P1-N1-P2 Complex and Acoustic Change Complex Elicited by Speech Sounds: Current Research and Applications

Woojae Han
Department of Speech and Hearing Science, College of Applied Health Science, University of Illinois at Urbana-Champaign, Champaign, IL 61820 US

ABSTRACT
Auditory cortical-evoked potentials reflect neural activity related to the detection, perception, discrimination, and cognition of incoming sounds from their timing (latency) and magnitude (amplitude) at the levels of primary auditory cortex and associated areas of the temporal lobe. Researchers have become interested in cortical-evoked potentials elicited by speech stimuli, referred to as speech-evoked potentials (SEPs). SEPs can provide information about a listener’s speech perception abilities at the auditory cortical level, primarily regarding features within speech signals that yield a greater chance of extracting essential cues for human auditory/oral communication. However, to date, relatively little data exist based on naturally produced speech stimuli, which include the listener’s perceptual information, such as specific acoustic/phonetic features, and unmodified spectral/temporal cues produced in natural speech. This tutorial paper reviews current SEP studies and provides feasible applications, focusing on two components of auditory potentials: P1-N1-P2 complex and acoustic change complex (ACC). SEPs offer a powerful objective technique for understanding speech perception of both normal and hearing-impaired listeners, particularly infants and elderly patients who cannot be assessed reliably using typical behavioral speech perception measures; SEPs also provide a means to investigate how acoustical characteristics of various naturally spoken speech stimuli are reflected in the evoked responses for both populations.

KEY WORDS: Acoustic change complex · Natural speech stimulus · P1–N1–P2 complex · Speech-evoked potentials · Voice onset time.

INTRODUCTION
Electrical measures from the human nervous system, called event-related potentials (ERPs), are used to study dynamic brain activity in response to stimulation from any of the sensory modalities (i.e., auditory, visual, somatosensory, olfactory, and gustatory). ERP recordings are less expensive than magnetic resonance imaging (MRI) and relatively noninvasive because they use surface electrodes to record electrical activity generated by the brain in response to sensory stimuli.21) One type of ERP, auditory-evoked potentials (AEPs), has provided useful information about the auditory system regarding detection, perception, discrimination, pre-cognition, and processing contingent properties of both simple and complex sounds.5) Generally, AEPs are divided into five groups according to the latency (or timing) of their response following stimulus onset.18) AEP components have been labeled: first (0–5 ms), fast (2–20 ms), middle (10–100 ms), slow (50–300 ms), and late (150–300 ms).18) Among them, the middle, slow, and late response groups, which this paper reviews, have been referred to as auditory cortical potentials because of the specific area of the brain from which responses originate.17)

Many contemporary researchers have become interested in using speech stimuli, rather than a click or tone-burst stimulus, to elicit auditory cortical potentials. These stimuli are called speech-evoked potentials (SEPs) and their possible clinical uses have been investigated.17) One advantage of SEPs over simpler stimuli is that they may better reflect the acoustic information available for speech processing at the auditory cortical level. This information includes acoustic features related to spectral (frequency) and temporal (timing) cues within the speech signals that may contribute to speech perception and, by extension, be essential for human communication.17,26) Furthermore, SEP measures, which are influenced by frequency and timing features, may provide more objective descriptions of hearing-impaired listeners’ speech perception as well as more
valuable information about biological and neural processes underlying speech processing.17)

MATERIALS AND METHODS: REVIEW OF LITERATURE RELATED TO CURRENT SEP STUDIES

To (a) introduce recent studies about auditory evoked potentials with natural speech stimuli, (b) provide a tutorial for interpretation of SEPs, including the components of P1-N1-P2 complex and acoustic change complex (ACC), and (c) provide clinicians and researchers with future potential applications for SEPs, especially with naturally produced speech stimuli, which overcome some of the limitations of widely used synthetic speech stimuli.

RESULTS: OVERVIEW OF CURRENT SEP STUDIES BY COMPONENT

P1-N1-P2 complex

The P1-N1-P2 complex is associated with detection of sound17): it is thought to reflect underlying neural encoding of temporal and spectral features of sound at the auditory cortex.18) P1 is the first positive component: its neural generator includes the primary auditory cortex (Heschl’s gyrus), the hippocampus, the planum temporale posterior to the auditory cortex, and the lateral temporal cortex.17) N1 or N100 is the first negative component and demonstrates a large amplitude relative to P1 in adults.17)

Although significant ongoing debate has emerged regarding the sources of N1, the primary and secondary auditory cortices are commonly considered to be its generators.17) Following the N1 response, P2 is the second positive peak. The major generators of P2 are the primary and secondary auditory cortices.17)

In adults who have normal hearing, the P1, N1, and P2 components appear at approximately 50 ms, 100 ms, and 180 ms, respectively, after the onset of the stimulus. P1 shows very large amplitude in young children who have an immature central nervous system: however, P1 amplitude is smaller in adults while N1 and P2 amplitude is larger in adults than in children.17) Compared to other cortical auditory evoked responses, the P1-N1-P2 complex is regarded as an early detection response and is passively measured, without requiring the listener’s attention. In the early 1960s, the P1-N1-P2 complex was used primarily to estimate hearing threshold sensitivity,11 but today it is often recorded to evaluate neural function at supra-threshold levels.17) In addition, application of speech stimuli at the supra-threshold has revealed that the P1-N1-P2 complex may represent the neural encoding of acoustic features and that these SEPs may reflect neural processes that underlie speech perception.22)

In an early study, synthetic speech stimuli were used by Dorman9) to observe whether differences in the amplitude of the N1-P2 response would reflect acoustic and phonetic differences in voice onsets time (VOT). Two synthetic syllables with three different lengths of VOT were presented to normal-hearing listeners: 0 and 20 ms as /ba/ and 40 ms as /pa/. The three stimuli were presented in various patterns involving the pairing of syllables within (0 vs. 20 ms VOT) and across (20 vs. 40 ms VOT) phonetic category in various orders and under various conditions in terms of the listeners’ attention to the VOT difference. The results showed that differences in amplitude of these SEPs reflected the differences in phonetic category. For a given difference in VOT, the evoked response was larger for the stimuli presented across rather than within the phonetic category, indicating that the N1-P2 amplitude may reflect neural encoding that contributes to categorical perception.6) While this study did not directly assess the encoding of acoustic features within the stimuli (i.e., a difference between 0 and 20 ms VOT), it showed that N1-P2 amplitude may reflect differences in the phonetic category under the condition of sufficient attention in normal-hearing listeners.

Sharma et al. demonstrated a more specific approach that is sensitive to acoustic features within speech.22) They investigated how N1 morphology would be transformed to modify VOT in 10-ms intervals between 10 ms and 70 ms for two-voice and voiceless continua, /ba/-/pa/ and /ga/-/ka/.

Results of this study revealed that neural representation of voicing onset with increasing VOT produced a change in the morphology of the P1-N1-P2 complex such that a second negative peak [labeled N1 in the bottom panels of Figure 3 and Figure 4 of Sharma et al. (2000) pp. 3031-3032] appears for longer VOT.22) The evoked potentials demonstrated a single N1 for stimuli with VOT of 30 ms or less, but two negative components (N1’ and N1) in response to stimuli with longer VOT (40 ms to 70 ms) in both continua. However, the appearance of the second negative peak in the evoked response (at 40 ms in both continua) coincided with a change in the categorical perception from voiced to voiceless for the /ba/-/pa/ stimulus continuum (at 40 ms), but not for the /ga/-/ka/ stimulus continuum (at 46 ms). Thus, the authors suggested that N1 morphology did not reliably predict phonetic identification of stimuli varying VOT.22) Careful evaluation of Sharma et al.’s Figures 3 and 4 provides the broad interpretation that the second negative component represents the onset of voicing for the vowel portion of stimuli with longer VOT, while the first negative component repre-
sents a combination of the consonant and vowel parts for stimuli with shorter VOT. This finding links to produce ACC responses, discussed in the next section.

Ostroff et al.’s research supported the notion that those multiple N1 components appear to derive from responses to distinct acoustic events within complex speech sounds. They observed the evoked response of a consonant-vowel syllable /sei/ and also studied whether acoustic events within the syllable would be evident in the components of the evoked response for the isolated consonant and vowel elements. Thus, to isolate the contribution of each acoustic event, the N1 and P2 evoked responses were measured in response to the entire syllable /sei/ as well as to its extracted sibilant /s/ and vowel /ei/. The results showed that the evoked responses elicited by each consonant and vowel part demonstrated sound onset consistent with the entire syllable in latencies. However, compared to the consonant part, the response in the vowel part showed smaller amplitude. The evoked response of the vowel within the entire syllable carried larger amplitude than the response elicited by the isolated vowel stimulus. The authors proposed that SEPs reflect the phonetic contrast effectively, and thus they could be useful for evaluation of speech perception of listeners with hearing loss.

In the studies described above, only relatively large temporal differences within speech stimuli generated distinct components within the evoked response. However, Tremblay et al. studied the possibility that smaller differences in the interval between acoustic events could be registered in the SEPs after perceptual training. Prior to training, the research participants did not recognize any difference between the two durations of VOT, specifically, /ba/ (−10 ms) and /mba/ (−20 ms). After 10−day training that involved practice in identifying temporal cues (i.e., VOT), however, participants evoked measures clearly exhibited increased N1-P2 amplitude and decreased latencies. The increased N1-P2 amplitude could indicate an increase in neural synchrony and may be associated with improved speech perception after auditory training. Those findings might verify the suitability of the P1-N1-P2 complex in terms of central auditory plasticity for learning speech discrimination.

Aging has also been shown to affect discrimination of temporal features within speech, and its effect is evident in SEPs. Specifically, Tremblay et al. recognized that older adults often demonstrate poorer behavioral discrimination of 10−ms VOT differences in the /ba/-/pa/ continuum than younger adults. To isolate the effects of temporal processing from other effects of aging, the authors compared the discrimination of VOT with the SEP responses to their results of behavioral VOT discrimination in normal-hearing older adults. Based on the results of behavioral discrimination tests, older adults showed significantly poorer discrimination performance than young adults. However, in the evoked response, there was no significant difference by aging effect in the amplitude of N1 and P2. All listeners showed delayed N1 and P2 latencies as VOT increased; however, older listeners showed significantly prolonged N1 latencies only for longer VOTs of 30 ms, 40 ms, and 60 ms. The authors suggested that the prolonged N1 latencies for older adults reflect extended neural recovery times for response to the VOT following the response to the consonant burst. For shorter VOTs, recovery time is not measurable because the N1 is dominated by the response to the consonant burst. Such a result suggests that older listeners have more difficulty than younger listeners in terms of perceiving temporal cues, and thus difficulties in speech perception among the older population might be due to age-related reductions in temporal acuity.

Spectral properties of speech stimuli also play a critical role in human speech perception. Digerer et al. studied the influence of spectro-temporal properties on evoked N1 and P2 responses using two naturally produced speech syllables, /da/ and /ta/, and their components. For each syllable, the entire consonant-vowel syllable, as well as its isolated vowel and consonant (i.e., /da/, /ta/, /ad/, /at/, /d/ and /t/), were presented under two conditions: (a) with the original spectrum and (b) with white noise substituted for the original spectrum but with the envelope maintained. Thus, 12 stimuli were presented to measure the evoked responses. The results showed that N1 and P2 responses had significantly shorter latency and larger amplitude to the voiceless /ta/ and isolated consonant /t/ than the voiced stimuli. The responses to the vowels, /ad/ and /at/, did not differ in terms of the evoked responses. When noise was substituted for the original spectrum, there was no significant difference in N1 and P2 latency or N1-P2 amplitude between stimuli associated with the voiced and voiceless consonants. The authors concluded that even small changes in the short-term spectro-temporal content can modify the evoked responses, despite stimuli having similar or identical wide-band envelopes.

With respect to the P1-N1-P2 complex with speech stimuli, the responses were well represented in their evoked latency, but did not reflect much amplitude. This demonstrates an acceptable correlation between acoustic features of the speech stimulus (i.e., in the voice-voiceless continuum) or detection of VOT in the consonant-vowel syllable and auditory cortical evoked responses. This implies that SEPs could be an effective tool for investigating speech perception capacity for normal-hearing and
hearing-impaired listeners underlying audiological diagnosis and rehabilitation.\(^{20}\)

**Acoustic change complex**

When very brief stimuli, such as clicks, tone bursts, and synthesized short speech, are used to elicit the P1-N1-P2 complex, the components of the waveform show a clear morphology.\(^{17}\) Longer duration stimuli, including naturally spoken speech, often elicit waveforms with a double peak (or second peak).\(^{20}\) The morphology of the cortical auditory evoked potentials consistently demonstrates multiple N1 peaks or multiple P1-N1-P2 complexes associated with the temporal sequence of acoustic elements in the stimulus.

A possible explanation for the complex morphology of the evoked potentials generated by long-duration speech was provided by Kaukoranta et al., who observed a double peak in the magnetoencephalographic (MEG) responses to synthetic speech syllables. The MEG is an evoked response measure of electromagnetic fields generated by the brain in response to a stimulus. The responses recorded by MEG were produced by the same neural sources as the AEPs.\(^{10}\) According to the authors, humans can recognize a difference between noise bursts and actual phonemes because they are sensitive to acoustic features such as consonant release and vowel onset.\(^{10}\) To evaluate the ability of the human auditory cortex to provide distinguishable responses to different speech sounds, Kaukoranta et al. measured the MEG response to the Finnish syllable /hei/. The MEG response showed two remarkable peaks at latencies of 100 ms and 200 ms in response to the Finnish word /hei/. The VOT of the word corresponded to the peak at a latency of 200 ms, which had never before been seen in response to a tone-burst stimulus.\(^{10}\) The authors pointed out that the acoustic/phonetic features of speech, in this case VOT, might evoke a prominent complex response generated by the human supratemporal auditory cortex.\(^{10}\) They suggested that such robust acoustic features would reflect underlying neural encoding of acoustic elements of speech.\(^{10}\)

Whereas multiple peaks had been observed when recording P1-N1-P2 SEPs, Martin and Boothroyd first named this pair of sequential peaks in the AEP waveform the acoustic change complex (ACC)\(^{14}\): “That is, ACC refers to a scalp-recorded negative-positive voltage swing elicited by a change during an ongoing acoustic stimulus [studies in original].”\(^{15}\) For example, the natural /ʃi/ stimulus as in shee shows two prominent N1 peaks in the evoked response: the first peak for fricative onset [consonant portion, first N1, peak 1 in Figure 1 of Tremblay et al. (2006), p. 94] and the second peak for vowel onset or VOT [vo-

The previously described study by Ostroff et al. presented the multiple P1-N1-P2 basis for the ACC.\(^{20}\) Ostroff et al. demonstrated that the response to the syllable /sei/ consisted of a series of two overlapping P1-N1-P2 responses generated by the onsets of the fricative /s/ and the vowel /ei/, respectively. The consonant and the vowel each evoked different cortical activation patterns. The complex SEP of the entire syllable represented cortical activation, including the response to the transition from the consonant to the vowel.\(^{20}\) That is, the ACC may signify which details of rapidly changing sounds such as speech are encoded in the human auditory cortex. Thus, the ACC response may be a useful physiological marker for relatively long duration and time-varying stimuli such as natural speech.

The reliability of the ACC as a marker of acoustic change in speech syllables was addressed by Tremblay et al. in 2003.\(^{20}\) When four naturally spoken stimuli were used /bi/, /pi/, /ʃi/ and /si/ the ACC response evoked by the /bi/ stimulus had a significantly higher amplitude than the response evoked by /pi/, and the response to /ʃi/ had a shorter latency than the /si/ stimulus. The evoked ACC for each syllable, however, revealed consistently repeated patterns among participants across several trials.\(^{24}\) These results indicate that the auditory cortical evoked responses consistently reflect detection of the change in the acoustic characteristics of naturally produced speech syllables.\(^{24}\)

Through an ACC study, the normal auditory cortical maturation was explained by Kummer et al., who examined children ages 4 to 14.\(^{12}\) The researchers observed two evoked responses: N100 and N170. The N100 component was elicited by a tone, and the N170 component was derived from the wave differences of monosyllabic speech and noise stimuli within the same temporal envelope and occurring approximately 170 ms after the onset of a stimulus.\(^{12}\) The authors wanted to see whether the N170 component might be useful in assessing the cortical auditory maturation of children. They found that the N100 and N170 latency showed a similar maturation pattern latency in children, as latency gradually decreased un-
Such similar results between tone and speech stimuli demonstrate a clear correlation between electrophysiological measures and maturation of evoked response to acoustical events in spoken syllables.  

To examine the effects of amplification on reliability of evoked responses to speech, Tremblay et al. measured the ACC from seven normal-hearing listeners under unaided and aided conditions. The syllables /i/ and /si/ were presented from loudspeakers and participants listened to stimuli binaurally. In the unaided condition, N1 corresponding to the onset of vowel /i/ was recorded approximately 30 ms earlier in response to /fi/ than to /si/, reflecting the difference in VOT between the two syllables [see Figures 2 and 3 of Tremblay et al. (2006), p. 94]. The second N1 to each syllable in the binaurally aided condition (mild high-frequency gain) had latencies similar to the unaided condition. A small but significant increase in the amplitude of the ACC in the aided condition was demonstrated in response to the /si/ stimulus. Overall, this study revealed that distinctions in the ACC between the two syllables are maintained under mild amplification in normal-hearing listeners. Therefore, because cortical AEPs can be altered by hearing loss, this study provided a means to distinguish the effects of amplification and peripheral hearing loss on AEPs.

In a companion study, Tremblay et al. observed ACC responses of hard-of-hearing listeners with mild to severe hearing loss under aided conditions to /fi/ and /si/. Similar to the results of normal-hearing listeners fit with hearing aids, the responses of hard-of-hearing listeners who wore their own hearing aids showed that the second N1 peak occurred about 30 ms earlier for /fi/ than for /si/. Latencies and amplitude of the first N1 (onset to each consonant) showed no changes as a result of amplification in hard-of-hearing listeners. The results of these studies support the concept that neural encoding of temporal components of speech is maintained under amplification, as evidenced by the presence of intact ACC waveforms recorded under aided conditions.

In comparison to hearing aids, the cochlear implant (CI) provides a more extreme modification of speech for presentation to the auditory system. To better understand the extent to which speech processing by CIs can maintain temporal features of speech available to the auditory cortex, CI users were also tested using naturally produced /fi/ and /si/ stimuli in sound fields. The evoked responses to the /fi/ stimuli were recorded 20 ms to 30 ms earlier than to the /si/ stimuli in CI users at both the first and second N1. The CI users revealed small individual variance in their responses, similar to their normal-hearing counterparts. Although the stimulus set is extremely limited, the results suggest that neural encoding of cues at the auditory cortex that underlie speech perception can be maintained by CIs. Other electrically evoked acoustic change complex (EACC) studies with CI users have been conducted by several researchers, but they are not reviewed here because they used synthetic speech as the stimuli and produced slightly different evoked responses than other ACC studies.

The ACC response provides specific advantages, including an index reflecting the neural encoding of temporal and spectral elements of speech that may underlie speech discrimination, a better signal-to-noise ratio than other auditory cortical responses, and test-retest reliability. To reduce testing time, Martin et al. proposed the efficiency of ACC stimuli, creating an experimental manipulation (i.e., interruption and continuous alternation) between two vowels, /u/ and /i/. Their results showed that the continuous stimuli are higher in efficiency than interrupted stimuli because of an enhanced elicited ACC response. The ACC response is more sensitive than other cortical potentials to changes within the signal, such as time-varying acoustic cues. Nevertheless, little research on the use of natural speech stimuli to identify distinct components of ACC exists. When systematically studied with various naturally spoken stimuli, the ACC has been useful in diagnosing and rehabilitating speech perception difficulties while also reflecting the neural representation of reliably consistent waveform patterns.

DISCUSSIONS AND CONCLUSIONS

A healthy ear effectively generates a neural representation with fidelity when it hears a speech sound. The speech signal is processed in the peripheral and at different levels of the brainstem and is subsequently conveyed to the auditory cortices via several stages of auditory pathway. Unlike those with healthy auditory systems, individuals with sensorineural hearing loss have elevated thresholds, broadened psychophysical tuning curves at the cochlea, and a reduction in phase locking, as well as less precision of neural synchrony at the level of the auditory nerve. Consequently, while traveling from the impaired ear to the auditory cortex, many spectral and temporal cues found in speech are distorted. Such processing makes it more challenging for hearing-impaired listeners to cope with incoming speech information and understand its meaning. Furthermore, since unexpected environmental noise masks acoustic features, speech perception in noisy environments contributes to hearing-impaired listeners misunderstanding of the cues. Therefore, researchers need to be aware of the relationship between a possible distortion...
tion in neural processing of spectral and temporal features in hearing-impaired listeners and their signal processing and speech perception abilities. A full understanding of this relationship will enable audiologists and scientists to seek a clearer explanation of why individuals with hearing loss have difficulty with speech perception in everyday listening conditions. When used for this purpose, SEPs offer a powerful technique for more effectively understanding speech perception of hearing-impaired listeners, especially in young infants and elderly patients who cannot be assessed reliably using standard behavioral measures. The clinical use of SEPs with these populations may be both feasible and informative.

The SEP literature suggests that the P1-N1-P2 complex and the ACC may be assessed for speech detection at the level of the auditory cortex as well as subcortical areas. The responses of two components have the advantage of estimating the listeners’ speech perception abilities at the cortical auditory level, while using the objective measures via evoked amplitude and latency as a marker for the neural coding of briefly varying timing cues and relatively salient features within speech signals.

SEPs are presently being extended for application to hearing-aid users and those who use CI as well as for applications of naturally produced and synthetic speech. Such an objective evaluation of the central neurophysiology process underlying speech perceptions of hearing-impaired listeners will be useful to audiologists and researchers when evoked potentials with naturally produced speech are reliably recorded. In addition, with natural speech stimuli for the SEPs measurement, the waveform of the cortical response is evoked by time-varying changes with a speech signal, indicating multiple positive and negative peaks that reflect acoustic changes across the speech stimulus.

Although speech discrimination tests by AEP is developing with a challenge, the SEPs literature is in its infancy, with relatively few studies to date. The limited data of current SEP studies (i.e., few speech stimuli) make it difficult to understand the overall speech perception of listeners with hearing loss. Therefore, future studies should consider the following four issues with regard to natural speech stimuli for SEPs to scrutinize speech cortical responses in hearing-impaired listeners:

1. Various speech stimuli such as consonant-vowel (CV), vowel-consonant (VC), and consonant-vowel-consonant (CVC) may be used to determine the SEP responses (i.e., latency, amplitude, and waveform) among both hearing-impaired and normal-hearing listeners.

2. The responses determined by talker differences due to gender (e.g., pitch) and clearness (i.e., VOT) in same speech stimuli should be examined.

3. Use of listeners’ temporal versus spectral cues for the SEP measure should be addressed.¹⁵

4. SEPs should be studied for differences in auditory cortical neural coding in different languages.

Note: Four published figures were mentioned in this paper to help the reader better understand the material. However, due to a permission issue, the author could not insert the figures directly into the paper.

REFERENCES


17. Martin BA, Tremblay KL, Korczak P. Speech-evoked potentials:
From the laboratory to the clinic. Ear Hear. 2008;29:285-313.


