The periodic structure of vowel sounds is reflected in human electromagnetic brain responses

Paavo Alku[a],*, Päivi Sivonen[b], Kalle Palomäki[a], Hannu Tiitinen[c,d]

[a]Acoustics Laboratory, Helsinki University of Technology, Otakaari 5 A, Fin-02015 TKK, Finland
[b]Cognitive Brain Research Unit, University of Helsinki, Finland
[c]Department of Psychology, Box 13, Fin-00014 University of Helsinki, Finland
[d]BioMag Laboratory, Medical Engineering Centre, Helsinki University Central Hospital, Box 442, Fin-00029 HUS, Finland

Abstract

Periodicity, which is caused by the vibration of the vocal folds, is an inherent feature of vowel sounds. Whether this periodic structure is reflected in cerebral processing of vowels was addressed via the use of non-invasive brain research methods combined with advanced stimulus production methodology. We removed the contribution of the source of the periodic structure, the glottal excitation produced by the vocal folds, from vowel stimuli and found that electromagnetic responses generated in the auditory cortex reflect this removal. The N1(m) amplitude decreased even though the rest of the acoustical features of the stimuli were identical. Thus, we conclude that speech production mechanisms have significant effects on human brain dynamics as reflected by magnetoencephalography and electroencephalograph. © 2001 Published by Elsevier Science Ireland Ltd.

Whether speech is processed through acoustic and phonetic transformations, or whether voice production mechanisms are crucial in this process has remained unresolved [11]. Several attempts have been made towards solving this issue by using auditory event-related potentials (ERPs) and magnetic fields (EMFs). The most prominent cortical deflection, the N1, has received considerable attention in auditory brain studies [13]. The amplitude, latency and source origin of the N1(m) has been analyzed by using tones [4], noise bursts [8], consonant-vowel-syllables [10,15], and words [7]. However, N1(m) studies of brain activity in humans evoked by sustained vowel sounds (i.e. produced with a constant fundamental frequency) are relatively sparse [3,5].

The lack of experimental data on cerebral processing of sustained vowels can be contrasted with their essential role in speech sounds. From the acoustical point of view, the importance of vowels lies in their long duration and larger energy in comparison to other utterances. Vowels are produced by exciting the vocal tract with the glottal excitation, a periodic waveform generated by the vibrating vocal

folds [6]. By changing the resonant frequencies of the vocal tract, termed the formants, different vowels are produced.

Phonetically, vowels form the most important category of voiced sounds and they have a crucial role in most languages. Voiced sounds are more frequent than voiceless in most West European languages: in English, for example, 78% of phonemes are voiced [2]. From the physiological point of view, formants are determined primarily by the position of the tongue, the jaw and the lips. By using the lowest two resonances, the first (F1) and the second formant (F2), it is possible to express vowels in the F1–F2-space [14]. The F1–F2 space is a widely used tool for quantifying vowel sounds. However, characterizing vowels by using their formant information alone might be an oversimplification as it ignores the glottal excitation, which has a crucial role in the production of vowels [6].

Previous brain research has turned out to be of limited value in showing that the N1(m) is passive in reflecting differences in the processing of vowels with different formant patterns. Despite repeated demonstrations of the dependence of N1(m) latency on stimulus complexity [3,5,17] neither the amplitude nor the latency of N1(m) differentiates whether we hear, for example, the vowel /a/ or /o/. In the present study, our goal was to find out whether

* Corresponding author. Tel.: +358-9-451-5680; fax: +358-9-460-224.
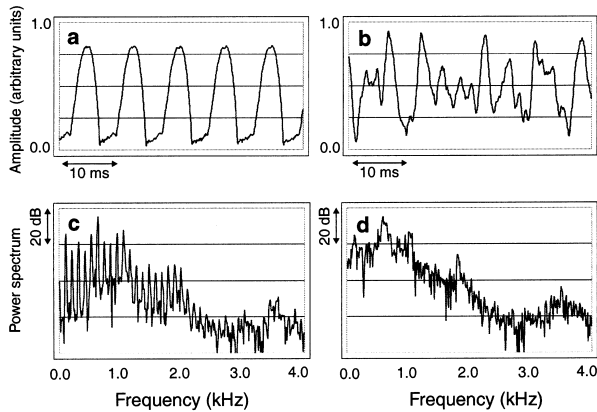E-mail address: paavo.alku@hut.fi (P. Alku).

Fig. 1. The stimuli used in the experiment. (a) Periodic glottal excitation (denoted by $g(n)$) occurring in natural production of vowels, (b) aperiodic excitation waveform ($n(n)$), (c) the spectrum of the vowel /a/ excited by the periodic glottal waveform (/a/$_g$), and (d) the spectrum of the vowel /a/ excited by the aperiodic waveform (/a/$_n$). Excitation waveform $g(n)$ has a periodic structure due to the fluctuation of the vocal folds that occurs in natural production of voiced speech. This periodicity is absent from waveform $n(n)$. The spectra of the two /a/-vowels have equal spectral envelopes and their formants are in the same positions (i.e. /a/$_g$ and /a/$_n$ are equal when expressed in the F1–F2 space). The periodicity of $g(n)$ causes the spectrum of /a/$_g$ to have a comb structure which is absent from the spectrum of /a/$_n$.

two vowel sounds of an equal formant structure, both easily recognizable as the same phoneme, would result in differences in brain activity when the natural glottal waveform with a periodic structure is replaced by an ecologically invalid aperiodic counterpart. Possible changes in the N1(m) behaviour would, consequently, shed light on the role of speech production mechanisms in the cortical processes underlying human speech perception.

The vowel stimuli were generated by using semi-synthetic speech generation (SSG), which produces synthetic vowels from natural glottal excitation in conjunction with an artificial vocal tract model [1]. The stimuli consisted of two representations of the vowel /a/ and /o/, synthesized with SSG using glottal excitations of different characteristics. Firstly, a glottal waveform ($g(n)$, where n denotes the time variable) was extracted from real speech. The waveform obtained corresponds to the natural glottal excitation in the production of vowels due to the vibrating vocal folds. Secondly, an aperiodic glottal excitation, denoted by $n(n)$, was produced using random noise. This waveform served as an excitation that differed remarkably from $g(n)$. In particular, the periodic structure, which is a characteristic feature of natural glottal waveform, is absent from $n(n)$. (In speech communication, replacing $g(n)$ by $n(n)$ corresponds to changing normal phonation to whisper or to changing voice production from healthy to disordered). The spectral envelopes of $g(n)$ and $n(n)$ were equalized by filtering the random sequence through a 6th-order linear predictive coding (LPC)-filter [16] that models the power spectrum of $g(n)$.

A built-in replication was included by synthesizing two Finnish vowels /a/ and /o/. The vocal tract filtering effects were modelled with digital filters the transfer functions of which are denoted by $V_a(z)$ and $V_o(z)$ for the vowel /a/ and /o/, respectively. Four representations of vowel stimuli were synthesized by using the periodic and aperiodic excitation waveforms as an input to both $V_a(z)$ and $V_o(z)$. The speech sounds produced by using periodic $g(n)$ as an input to $V_a(z)$ and $V_o(z)$ are denoted by /a/$_g$ and /o/$_g$, respectively. Notations /a/$_n$ and /o/$_n$ denote sounds generated by using aperiodic $n(n)$ as an excitation to $V_a(z)$ and $V_o(z)$, respectively. Energies of the four stimuli were normalized and, consequently, their intensities were equalized sound pressure level (SPL) with A-weighting at the output of the sound delivery system = 70 dB). Since the stimuli were of different spectral characteristics, we also measured psychoacoustically the loudness level of the stimuli using a loudness model [12]. It was found that the intensity normalization corresponded well with the results obtained using the loudness model: the difference in the loudness level between the periodic and aperiodic vowel was only 1.3 phons and 1.5 phons for the vowels /a/ and /o/, respectively.

Fig. 1 demonstrates that the periodic structure is present in the naturally occurring glottal waveform (Fig. 1a) but absent from its unnatural, aperiodic counterpart (Fig. 1b). The spectra of the vowel /a/ excited using periodic and aperiodic excitations are shown in Fig. 1c,d, respectively. While the formants of the two /a/-vowels are in equal positions, the spectrum of /a/ produced by the periodic glottal waveform, however, comprises a comb structure which results from the regular oscillation of the vocal folds. As can be observed in Fig. 1d, this comb structure is absent from the spectrum of the /a/-vowel produced using the aperiodic glottal excitation. Importantly, the application of the classical F1–F2 space approach would not reveal any differences between the /a/-vowels produced by the periodic and aperiodic excitations because these two /a/-vowels have exactly the same formant values.

Fourteen right-handed, normal-hearing volunteers participated in the experiment. The subjects (12 female; mean age 24 years) provided informed consent and the experiment was approved by the Ethical Committee of Helsinki University Central Hospital. The EMFs elicited by auditory stimuli were recorded with a 122-channel whole-head magnetometer [9]. The stimuli (duration 200 ms, including 10-ms rise- and fall-times) were binaurally delivered through plastic tubes using a constant (onset-to-onset) interstimulus interval of 1500 ms. Electrodes monitoring both horizontal electro-oculogram (HEOG) and vertical electro-oculogram (VEOG) eye movements were used in removing artefacts, defined as activity in excess of $\pm 150 \, \mu V$. Simultaneous electroencephalograph (EEG)-recordings referenced to the subject's nose, were performed at electrode locations Fz, Cz, Pz, F3, F4 and the left mastoid (LM). The responses obtained via magnetoencephalography (MEG) were quantified at N1(m) response maxima with
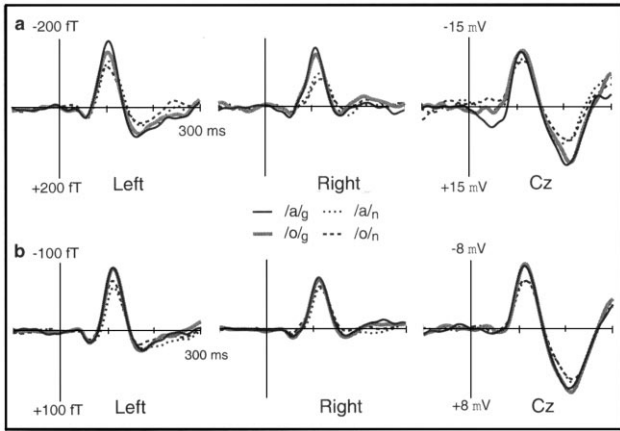
Fig. 2. The responses obtained in the experiment. (a) Magnetic and electric single-subject and (b) grand-averaged responses elicited by periodic (/a/$_g$ and /o/$_g$; thick and thin lines, respectively) and aperiodic (/a/$_n$ and /o/$_n$; dashed and dotted lines, respectively) stimulation, displayed at sensors where the most prominent activity was observed. Both periodic and aperiodic stimuli elicited N1(m) responses peaking at 100 ms. In all the conditions, the N1(m) and the N1 amplitudes attenuated when the contribution of the natural periodic glottal excitation was removed.

the use of unrestricted equivalent current dipoles (ECDs). A subset of 34 channels over either the left or right temporal brain areas was separately used in the ECD fitting. Statistical analyses were performed using a repeated measures ANOVA and Newman–Keuls post-tests.

Fig. 2 shows EMFs and ERPs of a single subject (top) and those grand-averaged across 14 subjects (bottom). Both the natural periodic stimuli (/a/$_g$ and (/o/$_g$) and their aperiodic counterparts (/a/$_n$ and /o/$_n$) elicited prominent N1(m) responses. Importantly, changes in the excitation type (i.e. periodic vs. aperiodic stimulation) had a profound effect on

the N1(m) amplitude: the magnetic N1(m) amplitudes were larger for the vowels excited by the natural periodic glottal pulseform than for those generated by their aperiodic counterparts. This response behaviour was replicated by the ERPs, with the N1 amplitudes being always larger for periodic than for aperiodic stimulation. Statistical analyses (ANOVA, excitation type × hemisphere) revealed that this amplitude difference was significant for both the vowel /a/ ($F(1,11) = 19.30$, $P < 0.01$) and /o/ ($F(1,13) = 17.50$, $P < 0.01$). Post-hoc analyses showed that the N1(m) amplitudes were always larger for periodic than for aperiodic stimulation (/a/$_g = 40.6$ nAm and /a/$_n = 28.6$ nAm, Newman–Keuls $P < 0.01$; /o/$_g = 39.0$ nAm and /o/$_n = 30.4$ nAm; $P < 0.01$). In the electric responses, the N1 amplitudes were also different as a function of the excitation type (for /a/$_g$ vs. /a/$_n$: $F(1,13) = 8.53$, $P < 0.05$; for /o/$_g$ vs. /o/$_n$: $F(1,13) = 7.00$, $P < 0.05$). The N1 amplitudes were always larger for periodic stimulation (both /a/$_g$ and /o/$_g = -8.0$ μV) than for a periodic stimulation (/a/$_n = -6.3$ μV; /o/$_n = -6.1$ μV).

ECD modelling revealed that the N1(m) was generated in the auditory cortex and, corroborating previous observations [3,17], the ECDs of the N1(m) in the right hemisphere were anterior to those in the left hemisphere. The excitation type also caused a small, albeit significant, location difference (Fig. 3). The ECDs of the N1(m) responses elicited by the periodic stimuli were anterior than those for aperiodic stimuli (/a/$_g$ vs. /a/$_n$: $F(1,11) = 9.84$, $P < 0.01$; /o/$_g$ vs. /o/$_n$: $F(1,13) = 18.97$, $P < 0.001$), this anterior–posterior location difference was 3.6 and 1.6 mm between /a/$_g$ and /a/$_n$, ($P < 0.01$) and 2.8 and 2.3 mm between /o/$_g$ and /o/$_n$ ($P < 0.001$) in the left and right hemispheres, respectively.

Our observations are summarized in Fig. 4, which shows the response amplitude as a function of stimulus periodicity. In our understanding, the present observations constitute the first demonstration that cortical processing of speech at the fundamental level of sustained vowels has a clear corre-
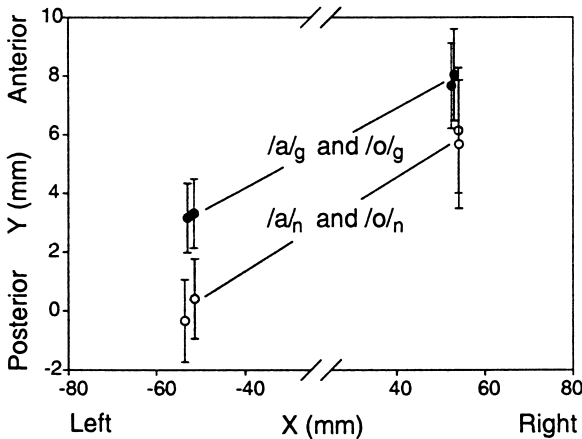

Fig. 3. Mean dipole locations of the N1(m) elicited by periodic (black circles; bars indicate the standard error of the mean) and aperiodic (white circles) stimulation. A slight anterior–posterior shift of approximately 2 mm (with the periodic stimuli resulting in a more anterior ECD) was observed in both the left and right hemispheres.
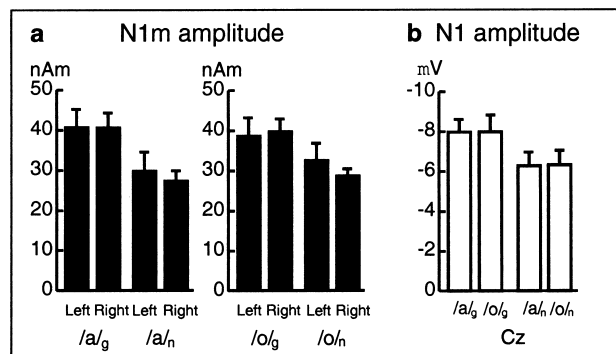

Fig. 4. Summary of the amplitude behaviour of (a) the magnetic N1(m) and (b) the electric N1. The mean amplitude of the N1(m) (black bars) and the N1 (white bars) elicited by periodic (/a/$_g$ and /o/$_g$) and aperiodic (/a/$_n$ and /o/$_n$) stimulation. The removal of the contribution of periodic glottal excitation occurring in natural production of voiced speech significantly attenuated the amplitude of both the N1(m) and the N1.

spondence with the production of these sounds in the human larynx: replacing the unnatural aperiodic excitation of the vowel with the periodic glottal waveform extracted from natural speech increased brain activity indexed by the N1(m) amplitude. This occurred even though the formant frequencies, the intensity and the duration of the utterance were kept constant. That is, sounds with equal positions in the F1–F2 space can nevertheless lead to different levels of cortical activation. These observations demonstrate, firstly, that speech production mechanisms have profound effects on human brain dynamics as reflected by MEG and EEG. Secondly, they show that characterization of vowels by only their positions in the F1–F2 space and ignoring the role of the glottal excitation is an oversimplification for understanding speech perception.

The present results can be related to previous studies on the cerebral processing of different vowels [5] as well as to studies comparing the processing of sinusoidal and vowel stimuli [3,10,17]. These studies have repeatedly demonstrated that the N1(m) amplitude shows no differences between the processing of different sustained vowels or between the processing of tones and vowels. It is particularly striking that tones and vowels that are perceptually very different can not be distinguished in terms of the N1(m) amplitude [3,17]. However, in the present study, a clearly smaller perceptual difference between instances of the same phoneme resulted in a significant decrease in the N1(m) amplitude when the periodic structure of the vowel was violated. Thus, we conclude that the N1(m) amplitude seems to index a much more subtle difference, the presence of natural glottal excitation, which, in turn, relates our findings to the motor theory of speech perception [11]. This previously overlooked link between the production and perception of speech can be found already at the level of sustained vowel sound processing.

[1] Alku, P., Tiitinen, H. and Näätänen, R., A method for generating natural-sounding speech stimuli for cognitive brain research, Clin. Neurophysiol., 110 (1999) 1329–1333.

[2] Catford, J.C., Fundamental Problems in Phonetics, Edinburgh University Press, Edinburgh, 1977, pp. 1–107.

[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., Magnetic auditory responses from the human brain, Scand. Audiol., 9 (1980) 185–190.

[5] Eulitz, C., Diesch, E., Pantev, C., Hampson, C. and Elbert, T., Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli, J. Neurosci., 15 (1995) 2748–2755.

[6] Fant, G., Acoustic Theory of Speech Production, Mouton, The Hague, 1970, pp. 265–280.

[7] Hari, R., Activation of the human auditory cortex by speech sounds, Acta Otolaryngol., Suppl. 491 (1991) 132–138.

[8] Hari, R., Pelizzone, M., Mäkelä, J.P., Hällström, J., Leinonen, L. and Lounasmaa, O.V., Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts, Audiology, 26 (1987) 31–43.

[9] Knuutila, J., Ahonen, A., Hämäläinen, M., Kajola, M., Laine, P., Lounasmaa, O.V., Parkkonen, L., Simola, J. and Tesche, C., A 122-channel whole-cortex SQUID system for measuring the brain's magnetic fields, IEEE Trans. Magn., 29 (1993) 3315–3320.

[10] 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.

[11] Liberman, A.M. and Mattingly, I.G., The motor theory of speech perception revised, Cognition, 21 (1985) 1–36.

[12] Moore, B.C. and Glasberg, B.R., A revision of Zwicker's loudness model, Acta Acustica, 82 (1996) 335–345.

[13] 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 (1987) 375–425.

[14] Peterson, G.E. and Barney, H.L., Control methods used in a study of the vowels, J. Acoust. Soc. Am., 24 (1952) 175–184.

[15] Poeppel, D., Yellin, E., Phillips, C., Roberts, T., Rowley, H., Wexler, K. and Marantz, A., Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds, Cogn. Brain Res., 4 (1996) 231–242.

[16] Rabiner, L.R. and Schafer, R.W., Digital Processing of Speech Signals, Prentice Hall, Englewood Cliffs, 1978, pp. 396–404.

[17] Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J. and Näätänen, R., Electromagnetic recordings reveal latency differences in speech and tone processing in humans, Cogn. Brain Res., 8 (1999) 355–363.