Neuroscience Letters 614 (2016) 119–126

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research paper

Representation of spectro-temporal features of spoken words within the P1-N1-P2 and T-complex of the auditory evoked potentials (AEP)

Monica Wagner[a,*], Arindam Roychoudhury[b], Luca Campanelli[c], Valerie L Shafer[c], Brett Martin[c], Mitchell Steinschneider[d]

[a] St John's University, Queens, NY, USA
[b] Columbia University, New York, NY, USA
[c] The Graduate Center of the City University of New York, New York, NY, USA
[d] Albert Einstein College of Medicine, Bronx, NY, USA

HIGHLIGHTS

- P1-N1-P2 and T-complex reflect spectral and temporal features of natural speech.
- P1-N1-P2 and T-complex are reliable despite multiple productions of natural speech.
- Single trial responses within sensory waveforms are stable between 50 through 600 ms.

ARTICLE INFO

Article history:
Received 26 March 2015
Received in revised form 22 October 2015
Accepted 10 December 2015
Available online 14 December 2015

Keywords:
P1-N1-P2
T-complex
Auditory evoked potentials (AEP)
Spectral processing
Temporal processing
Single trial analysis

ABSTRACT

The purpose of the study was to determine whether P1-N1-P2 and T-complex morphology reflect spectro-temporal features within spoken words that approximate the natural variation of a speaker and whether waveform morphology is reliable at group and individual levels, necessary for probing auditory deficits. The P1-N1-P2 and T-complex to the syllables /pət/ and /sət/ within 70 natural word productions each were examined. EEG was recorded while participants heard nonsense word pairs and performed a syllable identification task to the second word in the pairs. Single trial auditory evoked potentials (AEP) to the first words were analyzed. Results found P1-N1-P2 and T-complex to reflect spectral and temporal feature processing. Also, results identified preliminary benchmarks for single trial response variability for individual subjects for sensory processing between 50 and 600 ms. P1-N1-P2 and T-complex, at least at group level, may serve as phenotypic signatures to identify deficits in spectro-temporal feature recognition and to determine area of deficit, the superior temporal plane or lateral superior temporal gyrus.

1. Introduction

The phoneme sequence that constitutes a spoken word consists of a unique series of time varying spectral and temporal features. Cortical decoding of these phoneme sequences is fundamental to speech perception and subsequent word recognition. A reliable and sensitive index of cortical representation of spectral and temporal features within spoken words would be beneficial for identifying deficits in auditory processing in various patient populations.

Neural ensembles within auditory cortex respond to acoustic features within the spoken word [22,35]. These cortical responses are reflected within the P1-N1-P2 and T-complex waveforms of the auditory evoked potentials (AEP). The P1-N1-P2 can be detected on the scalp surface overlying fronto-central brain regions and the superior temporal plane is the primary generator of activity reflected within the waveform. The T-complex is detected on the scalp surface overlying lateral posterior temporal brain regions and has as its primary generator, posterior superior temporal gyrus [15,25,34,35]. The morphology of these sensory waveforms is modulated by the acoustic characteristics of sound (e.g., frequency, intensity, duration) [21,24]. Thus, the P1-N1-P2 and T-complex to each acoustic feature change within a spoken word overlap and form a unique signature waveform to the word. We use the terms P1-N1-P2 and T-complex to describe cortical sensory responses to the onset (and offset) of acoustic feature changes within the spoken word [19,27].

Neurophysiological studies examining adults and children experiencing speech and language impairment, dyslexia and autism have found atypical P1-N1-P2 and T-complex patterns to speech stimuli [4,7,11,13,31,33]. These aberrant patterns,

* Corresponding author. +1 718 990 2435.
E-mail address: wagnerm@stjohns.edu (M. Wagner).

http://dx.doi.org/10.1016/j.neulet.2015.12.020
0304-3940/

however, have rarely been interpreted relative to spectral and temporal features of the acoustic stimuli used within the experiments. An exception is the work of Giraud and colleagues (2005), who identified atypical temporal feature processing within the P1-N1-P2 complex to the syllables /ba/ and /pa/ in adults with dyslexia relative to a control group. The CV syllables consisted of naturally recorded speech synthetically modified to create endpoints on a /p-b/ voice onset time (VOT) continuum. Adults without impairment showed an initial onset N1/P2 to French pre-voicing for /ba/ followed by a second evoked complex to the burst for /b/. The delay between the first and second complex paralleled the VOT of the stimuli. In four adults with dyslexia, the response to the burst for /b/ (the second N1/P2 complex) was absent within the sensory waveform, suggesting an absence of a synchronous response to the burst for /b/.

Spectral characteristics of tone stimuli have been found to influence the amplitude and latency of the N1 and P2 deflections in the AEP waveform [24,39]. For example, N1 and P2 amplitudes were found to be smaller for mid (i.e., 1500 and 1650 Hz) and high frequency tones (i.e., 3000 and 3300 Hz) relative to low frequency tones (i.e., 400 and 440 Hz) and latencies were shorter for the mid and high frequency tones relative to low tones for N1, but not for P2 [39]. Also, some studies have demonstrated that spectral features within speech affect the N1-P2 response. For example, spectral changes within synthetically created vowel stimuli (i.e., /ui/) evoked an N1-P2 response [20].

Atypical P1-N1-P2 and T-complex waveform patterns found for participant groups with auditory deficits may result from absence of a synchronized neural response to a stimulus characteristic. Alternately, the atypical sensory patterns may be caused by instability in neural responding to the acoustic features of speech [10]. Therefore, an assessment of trial-to-trial stability in sensory responses to naturally spoken words is necessary to provide benchmarks for investigating abnormalities in cortical feature processing in future studies. Trial-to-trial cortical sensory responses show fluctuations caused, in part, by the variation in production of naturally spoken words, external noise within the environment or noise at the cellular and sensory levels of the nervous system [9]. Cortical sensory processing, even to an identical stimulus, shows significant variability within normal subjects [2].

As cortical activation spreads to multiple regions within 30–50 ms of auditory stimulus presentation, with source contributions beyond auditory cortex increasing with time from stimulus onset [34], we explore whether variability within single trial responses to natural speech increases with increasing time from stimulus onset, possibly reflecting multiple source contributions, from both within and beyond auditory cortex. We are unaware of other studies examining response variability in this manner.

Studies investigating the relationship between stimulus characteristics and sensory waveform morphology have used tonal stimuli [14,39], synthetically created or synthetically modified speech [11,20] or single productions of natural speech [8]. The purpose of the current study is to determine whether spectral and temporal characteristics within spoken words that approximate the natural variation from a single speaker can be identified within the P1-N1-P2 and T-complex waveforms as has been previously demonstrated for tones and synthetic speech. We examine whether these characteristics can be identified at group and single subject levels. We compare sensory waveforms to spoken words containing highly contrastive spectro-temporal characteristics to facilitate identification of feature processing within individual participant waveforms. If demonstrated, clinical and research application of the P1-N1-P2 and T-complex for probing auditory processing deficits in patient populations will have ecological validity.

We hypothesize that spectral and temporal characteristics within 70 natural productions of spoken words from a single speaker will be reflected within the P1-N1-P2 and T-complex at group and single subject levels. Further, we hypothesize that variability within single trial cortical responses will increase with increasing time from stimulus onset, reflecting higher level processing of the stimulus and more contributions from multiple generators within the composite AEP waveforms.

2. Method

2.1. Participants

Participants were recruited for two larger projects investigating perception of native and non-native phoneme sequence contrasts in Polish and English listeners; these two projects had slightly different experimental methods. Therefore, data for each group (Groups 1 and 2) of 24 subjects was analyzed separately. Each group of 24 subjects included 12 native-English and 12 native-Polish speakers. These forty-eight participants (16 females in Group 1; 20 females in Group 2) were between 21 and 39 years, demonstrated normal hearing and were without a history of speech, language or cognitive impairment (for detail see Supplementary content Text and Table S1).

This project was conducted in accordance with The Code of Ethics of the World Medical Association, (Declaration of Helsinki) for experiments involving humans. All procedures were approved by the Internal Review Boards of The Graduate Center of the City University of New York and St. John's University. Prior to testing, all subjects provided written and oral consent to participate in the study.

2.2. Stimuli

Participants were presented with same (e.g., "petisa-petisa") and different (e.g., "setisa-stisa") nonsense word pairs. AEP epochs to the first word in the word pairs, which contained /pət/or /sət/ syllable onsets were analyzed. Phoneme sequences that followed the syllable onsets varied and were matched for rhyme. Each word type (e.g., /sət/) consisted of two natural productions of 35 non-words (e.g., "setila") produced by a male speaker. Each of these 70 productions occurred twice for a total of 140 stimuli per word type. Two participants in Group 1 and four participants in Group 2 were presented with additional productions of each word type (maximum 20 productions), presented randomly from the 140 stimuli.

The syllables contained either a fricative /s/ onset (i.e., /sət/) or a stop-consonant /p/ onset (i.e., /pət/), which are highly contrastive in spectral (frequency) and temporal (duration) features [5,30]. Fig. 1 displays a spectrogram and waveform for sample stimuli. Notice that the fricative /s/ consonant has a long duration (mean = 124 ms) and contains a concentration of high frequency energy above 4 kHz. In contrast, the stop-consonant /p/ has a short duration (mean = 26 ms) prior to voicing of the vowel and contains a concentration of low-to-mid frequency energy. The average duration of the syllables /pət/ and /sət/ measured from word onset to the onset of the burst for /t/ was 149 ms (range 123 to 174 ms) and 258 ms (range 210–299 ms), respectively. The average durations of the whole words, containing /sət/ and /pət/ onsets, were 698 ms (range 633–801 ms) and 550 ms (range 481–671 ms), respectively.

Words were presented in the free field in a sound-treated electrically shielded room using E-prime software (version 1.1). Overhead speakers (Realistic Minimus-7) placed approximately 1 meter from the speaker were used for Group 1 and diagonal left and right speakers (Realistic Minimus-7) placed approximately 1.5 meters from the speaker were used for Group 2. Stimuli were equalized for root mean square (rms) amplitude and then played at 55.5 dB SPL (Group 1) or 62.5 dB SPL (Group 2). An inter-stimulus