ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

NEUROMAGNETIC STUDY ON LOCALIZATION OF SPEECH SOUNDS

Kalle J. Palomäki[1], Paavo Alku[1], Ville Mäkinen[2,3], Patrick May[2] and Hannu Tiitinen[2,3]

[1]Helsinki University of Technology, Laboratory of Acoustics and Audio
Signal Processing, PO Box 3000, Fin-02015 TKK, Finland
email: Kalle.Palomaki@hut.fi
[2]University of Helsinki, Dept. of Psychology, Finland
[3] BioMag Laboratory, Medical Engineering Centre, Helsinki University Central Hospital, Finland

ABSTRACT

Spatial processing of speech sounds by the human auditory cortex was studied measuring neuromagnetic responses utilizing magnetoencephalography (MEG). Realistic spatial sound environment was produced using modern stimulus generation methodology utilizing head-related transfer functions (HRTFs). In order to compare localization of speech sounds to that of non-speech, the stimulus set involved three different stimulus types: 1) a semi-synthetic /a/-vowel, 2) a pseudo-vowel composed as a sum of sinusoids and 3) a wide band noise burst. Stimuli were filtered through HRTFs of eight horizontal equally spaced directions. The most prominent response, the cortically generated N1m, was investigated above the left and right hemisphere. We found, firstly, that cortical activity reflecting the processing of spatial sound stimuli was more pronounced in the right than in the left hemisphere. Secondly, we found that N1m amplitudes were largest for the /a/-vowel. However, behaviour of the N1m amplitude elicited by the pseudo-vowel was relatively similar to that of the /a/-vowel.

1. INTRODUCTION

In processing of speech sounds, the human hearing mechanism analyzes many physical properties of speech such as spectral and temporal structure and spatial location of the sound. Spectral and temporal properties of speech sounds play the fundamental role in recognising different phonemes and words. Spatial location of the source of speech becomes important, for example, in the so-called cocktail party situation, where there are multiple speakers in different directions [1]. The ability to discriminate spatial locations of the source of speech is important for locating speakers and thus helping in concentrating on a specific speaker as well as helping in ignoring unwanted sound sources. As described in [2], perception of sound location utilizes binaural cues from interaural time and level differences (ITD and ILD, respectively) and monaural spectral cues from the filtering effects of the pinna, head and body.

The way the human brain processes auditory stimuli has been investigated non-invasively by measuring, for example, the electric and magnetic fields outside head, which are generated by the auditory brain activity. These electric and magnetic responses are measured with Electroenchephalography (EEG) and Magnetoenchephalography (MEG), respectively. In speech perception, these methologies have yielded information on, for example, how different vowel sounds [3] and words [4] are represented in the human brain. However, with studies on sound localization these measurement methodologies have been applied only together with non-speech stimuli (e.g., [5,6]). Therefore, previous studies do not answer to the question of how the brain processes speech sounds appearing naturally in three dimensional spatial locations, which is the main focus of the present study.

In order to investigate speech sound localization we used the neuromagnetic measurement technique, MEG, and investigated the event related magnetic field component, the N1m, which is a prominent response of the auditory cortex after 100ms from stimulus onset (for a review see [7]). The behaviour of the N1m reflects how strongly the auditory cortex responds at a sound onset.

Previously, neuromagnetic measurements of sound localization have been impossible because loudspeakers containing magnetic parts cannot be placed into a magnetically shielded measurement room. However, the latest developments in audio technology, especially the HRTFs [8] now allow the use of natural spatial sounds even in the very sensitive measurement conditions of MEG. In the present study, we used HRTFs to represent speech stimuli naturally in three dimensional space surrounding the subject. For the first time, the HRTF-stimulus production method was combined with MEG in [6]. Similar HRTF sound delivery system was used together with positron emission tomography (PET) brain imaging in [5].

2. MATERIAL AND METHODS

A total of eleven volunteers (all right-handed, 2 female, mean age 28 years) served as subjects with informed consent and the approval of the Ethical Committee of Helsinki University Central Hospital (HUCH).

Altogether three different stimulus types were applied in the production of the spatial stimuli: 1) a semi-synthetic Finnish /a/-vowel, 2) a pseudo-vowel composed as the sum of sinusoids (see Figure 2) and 3) a wideband noise burst. The vowel sound was synthesized using a method described in [9]. This method synthesizes voiced speech as a combination of glottal flow, computed from a natural utterance, and a digital all-pole filter, which models the vocal tract. The pseudo-vowel consisted of ten sinusoids. The frequency and the level of these sinusoids were selected to match the harmonics in the spectrum of the vowel /a/ as follows: The first sinusoid corresponded to the fundamental of /a/ and the rest of the sinusoids were selected to match the harmonic in vicinity of each formant.

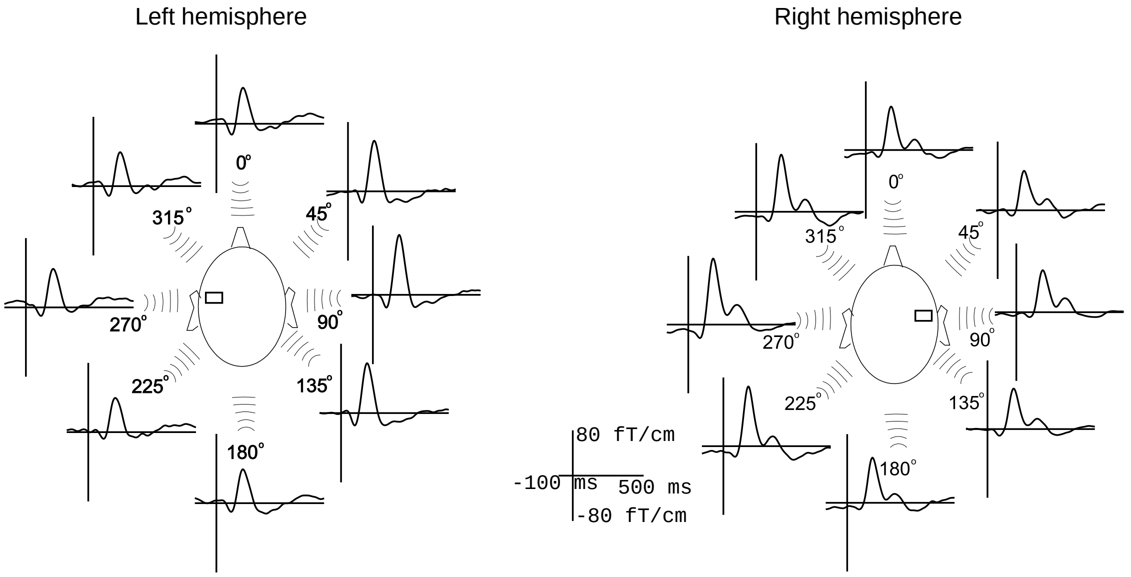

Figure 1: N1m responses elicited by the vowel /a/ for both hemispheres (grand-averaged over 11 subjects) from the sensor maximally detecting N1m activity shown for each of the eight direction angles. MEG measurements were conducted by using spatial sounds corresponding to eight different source locations as stimuli. The stimuli comprised HRTF-filtered sounds whose direction angle was varied between 0° and 315° in the azimuthal plane. The obtained magnetic responses were quantified by using the channels with the largest N1m amplitude over the left and the right hemisphere.

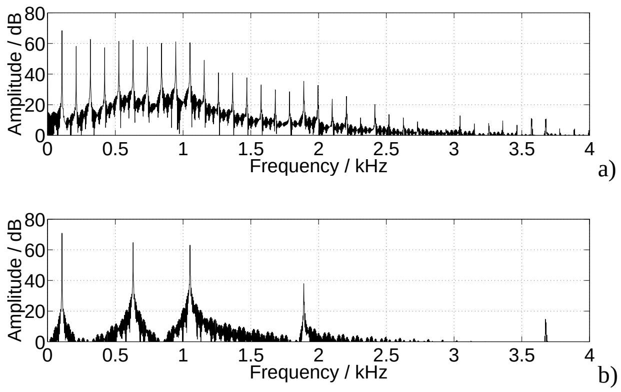

Figure 2: Spectra of a) the semi-synthetic /a/-vowel and b) the pseudo-vowel. Figure represents 0-4 kHz of the whole 0-11 kHz band of the stimuli.

Stimuli were presented in eight equally spaced directions (See Figure 1) in horizontal plane (0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°) using the HRTFs provided by the University of Wisconsin [8]. Bandwidths of the stimuli were 11 kHz, stimulus duration 100 ms and onset-to-onset interstimulus interval 800 ms. In the MEG experiment, the stimuli were played by randomizing their direction angle. The intensity of each stimulus type was scaled by adjusting the SPL-value of the sound with 0° azimuth to 75 dB.

The EMFs elicited by auditory stimuli were recorded (passband 0.03-100 Hz, sampling rate 400 Hz) with a 122-channel whole-head magnetometer (Neuromag-122) which measures the gradients $\partial Bz/\partial x$ and $\partial Bz/\partial y$ of the magnetic field component Bz at 61 locations over the head. The subject, sitting in a reclining chair, concentrated on watching a silent movie and was instructed not to pay attention to the auditory stimuli. Brain activity time locked to stimulus onset was averaged over a 500-ms post-stimulus period (baseline-corrected with respect to a 100-ms pre-stimulus period) and filtered with a passband of 1-30 Hz. Electrodes monitoring both horizontal (HEOG) and vertical (VEOG) eye movements were used in removing artefacts, defined as activity in excess of +150 μV. Over 100 instances of each stimulus type were presented to each subject.

Data from channel pairs above the temporal lobes with the largest N1m amplitude were analyzed separately for both hemispheres, and the N1m amplitude was determined as a vector sum from these channel pairs. As MEG sensors pick up brain activity maximally directly above the source [10], the channel pair at which the maximum was obtained indicates the approximate location of the underlying source.

3. RESULTS

In the MEG recordings, all stimuli elicited prominent N1m responses. Figure 1 shows grand-averaged N1m responses for the /a/-vowel from eight directions. The grand-averaged N1m amplitude is shown in Figure 3 for the /a/-vowel, the pseudo-vowel and the wideband noise burst in top, middle and bottom panels respectively.

For all the three stimulus types both hemispheres exhibited tuning to the direction angle, with a contralateral maximum and an ipsilateral minimum in the N1m amplitude. For the vowel /a/ this implies that the N1m amplitude reaches the maximum value in the direction angles 90° and 270° for the left and right hemisphere, respectively and the minimum value in the direction angles 225° and 135° for the left and right hemisphere, respectively.

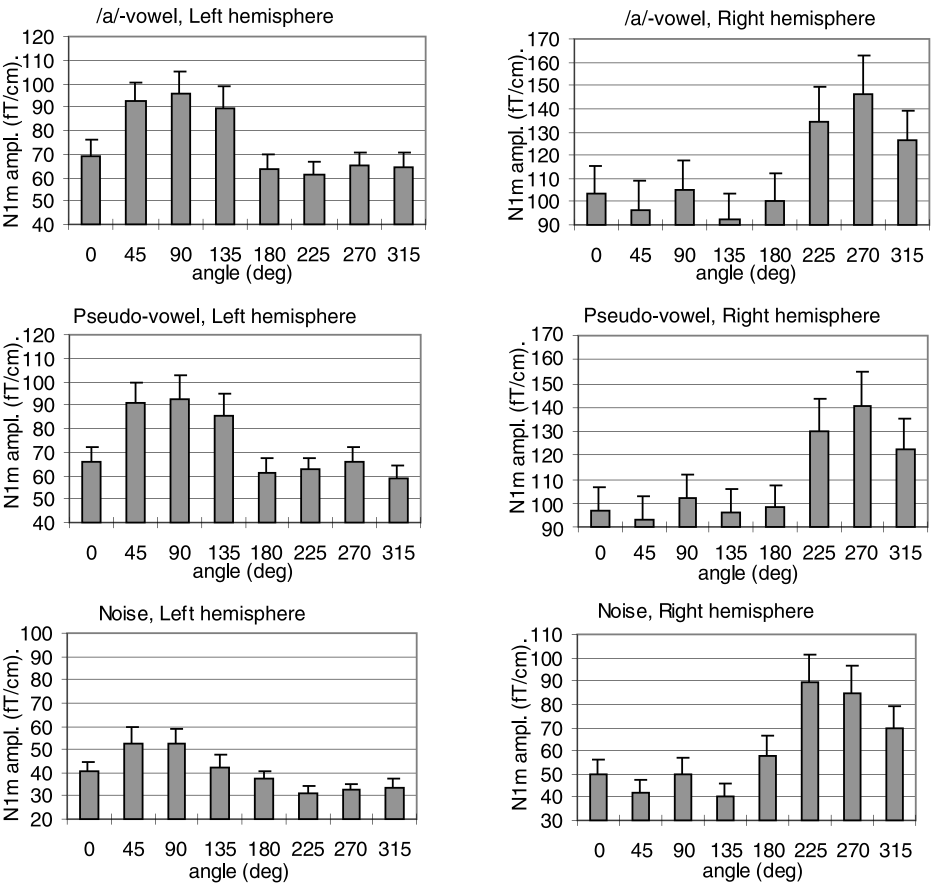

Figure 3: The N1m amplitude (grand-averaged over 11 subjects) as a function of direction angle calculated as the vector sum from the channel pair maximally detecting N1m activity over the left and right hemisphere. Stimulus types from top to down are: the /a/-vowel, the pseudo-vowel and the noise burst. Error bars indicate standard error of the mean.

The current results suggest the right-hemispheric preponderance in processing of spatial sound. The N1m responses were considerably larger in amplitude over the right than over the left hemisphere with all three stimulus types. The mean N1m amplitude across the eight direction angles was larger in the right hemisphere (113.19, 109.94 and 60.41 fT/cm, for the /a/-vowel, pseudo-vowel and noise, respectively) than in the left (75.14, 72.86 and 40.35 fT/cm, for the /a/-vowel, pseudo-vowel and noise, respectively). Interestingly, these numerical data indicate that the ratio in N1m activity between the right and the left hemisphere was 1.5 for all three stimulus types. Further, the N1m amplitude variation was larger in the right than in the left hemisphere.

Comparing brain activity between each stimulus type shows that the /a/-vowel exhibits the largest N1m amplitude in both hemispheres. This is seen in both the peak amplitude of the N1m (occurring in 270°) and the N1m amplitude averaged over all directions. Furthermore, the N1m amplitude elicited by both the /a/-vowel and the pseudo-vowel was considerably larger than that of the noise burst.

4. DISCUSSION

Our results show that both cortical hemispheres are sensitive to sound location in three-dimensional space. The amplitude of the N1m exhibited tuning to the sound direction angle in both hemispheres, with response maxima and minima occurring for contralateral and ipsilateral stimuli, respectively. This contralateral preponderance was expectable according to previous studies (e.g., [11,12,13]).

The present results suggest a degree of right-hemispheric specialization in the processing of auditory space as previously observed in [6] where spatial stimuli were produced using HRTF-filtered noise. The current results demonstrate that this right hemispheric preponderance in the N1m remained when spatial stimuli consisted of different kinds of periodic sounds (the vowel /a/ and the pseudo-vowel). Previous studies reveal that there is no significant difference in the N1m amplitude between the hemispheres in processing of vowels [3]. Hence, vowels presented in different spatial locations seem to exhibit hemispheric asymmetry in the N1m amplitude. In addition, the right hemisphere appeared to be more sensitive to changes in the direction angle, with the variation of the amplitude of the N1m across direction angle being almost twice as large in the right hemisphere than in the left.

The current study shows, interestingly, that N1m behaviour was suprisingly similar for two periodic stimuli (the vowel /a/ and the pseudo-vowel) that were perceptually clearly different. This may be related to the fact that the spectral partials of the pseudo-vowel were seleceted among the strongest harmonics in the vicinity of the formants of /a/. Comparing the N1m amplitude of periodic sounds to the that of noise stimulus it was observed that the N1m amplitude reduced remarkably when using noise as stimuli. Similar reduction of the N1m amplitude was observed in [3] when a periodical vowel generated by a natural glottal excitation was compared to its aperiodic, noise-like counterpart. However, in [3] the periodic and aperiodic stimuli were easily recognizable as the same vowel because they were generated with an equal formant structure, which was not the case in the current experiment.

Previous behavioural studies have shown that natural-sounding spatial auditory environments can be created by using non-individualized HRTFs [15]. This is supported by the behavioural test of our previous study for the [6]. Hence, already the use of non-individualized HRTFs provide realistic enough spatial sound conditions for MEG measurement purposes.

5. ACKNOWLEDGEMENTS

This study was supported by the Academy of Finland (Proj. no. 1168030).

6. REFERENCES

1. Yost, W.A. “The cocktail party problem: Forty years later”, in Gilkey, R.H., and Anderson, T.R. eds., *Binaural hearing in real and virtual environments*, Lawrence Erlbaum associates, Publishers, Mahwah, New Jersey, 329-347, 1997.

2. Blauert, J., *Spatial Hearing*, The MIT Press, Cambridge, 1997.

3. Alku, P., Tiitinen, H., Palomäki, K., and Sivonen, P. ”MEG-measurements of brain activity reveal the link between human speech production and perception”, *Proc. of ICSLP 2000.*

4. Hari, R. “Activation of the human auditory cortex by speech sounds”, *Acta Otolarynglo. Suppl.* 491:132-138, 1991.

5. Bushara, K.O., Weeks, R.A., Ishii, K., Catalan, M.-J., Tian, B., Rauschecker, J.P., and Hallet, M. “Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans”, *Nature neurosci.* 2(8):759-766, 1999.

6. Palomäki, K., Alku, P., Mäkinen, V., May, P., and Tiitinen, H. “Sound localization in the human brain: neuromagnetic observations”, *NeuroReport* 7(11):1535-1538, 2000.

7. Näätänen, R., and Picton, T. “The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure”, *Psychophysiology* 24:375-425, 1987.

8. Wightman, F.L., and Kistler, D.J. “Headphone simulation of freefield listening. I: Stimulus synthesis. II: Psychophysical validation”, *J. Acoust. Soc. Am.* 85:858-878, 1989.

9. Alku, P., Tiitinen, H., and Näätänen, R. “A Method for generating natural sounding speech stimuli for cognitive brain research”, *Clin. Neurophys.*, 110:1329-1333, 1999.

10. Knuutila, J.E.T., Ahonen, A.I., Hämäläinen, M.S., Kajola, M.J., Laine, P., Lounasmaa, O.V., Parkkonen, L.T., Simola, J.T.A., and Tesche C.D. “A 122-channel whole-cortex SQUID system for measuring the brain’s magnetic fields”, *IEEE Trans. on magn.*, 29(6):3315-3320, 1993.

11. Aitkin, L.M., *The Auditory Cortex: Structural and functional bases of auditory perception*, Chapman and Hall, London, 1990.

12. Reite, M., Zimmerman, J.T., and Zimmerman, J.E. ”Magnetic auditory evoked fields: interhemispheric asymmetry”, *Electroencephalogr. and Clin. Neurophysiol.*, 51:388-392, 1981.

13. Pantev, C., Ross, B., Berg, P., Elbert, T., and Rockstroh, B. “Study of the human auditory cortices using a whole-head magnetometer: left vs. right hemisphere and ipsilateral vs. contralateral stimulation”, *Audiol. Neurootol.* 3:183-190, 1998.

14. 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(4):2748-2755, 1995.

15. Wenzel, E.M., Arruda, M., Kistler, D.J., and Wightman, F.L. “Localization using nonindividualized head-related transfer functions”, *J. Acoust. Soc. Am.* 94:111-123, 1993.