Published in final edited form as:

Neurosci Lett. 2016 February 12; 614: 119–126. doi:10.1016/j.neulet.2015.12.020.

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, and Mitchell Steinschneider^d

Monica Wagner: wagnerm@stjohns.edu

^aSt John's University, Queens, NY, USA

^bColumbia University, New York, NY, USA

^cThe Graduate Center of the City University of New York, New York, NY, USA

^dAlbert Einstein College of Medicine, Bronx, NY, USA

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.

Keywords

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

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

*Corresponding author. +1 718 990 2435.

Appendix A. Supplementary data: Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.neulet.2015.12.020>

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, 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 interval (ISI) of 250 ms separated words in the word pairs and the inter-trial interval (ITI) was 2 s.

2.3. Procedure

EEG was recorded while participants listened to nonsense word pairs and identified whether the second word in the pairs had two or three syllables. This syllable identification task determined whether participants could distinguish words having /pt/ vs /pət/ or /st/ vs /sət/ onsets. All participants included for this study performed the behavioral task in a manner consistent with their language group [37].

A 65 channel net was used for data collection (Electrical Geodesic Inc.). Fig. 2 shows the Geodesic sensor map of electrodes. The EEG was recorded at a sampling rate of 250 Hz (Group 1) or 500 Hz (Group 2) and bandpass filtered between 0.1 and 30 Hz (Group 1) or between 0.1 and 100 Hz and low pass filtered at 30 Hz using Net station software (version 4.5.1) (Group 2). The EEG was referenced to Cz for data acquisition. BESA Research

(version 5.3.7) was used for processing. Eye blinks and movements were identified and then removed using an adaptive artifact correction function [16]. An artifact scan, using amplitude and signal gradient criteria, identified remaining artifacts. If an excessive number of trials from a channel contained artifacts, data from that channel was interpolated using spherical spline interpolation [28]. Epochs were segmented at 100 ms pre- and 800 ms post-stimulus, data was re-referenced for average referencing and, single trials were exported without averaging for analysis. Finally, using R programming (R core team, 2014), single trial amplitude values for each channel (each trial) were baseline corrected using mean amplitude for the time interval -100–0 ms.

2.4. Statistical analyses

P1-N1-P2 and T-complex for the stimuli used in the study [36] were identified as amplitude values between 50 and 600 ms at fronto-central (FC) and posterior-temporal (PT) electrode sites, respectively. P1-N1-P2 amplitude values were analyzed at FC site 4 and averaged FC sites 4, 5, 55, Cz. T-complex amplitude values were analyzed at left PT site 24 and averaged left and right PT sites 24, 52.

The mean number of accepted single-response trials to the /pət/ and /sət/ onsets for Group 1 were 134 (range 119–153) and 136 (range 122–157), respectively. A mean number of 136 (range for /pət/: 117–154; /sət/: 112–156) accepted single-response trials were included for analyses for both /pət/ and /sət/ for participants in Group 2.

Our previous study [36] examining native-English and native Polish listeners revealed P1-N1-P2 and T-complex morphology in response to native and non-native phoneme sequence contrasts to reflect acoustic, rather than linguistic feature processing. The syllable onsets /pət/ and /sət/ occur in both the English and Polish languages, thus, language is not the focus of the current study. However, prior to combining an equal number of Polish and English listeners within Group 1 and Group 2 for the current study, we confirmed that waveform patterns were highly similar for the language groups (see Supplementary content, Text).

For the current study, reliability of the P1-N1-P2 and T-complex morphology to the syllables was examined using Pearson's product moment correlation coefficient. We determined the association between Group 1 and 2's waveform to /pət/ and /sət/ onsets and the association between individual subjects' waveform patterns and their grand mean patterns. Waveform patterns between 50 and 600 ms were assessed and averaged in 24 ms time intervals (23 data points). Significance for the obtained r values (degree of freedom = 21, critical value for Pearson's $r = .53$) was determined using alpha .01.

A mixed effect linear regression model (MEM) [18] was then used to determine whether apparent differences in the P1-N1-P2 and T-complex morphology to /pət/ and /sət/ onsets at averaged FC and averaged PT sites reflected spectro-temporal characteristics of the stimuli (significant at alpha .01). This model controls for the contribution of individual subject means and, thus, is suitable for the current study in that we are examining the potential use of sensory waveforms for probing auditory deficits in individual subjects. The model examined the fixed effects of (1) word type (/pət/ and /sət/), (2) language group (Polish and English), and (3) the interaction of word type and language group on the P1-N1-P2 and T-

complex amplitude values. All models included random intercept for subject. Amplitudes were averaged in 100 ms intervals between 100 and 600 ms. Approximately 6500 single trials for each group were included for analysis and outliers >99.75 percentile and <.25 percentile of the data were eliminated prior to analysis. The response to /pət/ was the reference and the coefficient estimates for the fixed effects are interpreted as the effect of that variable when all other variables are zero or at reference level (categorical variables). Thus, word type has two levels, /pət/ and /sət/ and with /pət/ as the reference, the coefficient of word type indicates the difference between /pət/ and /sət/ when the other variables are at their reference level.

Lastly, we assessed whether variability of single trial responses increased with increasing time. The change in standard deviation (SD) for each participant's single trial amplitude values between 50 and 600 ms to /sət/ and /pət/ at averaged FC and PT sites was assessed using a mixed effect linear regression analysis. Data were analyzed with R version 3.1.2 (R Core Team, 2014) using the lmer function from the lme4 package, version 1.1–7 (see Supplementary content, Text).

3. Results

Prior to establishing whether P1-N1-P2 and T-complex morphology was consistent with spectral and temporal characteristics of the stimuli, we determined whether waveforms were reliable at the group and single subject levels.

3.1. Reliability of P1-N1-P2 and T-complex morphology in Group 1 and Group 2

AEP sensory waveforms for Group 1 and Group 2 were highly similar (Fig. 3). The association between Group 1 and 2 waveforms to the /pət/ onsets at site 4 (P1-N1-P2) was .90 and to the /sət/ onsets was .95 between 50 and 600 ms; at site 24 (T-complex) the association between Group 1 and 2 waveforms for /pət/ onsets was .83 and for /sət/ onsets was .87. Reliable sensory waveform patterns were further demonstrated by showing that 95% confidence intervals did not vary widely from the means of the P1-N1-P2 and T-complex (Fig. 3), as well as a small SE for the grand means (Fig. 4).

3.2. Reliability of P1-N1-P2 morphology in individual subjects

In comparisons of each individual participant's P1-N1-P2 complex to the grand mean at FC sites, we found high similarity for all, but two participants to the /pət/ onset and one other participant to the /sət/ onset. Thus, 97% (93 of 96 comparisons) of the P1-N1-P2 waveform associations for individual subjects were significant ($p < .01$). The mean r value (range) demonstrating association between each individual's P1-N1-P2 waveform to their grand mean to /pət/ for Group 1 was .84 (range .30–.97) and for Group 2 was .94 (range .82–.99) and to /sət/ for Group 1 was .83 (range .39–.97) and for Group 2 was .93 (range .82–.98). Only two participants' waveforms in response to the /pət/ onsets showed poor association to the grand mean waveform (r values = .30; r value = .51) and one additional participant's pattern showed poor correlation in response to /sət/ (r value = .39) (See Supplementary content, Table S2 for all values).

3.3. Reliability of T-complex morphology in individual subjects

Individual participant T-complex patterns showed less consistency in morphology than P1-N1-P2 patterns at averaged PT sites. Sixty eight percent of individual waveform patterns (i.e., 31 out of 48 participant waveforms to the /pət/ onset and 34 of 48 participant waveforms to the /sət/ onset) at averaged posterior temporal sites significantly correlated with the grand mean ($p < .01$), suggesting greater variability for T-complex morphology at the single subject level. The mean r value (range) demonstrating association between each individual's T-complex waveform to their grand mean to /pət/ for Group 1 was .56 (range $-.12$ – $-.93$) and for Group 2 was .58 (range $-.26$ – $-.92$) and to /sət/ for Group 1 was .61 (range $-.36$ – $-.94$) and for Group 2 was .59 (range $-.46$ – $-.96$) (see Supplementary content Table S2, for all values).

3.4. Spectro-temporal features reflected within P1-N1-P2 and T-complex

Representation of the spectro-temporal distinctions within the syllables /sət/ and /pət/ were evident within the P1-N1-P2 and T-complex morphology. Table 1, displays the effects of word type, language group and the interaction of word type and language (MEM) on the P1-N1-P2 amplitudes at FC sites and includes coefficient estimates (standard error) and p values. The apparent differences between 100 and 200 ms to /sət/ and /pət/ evident within the P1-N1-P2 waveforms as shown in Fig. 4 (top), consisting of an earlier direction change between the first negative deflection (N1) and the second positive deflection (P2) and smaller amplitudes for the N1 to /sət/ relative to /pət/ were significant for word type for both Groups 1 and 2. This suggests an effect of frequency differences within the onset syllables as well as other acoustic characteristics within the natural stimuli, including rise time and duration differences as detailed below. Word type also accounted for differences in the waveforms between 300 and 400 ms for Groups 1 and 2 and between 500 and 600 ms for Group 1 and approached significance for Group 2. These significant effects of word type (i.e., /pət/ and /sət/ onsets) reflect differing spectral and temporal characteristics of the stimuli. An additional positive peak within the P1-N1-P2 waveforms between 500 and 600 ms only to the /sət/ onsets, as shown in Fig. 4, highlights the temporal distinction of approximately 100 between the /sət/ and /pət/ onsets.

Also, an effect of language group for only Group 1 at all time intervals except for the 100–200 ms interval, reflects a more negative waveform for Polish participants relative to English participants (Table 1). Importantly, the interaction between word type and language group was not significant at any time period, which indicates that waveform morphology does not differ for the language groups.

The T-complex, shown within Fig. 4 (bottom), at averaged PT sites reflects similar acoustic feature representations to naturally spoken words evident at FC sites, however, amplitude responses were smaller. As shown in Table 2, an earlier direction change between the first positive deflection (Ta) and the second negative deflection (Tb) and smaller amplitude for the first positive deflection (Ta) to “set” relative to “pet” between 100 and 200 ms, highlighting spectral distinctions within the stimuli, was significant for both Groups 1 and 2 for word type. Word type also affected the T-complex between 300 and 400 ms for both Groups 1 and 2, highlighting temporal distinctions within the syllable onsets. In contrast to

the effects of word type on the P1-N1-P2 response at 500 to 600 ms, word type was not found to affect the T-complex at PT sites beyond 400 ms.

A significant effect of language group on the T-complex between 200 to 300 ms and 300 to 400 ms, resulting from a more negative waveform for the Polish listeners relative to the English listeners, was found for only Group 1 (Table 2). No interaction of word type and language group within the T-complex suggests a similar morphology for both English and Polish listeners.

3.5. Stability of single response trials

The prediction that variability for neural responses to naturally spoken words would increase as time increased between 50 and 600 ms was not supported by our results, rather single trial responses for both the P1-N1-P2 and T-complex remained stable or standard deviation (SD) decreased as time increased to 600 ms post stimulus onset.

P1-N1-P2 single trial responses were found to remain constant between 50 and 600 ms for two of the four analyses (Group 1: FC /sət/ $p = .82$ Group 2: FC /pət/ $p = .14$) and for two analyses, SD decreased as time increased (Group 1: FC /pət/ $p < .01$ Group 2: FC /sət/ $p = .04$). For the T-complex, SD remained constant for three of the analyses between 50 and 600 ms (Group 1: PT /pət/ $p = .76$; PT /sət/ $p = .31$ Group 2: PT /sət/ $p = .95$). For Group 2, however, in response to the /pət/ onset, variability decreased as time increased (Group 2: PT /pət/ $p = .05$) (Supplementary content, Fig. S1).

4. Discussion

This study has demonstrated that P1-N1-P2 morphology, at group and individual levels, and T-complex morphology, at group levels, to spoken words that approximate the natural variability of a single speaker reflect the spectral and temporal features within the words and these waveform patterns are sufficiently reliable for research and clinical application. Highly similar P1-N1-P2 and T-complex patterns were found in two independent groups of subjects even though participants heard 70 natural productions of each word type within experiments that differed slightly in stimulus delivery and intensity. P1-N1-P2 and T-complex morphology were modulated by the onset syllables /pət/ and /sət/ within waveforms of individual subjects at fronto-central sites and, to a lesser degree, at posterior temporal sites. Thus, the current findings suggest that P1-N1-P2 complex and T-complex, at least at group levels, may serve as phenotypic signatures that can be used for assessing cortical deficits in spoken word recognition and for identifying areas of impairment, superior temporal plane for the P1-N1-P2 complex or lateral superior temporal gyrus for the T-complex [15,34].

Spectro-temporal features elicited by the onsets of the syllables /pət/ and /sət/ were represented within the P1-N1-P2 and T-complex, although representations were more distinct for P1-N1-P2 waveforms. Previous studies using stimuli (i.e., tones and synthetic speech) that allowed for isolation of spectral [20,24,29,39] and temporal characteristics [11] have demonstrated that spectral and temporal characteristics are represented within the P1-N1-P2 waveforms, although we are unaware of similar studies examining T-complex

morphology. The current results have demonstrated that spectral and temporal characteristics within multiple natural productions of speech can be similarly identified.

Syllable onsets presented within words were highly contrastive in spectral and temporal characteristics. For example, the difference in (mean) duration between the onset syllables /pət/ and /sət/ is approximately 100 ms and this, in large part, is due to the 100 ms difference in duration of the onset consonants (/p/ and /s/). This temporal difference in the acoustics of the onset syllables was found to correspond to a 100 ms increase in duration for the sensory waveforms for /sət/ relative to /pət/ that was represented by an additional peak between 500 and 600 ms. It is unlikely that this increased duration of the sensory waveform for /sət/ was due to the difference in whole word duration because the phoneme sequences that followed the onset syllables differed (i.e., 35 different words) and sensory responses to these varied acoustic changes would likely cancel. This view is supported by our results in that the mean duration for the words having /sət/ onsets was 698 ms, which exceeds the sensory waveform duration that extended to 600 ms for the /sət/ onsets. Furthermore, the mean difference for the whole words containing /sət/ and /pət/ onsets was 148 ms, which did not correspond to the difference in duration of the sensory waveforms. The fricative /s/ in /sət/ contains a concentration of high frequency energy above 4 kHz, whereas, the stop-consonant /p/ contains a concentration of low-to-mid frequency energy. While spectral and temporal characteristics of speech cannot be isolated using natural speech, the spectral differences within the onset consonants reflected in the sensory waveforms are consistent with that demonstrated for tones and synthetic speech stimuli [20,24,29,39] for which acoustic features can be isolated. Picton et al. [29] found N1 amplitude to be significantly smaller to 4 kHz and 8 kHz tone bursts (1 s duration) relative to all tone bursts of lower frequency (i.e., 250 Hz, 500 Hz, 2 kHz). Also, Beukes et al. [3] examined the cortical sensory responses to four stimuli, including single productions of the natural speech sounds /m/ (low frequency) and /sh/ (mid to high frequency) and low (500 Hz) and high (4000 Hz) frequency tones. While not the focus of their study, amplitudes described in their Tables 1 and 2 reveal that for both tones and speech stimuli of 100 ms duration, the stimulus having higher frequency (tone or speech) results in a smaller and earlier N1 relative to the low frequency stimulus, consistent with our result. However, their results also found that extending the duration of the same stimulus (e.g., /m/) from 100 ms duration to 500 ms duration resulted in reduced N1 amplitude [3]. N1 amplitude increases as sound duration increases through approximately 30 to 50 ms [26]. Hence, Beukes et al. [3] findings suggest that speech sound durations that exceed 50 ms may contribute to smaller N1 amplitude. These finding exemplify that while the spectral feature contrasts are prominent within our phoneme onsets, additional acoustic characteristics including duration and rise time within the onset consonants, likely affected the amplitudes within the N1-P2 sensory responses between 100 and 200 ms.

A slower rise time from onset to peak energy for the fricative /s/ relative to the stop consonant /p/ may also have contributed to the smaller N1 amplitude. A smaller N1-P2 amplitude was found to be elicited to tones (1000 Hz), which consisted of a slower rise from onset to peak energy (i.e., shallow slope) relative to those that contained a sharper rise from onset to peak energy [17,26], consistent with our result. However, a slower rise time resulted in an extended N1-P2 latency [17], which contrasts with our result that found an earlier N1-

P2 latency for /sət/ relative to /pət/. Finally, intensity differences within the onset syllables cannot explain our results because the increased amplitude for the /s/ relative to the /p/ consonant would have resulted in a larger, rather than smaller N1 [6].

The onset consonants (/p/ and /s/) within spoken words reflect phoneme category distinctions that differ in both manner and place of articulation, however, language can affect categorical speech perception [32,36,37]. Even though language was not the focus of the current study because the syllable onsets (i.e., /sət/ and /pət/) occur within both Polish and English, we examined the effects of language on cortical sensory waveform morphology. P1-N1-P2 and T-complex waveform morphology representing spectro-temporal feature processing for spoken words remained unchanged for English and Polish listeners within Group 1 and 2 for all time intervals. This result provides additional support for the claim that the P1-N1-P2 and T-complex morphology reflect acoustic, rather than linguistic features of speech [36] (see Supplementary content).

Cortical activation in response to an auditory stimulus will extend to multiple sources beyond auditory cortex as time increases [34]. Therefore, we tested an exploratory hypothesis that variability in individuals' single trial responses would increase as time increased. This hypothesis was not supported. We used only nonsense word stimuli, therefore, we cannot rule out the possibility that greater variability in individual responses would increase with increasing time from stimulus onset using real words as stimuli, possibly reflecting additional cognitive processing or for more complex behavioral tasks. We are unaware of other studies examining variability using a similar method, thus, subsequent studies will be necessary to confirm the current finding. These results provide preliminary benchmarks for stability of single trials responses to natural speech that may be used for future comparisons.

Together, these results suggest that P1-N1-P2 at group and individual levels, and T-complex at the group levels, may be used to probe spectral and temporal feature recognition at cortical levels in individuals with language impairment, dyslexia, autism and auditory processing disorders [1,4,7,31,33]. It has been suggested that weak neural representation of phoneme sequences or access to representation may underlie speech sound disorder [23], specific language impairment [38] and dyslexia [12]. The present study has identified representation of spectral and temporal features of naturally varying speech within the P1-N1-P2 complex and T-complex as has been previously demonstrated for tones and synthetic speech, providing ecological validity for use of these waveforms for probing cortical level auditory deficits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This publication was partially supported by a grant from the National Institute of Child Health and Human Development (NICHD) at the National Institutes of Health (NIH) to Valerie Shafer (grant number HD-46193). The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views

of NIH. We acknowledge Jungmee Lee, Gary Chant, and Patrick Walden for providing assistance on this project. We are grateful to the reviewers for their excellent suggestions.

References

1. Benasich A, Tallal P. Infant discrimination of rapid auditory cures predicts later language impairment. *Behav Brain Res.* 2002; 136:31–49. [PubMed: 12385788]
2. Bernasconi F, De Lucia M, Tzovara A, Manuel A, Murray M, Spierer L. Noise in brain activity engenders perception and influences discrimination sensitivity. *J Neurosci.* 2011; 31:17971–17981. [PubMed: 22159111]
3. Beukes E, Munro K, Purdy S. Duration-sensitive neurons in the auditory cortex. *Neuroreport.* 2009; 20(13):1129–1133. [PubMed: 19597375]
4. Bishop D, Hardiman M, Uwer R, Suchodoletz W. Atypical long-latency auditory event-related potentials in a subset of children with specific language impairment. *Dev Sci.* 2007; 10:576–587. [PubMed: 17683344]
5. Bordon, G.; Harris, K.; Raphael, L. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech.* 4th. Lippincott, Williams & Wilkins; Baltimore, MD: 2003.
6. Carillo-de-la-Pena M. Effects of intensity and order of stimuli presentation on AEPs: an analysis of the consistency of ER augmenting/reducing in the auditory modality. *Clin Neurophysiol.* 1999; 110:924–932. [PubMed: 10400207]
7. Ceponiene R, Cummings A, Wulfeck B, Ballantyne A, Townsend J. Spectral vs temporal auditory processing in specific language impairment: a developmental ERP study. *Brain Lang.* 2009; 110:107–120. [PubMed: 19457549]
8. Digeser F, Wohlberedt T, Hoppe U. Contributions of spectrotemporal features on auditory event-related potentials elicited by consonant-vowel syllables. *Ear Hear.* 2009; 30:704–712. [PubMed: 19672195]
9. Faisal A, Selen L, Wolpert D. Noise in the nervous system. *Nat Rev Neurosci.* 2008; 9:292–303. [PubMed: 18319728]
10. Gilley P, Sharma A. Functional brain dynamics of evoked and event-related potentials from the central auditory. *Perspect Hearing Hearing Disorders: Res Diagnosis.* 2010; 14:12–20.
11. Giraud K, Demonet J, Habib M, Marquis P, Chauvel P, Liegeois-Chauvel C. Auditory evoked potential patterns to voiced and voiceless speech sounds in adult developmental dyslexics with persistent deficits. *Cereb Cortex.* 2005; 15:1524–1534. [PubMed: 15689520]
12. Giraud A, Ramus R. Neurogenetics and auditory processing in developmental dyslexia. *Curr Opin Neurobiol.* 2013; 23:37–42. [PubMed: 23040541]
13. Guttorm T, Leppanen P, Hamalainen J, Eklund K, Lyytinen H. New event-related potentials predict poorer pre-reading skills in children at risk for dyslexia. *J Learn Disabil.* 2010; 45:391–401. [PubMed: 19890075]
14. Hämäläinen J, Fosker T, Szucs D, Gowami U. N1, P2 and T-complex of the auditory brain event-related potentials to tones with varying rise times in adults with and without dyslexia. *Int J Psychophysiol.* 2011; 81:51–59. [PubMed: 21565226]
15. Howard M, Volkov I, Mirsky R, Garell P, Noh M, Granner M, Damasio H, Steinschneider M, Reale R, Hind J, Brugge J. Auditory cortex on the human posterior superior temporal gyrus. *J Comp Neurol.* 2000; 416:79–92. [PubMed: 10578103]
16. Ille N, Berg P, Scherg M. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J Clin Neurophysiol.* 2002; 19:113–124. [PubMed: 11997722]
17. Kodera K, Hink R, Yamada O, Suzuki J. Effects of rise time on simultaneously recorded auditory evoked potentials from the early, middle and late ranges. *Audiology.* 1979; 18:395–402. [PubMed: 496722]
18. Luke, D. *Multilevel Modeling, Series: Quantitative Applications in the Social Sciences.* Sage Publication, Inc.; California: 2004.
19. Martin B, Boothroyd A. Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear.* 1999; 20:33–44. [PubMed: 10037064]

20. Martin B, Boothroyd A. Cortical, auditory, event-related potentials in response to changes of spectrum and amplitude. *J Acoust Soc Am*. 2000; 107:2155–2161. [PubMed: 10790041]
21. Martin B, Tremblay K, orczak PK. Speech evoked potentials: From the laboratory to the clinic. *Ear Hear*. 2008; 29:285–313. [PubMed: 18453883]
22. Mesgarani CN, Cheung K, Chang Johnson E. Phonetic feature encoding in human superior temporal gyrus. *Science*. 2014; 343:1006–1010. [PubMed: 24482117]
23. Munson, B.; Edwards, J.; Beckman, M. Phonological representations in language acquisition: Climbing the ladder of abstraction; *Handbook of Laboratory Phonology*. 2011. p. 288-309.http://learningtotalk.org/sites/learningtotalk.org/files/MunsonEdwardsBeckman_LabPhonHandbook_Revision_08January2010.pdf
24. Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*. 1987; 24:375–425. [PubMed: 3615753]
25. Ponton C, Eggermont J, Khosla D, Dwong B, Don M. Maturation of human central auditory system activity: Separating auditory evoked potentials by dipole source modeling. *Clin Neurophysiol*. 2002; 113:407–420. [PubMed: 11897541]
26. Onishi S, Davis H. Effects of duration and rise time of tone bursts on evoked V potentials. *J Acoust Soc Am*. 1968; 44:582–591. [PubMed: 5665530]
27. Ostroff J, Martin B, Boothroyd A. Cortical evoked response to acoustic change within a syllable. *Ear Hear*. 1998; 19:290–297. [PubMed: 9728724]
28. Perrin F, Pernier J, Bertrand O, Echallier J. Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophys*. 1989; 72:184–187.
29. Picton T, Woods D, Proulx G. Human auditory sustained potentials. II. Stimulus relationships. *Electroencephalogr Clin Neurophys*. 1978; 45:198–210.
30. Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Phil Trans R Soc Lond B*. 1992; 336:367–373. [PubMed: 1354376]
31. Seery A, Vogel-Farley V, Tager-Flusberg H, Nelson C. Atypical lateralization of ERP response to native and non-native speech in infants at risk for autism spectrum disorder. *Dev Cogn Neurosci*. 2013; 5:10–24. [PubMed: 23287023]
32. Shafer V, Schwartz R, Kurtzberg D. Language-specific memory traces of consonants in the brain. *Cogn Brain Res*. 2004; 18:242–254.
33. Shafer V, Schwartz R, Martin B. Evidence of deficient central speech processing in children with specific language impairment: the T-complex. *Clin Neurophysiol*. 2011; 122:1137–1155. [PubMed: 21147550]
34. Steinschneider, M.; Liegeois-Chauvel, C.; Brugge, J. Auditory evoked potentials and their utility in the assessment of complex sound processing. In: Winer, JA.; Schreiner, CE., editors. *The Auditory Cortex*. Springer Science and Business Media LLC; New York: 2011. p. 535-559.
35. Steinschneider M, Nourski K, Fishman Y. Representation of speech in human auditory cortex: is it special? *Hear Res*. 2013; 305:57–73. [PubMed: 23792076]
36. Wagner M, Shafer V, Martin B, Steinschneider M. The effect of native-language experience on the sensory-obligatory components, the P1N1P2 and the T-complex. *Brain Res*. 2013; 1522:31–37. [PubMed: 23643857]
37. Wagner M, Shafer V, Martin B, Steinschneider M. The phonotactic influence on perception of a consonant cluster/pt/by native-English and native-Polish listeners: a behavioral and event-related potential (ERP) study. *Brain Lang*. 2012; 123:30–41. [PubMed: 22867752]
38. Williams D, Payne H, Marshall C. Non-word repetition impairment in Autism and Specific Language Impairment: Evidence for distinct underlying cognitive causes. *J Autism Dev Disord*. 2013; 43:404–417. [PubMed: 22733298]
39. Wunderlich J, Cone-Wesson B. Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *J Acoust Soc Am*. 2001; 109:1526–1537. [PubMed: 11325124]

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.

Fig. 1. Acoustic waveforms and spectrograms for sample stimuli having “pet” and “set” onsets. The final number on the *x*-axis indicates whole word duration and syllable duration measured from word onset to the burst for /t/ is in small print.

Fig. 2. Geodesic Sensor Map. A map of 64 channel locations and the vertex reference (VREF) channel location for the Geodesic Sensor Net (Electrical Geodesics, Inc.). Averaged fronto-central electrode sites are circled and arrows indicate averaged left (site 24) and right (site 52) posterior temporal sites.

Fig. 3.

Waveform reliability. Figure illustrates highly similar P1-N1-P2 (FC site 4) and T-complex (left PT site 24) waveforms for Groups 1 (top) and 2 (bottom) $\pm$ 95% confidence intervals to "pet" (red) and "set" (blue). Grand mean data at each time point shown (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Fig. 4.

Spectro-temporal features represented within waveform morphology. Overlay of P1-N1-P2 and T-complex waveforms to “pet” (red) and “set” (blue) $\pm$ SEM from FC site 4 (top) and left PT site 24 (bottom) for Groups 1 (left) and 2 (right). Grand mean data at each time point shown (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Table 1

Summary of linear mixed-effects regression analyses (fixed effects only). The table shows the effect of word type, language group and the interaction of word type and language group on amplitude values in response to “pet” and “set” between 100 and 600 ms at averaged FC sites (P1-N1-P2). Coefficient estimates (SE) and *p* values are shown for Group 1 and Group 2.

Time	Word	Language		Word* language	
		Group 1	Group 2	Group 1	Group 2
100–200	.53 (.14)	.48 (.14)	–.48 (.26)	.34 (.20)	.09 (.19)
	<i>p</i> < .001	<i>p</i> < .001	<i>p</i> = .07	<i>p</i> = .10	<i>p</i> = .65
200–300	–.49 (.16)	–.01 (.15)	–.76 (.30)	.01 (.22)	.10 (.21)
	<i>p</i> < .01	<i>p</i> = .93	<i>p</i> = .02	<i>p</i> = .96	<i>p</i> = .65
300–400	.98 (.17)	.65 (.16)	–1.10 (.33)	.12 (.24)	–.01 (.23)
	<i>p</i> < .001	<i>p</i> < .001	<i>p</i> < .01	<i>p</i> = .61	<i>p</i> = .96
400–500	–.34 (.18)	.34 (.17)	–1.37 (.45)	–.03 (.25)	–.02 (.24)
	<i>p</i> = .05	<i>p</i> = .05	<i>p</i> < .01	<i>p</i> = .90	<i>p</i> = .94
500–600	.46 (.18)	.37 (.18)	–1.41 (.47)	.38 (.26)	.49 (.25)
	<i>p</i> = .01	<i>p</i> = .04	<i>p</i> < .01	<i>p</i> = .14	<i>p</i> = .05

Table 2

Summary of linear mixed-effects regression analyses (fixed effects only). The table shows the effect of word type, language group and the interaction of word type and language group on amplitude values in response to “pet” and “set” between 100 and 600 ms at averaged PT sites (T-complex). Coefficient estimates (SE) and *p* values are shown for Group 1 and Group 2.

Time	Word	Language		Word* language	
		Group 1	Group 2	Group 1	Group 2
100–200	–.34 (.10) <i>p</i> < .001	–.42 (.11) <i>p</i> < .001	.03 (.14) <i>p</i> = .81	–.11 (.22) <i>p</i> = .62	–.08 (.14) <i>p</i> = .55
200–300	.15 (.11) <i>p</i> = .21	.10 (.14) <i>p</i> = .47	.58 (.17) <i>p</i> = .001	–.15 (.24) <i>p</i> = .53	–.13 (.20) <i>p</i> = .51
300–400	–.53 (.13) <i>p</i> < .001	–.39 (.17) <i>p</i> = .02	.42 (.20) <i>p</i> = .04	–.36 (.26) <i>p</i> = .17	.16 (.18) <i>p</i> = .37
400–500	.18 (.14) <i>p</i> = .20	–.08 (.19) <i>p</i> = .65	.41 (.23) <i>p</i> = .09	.16 (.20) <i>p</i> = .26	.21 (.26) <i>p</i> = .43
500–600	–.25 (.15) <i>p</i> = .09	–.04 (.20) <i>p</i> = .83	.25 (.27) <i>p</i> = .37	–.15 (.35) <i>p</i> = .66	.23 (.21) <i>p</i> = .28
					–.16 (.29) <i>p</i> = .57