

Temporal Encoding of Voice Onset Time in Young Children

Aseel Almeqbel*

Department of Hearing and Speech Sciences, Faculty of Allied Health Sciences, Health Sciences Center, Kuwait University, Kuwait

*Corresponding author: Aseel Almeqbel, Department of Hearing and Speech Sciences, Faculty of Allied Health Sciences, Health Sciences Center, Kuwait University, P.O. Box 31470, 90805, Sulaibikat, Kuwait, Tel: 0096524633412; E-mail: aseel.m@hsc.edu.kw

Received date: June 06, 2016; Accepted date: June 20, 2016; Published date: June 27, 2016

Copyright: © 2016 Almeqbel A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Objective: Voice onset time (VOT) is an important parameter of speech that denotes the time interval between consonant onset and the onset of low frequency periodicity generated. In this study we examined the temporal encoding of voice onset time in the cortical auditory evoked response (CAEP) of young children.

Methods: Scalp recorded CAEP were measured in 18 children aged from 5-8 participated (n=18). The N2 latency was evoked using differences in voice-onset-times (VOTs) using stop consonant-vowel syllables /ga/-/ka/.

Results: A significant and systematic shift in the N2 latency was observed for differences in VOT.

Conclusions: Our results demonstrate that temporal encoding of VOT exists in the developing cortical evoked response.

Keywords: Temporal processing; Cortical auditory evoked potentials (CAEP); Voice-Onset-Time (VOT); N2 latency

Introduction

Voice onset time (VOT) is an important temporal cue for speech perception. This important speech parameter signifies the interval between consonant release (onset) and the start of rhythmic vocal-cord vibrations (voicing) [1]. In English, the VOT for voiced stop consonants such as /ga/ is short, voiceless stop consonants such as /ka/ have a much longer VOT duration. For speech perception, rapidly changing acoustic cues such as VOT are used to discriminate stop consonants while steady-state cues such as formant transition provide information on vowel discrimination [2].

In individuals with normal hearing the auditory cortical area plays a major role in encoding temporal acoustic cues of the speech signal such as VOT [3], which are essential to language processing. That means when recognising words or sentences, listeners make use of a temporal cue such as VOT to differentiate between stop consonants that differ in onset time. Specifically, VOT is partly represented in the primary and secondary auditory cortical field by synchronized activity that is time-locked to consonant release and voicing onset [4].

Understanding how voicing information is encoded at the cortical level is a fundamental goal of auditory neurophysiology field. The research interest in cortical auditory evoked potentials (CAEP) has expanded during the last two decades as a result of increasing understanding of the underlying neuronal encoding of speech sounds at the cortical level in children and infants.

Cortical responses to consonant-vowel (CV) syllables with different VOTs have been well studied in listeners with normal-hearing (NH) [4-7]. Some investigators have shown that N1 latency prolongs with the duration of VOT and that a second N1 peak emerges that follows the voicing onset time.

However, In the absence of a consistent N1 component of the CAEPs in children under age 9 years, it is likely that another component of the CAEP complex, such as the P1 or N2, functions as a cortical marker of VOT in children. Possibly is that the developing nervous system may have a completely different cortical representation of VOT, which is not reflected in the CAEP. For this reason, we recorded CAEPs in response to stimuli differing along a /ga/-/ka/ VOT in children aged 5-8 years. Our goal was to examine whether N2 components of the CAEP would reflect changes in VOT in a manner similar to the adult N1, Therefore, we used N2, a negativity following the P1 at about 200 or 250 ms as a function of VOT in young children [8,9]

Method

Participants

18 children (7 males, 11 females) aged 5-8 years (mean 6.4, SD 1.5) were recruited. All had pure-tone air conduction thresholds \leq 15 dB HL at octave frequencies from 250 Hz to 8 kHz with type (A) tympanograms. They had no history of hearing or speech problems, or significant noise exposure, and no reported previous history of reading or learning problems.

Air conduction and bone conduction pure-tone thresholds were determined using a calibrated clinical audiometer (AC33 Interacoustics 2 channel Audiometer) using a modified version of the Hughson and Westlake procedure [10]. Tympanometry was undertaken using a calibrated immittance audiometer (GSI-Tympstar V2, calibrated as per ANSI, 1987) using a standard 226 Hz probe tone.

Stimuli

Two naturally produced stop consonant-vowel-syllables /ga/-/ka/ were recorded. These speech stimuli differ in VOT, for example the

VOT for /ga/=0.05 ms and for /ka/=0.90 ms. Speech stimuli /ga/ and /ka/ were recorded using an AKG C535 condenser microphone connected to a Mackie sound mixer, with the microphone positioned 150 mm in front and at 45 degrees to the speaker's mouth. This position avoids air turbulence and popping. The mixer output was connected via an M-Audio Delta 66 USB sound device to a Windows computer running Cool Edit audio recording software and captured at 44.1 KHz 16 bit wave format. All speech stimuli were collected in a single session to maintain consistency of voice quality. Speech stimuli were modified after selection and recording using Cool Edit 2000 software. All speech stimuli of 200 ms duration were ramped with 20 ms rise and fall time to prevent any audible click arising from the rapid onset or offset of the waveform. The inter-stimulus interval (ISI), calculated from the onset of the preceding stimulus to the onset of the next stimulus was 1207 ms, as it has been shown that a slower stimulation rate results in more robust CAEP waveforms in immature auditory nervous systems [8].

Stimulus presentation

All stimuli used in the electrophysiology procedure were presented at 65 dB SPL (as measured at the participant's head), which approximates normal conversational level. All sound levels presented in the current study were calibrated using Bruel and Kjaer SLM 2. Presentation was via a loudspeaker speaker placed at 1 m from the participant's seat at 0 azimuth.

Electrophysiological procedure

Participants sat on a comfortable chair in a quiet electrophysiology clinic and watched a DVD of their own choice. The volume was muted but subtitles were activated to ensure that participants would be engaged with the movie rather than attending to the stimuli. All participants were instructed to remain relaxed and awake, and not to pay attention to the sounds being presented.

Data acquisition

A NeuroScan and 32-channel NeuAmps evoked potential system was used for evoked potential recording. All sounds were presented using the Neuroscan STIM 2 stimulus presentation system.

Recording

Evoked potentials were recorded in continuous mode (gain 500, filter 0.1–100 Hz) and converted using an analog-to-digital sampling rate of 1000 Hz using scan (version 4.3) via gold electrodes placed at C3, C4 and Cz, with reference electrode A2 on the right mastoid bone, and ground on the contralateral ear on the mastoid bone. Electrode impedances were maintained below 5 kΩ. Children participated in a single 1-hour recording session, including the electrode application and CAEP recording. None of the participants showed signs of fatigue during the testing.

Off-line data analysis

EEG files with a -100 to 500 ms time window were obtained from the continuous file relative to the stimulus onset. Any responses on scalp electrodes exceeding $\pm 100 \mu\text{V}$ were rejected. Prior to averaging, EEG files were baseline corrected using a pre-stimulus period (-100 ms), in order to establish a baseline to correct for the DC level of background EEG activation. Averaging was digitally band pass filtered

from 1 to 30 Hz (zeroshift, 12 dB/octave) to smooth the waves for the final figures.

Peak latencies were detected based on the recordings from the Cz electrode site. Individual participants had two averaged CAEP waveforms (n=100), resulting in a grand average waveform of 200 sweeps for each stimulus presentation. The N2 latency in the participants was measured as the immediate negative component occurring after the P1, usually present around 200 to 250 ms after stimulus presentation. For each participant, the individual grand average waveform was computed, visually identified and subjected to suitable statistical analyses using SPSS (version 21) to investigate the aims of the current study.

Results

All participants had pure-tone air conduction thresholds ≤ 15 dB HL at octave frequencies from 250 Hz to 8 kHz with type (A) tympanograms which consistent with normal middle ear function. All children had no history of hearing or speech problems, or significant noise exposure, and no reported previous history of reading or learning problems.

Results from CAEP show that a robust N2 peak for both stimuli. The Wilcoxon Signed-Rank test was performed to compare the mean N2 latency of the stop CV voiced /ga/ vs. voiceless /ka/. Also referred to as the Wilcoxon matched pairs signed rank test, this is the non-parametric alternative to the repeated measures t-test. Results show a significant mean difference in N2 latency between the two conditions ($Z=-2.811$, $p<0.05$). There was an early N2 latency for the stop CV voiced /ga/ and later N2 latency for the stop CV voiceless /ka/ as shown in Figure 1. For example, the mean N2 latency for /ga/ was 202 ms (± 4.14 ms) and the mean N1 latency for /ka/ was 209 ms (± 5.81 ms).

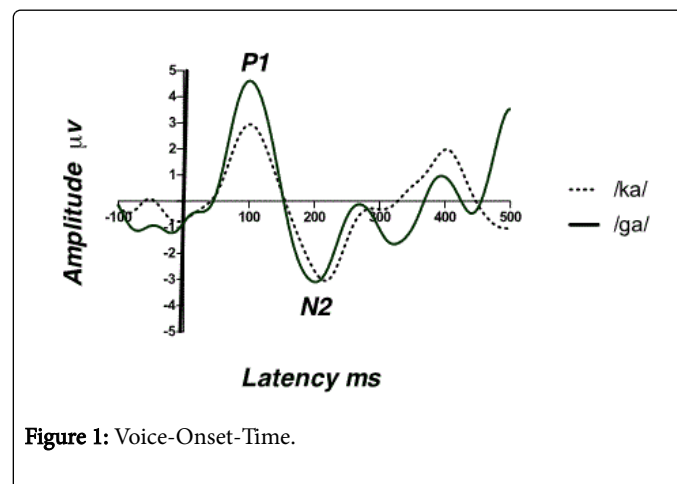


Figure 1: Voice-Onset-Time.

Discussion

The main aim of this study was to determine whether N2 components of the CAEP in children would reflect changes in VOT in a manner similar to the adult N1. Overall, the results from this study indicate that a robust N2 peak for both speech stimuli in children. Specifically, significant and systematic differences in N2 latency were measurable for varying VOTs. Voice-onset time has previously been measured objectively and subjectively in children and adults [6] as a measure of cortical temporal processing. While previous studies have

used synthesised speech tokens to correlate objective and subjective measures, we have used naturally produced speech tokens to better reflect difficulties that children might have in perceiving speech within a real-world environment. Our results show that the N2 latency for the voiced token /ga/ was significantly shorter than for the voiceless token /ka/. This difference might represent distinct responses that are time-locked to the onset of voicing at the cortical level of the auditory system.

Of critical relevance to the present study is that a disruption in sensitivity to the voicing onset cue could lead to a disruption in speech processing, which relies upon identification and/or discrimination of stop consonants and which has in turn been associated with inadequate phonological awareness skills and word reading skills. N2 latency could therefore be used as an objective measure to evaluate the perceptual dysfunction in a disordered paediatric population with poor temporal processing ability.

Conclusion

Our findings suggest that, in school-aged children with normal hearing the N2 component of the CAEP can be evoked in response to voice –onset-time paradigm. N2 latency is sensitive to the temporal cues of the acoustic stimuli such as voice-onset-time, therefore could be used as an objective measure to evaluate the temporal encoding in young children objectively. This finding could have clinical significance in allowing for a non-invasive objective examination of temporal processing ability using VOTs.

References

1. Lisker L, Abramson A (1967) A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384-422.
2. Mirman D, Holt LL, McClelland JL (2004) Categorization and discrimination of nonspeech sounds: differences between steady-state and rapidly-changing acoustic cues. *J Acoust Soc Am* 116: 1198-1207.
3. Roman S, Canévet G, Lorenzi C, Triglia JM, Liégeois-Chauvel C (2004) Voice onset time encoding in patients with left and right cochlear implants. *Neuroreport* 15: 601-605.
4. Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA 3rd (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 82: 2346-2357.
5. Kuruvilla-Mathew A, Purdy SC, Welch D (2015) Cortical encoding of speech acoustics: Effects of noise and amplification. *International Journal of Audiology* 54: 852-864.
6. Sharma A, Dorman MF (1999) Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am* 106: 1078-1083.
7. Sharma A, Marsh CM, Dorman MF (2000) Relationship between N1 evoked potential morphology and the perception of voicing. *J Acoust Soc Am* 108: 3030-3035.
8. Gilley PM, Sharma A, Dorman M, Martin K (2005) Developmental changes in refractoriness of the cortical auditory evoked potential. *Clinical Neurophysiology* 116: 648-657.
9. Sussman E, Steinschneider M, Gumenyuk V, Grushko J, Lawson K (2008) The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hear Res* 236: 61-79.
10. Cahart R, Jerger J (1959) Preferred method for clinical determination of pure tone thresholds. *J Speech Hear Disord* 24: 330-345.