Periodic Glottal Excitation and Formant Frequencies in the Perception of Vowels

Tiitinen, H\textsuperscript{1,2}, Mäkelä, AM\textsuperscript{1,2}, Mäkinen, V\textsuperscript{1,2}, May, PJC\textsuperscript{1,2}, Alku, P\textsuperscript{3}

\textsuperscript{1}Apperception & Cortical Dynamics (ACD), Department of Psychology, University of Helsinki, Finland; \textsuperscript{2}BioMag Laboratory, Engineering Centre, Helsinki University Central Hospital, Finland; \textsuperscript{3}Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland

Corresponding Author: Hannu Tiitinen, Apperception & Cortical Dynamics (ACD), Department of Psychology, P.O. Box 9, FIN-00014 University of Helsinki, Helsinki, Finland
Phone: +358–9–19129482; Email: hannu.tiitinen@helsinki.fi

ABSTRACT

Voiced speech is created by the fluctuating vocal folds generating the glottal pulseform. This excitation signal is the source of the speech fundamental frequency and its harmonic integer multiples. Periodic glottal excitation is required for the elicitation of speech-specific cortical processes indexed by the auditory N100m response. Here, we studied the cortical processing underlying the perception of the vowels /a/ and /u/ produced using normal and aperiodic phonation. The behavior of the N100m, registered with magnetoencephalography (MEG), was studied in 10 subjects. The amplitude and latency of the N100m as well as the center of gravity of the activated cortical areas varied as a function of stimulus periodicity. Further, the presence of glottal excitation had differential effects on the latency of the N100m elicited by the vowels /a/ and /u/. Thus, changes affecting the perceptual quality of speech signals without changing their phonetic content modify the dynamics of human auditory cortex.

KEY WORDS

Human auditory cortex, Speech perception, Fundamental frequency, Vowels, MEG, N100m, Magnetoencephalography.

INTRODUCTION

The acoustic content of speech is determined by the structure and functioning of the human voice production mechanism. The length of the vocal folds is the primary physiological variable underlying adjustment of the speech fundamental frequency (F0) and, consequently, its harmonic integer multiples (2 \cdot F0, 3 \cdot F0 etc.). The shape and the length of the vocal tract, then again, determine the locations of the formant frequencies (F1, F2, etc.). Furthermore, as the length of the vocal folds is proportional to the length of the vocal tract, F0 and the positions of the formant frequencies are also proportional: long vocal folds, typical for males, generally imply low formant frequencies.

By using synthetic stimulation it has been shown that the vowel /u/, which has relatively low F1 and F2 values (310 & 870 Hz), activates the auditory cortex at a longer latency than the vowel /a/, which has higher F1 and F2 values (approx. 710 & 1100Hz) [Poeppel, 1997]. This latency difference is similar to that obtained with pure tones and tone complexes [Roberts, 1996] [Ragot, 1996] and could be explained by the human auditory cortex responding to frequencies in the 500 - 4000 Hz range at a relatively invariant latency while stimuli at lower frequencies result in delayed responses. However, in our recent study we used the vowels /a/ and /u/ with natural periodic glottal excitation and observed no such latency differences [Mäkelä, 2003]. In a related study, we found that the vowel /a/ comprising a natural periodic glottal excitation structure elicited significantly larger N100m responses than the same phoneme with an aperiodic structure [Alku, 2001]. Furthermore, the violation of the glottal periodicity was accompanied with the cortical activation shifting to a more posterior location in both the right and the left hemisphere.
Moreover, we recently found that the vowel /a/ and a pseudo-vowel lacking in phonetic content elicited, both with natural male and female F0s, shorter-latency N100m responses than pure tones of corresponding frequencies [Mäkelä, 2002]. These observations have led us to propose that the presence of the periodic glottal excitation in the stimulus is a prerequisite for the elicitation of speech-specific cortical activity.

Here, we report our preliminary results of a study aimed at clarifying the effects of the periodic glottal excitation and the F1 and F2 formant frequencies in the cortical processing of two Finnish vowels, /a/ and /u/. We presented the subjects with two variants of each vowel, one with a periodic structure generated by a natural glottal excitation and the other with the aperiodic structure produced with a noise excitation.

METHODS

Ten right-handed, normal-hearing subjects (age 20-44 years, 5 females) participated in the study with informed consent. The subjects were instructed not to pay attention to auditory stimuli and to concentrate on reading a self-selected book or to watch a silent video. The stimuli were created by using the Semi-synthetic Speech Generation method [Alku, 1999], where the vowels /a/ and /u/ were synthesized using a natural, periodic glottal excitation produced by the vibrating vocal folds (F0 = 115 Hz). The lowest four formant frequencies of the vowels were adjusted as follows: /a/: F1 = 670 Hz, F2 = 1000 Hz, F3 = 1950 Hz and F4 = 3440 Hz, /u/: F1 = 330 Hz, F2 = 580 Hz, F3 = 1900 Hz and F4 = 2900 Hz. The aperiodic counterparts of the vowels were produced by replacing the natural glottal excitation with a noise sequence whose spectral envelope was matched to that of the glottal excitation. Sound intensity (computed as the root mean square of the digital time-domain signals) was equalized across the stimuli, and, for each subject, the sound pressure level was adjusted to a comfortable level by using the vowel /a/ as a reference stimulus (intensity range 70-75 dB SPL). The 200-ms stimuli were binaurally delivered to the subject through plastic tubes and ear pieces at an interstimulus interval of 800 ms. The stimuli (N > 150) were presented in a pseudorandom order, counterbalanced across subjects.

Cortical activation elicited by the stimuli was registered by using a 306-channel MEG measurement device (Vectorview, Elekta Neuromag Oy, Finland) in a magnetically shielded room. The evoked responses were averaged over a period of 600 ms including a 100-ms pre-stimulus baseline and 1-30 Hz passband filtered. Epochs exceeding an absolute amplitude variation of 3000 fT/cm were excluded online, and electrodes monitoring horizontal and vertical eye movements were used in removing artefacts (>150 µV). The N100m was studied for effects in amplitude, latency, and source location. The waveform analysis was carried out for the field gradient vector sums from three pairs of planar gradiometers displaying maximum N100m responses, separately in each hemisphere and for each subject. Equivalent current dipoles (ECD) were fitted to a time point defined as the moment of the N100m reaching its peak amplitude in the waveform calculated over the 66 sensors located over either the left or right temporal brain areas. Statistical analyses were performed by using repeated measures-ANOVA and Newman-Keuls post hoc tests.

RESULTS

Fig. 1 shows the results. The analysis of the N100m waveforms revealed that glottal periodicity in stimulation had distinctly different effects on the latency of the N100m elicited by the vowels /a/ and /u/ (F[3,27] = 6.06, P < 0.01). The human auditory cortex responded to the periodic /a/ and its aperiodic counterpart at the same latency (124 & 123 ms; P = n.s.). However, the latency of the N100m elicited by the periodic vowel /a/ was considerably shorter than the latency of the N100m elicited by its aperiodic counterpart (128 & 137 ms; P < 0.05). The N100m latencies for the periodic vowels /a/ and /u/ were
similar (124 & 128 ms, \( P = \text{n.s.} \)), while the latency of the N100m was significantly shorter for the aperiodic /a/ than for the aperiodic /u/ (123 & 137 ms; \( P < 0.01 \)).

The right-hemispheric N100m responses were more prominent than the left-hemispheric responses (37 & 22 fT/cm; \( F[1,9] = 11.10, P < 0.01 \)). Moreover, the spectral content of the stimulus was reflected in the amplitude of the N100m: The periodic vowels /a/ and /u/ elicited larger N100m responses than their aperiodic counterparts (39 & 27 fT/cm for the periodic and aperiodic /a/, \( P < 0.01 \); 31 & 22 fT/cm for the periodic and aperiodic /u/, \( P < 0.01 \)), with also the amplitude difference for the periodic /a/ and /u/ and the aperiodic /a/ and /u/ being statistically significant (\( P < 0.01 \) & \( P < 0.05 \), respectively).

The source localization results for the N100m were similar to those reported in [Alku, 2001] [Mäkelä, 2003] with the sources of the N100m elicited by the periodic stimuli being anterior to those elicited by the aperiodic stimuli.

**DISCUSSION**

Here we studied the cortical processing of periodic and aperiodic vowels /a/ and /u/, with the former having higher F1 and F2 values than the latter. We found that replacing the periodic glottal excitation with an aperiodic excitation had distinctly different effects on the temporal dynamics of cortical activation elicited by these vowels: while /a/ elicited the N100m at an invariant latency, aperiodic excitation in /u/ led to a considerably delayed N100m. Furthermore, for both /a/ and /u/, the presence of natural glottal excitation was reflected in the amplitude and location of the ECDs for the N100m, with the vowels comprising harmonic structure resulting in larger and more anterior responses than their aperiodic counterparts.

The present results corroborate previous observations showing that low-frequency pure tones (<500 Hz) [Roberts, 1996] [Mäkelä, 2002] elicit delayed cortical activation as reflected by the latency of the N100m: Here the aperiodic vowel /u/ with relatively low F1 and F2 values (330 & 580 Hz) activated the human auditory cortex considerably later than the aperiodic vowel /a/ with high F1 and F2 frequencies.
(670 & 1000 Hz). However, the periodic vowels /a/ and /u/ activated the auditory cortex at an identical latency, which appears to provide reasons for attributing the N100m latency effects to glottal excitation.

The above observations can be explained in terms of the acoustical properties of the stimuli: For a low-pitched tone to be perceived as equally loud as a high-pitched one, the intensity of the low-pitched tone has to be amplified. Conceivably, in human speech production and perception, this amplification is provided by the integer multiples of the F0, which increase the sound energy at frequencies above the F0. The human perceptual system might be tuned to the simultaneous presence of the F0 and the F1 and F2 frequencies, and, consequently, this tuning might lead to delayed cortical activity when the F0 information is removed. While emphasizing the link between speech production and perception [Liberman, 1989], the present results indicate that caution is advocated if results gained with other than natural speech stimulation are used for drawing inferences about the brain dynamics of speech perception, as brain responses are highly sensitive to minute variations in the stimulus material and the production methods used.

ACKNOWLEDGEMENTS

This study was supported by the Academy of Finland (proj. no. 1201602 & 205962).

REFERENCES


