Neuroscience Letters 348 (2003) 13–16

Neuroscience Letters

www.elsevier.com/locate/neulet

Magnetoencephalographic study of the cortical activity elicited by human voice

Atsuko Gunji[a,b,c], Sachiko Koyama[c,d], Ryouhei Ishii[a,e], Daniel Levy[f,g], Hidehiko Okamoto[a,c], Ryusuke Kakigi[c], Christo Pantev[a,*]

[a]The Rotman Research Institute for Neuroscience, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario, M6A 2E1, Canada
[b]The Japan Society for the Promotion of Science, Tokyo, Japan
[c]Department of Integrative Physiology, National Institute for Physiological Sciences, Okazaki, Japan
[d]Keck Center for Integrative Neuroscience, University of California, San Francisco, CA, USA
[e]Department of Post-Genomics and Diseases, Division of Psychiatry and Behavioral Proteomics, Osaka University, Osaka, Japan
[f]Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel
[g]Department of Psychiatry, University of California, and VA Healthcare System, San Diego, CA, USA

Received 22 January 2003; received in revised form 9 May 2003; accepted 15 May 2003

Abstract

In an attempt to identify voice-specific neural activities in auditory cortex in humans, we recorded cortical magnetic responses. Volunteers were instructed to listen to vocal and instrumental sounds matched in fundamental-frequency, duration, temporal envelope and average root mean square power. The stimuli were sounds produced by four singers and four musical instruments at each of two fundamental frequencies: 220 Hz (musical note A3) and 261.9 Hz (C3). Two components of the evoked responses were analyzed, one at approximately 100 ms (N1m) and the other 400 ms after the stimulus onset (sustained field, SF). The source locations of equivalent current dipoles for both components were estimated around the Heschl's gyrus in both hemispheres. Compared with the instrumental sound, the source strength of the SF component for the voice was significantly larger.

Keywords: Voice; Musical instruments; Auditory cortex; Magnetoencephalography; Auditory evoked magnetic field; Sustained field

Aside from phonological information, human voice sounds contain a wealth of non-linguistic information as well, including indicators of a speaker's identity and emotional state, which might be processed prior to, or in parallel with phonology [4,12]. Recent neuroimaging studies have shown that the cortical region around the superior temporal sulcus (STS) might be a 'human voice' area [1–3]. On the other hand, neurons in the auditory cortex of rhesus monkeys showed selectivity for species-specific communication calls [16]. Although the anatomical comparability of the auditory cortices between humans and monkeys still remains unresolved, voice-sensitive activity in humans might also be found in the auditory cortex other than the STS region.

Previous magnetoencephalographic (MEG) studies have shown that a stimulus locked DC shift, the sustained field (SF), lasts for the duration of auditory stimuli longer than 150–200 ms [5,10,15]. Eulitz et al. [5] reported that the SF is larger for synthesized vowel sounds than for pure tones of 1000 Hz. Furthermore, in contrast to the SF elicited by the pure tones, the SF to vowels was stronger in the left hemisphere. The authors interpreted this finding as a manifestation of left hemisphere dominance of the linguistic processing. Synthesized vowels possess formant structure, and are acoustically more complex than pure tones; this might account for the larger SF they elicited. Alternatively, the larger SF might be a function of the phonetic character of the vowels. Yet another possibility is that a stronger SF is evoked by the human voice, regardless of the phonetic information it carries.

In the present MEG study, we examined how human vocal sounds without phonetic information affect the SF. We used the same stimuli as Levy et al. [7]: vocal sounds sung by professional singers and the instrumental sounds

* Corresponding author. Fax: +1-416-785-2862.
E-mail address: pantev@rotman-baycrest.on.ca (C. Pantev).

0304-3940/03/$ - see front matter © 2003 Elsevier Science Ireland Ltd. All rights reserved.
doi:10.1016/S0304-3940(03)00640-2

sharing the characteristics of a harmonic structure and a dynamic course of the amplitude changes of their harmonic components. If the increased left hemispheric SF strength is due to phonetic information, then no clear hemispheric difference would be obtained in response to either vocal or instrumental stimuli. On the other hand, if the increased left hemispheric SF strength depends on stimulus complexity, then it would be observed for both types of stimuli.

Eleven healthy volunteers with normal hearing (five females; mean age 31 years, range 26–39 years) participated in this study. Ten subjects were right-handed and one left-handed. The subjects consented to their participation after they were completely informed about the nature of the study. The Ethics Commission of the Baycrest Centre for Geriatric Care approved all experimental procedures, which are in accordance with the Declaration of Helsinki. MEG data were recorded in a magnetically shielded room using a whole-head helmet-shaped 151-channel SQUID (Superconducting Quantum Interference Device) sensor array (Omega 151, CTF Systems Inc., Canada). The magnetic responses were digitized at 312.5 Hz and then filtered with a 60 Hz notch filter and a 100 Hz low-pass filter. Stimulus related epochs in which signal variations were larger than 3 pT were excluded from averaging (at least 165 artifact-free epochs for averaging). The DC offset was corrected for each channel according to the mean value of the 500 ms pre-stimulus signal.

Sixteen sounds were applied in the present study (Fig. 1) [7]: (1) human voice sounds (having the general character of a sung neutral vowel) produced by two singers of each gender (mezzo soprano, alto, bass and baritone); and (2) musical sounds produced by four string instruments (violin, viola, cello and bass), at each of two fundamental frequencies: A3 (220 Hz) and C4 (261.9 Hz). The stimuli were delivered binaurally to the subject's ears at 60 dB SL (sensation level) through a pair of plastic tubes and ear-pieces at a random inter-stimulus interval (ISI) between 1000 and 2000 ms. In order to prevent the perception of a pseudo-melody, the stimuli were presented separately in two blocks for the same fundamental frequency (block A3 and block C4). Sound stimuli with the same fundamental frequency were presented in random order for each block. The number of trials was 200 for each stimulus category (voice and instrument); 25 trials for each of the four voices and 25 trials for each of the four instruments in each block. In order to keep the vigilance level constant, subjects were instructed to watch a silent movie and to not pay attention to the auditory stimulation.

Source analysis based on a single moving equivalent current dipole (ECD) [13] was used. Source locations, their confidence volumes and dipole moments were estimated for each sampling point. The locations of the estimated sources were described in a head-based coordinate system. The origin of the system was set at the midpoint between preauricular points. The x-axis joins the nasion and the origin. The y-axis extends from left to right preauricular points through the origin such that it is perpendicular to the x-axis. The z-axis indicates the level of an axial plane with positive values directed upward such that it is perpendicular to the x- and y-axes.

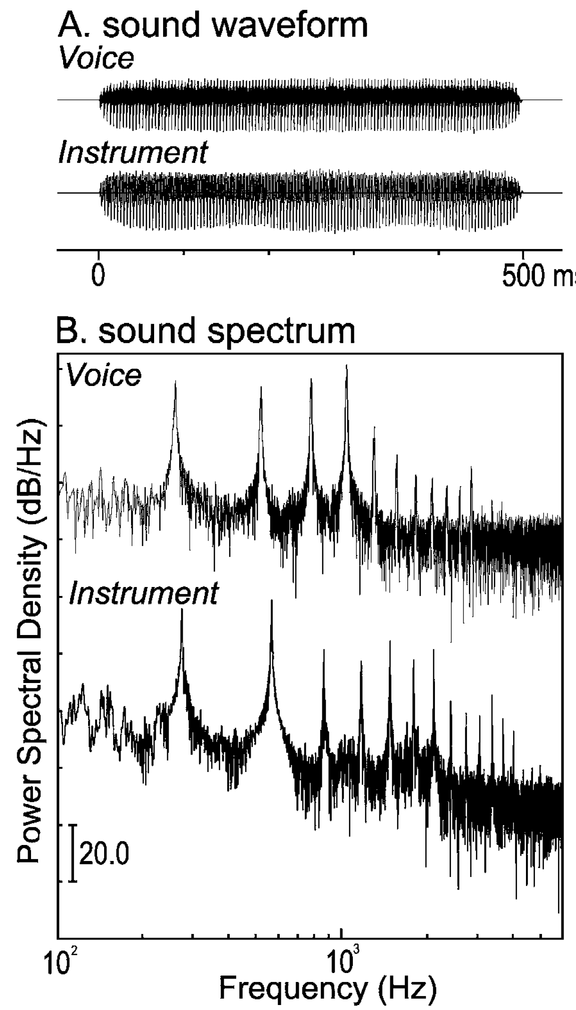

Fig. 1. Sound waveform (A) and sound spectrum (B) of stimuli in the voice and instrument conditions (from Levy et al. [7]). The stimuli were sampled from the central portion of the source sound with a duration of about 1 s, so that the original attack and the decay portions were not included. The duration of the stimuli was 500 ms with rise and fall times of 10 ms. All stimuli were adjusted to yield equivalent root mean square (RMS) power.

The peak of the N1m component was identified as the time point where the root mean square (RMS) value has a maximum about 100 ms after the stimulus onset. The N1m source strength was calculated as the mean value of dipole moments of a 32 ms time interval (ten sampling periods), 16 ms before and after the N1m peak for both stimulus types (vocal and instrumental). A mean value of dipole moments of the time interval between 300 and 500 ms after the stimulus onset was used as the strength of SF source for both stimulus types [5,7]. The dipole location (x, y and z co-ordinates), orientation (x-, y- and z-orientation vectors) and source strength (dipole moment) were analyzed by a repeated measures analysis of variance (ANOVA) with factors of stimulus type (vocal vs. instrumental) and hemisphere (left/right) for N1m and SF components, respectively. Post-hoc tests employed Fisher's protected least significant difference (PLSD).

For both types of stimuli and all subjects, clear N1m (around 100 ms after stimulus onset) and SF components of the auditory evoked field (AEF) were observed. Fig. 2 shows contour maps of N1m and SF components recorded for the vocal and the instrumental sounds. It also shows the time behavior of magnetic responses from two sensors in one subject. The iso-contour map pattern suggests that a single ECD source was located in each hemisphere both for N1m and SF.

The mean value of the correlation between the measured data and those predicted by the ECD model was 0.90 for N1m and 0.85 for SF. The x values of dipole locations for N1m and SF were significantly larger in the right than in the left hemisphere ($P < 0.05$), indicating that their equivalent sources are located more anteriorly in the right hemisphere (cf. Table 1). However, source locations and orientations did not differ between stimulus conditions, as shown in Fig. 3A. Repeated measures ANOVA with the factors of stimulus type, hemisphere and component was carried out to compare the dipole location and orientation between N1m and SF components as additional analysis. There were no significant differences between the N1m and SF components.

The source strength for the N1m component was 47.7 ± 28.8 nAm for the vocal and 46.5 ± 30.0 nAm for the instrumental conditions (Fig. 3B). They did not differ significantly. In contrast, the source strength for the SF for the vocal sound (29.8 ± 21.8 nAm) was significantly larger than that for the instrumental sound (25.7 ± 21.6 nAm) ($P < 0.05$, main effect of the stimulus condition) (Fig. 3B). The main effect of the hemisphere or the interaction of condition and hemisphere was not significant.

In the present study, cortical magnetic responses elicited by human vocal sounds and by instrumental sounds (with matched fundamental frequencies) were recorded. As in a previous study [5], both the N1m and the following SF were

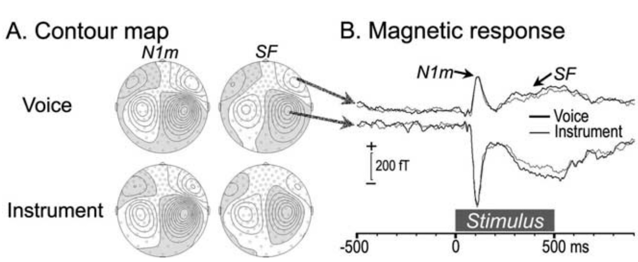

Fig. 2. (A) Iso-contour maps of magnetic responses recorded for the vocal and the instrumental sounds in one subject. The shadowed area illustrates the ingoing magnetic field whereas the white area indicates the outgoing magnetic field. Two peaks of opposite phase are recognized at approximately 100 ms (N1m) and 400 ms (SF) after the stimulus onset for both sounds. (B) Time series of the auditory evoked response to the vocal and instrumental sounds as seen by two MEG sensors. The sensor location is illustrated by filled circles in (A). The arrows indicate the time point of the presented iso-contour maps.

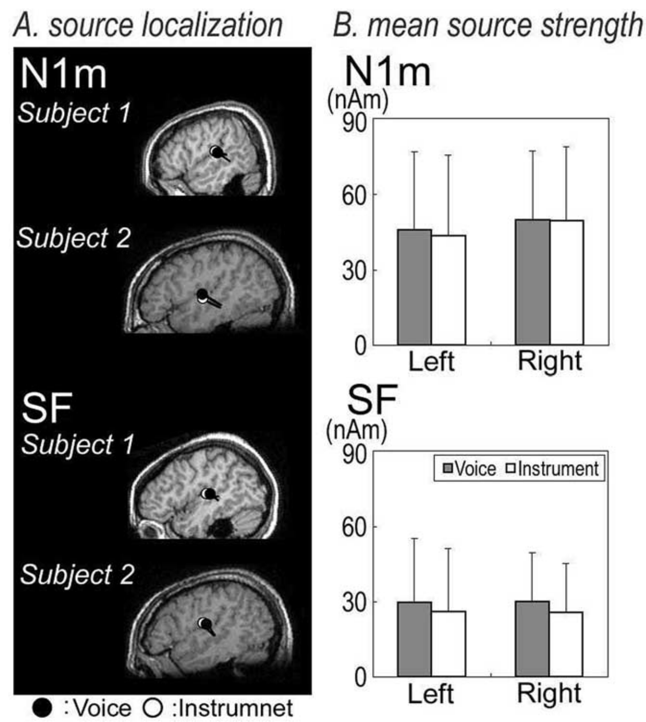

Fig. 3. (A) ECD locations of N1m and SF components for the vocal (filled circle) and the instrument (opened circle) sounds in two subjects. The equivalent sources of N1m and SF are located similarly in the auditory cortex. (B) Mean source strength of N1m and SF components with their standard deviations for the vocal (hatched columns) and instrumental (opened columns) conditions.

observed and estimated sources for both components were located in the auditory cortex [5,10,12]. The source strength of the SF was significantly larger for the human vocal sounds than for the instrumental sounds, whereas for the N1m there was no significant difference between the sounds. Thus, we found human voice-related activities in the auditory cortex.

The SF did not show clear laterality in source activity in the present study, although Eulitz et al. [5] have shown larger SF source activity for the vowel sound in the left than in the right hemisphere. It is thus suggested that vocal sounds without phonetic information are processed bilaterally without a clear hemispheric dominance, whereas the phonetic information is processed bilaterally but more dominantly in the left hemisphere (see also recent reviews on neural correlates of speech perception [1,6,18]).

Considering the late latency of the SF, it is plausible that the process reflected by SF interacts with feedback information from higher cortical areas rather than peripheral inputs. Nakamura et al. [9] reported that the temporal pole plays an important role in identifying familiar people by their voices. The studies using monkeys suggest that 'what' and 'where' streams exit the auditory system in anatomically different locations [11,14] as is the case in the visual system [17]. The 'what' stream is associated with vocalization, and presumably subserves auditory object identification. This stream originates from the anterior part of the

Table 1
N1m and SF means and standard deviations of the dipole locations (*x*, *y* and *z* co-ordinates) for the vocal and instrumental sounds as found for the left and the right hemispheres

Dipole location (cm)	N1m				SF			
	Left		Right		Left		Right	
	Voice	Instrument	Voice	Instrument	Voice	Instrument	Voice	Instrument
x	2.1 ± 0.8	2.1 ± 0.8	2.6 ± 0.7	2.6 ± 0.7	2.0 ± 0.7	2.3 ± 0.6	2.6 ± 0.8	2.7 ± 0.5
y	5.2 ± 0.8	5.3 ± 0.7	−4.9 ± 0.6	−5.1 ± 0.5	5.1 ± 0.8	5.0 ± 1.0	−4.9 ± 0.7	−4.9 ± 0.8
z	6.0 ± 0.6	6.1 ± 0.8	5.9 ± 0.8	6.0 ± 0.7	6.0 ± 0.7	6.1 ± 0.7	6.0 ± 0.6	6.0 ± 0.6

lateral belt area and reaches the anterior temporal area (corresponding to the human temporal pole) and terminates in the dorsolateral frontal cortex. Feedback information from cortical areas involved in the 'what' system might thus influence the SF development.

Although Levy et al. [7] found a positive component peaking at 320 ms (P320) to the vocal sounds but not to the instrumental sounds, the magnetic counterpart of the electric P320 was not observed in the present study. We did not observe activity in the STS either [1–3]. One possible explanation is that the neuronal electromagnetic sources of these activities are mainly radially oriented and therefore not detected by MEG measurements because they do not generate significant extracranial magnetic fields. Another possibility is that because the participants in the current study were watching a silent film while listening to the sounds, the differential responses to the sounds of human voices were diminished, as was reported in a subsequent study by Levy et al. [8].

Acknowledgements

This study was supported by the CIHR, the Ontario Innovation Trust and the Canadian Foundation for Innovation, a Grant-in-Aid for Scientific Research (07335), and the Ministry of Public Management, Home Affairs, Posts and Telecommunications, Japan and the Support Center for Advanced Telecommunications Technology Research, Foundation 'Break Through 21'. We would like to thank Dr Herdman for suggestions.

References

[1] P. Belin, R.J. Zatorre, P. Lafaille, P. Ahad, B. Pike, Voice-selective areas in human auditory cortex sounds, Nature 403 (2000) 309–312.
[2] P. Belin, R.J. Zatorre, P. Ahad, Human temporal-lobe response to vocal sounds, Brain Res. Cogn. Brain Res. 13 (2002) 17–26.
[3] J.R. Binder, J.A. Frost, T.A. Hammeke, P.S. Bellgowan, J.A. Springer, J.N. Kaufman, E.T. Possing, Human temporal lobe activation by speech and non-speech sounds, Cereb. Cortex 10 (2000) 512–528.
[4] B. de Gelder, J. Vroomen, The perception of emotions by ear and by eye, Cognit. Emot. 14 (2000) 289–311.
[5] C. Eulitz, E. Diesch, C. Pantev, S. Hampson, T. Elbert, Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli, J. Neurosci. 15 (1995) 2748–2755.
[6] G. Hickok, D. Poeppel, Towards a functional neuroanatomy of speech perception, Trends Cogn. Sci. 4 (2000) 131–138.
[7] D.A. Levy, R. Granot, S. Bentin, Processing specificity for human voice stimuli: electrophysiological evidence, NeuroReport 12 (2001) 2653–2657.
[8] D.A. Levy, R. Granot, S. Bentin, Neural sensitivity to human voices: ERP evidence of task and attentional influences, Psychophysiology 40 (2003) 291–305.
[9] K. Nakamura, R. Kawashima, M. Sugiura, T. Kato, A. Nakamura, K. Hatano, S. Nagumo, K. Kubota, H. Fukuda, K. Ito, S. Kojima, Neural substrates for recognition of familiar voices: a PET study, Neuropsychologia 39 (2001) 1047–1054.
[10] C. Pantev, C. Eulitz, T. Elbert, M. Hoke, The auditory evoked sustained field: origin and frequency dependence, Electroenceph. clin. Neurophysiol. 90 (1994) 82–90.
[11] J.P. Rauschecker, B. Tian, Mechanisms and streams for processing of "what" and "where" in auditory cortex, Proc. Natl. Acad. Sci. USA 97 (2000) 11800–11806.
[12] R. Salmelin, A. Schnitzler, L. Parkkonen, K. Biermann, P. Helenius, K. Kiviniemi, K. Kuukka, F. Schmitz, H. Freund, Native language, gender, and functional organization of the auditory cortex, Proc. Natl. Acad. Sci. USA 96 (1999) 10460–10465.
[13] J. Sarvas, Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem, Physiol. Med. Biol. 32 (1987) 11–12.
[14] S.K. Scott, I.S. Johnsrude, The neuroanatomical and functional organization of speech perception, Trends Neurosci. 26 (2003) 100–107.
[15] M.D. Szymanski, D.W. Perry, N.M. Gage, H.A. Rowley, J. Walker, M.S. Berger, T.P. Roberts, Magnetic source imaging of late evoked field responses to vowels: toward an assessment of hemispheric dominance for language, J. Neurosurg. 94 (2001) 445–453.
[16] B. Tian, D. Reser, A. Durham, A. Kustov, J.P. Rauschecker, Functional specialization in rhesus monkey auditory cortex, Science 292 (2001) 290–293.
[17] L.G. Ungerleider, M. Mishkin, Two cortical visual systems, in: D.J. Ingle, M.A. Goodale, R.J.W. Mansfield (Eds.), Analysis of Visual Behavior, MIT Press, Cambridge, MA, 1982, pp. 549–586.
[18] D.R. Van Lancker, J. Kreiman, J. Cummings, Voice perception deficits: neuroanatomical correlates of phonagnosia, J. Clin. Exp. Neuropsychol. 11 (1989) 665–674.