ELSEVIER

Cognitive Brain Research 14 (2002) 294–299

COGNITIVE BRAIN RESEARCH

www.elsevier.com/locate/bres

Research report

Cortical processing of speech sounds and their analogues in a spatial auditory environment

Kalle J. Palomäki[a,b,*], Hannu Tiitinen[c,d], Ville Mäkinen[c,d], Patrick May[c], Paavo Alku[b]

[a]*Speech and Hearing Research Group, Department of Computer Science, University of Sheffield, Sheffield, UK*
[b]*Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Helsinki, Finland*
[c]*Apperception & Cortical Dynamics (ACD), Department of Psychology, University of Helsinki, Helsinki, Finland*
[d]*BioMag Laboratory, Medical Engineering Centre, Helsinki University Central Hospital, Helsinki, Finland*

Accepted 8 March 2002

Abstract

We used magnetoencephalographic (MEG) measurements to study how speech sounds presented in a realistic spatial sound environment are processed in human cortex. A spatial sound environment was created by utilizing head-related transfer functions (HRTFs), and using a vowel, a pseudo-vowel, and a wide-band noise burst as stimuli. The behaviour of the most prominent auditory response, the cortically generated N1m, was investigated above the left and right hemisphere. We found that the N1m responses elicited by the vowel and by the pseudo-vowel were much larger in amplitude than those evoked by the noise burst. Corroborating previous observations, we also found that cortical activity reflecting the processing of spatial sound was more pronounced in the right than in the left hemisphere for all of the stimulus types and that both hemispheres exhibited contralateral tuning to sound direction.

Theme: Sensory systems

Topic: Auditory systems: central physiology

Keywords: Auditory; Vowel; Pseudo-vowel; Sound localization; Magnetoencephalography; N1m

1. Introduction

When the auditory system processes speech, it has to extract information from the various physical features of the acoustic speech waveform reaching our ears. The most important acoustic cues for speech intelligibility are extracted from the spectral and the temporal structure of the waveform. The spatial location of the source of speech or that of any type of sound becomes important when the subject's attention needs to be directed towards important sound events that occur in different regions of auditory space.

In the human brain, the cortical processing of speech has been investigated by utilizing event related potentials (ERPs) and magnetic fields (EMFs). The dynamics of the most prominent auditory ERP/EMF deflection, the N1m [18], has been in the focus of the research, and it has been successfully applied in studies of cortical processing of sustained vowels [1,10,11,23]. However, the cortical processing of speech sounds in natural spatial environments has remained largely unaddressed. Psychoacoustic data on human sound localization [3] indicates that in the perception of sound location, binaural cues from interaural time and level differences (ITD and ILD, respectively) as well as monaural spectral cues from the filtering effects of the pinna, head and body are utilized. While the cortical processing of interaural cues have been investigated using magnetoencephalography (MEG) [14,15,17] and electroencephalography (EEG) [19], however, these studies have not taken into account the spectral cues which are known to be crucial in sound localization.

Conventionally, realistic spatial auditory environments have been achieved only by presenting stimuli through loudspeakers situated around the subject (e.g., Refs.

*Corresponding author. Department of Computer Science, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK. Tel.: +44-114-222-1905; fax: +44-114-222-1810.

E-mail address: kalle.palomaki@hut.fi (K.J. Palomäki).

0926-6410/02/$ – see front matter
PII: S0926-6410(02)00132-5

[7,19,22]). This kind of experimental setup has precluded the use of MEG, with its superior spatial and temporal resolution, because of magnetic interference due to loudspeaker parts. The application of the novel audio-technological method of head-related transfer functions (HRTFs) [26] allows the representation of stimuli in natural three-dimensional space around the subject using acoustic tube earphones made of plastic which do not cause magnetic interference.

The cerebral processing of spatial sounds produced by HRTFs has been studied with positron emission tomography (PET) [5,24]. These studies have shown that blood-flow increases in superior parietal and prefrontal areas during spatial auditory tasks [5,24]. A recent MEG study [20] on the cortical processing of spatial broad-band noise bursts generated by HRTFs showed that the N1m in both the left and right cerebral hemispheres is maximal to contralaterally located sound sources and that the right hemisphere might be more sensitive in the processing of auditory spatial information. Evidence for the importance of the right hemisphere in auditory spatial processing has also been found in studies of auditory neglect following right-hemispheric damage [8,9]. In addition, previous psychophysical observations [4,6] have revealed that humans are more accurate in localizing events in the left hemi-field, which has been interpreted as evidence for right-hemispheric specialization of spatial processing. However, the above results [4,6,8,9] might be regarded as indirect evidence for the right-hemispheric specialization of auditory space, and they further accentuate the need to find an objective, direct measure for understanding the neuronal mechanisms involved in spatial sound processing. As tentatively addressed in Ref. [20], the N1m seems to be a promising candidate for these purposes.

Contrasting the evidence on right-hemispheric preponderance of processing of spatial sound features, there is ample documentation that speech and language processes occur predominantly in the left hemisphere [12]. However, studies exploiting the N1m have not revealed significant differences between the hemispheres in the processing of speech stimuli [1,10,11]. An important question, addressed in this study, is whether this hemispheric 'non-selectivity' of the N1m amplitude in vowel processing remains when the stimuli have a spatial quality or whether the N1m activity shifts to the right hemisphere, as was previously observed with noise stimuli [20].

In the present study, we use MEG, firstly, to investigate the cortical processing of a vowel stimulus (Finnish vowel /a/) presented in a realistic spatial environment. Secondly, we analyse to what extent the processing of this phonetic stimulus differs from the processing of the noise burst that provides localization cues over wider frequency band than the vowel. As an intermediate stimulus type between these two extremes we introduce a third stimulus, the pseudo-vowel, which shares the main spectral structure of the vowel but lacks in phonetic content.

2. Materials and methods

Ten volunteers (all right-handed, two female, mean age 28 years) with normal hearing served as subjects with informed consent and with the approval of the Ethical Committee of Helsinki University Central Hospital (HUCH).

The spatial stimuli (see Fig. 1) were of three different stimulus types: (1) Finnish vowel /a/, (2) a pseudo-vowel comprising the sum of sinusoids and (3) a wide-band noise burst. The vowel sound was produced using a method [2] which synthesizes voiced speech as a combination of glottal flow, computed from a natural utterance, and a digital all-pole filter modeling the vocal tract. The pseudo-vowel consisted of 11 sinusoids the frequency and the level of which were selected to match the harmonics in the spectrum of the vowel (the first sinusoid corresponded to the fundamental of the vowel and the rest matched the harmonics in the vicinity of each formant).

The use of a vowel to represent speech sound was motivated by two issues. Firstly, the application of the method [2], whereby the main spectral characteristics of the speech sound and the non-phonetic stimulus type (the pseudo-vowel) were matched, required the speech sound to be a vowel rather than an unvoiced utterance such as a sibilant. Secondly, the perceptual quality of an isolated unvoiced speech sound would not have been recognized as speech as easily as the vowel stimulus and, further, might have become too close to that of the noise stimulus.

The stimuli were presented in stimulus type blocks (vowels, pseudo-vowels & noise) whose order of presentation was counterbalanced across subjects. In each block, the stimuli were presented randomly from eight

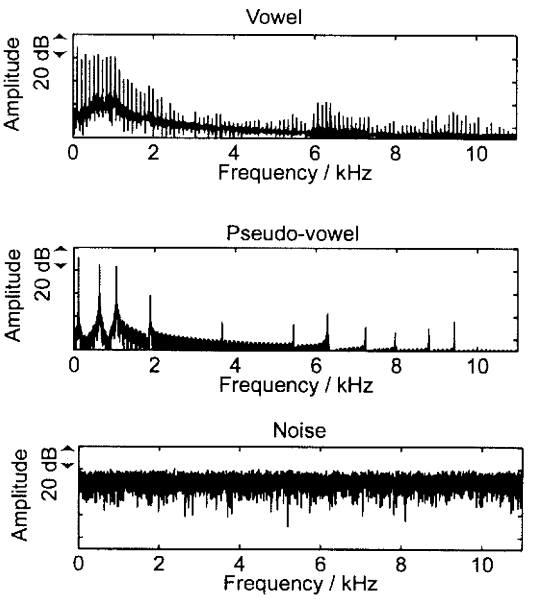

Fig. 1. The spectra of the vowel, the pseudo-vowel and the noise burst used as stimuli in the current experiment.

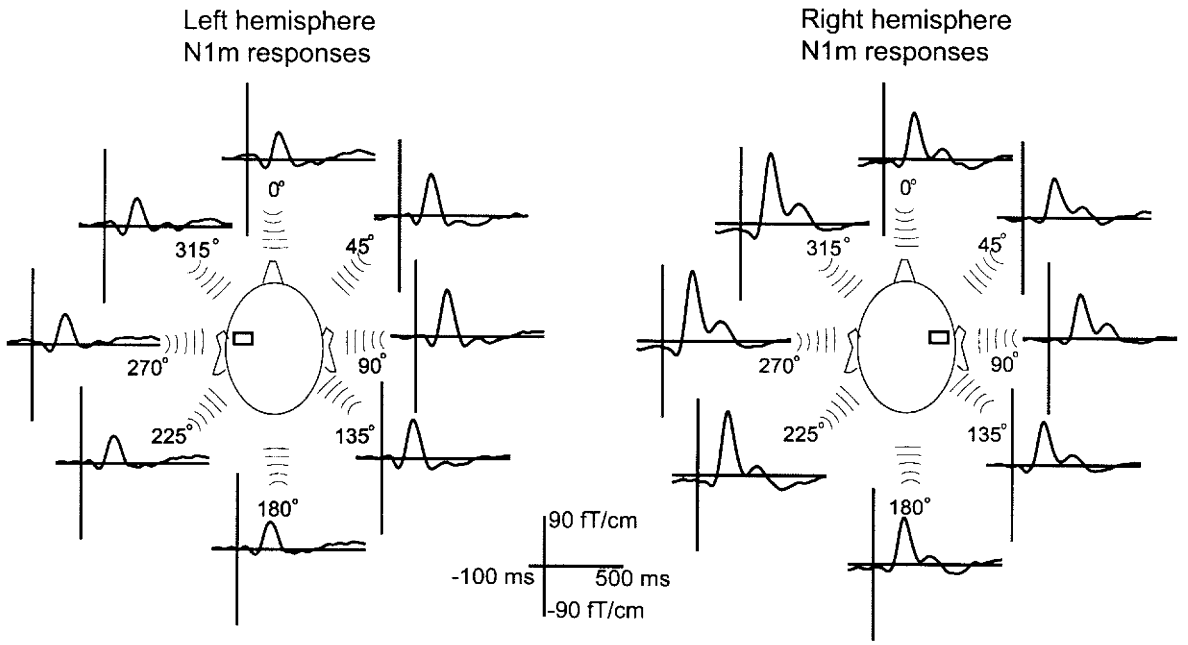

Fig. 2. The N1m responses elicited by the vowel for both hemispheres (grand-averaged over 10 subjects) from the sensor maximally detecting N1m activity shown for each of the eight direction angles. MEG measurements were conducted by using spatial sounds corresponding to eight different source locations. The stimuli comprised HRTF-filtered sounds whose direction angle was varied between 0 and 315° in the azimuthal plane.

equally spaced directions in the horizontal plane (0, 45, 90, 135, 180, 225, 270 and 315°; see Fig. 2) using HRTFs provided by the University of Wisconsin [26]. The HRTF stimulus condition was non-individualized [25]. The localization performance using the same set of HRTFs during stimulus generation was previously established in a simple behavioral test [20]. The stimulus bandwidth was 11 kHz, the stimulus duration was 100 ms, and the onset-to-onset interstimulus interval was 800 ms. The stimulus intensities for each stimulus type were normalized by scaling the sound pressure level (SPL) of the 0°-sound to 75 dB (A). The SPLs of the virtual sources were kept constant over azimuthal direction. The stimuli were delivered to the subject's ears with plastic tubes whose frequency response was equalized digitally up to 11 kHz.

The magnetic responses elicited by the auditory stimuli were recorded (passband 0.03–100 Hz, sampling rate 400 Hz) with a 122-channel whole-head magnetometer which measures the gradients $\partial B_z/\partial x$ and $\partial B_z/\partial y$ of the magnetic field component B_z at 61 locations over the head. The subject, sitting in a reclining chair, was instructed not to pay attention to the auditory stimuli and to concentrate on watching a self-selected silent film. For each of the eight directions, over 100 instances of each stimulus type were presented to each subject. Brain activity was baseline-corrected with respect to a 100-ms pre-stimulus period, averaged over a 500-ms post-stimulus period, and filtered with a passband of 1–30 Hz. Electrodes monitoring both horizontal (HEOG) and vertical (VEOG) eye movements were used for removing artefacts, defined as activity with an absolute amplitude greater than 150 μV.

Data from the channel pairs above the temporal lobes with the largest N1m amplitude, determined as the peak amplitude of the channel pair vector sum, were analyzed separately for both hemispheres. As gradient magnetometers pick up brain activity maximally directly above the source [16], the channel pair at which the maximum was obtained indicates the approximate location of the underlying source. The location of the source of the N1m responses was also estimated using unrestricted equivalent current dipoles (ECDs) [13]. A subset of 34 channels over the temporal areas of the left and right hemisphere were separately used in the ECD estimation. The head-based coordinate system was defined by the *x*-axis passing through the preauricular points (positive to right), the *y*-axis passing through the nasion, and the *z*-axis as the vector cross product of the *x* and *y* unit vectors. The results on N1m amplitudes and ECD locations with respect to the three stimulus types were tested with analyses of variance (ANOVAs).

The perceptual qualities of the three stimulus types were tested with a behavioral listening test. The subjects ($N=10$) were asked to listen to a sound sample of the vowel, pseudo vowel and noise from the 0° sound source direction freely, as many times as they wished. Their task was to categorize the signals into three different categories: vowel, harmonic signal, or noise.

3. Results

The behavioral listening test revealed that the subjects categorized the signals with a 100% accuracy, which shows that the signals were clearly perceptually different.

Fig. 2 shows the grand-averaged N1m responses elicited by the vowel presented from eight different directions. The amplitude variation of N1m as a function of the three stimulus types was statistically significant in both the left and the right hemisphere ($F[2,18]=11.12$, $P<0.001$ & $F[2,18]=24.86$, $P<0.001$, respectively). Post-hoc analyses (Newman–Keuls tests) revealed that the N1m responses elicited by both the vowel and the pseudo-vowel were significantly larger in amplitude than those elicited by the noise bursts ($P<0.01$ and $P<0.001$ in the left and right hemisphere, respectively; see also Fig. 3). The amplitudes of the N1m responses elicited by the vowel and the pseudo-vowel, however, did not differ from one another ($P=$n.s.).

For all three types of stimulus, the N1m responses were always larger in amplitude over the right than over the left hemisphere (Fig. 3). The respective mean N1m amplitudes (averaged across the eight direction angles) for the vowel, pseudo-vowel and noise, respectively, were 120.31, 114.02, and 63.34 fT/cm in the right hemisphere and 66.82, 61.95, and 33.50 fT/cm in the left. This right-hemispheric preponderance was statistically significant for the vowel ($F[1,9]=9.79$, $P<0.05$) and pseudo-vowel ($F[1,9]=8.32$, $P<0.05$) and also approached statistical significance for noise ($F[1,9]=4.27$, $P<0.07$). On the average, the N1m amplitudes of the right hemisphere were 1.80, 1.84 and 1.89 times larger than those observed in the left hemisphere for the vowel, pseudo-vowel and noise, respectively.

Table 1
ANOVA results for the N1m amplitude show that in both hemispheres the amplitude variance was statistically significant for the vowel, the pseudo-vowel and the noise stimuli

	Left hemisphere	Right hemisphere
Vowel	$F[9,63]=13.67$, $P<0.001$	$F[7,63]=14.96$, $P<0.001$
Pseudo-vowel	$F[7,63]=9.36$, $P<0.001$	$F[7,63]=11.56$, $P<0.001$
Noise	$F[7,63]=2.75$, $P<0.05$	$F[7,63]=10.00$, $P<0.001$

Both hemispheres exhibited tuning to the direction angle of all three stimulus types, with a contralateral maximum and an ipsilateral minimum in the N1m amplitude (Fig. 3). The amplitude variation as a function of direction angle was statistically significant over both hemispheres and for all stimulus types (see Table 1). On the average, this

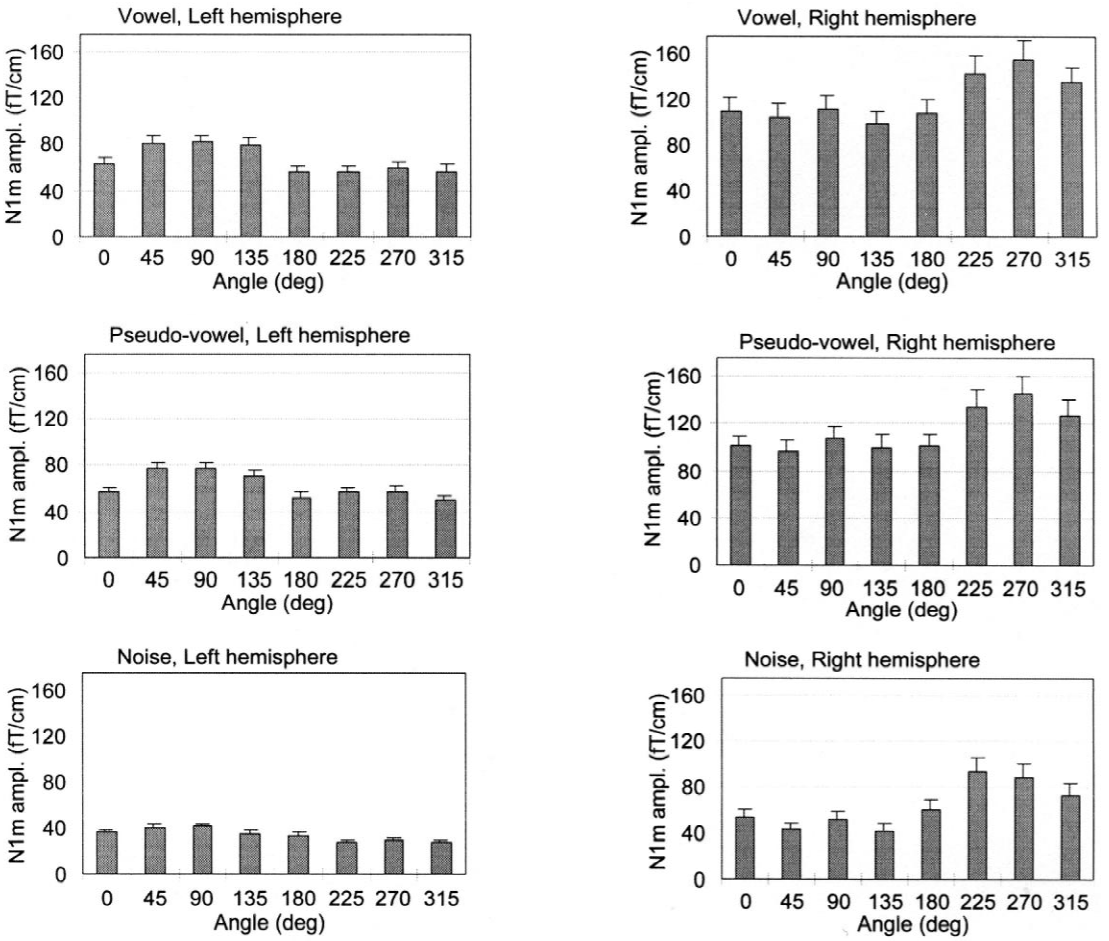

Fig. 3. The N1m amplitude (grand-averaged over 10 subjects) as a function of direction angle calculated as the vector sum from the channel pair maximally detecting N1m activity over the left and right hemisphere. The stimulus types from top to bottom are: the vowel, the pseudo-vowel and the noise burst. Error bars indicate standard error-of-the-mean.

amplitude variation (i.e., maximum minus minimum amplitude in each hemisphere) over direction angle was 2.18 (vowel), 1.86 (pseudo-vowel) and 3.81 (noise) times larger in the right than the in left hemisphere.

The mean goodness-of-fit of the ECD estimation of N1m was 78.4% in the left and 84.2% in the right hemisphere. While stimulus type did not have a statistically significant effect on the location of the ECDs describing the N1m responses ($F[2,68]=2.49$, $F[2,68]=0.20$, $F[2,68]=1.56$, for x, y, & z axes, respectively; P=n.s. in all cases), the stimulus direction angle affected the location of the ECDs. In the left hemisphere, the source location of the N1m averaged over stimulus type varied along the inferior–superior axis (z-axis of the coordinate system: $F[7,63]=2.21$, $P<0.05$), with the value of the z-coordinate varying over a range of 12.6 mm. Further, the sources of the N1m responses elicited by sounds from the front (315, 0, & 45°) were 6.4 mm inferior to those elicited by sounds from behind (135, 180, & 225°; $F[1,9]=9.73$, $P<0.05$). In the right hemisphere, an analysis of the N1m responses elicited by sounds from the right (45, 90, & 135°) and the left side of the subject (225, 270 & 315°) revealed that the generators of the responses of the right-side sound sources were 2.1 mm lateral to those of the left-side sound sources ($F[1,9]=7.46$, $P<0.05$).

4. Discussion

We studied the human cortical processing of vowel, pseudo-vowel and noise sounds presented in a realistic spatial environment. In both the left and right hemisphere, the N1m responses elicited by the vowel and pseudo-vowel were much larger in amplitude than those elicited by noise-burst stimuli. While both cortical hemispheres are sensitive to sound location, the right hemisphere seems to be highly specialized in the processing of auditory space. This specialization is indicated by the right-hemispheric preponderance of the N1m responses. The right hemisphere also appears to be much more sensitive to changes in the sound direction angle, as the overall variation of the amplitude of the N1m across direction angle ranged from two up to four times larger in the right hemisphere than in the left.

The observed right-hemispheric preponderance of the N1m elicited by the three perceptually very different stimulus types corroborates our previous results obtained using white noise stimuli only [20]. Our results are also in line with previous indirect observations emphasizing the importance of the right hemisphere in auditory spatial processing [4,6,8,9]. Interestingly, this right-hemispheric preponderance holds also for the vowel stimulus which, however, has not been reported to cause lateralization of the N1m when presented diotically [1,10,11]. Thus, on the basis of our present observations, vowels presented in a realistic spatial environment seem to activate the auditory cortex differently than do diotically presented vowel stimuli.

These results are paralleled by the commonly accepted view [12] of the functional division between the hemispheres where the left hemisphere is specialized in language processing and the right hemisphere in analyzing spatial and spatiotemporal information. In contrast to left-hemispheric language processing, however, the speech sound used in our experiment elicited activity in the right hemisphere which was almost double the strength to that elicited in the left hemisphere. Thus, in conjunction with previous research attempts [1,10,11], it seems that left-hemispheric speech specialization is not likely to be reflected in the N1m response dynamics under passive (no-task) recording conditions [21].

When comparing the N1m response elicited by the vowel and the pseudo-vowel to that elicited by the noise stimulus, we found that noise reduces markedly the amplitude of the N1m. A similar reduction of the N1m amplitude was observed by Alku et al. [1] when responses to a periodic vowel and to its aperiodic, noise-like counterpart (where the natural glottal excitation was replaced by noise excitation) were measured. Interestingly, the N1m amplitude patterns in the present study were rather similar for the vowel and the pseudo-vowel although the stimuli were perceptually clearly different. From this, one might propose that the early auditory cortical processes treat incoming stimuli either as ‘speech’ or ‘non-speech’. Speech sounds, in this case, would comprise both those sounds that are identified as vowels as well as their analogies which share a similar spectro-temporal main structure but are not recognized as vowels. Consequently, when processing speech or speech analogues, the auditory cortex responds unequivocally, and this is observable as prominent N1m responses without amplitude variation. This suggests that speech sounds receive a very early selective processing which occurs already in the sensory brain areas. This, in turn, might imply a higher-order mechanism responsible for the subjective differentiation between different vowel identities. Further experimental evidence supporting this possibility will be presented elsewhere (Mäkelä et al., submitted for publication).

Finally, the ECD analysis provided tentative evidence for a spatial coding of sound source direction in the auditory cortex. In the left hemisphere, the generator locations of the N1m responses appeared to vary along the superior–inferior axis according to sound source direction. The generators for responses to sounds coming from the front of and behind the subject were localized differently along this axis also. In addition, the ECDs in the right hemisphere for responses to sounds from the left and right side of the head were organized in different locations. However, more conclusive proof for a topographic mapping of auditory space onto auditory cortex is obviously needed.

Acknowledgements

This study was supported by the Academy of Finland (project No. 1168030), the University of Helsinki, Helsinki University of Technology and EC TMR SPHEAR project.

References

[1] P. Alku, P. Sivonen, K.J. Palomäki, H. Tiitinen, The periodic structure of vowel sounds is reflected in human electromagnetic brain responses, Neurosci. Lett. 298 (2001) 25–28.
[2] P. Alku, H. Tiitinen, R.A. Näätänen, A method for generating natural-sounding speech stimuli for cognitive brain research, Clin. Neurophys. 110 (1999) 1329–1333.
[3] J. Blauert, Spatial Hearing, MIT Press, Campridge, 1997.
[4] K.A. Burke, A. Letsos, R.A. Butler, Asymmetric performances in binaural localization of sound in space, Neuropsychologia 32 (1994) 1409–1417.
[5] K.O. Bushara, R.A. Weeks, K. Ishii, M.-J. Catalan, B. Tian, J.P. Rauschecker, M. Hallet, Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans, Nat. Neurosci. 2 (1999) 759–766.
[6] R.A. Butler, Asymmetric performances in monaural localization of sound in space, Neuropsychologia 32 (1994) 221–229.
[7] R.A. Butler, The influence of spatial separation of sound sources on the auditory evoked response, Neuropsychologia 10 (1972) 219–225.
[8] L.Y. Deouell, S. Bentin, N. Soroker, Electrophysiological evidence for an early (pre-attentive) information processing deficit in patients with right hemisphere damage and unilateral neglect, Brain 123 (2000) 353–365.
[9] L.Y. Deouell, N. Soroker, What is extinguished in auditory extinction?, Neuroreport 11 (2000) 3059–3062.
[10] E. Diesch, C. Eulitz, S. Hampson, B. Ross, The neurotopography of vowels as mirrored by evoked magnetic field measurements, Brain Lang. 53 (1996) 143–168.
[11] C. Eulitz, E. Diesch, C. Pantev, C. Hampson, T. Elbert, Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli, J. Neurosci. 15 (1995) 2748–2755.
[12] M.S. Gazzaniga, R.B. Ivry, G.R. Mangun, Cognitive Neuroscience: The Biology of the Mind, W.W. Norton, New York, 1998.
[13] M. Hämäläinen, R. Hari, J. Ilmoniemi, J. Knuutila, O.V. Lounasmaa, Magnetoencephalography—theory, instrumentation, and applications to non-invasive studies of the working human brain, Rev. Mod. Phys. 65 (1993) 413–497.
[14] K. Itoh, M. Yumoto, A. Uno, T. Kurauchi, K. Kaga, Temporal stream of cortical representation for auditory spatial localization in human hemispheres, Neurosci. Lett. 292 (2000) 215–219.
[15] J. Kaiser, W. Lutzenberger, H. Preissl, H. Ackermann, N. Birbaumer, Right-hemisphere dominance for the processing of sound-source lateralization, J. Neurosci. 20 (2000) 6631–6639.
[16] J.E.T. Knuutila, A.I. Ahonen, M.S. Hämäläinen, M. Kajola, P. Laine, O.V. Lounasmaa, L. Parkkonen, J. Simola, C. Tesche, A 122-channel whole-cortex SQUID system for measuring the brain's magnetic-fields, IEEE Trans. Magn. 29 (1993) 3315–3320.
[17] L. McEvoy, R. Hari, T. Imada, M. Sams, Human auditory cortical mechanisms of sound lateralization: II. Interaural time differences at sound onset, Hear. Res. 67 (1993) 98–109.
[18] R. Näätänen, T. Picton, The N1 wave of the human electric and magnetic response to sound: a review and analysis of the component structure, Psychophysiology 24 (1987) 375–425.
[19] P. Paavilainen, M.-L. Karlsson, K. Reinikainen, R. Näätänen, Mismatch negativity to change in spatial location of an auditory stimulus, Electroencephaolgr. Clin. Neurophysiol. 73 (1989) 129–141.
[20] K. Palomäki, P. Alku, V. Mäkinen, P. May, H. Tiitinen, Sound localization in the human brain: neuromagnetic observations, Neuroreport 11 (2000) 1535–1538.
[21] D. Poeppel, E. Yellin, C. Phillips, T.P.L. Roberts, H.A. Rowley, K. Wexler, A. Marantz, Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds, Cogn. Brain Res. 4 (1996) 231–242.
[22] W. Teder-Sälejärvi, S.A. Hillyard, The gradient of spatial auditory attention in free field: an event-related potential study, Percept. Psychophys. 60 (1998) 1228–1242.
[23] H. Tiitinen, P. Sivonen, P. Alku, J. Virtanen, R. Näätänen, Electromagnetic recordings reveal latency differences in speech and tone processing in humans, Cogn. Brain Res. 8 (1999) 355–363.
[24] R.A. Weeks, A. Aziz-Sultan, K.O. Bushara, B. Tian, C.M. Wessinger, N. Dang, J.P. Rauschecker, M. Hallett, A PET study of human auditory spatial processing, Neurosci. Lett. 262 (1999) 155–158.
[25] E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman, Localization using nonindividualized head-related transfer functions, J. Acoust. Soc. Am. 94 (1993) 111–123.
[26] F.L. Wightman, D.J. Kistler, Headphone simulation of free-field listening. I: Stimulus synthesis, II: Psychophysical validation, J. Acoust. Soc. Am. 85 (1989) 858–878, See also http://www.waisman.wisc.edu/hdrl/index.html.