auditory cortex can be obtained non-invasively using event-related potentials (ERPs) and event-related magnetic fields (ERFs) (e.g., Refs. [16,18,19,21]). Using carefully selected stimulus series, categorical-like changes in ERP parameters have been demonstrated in a variety of age groups [13–15]. ERFs, on the other hand, reflect transient changes in intracellular currents that develop in close synchrony within relatively small neuronal populations. The distribution of ERFs on the head surface lends itself readily to reliable mathematical estimates of the location of active neurons (e.g., Refs. [18,19]). ERF measures possess the spatial and temporal resolution required to demonstrate phenomena related to tonotopicity in the human auditory cortex (e.g., Ref. [17]). Although several studies have searched for MEG correlates of speech sound discrimination abilities [4,7,11], success in demonstrating categorical-like phenomena has been limited.

In this study, ERFs were recorded from the left hemisphere in 9 right-handed, neurologically normal volunteers, with no history of hearing deficit (3 males and 6 females; mean age: 30.3 years) using a whole-head neuromagnetometer (Biomagnetic Technologies Inc., Magnes

WH2500). Four CVs with VOT values ranging from 0 to +60 ms in 20-ms steps were selected from a voicing continuum, ranging perceptually from /ga/ to /ka/, to serve as stimuli. Syllables were prepared using the cascade branch of the Klatt88 software synthesizer. Variation of VOT was achieved by broadening the bandwidth of F1 and exciting F2 and F3 with a noise source during the period between stimulus onset and voicing source onset. The fundamental frequency (F0) was constant at 120 Hz for all syllables. The nominal formant frequencies at stimulus onset were 300 Hz, 1840 Hz, and 1960 Hz and changed linearly across a 55 ms transition period to 768 Hz, 1333 Hz, and 2522 Hz. The transitions were followed by a 200 ms steady-state segment resulting in a total stimulus length on 255 ms. In addition, the relative balance of low-frequency energy in the voicing source was enhanced by setting the spectral tilt parameter of the Klatt synthesizer to 10 along the entire length of the syllable. Finally, the voiced portions of the stimuli were low-pass filtered at 3200 Hz. Stimuli were presented in a random order at 80 dB SPL to the right ear, against a background noise level of less than 10 dB SPL. The subjects were instructed to pay attention to the speech sounds and determine how many instances of each CV (/ga/ or /ka/) were actually presented. They were asked to sign a consent form prior to participating in the study, after the nature of the procedures involved had been explained. Stimulation, recording, signal processing, and analysis parameters were identical to those described in Ref. [22]. Given that the N1m was the only ERF response that was consistently observed in all subjects and that provided equivalent current dipole (ECD) solutions meeting the stringent criteria necessary to ensure

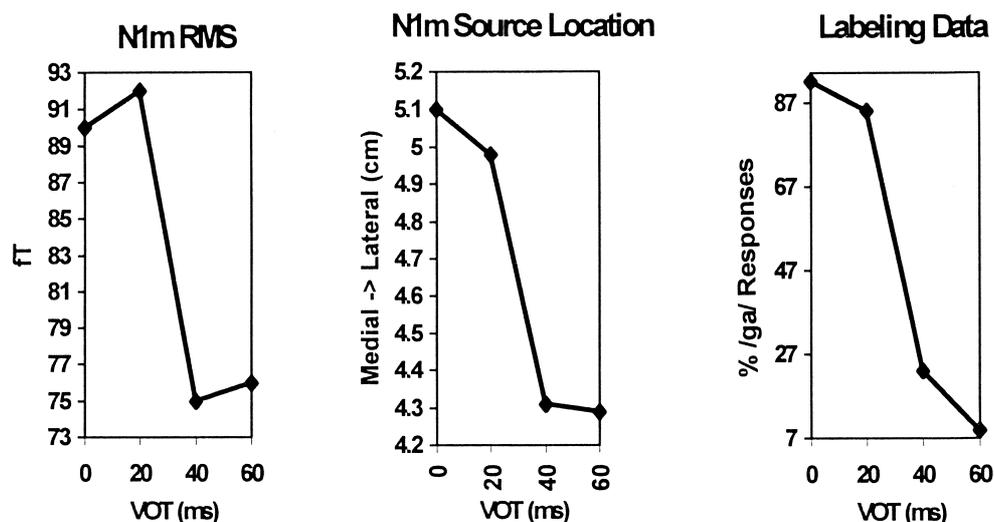


Fig. 1. (a) Mean peak-RMS values for the N1m response recorded over the left hemisphere ($n = 9$) for each of the four VOT stimuli; (b) Mean medial-lateral ECD coordinates computed near the peak of the N1m response for each of the four VOT stimuli. Notice the abrupt shift toward more medially located ECDs in the stimuli with long VOT values (+40 and +60 ms); (c) Group perceptual identification function suggesting a category boundary between +20 and +40 ms of VOT. The values on the vertical axis correspond to the proportion of times a given stimulus was identified as /ga/.

reliable localization estimates, we focused on this response for further analyses. A minimum of 120 MEG epochs were used to calculate each averaged waveform. Although a single ECD model was used for estimating ERF source parameters, it proved to be sufficient to account for at least 96% of the variance in the magnetic field distribution observed at the peak of the N1m response, in all cases. It should be noted, however, that the "point" dipole (a concept inherent in the ECD model) is a simplified representation of the summed intracranial current produced by a large number of synchronously active neurons. Therefore, the picture of auditory cortex activation derived with the use of such a measure can only serve as a gross index of neuronal activity that occurs within an intricately organized cortical field, such as the auditory cortex (e.g., Ref. [24]).

The percentage of /ga/ judgments was plotted against VOT (see Fig. 1c). Individual function curves indicate perceptual/phonetic boundaries (defined by the 50% '/ga/' response point) between +20 and +40 ms (i.e., as expected for velar syllable-initial consonants).

One-way repeated measures ANOVAs were used to examine the effect of VOT on five ECD parameters computed near the peak of the N1m response: root-mean square of magnetic flux strength (RMS), latency after stimulus onset, and three spatial coordinates (x , y , and z) that define the (estimated) ECD location with respect to the origin of the Cartesian coordinate system. Pairwise comparisons used for further examining significant ANOVA effects were performed using Bonferroni-adjusted t -tests and they always involved means from successive VOT conditions. The VOT main effect for peak N1m RMS was significant, $F_{3,24} = 8.79$, $p < 0.0001$ (see Fig. 1a). The effect was due to the fact that N1m RMS values were significantly smaller in response to the +40-ms stimulus than in response to the +20-ms token ($t_8 = 3.10$, $p = 0.007$, one-tailed). Differences between the 0- and +20-ms stimuli and between the +40- and +60-ms tokens were not significant ($p > 0.05$ one-tailed tests). The pattern of RMS change can be easily seen in individual waveforms (see Fig. 2). This finding is in agreement with our previous results using non-speech two-tone continua varying in tone-onset time (TOT) [22]. These results had revealed an abrupt drop in N1m RMS with TOT values exceeding +20 ms.

In addition, there was a significant main effect of VOT on one of the N1m ECD location parameters, the y coordinate, $F_{3,24} = 18.91$, $p < 0.0001$. Further tests indicated that the N1m ECD was found at significantly more medial locations in the +40-ms VOT condition than in the +20-ms VOT condition, ($t_8 = 7.31$, $p < 0.0001$, one-tailed). The mean distance between the two ECDs on the lateral-medial axis was 0.67 mm. No differences were found between the 0 and +20 and the +40 and +60-ms conditions (see Fig. 3). This finding suggests that the peak of the N1m response elicited by stimuli with short VOT

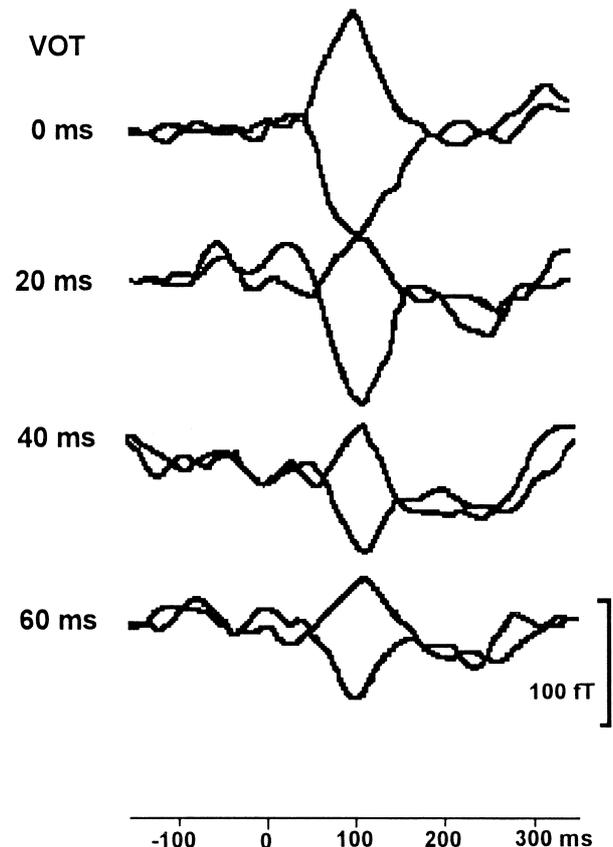


Fig. 2. Averaged ERFs from one subject (S1) recorded from the extrema of the distribution of the N1m response over the left hemisphere in response to the four CV tokens. The traces include the prestimulus period (150 ms) and 350 ms of post-stimulus recording.

values may reflect activity from, at least partially, distinct subpopulations of auditory cortex neurons compared to the N1m response elicited by stimuli with long VOT values.

Moreover, the change in ECD location as a function of VOT is in agreement with stimulus frequency-related variations in N1m ECD location demonstrated in earlier studies using pure tone stimuli (e.g., Ref. [17]). These studies have reported a regular progression of the N1m ECD toward more medial locations with increased stimulus frequency. In view of these findings we suggest that in the present study, the peak of the N1m response elicited by stimuli with long VOT values was primarily determined by the onset of F2 and F3, that preceded the onset of voicing by either 40 or 60 ms. The fact that neither N1m peak latency nor the duration of the N1m deflection varied significantly with VOT serves as additional evidence that the delayed onset of voicing did not contribute significantly to the peak of the N1m response elicited by these tokens.

The finding reported here is consistent with the results of multiunit invasive recordings from the auditory cortex of rhesus monkeys described above [25]. It appears that in the case of CVs with long VOT values that give rise to

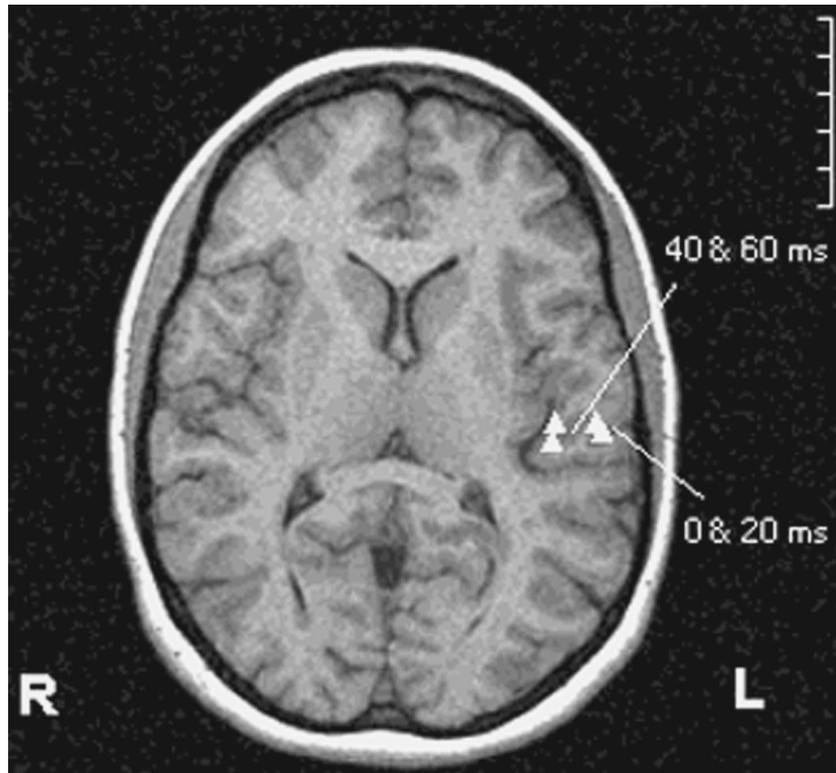


Fig. 3. Axial MRI slices from one subject (S1) at the level of the supratemporal plane showing the anatomical location of ECDs computed at the peak of the N1m response in the four VOT conditions. In order to demonstrate their relative position ECDs have been projected onto a single slice. Displayed ECDs actually span 3.5 mm in the inferior–superior dimension.

‘double-peaked’ responses, it is the first peak that primarily drives the activity that is later detected as the N1m response. On the other hand, CVs with short VOT values give rise to a single response peak, which is probably evident within a larger population of auditory cortex neurons with spectral preferences spanning a broader range of frequencies. The N1m peak in this case reflects activity associated with this single peak in the population response. The fact that all formants appear to contribute to the N1m in the 20-ms VOT condition might indicate that there is a brief window for temporal integration involved in the generation of this magnetic response.

The significance of the close correspondence between perceptual/phonetic sensitivity and N1m ECD location can only be speculated upon at present. From a practical standpoint, the ‘displacement’ of the N1m ECD could serve as a marker for the point along the VOT continuum at which a ‘double-peaked’ population response first becomes evident in the auditory cortex. Assuming that the latter event is indeed being utilized by the auditory system as a neural cue for voicing, then the behavior of the N1m ECD could eventually be used as a non-invasive index of a neurophysiological process that contributes to the perception of voicing contrasts. Such an index could potentially be useful for assessing the neurophysiological basis of deficits in phonological processes found in certain types of language impairment [3,23,26].

Considered in conjunction with previous data from our laboratory using non-speech, two-tone stimuli varying in tone-onset time [20,22], these findings support the view that the ability to perform phonetic distinctions based on VOT is, at least, partially determined by the capacity of the auditory system to discern temporal order between successive transient acoustic events [13,21]. In addition, our data further support the notion that phonetic perception (at least for VOT-based voicing contrasts) is determined by neurophysiological mechanisms common to the auditory system of several mammalian species including humans [5,6].

References

- [1] R.N. Aslin, D.B. Pisoni, B.L. Hennessy, A.L. Perey, Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience, *Child Devel.* 52 (1981) 1135–1145.
- [2] P.D. Eimas, E.R. Siqueland, P. Jusczyk, J. Vigorito, Speech perception in infants, *Science* 171 (1971) 303–306.
- [3] J.J. Godfrey, A.K. Syrdal-Lasky, K.K. Millay, M.K. Knox, Performance of dyslexic children on speech perception tests, *J. Exp. Child Psychol.* 32 (1981) 401–424.
- [4] E. Kaukoranta, R. Hari, O. Lounasmaa, Responses of the human auditory cortex to vowel onset after fricative consonants, *Exp. Brain Res.* 69 (1987) 19–23.
- [5] P.K. Kuhl, J.D. Miller, Speech perception by the chinchilla:

- voiced–voiceless distinction in alveolar plosive consonants, *Science* 190 (1978) 69–72.
- [6] P.K. Kuhl, D.M. Padden, Enhanced discriminability at the phonetic boundaries for the place feature in macaques, *J. Acoust. Soc. Am.* 73 (1983) 1003–1010.
- [7] S. Kuriki, Y. Okita, Y. Hirata, Source analysis of magnetic field responses from the human auditory cortex elicited by short speech sounds, *Exp. Brain Res.* 104 (1995) 144–152.
- [8] R.E. Lasky, A. Syrdal-Lasky, R.E. Klein, VOT discrimination by four and six and a half month old infants from Spanish environments, *J. Child Psychol.* 20 (1975) 215–225.
- [9] L. Lisker, A. Abramson, Some effects of context on voice onset time in English stops, *Lang. Speech* 10 (1967) 1–28.
- [10] L. Lisker, Is it VOT or a first-formant transition detector?, *J. Acoust. Soc. Am.* 57 (1975) 1547–1551.
- [11] J.P. Mäkelä, R. Hari, L. Leinonen, Magnetic responses of the human auditory cortex to noise/square wave transitions, *Electroencephal. Clin. Neurophysiol.* 69 (1988) 423–430.
- [12] T. McGee, N. Kraus, C. King, T. Nicol, Acoustic elements of speech-like stimuli are reflected in surface recorded responses over the guinea pig temporal lobe, *J. Acoust. Soc. Am.* 99 (1996) 3606–3614.
- [13] D.L. Molfese, Electrophysiological correlates of categorical speech perception in adults, *Brain Lang.* 5 (1978) 25–35.
- [14] D.L. Molfese, T. Hess, Hemispheric specialization for VOT perception in the preschool child, *J. Exp. Child Psychol.* 26 (1978) 71–84.
- [15] D.L. Molfese, V.J. Molfese, Right hemisphere responses from preschool children to temporal cues in speech and non-speech materials, *Brain Lang.* 33 (1988) 245–259.
- [16] N. Nakasato, S. Fujita, K. Seki, T. Kawamura, A. Matani, I. Tamura, S. Fujiwara, T. Yoshimoto, Functional localization of bilateral auditory cortices using an MRI-linked whole head magnetoencephalography (MEG) system, *Electroencephal. Clin. Neurophysiol.* 94 (1995) 183–190.
- [17] C. Pantev, Hoke, M. Lehnertz, B. Lütkenhöner, G. Anogianakis, W. Wittkowski, Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields, *Electroencephal. Clin. Neurophysiol.* 69 (1988) 160–170.
- [18] A.C. Papanicolaou, R.L. Rogers, S. Baumann, C. Saydjari, H.M. Eisenberg, Source localization of two evoked magnetic field components using two alternative procedures, *Exp. Brain Res.* 80 (1990) 44–48.
- [19] M. Reite, M. Adams, J. Simon, P. Teale, J. Sheeder, D. Richardson, R. Grabbe, Auditory N1m component: 1. Relationship to Heschl's gyri, *Cogn. Brain Res.* 2 (1994) 13–20.
- [20] P.G. Simos, D.L. Molfese, Electrophysiological responses from a temporal order continuum in the newborn infant, *Neuropsychologia* 35 (1996) 89–98.
- [21] P.G. Simos, D.L. Molfese, R.A. Brenden, Electrophysiological indices of voicing-cue discrimination in the newborn infant: laterality patterns and development, *Brain Lang.* 57 (1997) 122–150.
- [22] P.G. Simos, J.I. Breier, G. Zouridakis, A.C. Papanicolaou, Magnetic fields elicited by a tone onset time continuum in humans, *Cogn. Brain Res.* 6 (1998) 285–294.
- [23] M. Steffens L, R.E. Eilers, K. Gross-Glenn, B. Jallad, Speech perception in adult subjects with familial dyslexia, *J. Speech Hear. Res.* 35 (1992) 192–200.
- [24] C.E. Schreiner, Spatial distribution of responses to simple and complex sounds in the primary auditory cortex, *Audiol. Neurootol.* 3 (1998) 104–122.
- [25] M. Steinschneider, C.E. Schroeder, J.C. Arezzo, H.G. Vaughan Jr., Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns, *Brain Lang.* 48 (1995) 326–340.
- [26] P. Tallal, M. Piercy, Defects of non-verbal auditory perception in children with developmental dysphasia, *Nature* 241 (1973) 468–469.