The Use of Cortical Auditory Evoked Potentials to Evaluate Neural Encoding of Speech Sounds in Adults

Katrina Agung[*†]
Suzanne C. Purdy[*†‡]
Catherine M. McMahon[†]
Philip Newall[†]

Abstract

There has been considerable recent interest in the use of cortical auditory evoked potentials (CAEPs) as an electrophysiological measure of human speech encoding in individuals with normal as well as impaired auditory systems. The development of such electrophysiological measures such as CAEPs is important because they can be used to evaluate the benefits of hearing aids and cochlear implants in infants, young children, and adults that cannot cooperate for behavioral speech discrimination testing. The current study determined whether CAEPs produced by seven different speech sounds, which together cover a broad range of frequencies across the speech spectrum, could be differentiated from each other based on response latency and amplitude measures. CAEPs were recorded from ten adults with normal hearing in response to speech stimuli presented at a conversational level (65 dB SPL) via a loudspeaker. Cortical responses were reliably elicited by each of the speech sounds in all participants. CAEPs produced by speech sounds dominated by high-frequency energy were significantly different in amplitude from CAEPs produced by sounds dominated by lower-frequency energy. Significant effects of stimulus duration were also observed, with shorter duration stimuli producing larger amplitudes and earlier latencies than longer duration stimuli. This research demonstrates that CAEPs can be reliably evoked by sounds that encompass the entire speech frequency range. Further, CAEP latencies and amplitudes may provide an objective indication that spectrally different speech sounds are encoded differently at the cortical level.

Key Words: Cortical auditory evoked potentials, Ling sound test, speech evoked cortical potentials

Abbreviations: CAEPs = cortical auditory evoked potentials; VOT = voice onset time

Sumario

Ha existido un considerable interés reciente en el uso de los potenciales evocados auditivos corticales (CAEP) como una medida electrofisiológica de la codificación del lenguaje humano, en individuos con sistemas auditivos normales y alterados. El desarrollo de mediciones electrofisiológicas como los CAEP es importante, porque pueden ser usadas para evaluar los beneficios

[*]National Acoustic Laboratories, NSW, Australia; [†]Speech, Hearing and Language Research Centre, Division of Linguistics and Psychology, Macquarie University, NSW, Australia; [‡]Speech Science, Department of Psychology, University of Auckland, Auckland, New Zealand

Katrina Agung, Callier Advanced Hearing Research Center, Department of Brain and Behavioral Sciences, University of Texas at Dallas, 1966 Inwood Rd, Dallas, Texas 75235; Phone: 214-905-3104; E-mail: katrina.agung@utdallas.edu

de los audífonos o de los implantes cocleares en infantes, niños pequeños y adultos que no pueden cooperar en evaluaciones conductuales de discriminación del lenguaje. El presente estudio determinó si los CAEP producidos por siete diferentes sonidos de lenguaje, podrían diferenciarse entre sí con base en sus medidas de latencia y amplitud de la respuesta. Se registraron los CAEP de diez adultos con audición normal en respuesta a estímulos de lenguaje presentados a nivel de conversación (65 dB SPL) a través de altoparlantes. Se generaron respuestas corticales consistentes para cada uno de los sonidos lingüísticos en todos los participantes. Los CAEP producidos por estímulos de lenguaje dominados por energía de alta frecuencia fueron significativamente diferentes en amplitud que los CAEP producidos por sonidos dominados por bajas frecuencias. Se observaron también efectos significativos en la duración del estímulo, produciendo los estímulos de corta duración amplitudes mayores y latencias más tempranas que los estímulos de duración mayor. Esta investigación demuestra que los CAEP pueden ser evocados confiablemente por sonidos que comprenden todo el rango de frecuencias del lenguaje. Más aún, las latencias y amplitudes de los CAEP pueden aportar una indicación objetiva de que sonidos espectralmente diferentes están codificados en forma diferente a nivel cortical.

Palabras Clave: Potenciales evocados auditivos corticales, prueba de sonido de Ling, potenciales corticales evocados por lenguaje

Abreviaturas: CAEPs = potenciales evocados auditivos corticales; VOT = tiempo de iniciación de la voz

Cortical auditory evoked potentials (CAEPs) evoked by speech sounds have recently been investigated to determine the effect of phonologic and acoustic features on the cortical waveform (Crottaz-Herbette and Ragot, 2000) and to identify the cortical areas activated by these features (see Mäkela et al, 2003, for an example). This objective measure provides us with a tool to investigate the neurophysiological processes that underlie our ability to perceive speech (Purdy, Katsch, Sharma, et al, 2001; Purdy, Katsch, Storey, et al, 2001; Tremblay et al, 2003) and, ultimately, may allow us to better understand the neural encoding of speech in individuals with impaired auditory pathways (Eggermont and Ponton, 2003).

A number of studies have demonstrated that CAEPs can reliably be elicited by a variety of speech sounds including vowels (Obleser et al, 2003) and both synthetic (Martin and Boothroyd, 2000; Sharma et al, 2000; Tremblay et al, 2004) and naturally produced consonant-vowel syllables (Ostroff et al, 1998; Tremblay et al, 2003).[1] The morphology of these evoked potentials appears to correlate with acoustic features of speech. For example, in adults with normal hearing, increasing voice onset time (VOT) from 0–30 msec to 50–80 msec for the stimuli /da/ and /ta/ produces two negative peaks (N1 and N1') rather than the single negative-peaked CAEP observed for shorter VOTs (Sharma and Dorman, 1999). Further evidence for the impact of speech characteristics on CAEPs comes from the work of Ostroff et al (1998) who found that acoustic transitions from the consonant to the vowel in consonant-vowel (CV) monosyllables produce overlapping P1-N1-P2 complexes. Similarly, Tremblay et al (2003) found distinct cortical response patterns were elicited by syllables that differed in their initial phoneme /bi pi si ʃi/. Other studies have shown that cortical morphology correlates well with changes of spectral characteristics such as periodicity (Martin and Boothroyd, 1999) and amplitude (Martin and Boothroyd, 2000).

Another important finding is that the presence of CAEPs appears to correlate well with speech recognition ability in children with auditory neuropathy/dys-synchrony. In these children, pure-tone thresholds alone provide a poor indication of the ability to develop speech and language (Starr et al, 1996). Rance et al (2002) found that the development of "reasonable speech perception

performance" in children with auditory neuropathy was correlated with CAEPs of normal latency, amplitude, and morphology whereas the absence of the CAEP was associated with poor speech recognition scores. For these reasons, CAEPs are thought to reflect the functional integrity of the auditory pathways involved in processing of complex speech stimuli (Novak et al, 1989; Ostroff et al, 1998; Tremblay et al, 2003).

The development of spoken language in prelingually hearing-impaired individuals depends on the perception and discrimination of a broad range of speech sounds (Ling, 2002). In 1976, Ling identified a number of speech sounds with concentrations of energy that spanned the entire range of speech frequencies. Accordingly, it was suggested that the ability to identify and discriminate these sounds behaviorally was correlated with speech recognition and production (Cowan et al, 1990; Ertmer et al, 2002; Ling, 2002; Wei et al, 2000). Investigations of the speech of deaf adults have also shown correlations between the perception of speech sounds and speech production (e.g., Palethorpe et al, 2003) where hearing-impaired adults who are unable to identify and discriminate speech sounds behaviorally typically demonstrate poor speech recognition (Hornsby and Ricketts, 2003). The Ling sound test is typically used as a quick test to verify audiometric threshold assessment and/or the fitting of hearing aids or cochlear implants (Agung et al, 2005). It is also used to assess the ability of the listener to detect and discriminate between six different speech sounds that largely encompass the speech spectrum. In difficult-to-test populations, however, behavioral responses to assess discrimination ability cannot always be obtained. Therefore, the main objective of the current study was to determine whether CAEPs that do not require the active cooperation of the participant could be used as an electrophysiological test to evaluate encoding of the Ling sounds. The Ling six sounds comprise the vowels /a/ as in *car*, /u/ as in *two*, and /i/ as in *she*, along with the phonemes /s/ as in *us*, /ʃ/ as in *fish*, and /m/ as in *me*. The vowel /ɔ/ was added for the Australian population (see Agung et al, 2005, for a review). We hypothesized that because these sounds are spectrally distinct, they may evoke CAEPs with different morphological characteristics. To a first approximation, this might provide us with an objective measure of the ability to detect *and* discriminate between each of these different speech sounds.

Durations of speech stimuli used in previous studies vary widely, ranging from 90 to 600 msec for synthetic speech stimuli (e.g., Sharma et al, 1997; Obleser et al, 2001). Picton et al (2000) recommended that natural speech be used for evoked potential research, since the goal is to apply results to speech perception in everyday life. The durations of naturally produced speech stimuli can vary widely (Ladefoged, 1993) and would typically exceed the durations of tonal stimuli regarded as optimal for CAEP recordings. Very short stimulus durations (less than 100 msec) are regarded as optimal for tone-evoked CAEPs (Stapells, 2002). Stimulus durations are variable in CAEP studies using naturally produced stimuli, with durations ranging from 300 msec (Ostroff et al, 1998) to 756 msec (Tremblay et al, 2003). The influence of speech stimulus duration on CAEPs was also investigated in the current study, since there seems to be no consensus in the literature regarding optimal stimulus durations for speech-evoked CAEP recordings.

Thus, the aims of the current study were to determine if (1) CAEPs can be recorded in normally hearing, awake adults to a range of suprathreshold speech sounds that encompass the range of speech frequencies; (2) CAEPs show significant differences in response latencies and amplitudes between speech stimuli; and (3) short or long stimulus durations are optimal for speech-evoked CAEP recording.

METHODOLOGY

Participants

Cortical auditory evoked potentials were recorded from ten adults, five females, and five males, ranging in age from 20 to 29 years of age (mean 23.4, standard deviation 3.2 years). All participants had pure-tone air-conduction thresholds less than 20 dB HL at octave frequencies from 250–8000 Hz and bone-conduction thresholds less than 20 dB HL at octave frequencies from 500–4000 Hz. Participants reported no history of neurologic problems.
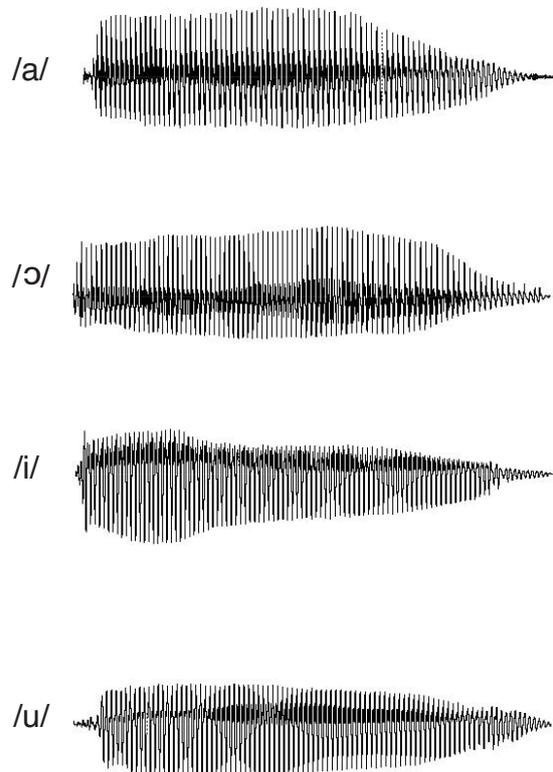
Stimuli

Speech stimuli used to elicit the CAEPs were naturally produced by a native female speaker of Australian English. The speaker was instructed to produce the same speech sounds with a range of durations. Stimuli consisted of four vowels: /i/ as in *heed*, /a/ as in *hard*, and /u/ as in *who'd* and /ɔ/ as in *hoard*. As shown in Table 1, generally all vowels had first and second formants that were within two standard deviations of the mean values for Australian English female speakers (Cox, 1996). Electric time waveforms of the stimuli are shown in Figure 1. Consonant speech stimuli included /m/, which is dominated by low-frequency spectral energy, /s/, which is dominated by high-frequency spectral energy, and /ʃ/, which is dominated by mid–high-frequency spectral energy (see Figure 2). All speech sounds were windowed using 20 msec linear rise and fall times.

Two stimulus durations of 500 msec and 100 msec (±5 msec) were used for each stimulus. Stimuli that were 100 msec (±5 msec) in duration were all shortened from their original length of approximately 200 msec, using a zero crossing technique to avoid audible clicks. Stimuli that were 500 msec (±5 msec) in duration were naturally produced. The interstimulus interval was 1125 msec, and stimuli were presented with alternating onset polarity. Stimuli were presented in two blocks of 100 per stimulus, with stimulus presentation order randomized.

Equipment and Procedure

The Neuroscan STIM and SCAN (version 4.2) evoked potential system was used for stimulus generation and CAEP recording. Stimuli were presented bilaterally at 65 dB


Figure 1. Electrical time waveforms of vowel stimuli for the 500 msec stimulus duration.

SPL via two loudspeakers at 45 and 315 degrees azimuth respectively and 1.5 m distance. Because larger amplitude responses are recorded from midline scalp locations and neural patterns to speech stimuli have been reported to be similar at electrode locations surrounding the midline (Tremblay et al, 2003), we recorded from the midline location, Cz. Specifically, gold cup electrodes were placed at Cz (vertex) and A2 (right earlobe, reference), with a ground electrode on the forehead. Eye blinks were recorded via

Table 1. Center Frequencies for Female Speakers of Australian English and the Female Talker in the Current Study

| Stimulus | Average center frequency for female speakers of Australian English | | Center Frequency for Female Talker in the Current Study | | | |
			500 msec duration		100 msec duration	
	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
/a/	955 (96)	1525 (105)	860	1423	856	1416
/i/	391 (43)	2729 (150)	440	2849	405	2823
/u/	399 (41)	2138 (153)	391	1342	352	1372
/ɔ/	494 (66)	954 (97)	507	944	473	843

Note: In the two left columns, standard deviations are in brackets (Cox, 1996). Generally vowel stimuli were within two standard deviations of the average F1 and F2 values.
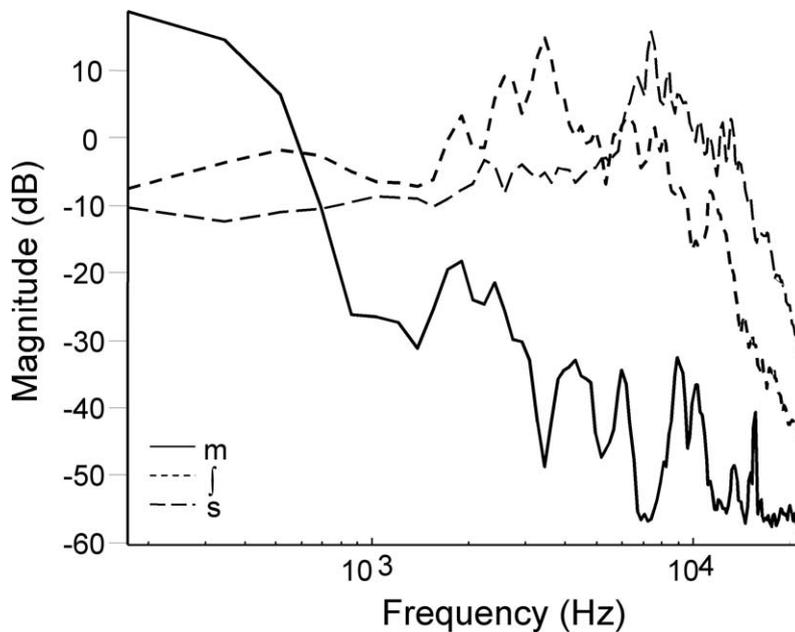
Figure 2. Power spectra of the consonant stimuli, showing the substantial differences in the region of dominant spectral energy across the consonants. The results are shown for the 500 msec duration stimuli, but the spectra were essentially identical for short and long consonants.

a noninverting electrode above the right eyebrow and an inverting electrode on the right earlobe (Kraus et al, 1993). Electrode impedances were maintained below 5 kohms. Participants were seated comfortably in a recliner chair in a sound-treated booth. To keep alert during testing, participants watched a self-chosen movie with subtitles but without sound.

Data Analysis

EEG files had a recording window of 700 msec (including a prestimulus time period of 100 msec). The EEG signals were filtered online (0.1–100 Hz, 24 dB/octave slope) and offline (0.1–30 Hz, 24 dB/octave slope). Artifact rejection was used online to exclude responses exceeding ±75 µV. Linear detrend and baseline correction was used for each response. Grand average CAEP waveforms for each subject were created from the two blocks for each stimulus type and duration. After artifact rejection, each subject's average waveforms were based on 200 responses per stimulus.

CAEP peak amplitudes and latencies were identified for each subject by two independent observers. Separate repeated-measures analyses of variance were performed to determine the effects of stimulus type and duration on the latencies and amplitudes of each peak (P1, N1, and P2). The amplitude of P1 was defined as the largest positive deflection occurring between 50–100 msec after stimulus onset. The amplitude of N1 was identified as the largest negative deflection between 80–120 msec after stimulus onset. P2 amplitude was defined as the largest peak occurring between 150–200 msec (Stapells, 2002). The latency of the peak was measured at the center of the peak. When the waveform contained a double peak of equal amplitude or a peak with a plateau, the latency was measured at the midpoint of the peak.

RESULTS

All subjects showed cortical responses to all stimuli. Grand averaged cortical responses elicited by each speech sound for the 100 and 500 msec stimulus durations are shown in Figure 3. Repeated-measures ANOVA showed no significant effect of duration for P1 amplitude [$F(1, 9) = 4.79$, $p < 0.0563$]. P1 latencies were significantly
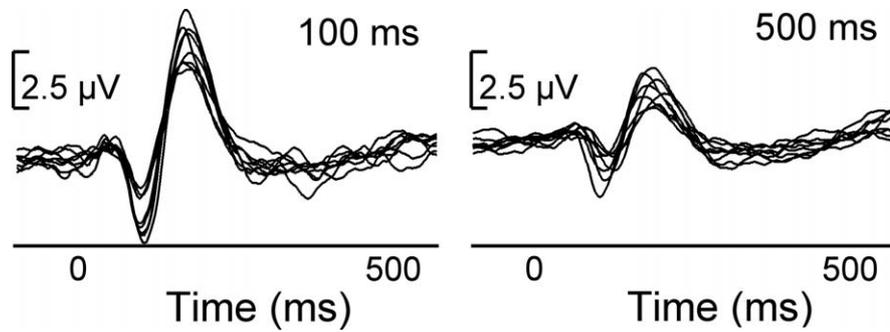
Figure 3. Grand average CAEP waveforms for all speech stimuli. N1 and P2 peaks were significantly larger in amplitude and earlier in latency for cortical responses elicited by shorter (100 msec) duration stimuli (left) compared to longer (500 msec) duration stimuli (right).

earlier for shorter compared to longer stimulus durations [$F(1, 9) = 47.20$, $p < 0.0001$].

As illustrated in Figure 3, there was a main effect of duration whereby shorter stimulus duration resulted in significantly larger N1 [$F(1, 9) = 17.06$, $p < 0.0026$] and P2 amplitudes [$F(1, 9) = 27.98$, $p < 0.0005$]. CAEP latencies also differed significantly between long and short durations for N1 [$F(1, 9) = 92.24$, $p < 0.0001$] and P2 [$F(1, 9) = 20.91$, $p < 0.0013$] whereby shorter stimulus durations elicited earlier N1 and P2 response latencies.

To determine whether responses differed between stimuli, short and long durations were analyzed separately. A one-way repeated measures ANOVA showed that N1 amplitudes were significantly different across stimuli; this occurred for both the 100 msec [$F(6,54) = 7.57$, $p < 0.0001$] and the 500 msec stimulus durations [$F(6,54) = 11.84$, $p < 0.0001$]. Similarly, repeated-measures ANOVA showed significant P2 amplitude differences across stimuli for both longer [$F(6,54) = 7.22$, $p < 0.0001$] and shorter [$F(6,54) = 2.97$, $p < 0.0139$] durations. There was no significant main effect of stimulus type on P1 amplitude for either the shorter [$F(6, 54) = 0.75$, $p < 0.6150$] or longer [$F(6, 54) = 0.95$, $p < 0.4688$] stimulus duration, however.

As documented in Table 2, Newman-Keuls post hoc comparisons for the longer stimulus durations showed that N1 and P2

Table 2. Results of Newman-Keuls Post Hoc Tests for N1 and P2 Amplitudes

N1	/m/	/u/	/a/	/i/	/sh/	/s/	/ɔ/
/m/		0.8701	0.2652	0.0023**	0.0326*	0.0289*	0.0377*
/u/	0.5623		0.4068	0.0026**	0.0274*	0.0288*	0.0217*
/a/	0.3732	0.8968		0.0180*	0.0021**	0.0016**	0.0033**
/i/	0.4562	0.6533	0.8302		0.0001**	0.0001**	0.0001**
/sh/	0.0005**	0.0048**	0.0052**	0.0093*		0.8316	0.7693
/s/	0.0002**	0.0013**	0.0012**	0.0031**	0.5527		0.8675
/ɔ/	0.1413	0.3823	0.4652	0.3812	0.0330*	0.0198*	

P2	/m/	/u/	/a/	/i/	/sh/	/s/	/ɔ/
/m/		0.5375	0.4756	0.4447	0.0593	0.0061**	0.0868
/u/	0.9969		0.6878	0.5493	0.0361*	0.0016**	0.0361*
/a/	0.5158	0.1970		0.8232	0.0116*	0.0003**	0.0084*
/i/	0.9986	0.9414	0.3583		0.0138*	0.0004**	0.0112*
/sh/	0.1994	0.3745	0.0325*	0.3119		0.2933	0.8139
/s/	0.0971	0.3013	0.0266*	0.2180	0.9584		0.2076
/ɔ/	0.3014	0.4493	0.0403*	0.4034	0.9817	0.9969	

Note: Upper table: Significant N1 amplitude differences evoked by the 500 msec stimuli are shown in the top, right-hand half of the table, while significant differences for the 100 msec stimuli are shown in the bottom, left-hand half of the table. Lower table: P2 amplitude differences evoked by the 500 msec stimuli are shown in the upper, right-hand half of the table, while the lower, left-hand half of the table shows P2 amplitude differences evoked by the 100 msec stimuli.

*A difference at the 0.05 significance level.

**A difference at the 0.007 significance level (Bonferroni adjustment to control for type I errors).

response amplitudes elicited by higher frequency speech stimuli (/s, ʃ/) produced significantly smaller amplitudes compared to stimuli that had dominant spectral energy in lower frequencies (/m, a, u, i/). To control for a type I error, Bonferroni adjustment was performed (see Table 2). Similarly, for the shorter stimulus duration, N1 amplitudes were significantly smaller for the higher-frequency sounds /s, ʃ/. P2 amplitudes elicited by the shorter stimulus durations also showed a similar trend, but only the CAEP produced by the vowel /a/ was significantly larger compared to the higher-frequency sounds.

A repeated measures ANOVA for N1 latencies yielded a main effect of stimulus type for both the 100 msec [$F_{(6,54)} = 4.41$, $p < 0.0011$] and 500 msec [$F_{(6,54)} = 19.82$, $p < 0.0001$] stimulus durations. A repeated-measures ANOVA for P2 latencies showed significant stimulus effects for P2 latency evoked by the 500 msec stimulus duration [$F_{(6,54)} = 9.77$, $p < 0.0001$] but not for the 100 msec stimulus duration [$F_{(6, 54)} = 1.04$, $p < 0.4121$]. Similarly, there was a significant main effect of stimulus for P1 latencies evoked by the 500 msec duration [$F_{(6, 54)} = 10.32$, $p < 0.0001$] but not for the 100 msec duration [$F_{(6, 54)} = 1.05$, $p < 0.4024$]. As shown in Table 3, Newman-Keuls tests (with Bonferroni adjustment) revealed that P1, N1, and P2 latencies evoked by the higher-frequency sounds /s, ʃ/ occurred significantly later than all other sounds with the exception of /u/ for the longer stimulus duration (illustrated in Figure 4). In addition, there were significant differences in N1 latency for all vowels ($p \leq 0.0500$), whereby N1 decreased in latency systematically when elicited by /u/, /ɔ/, /a/, and /i/ (see Figure 5). There was no significant difference between N1 latencies evoked by /a/ and /ɔ/, however. The same trend was also observed for N1 evoked by the shorter duration vowels; however, only /u/ and /i/ were significantly different in response latency (see Figure 5). Similarly, P1 and P2 elicited by the longer duration vowels /u/, /a/, /ɔ/, and /i/ decreased in latency in this order (see Table 3).

DISCUSSION

The main aim of this study was to determine whether different speech phonemes produced significant differences in CAEP morphology in adults with normal hearing. Robust CAEP recordings were obtained in all participants for both stimulus durations and the seven different speech sounds spanning the speech frequency range. The speech-evoked CAEP waveforms were dominated by N1 occurring at about 125 msec and P2 occurring at about 180 msec.
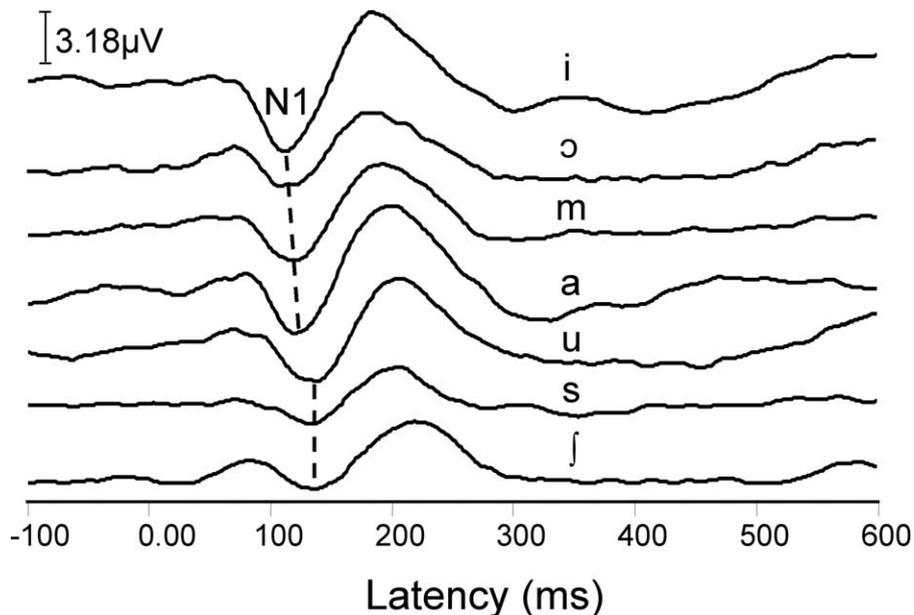
Figure 4. Grand average waveforms for the 500 msec stimuli showing the shift in latency across vowels and between consonants with low- and high-frequency emphasis.
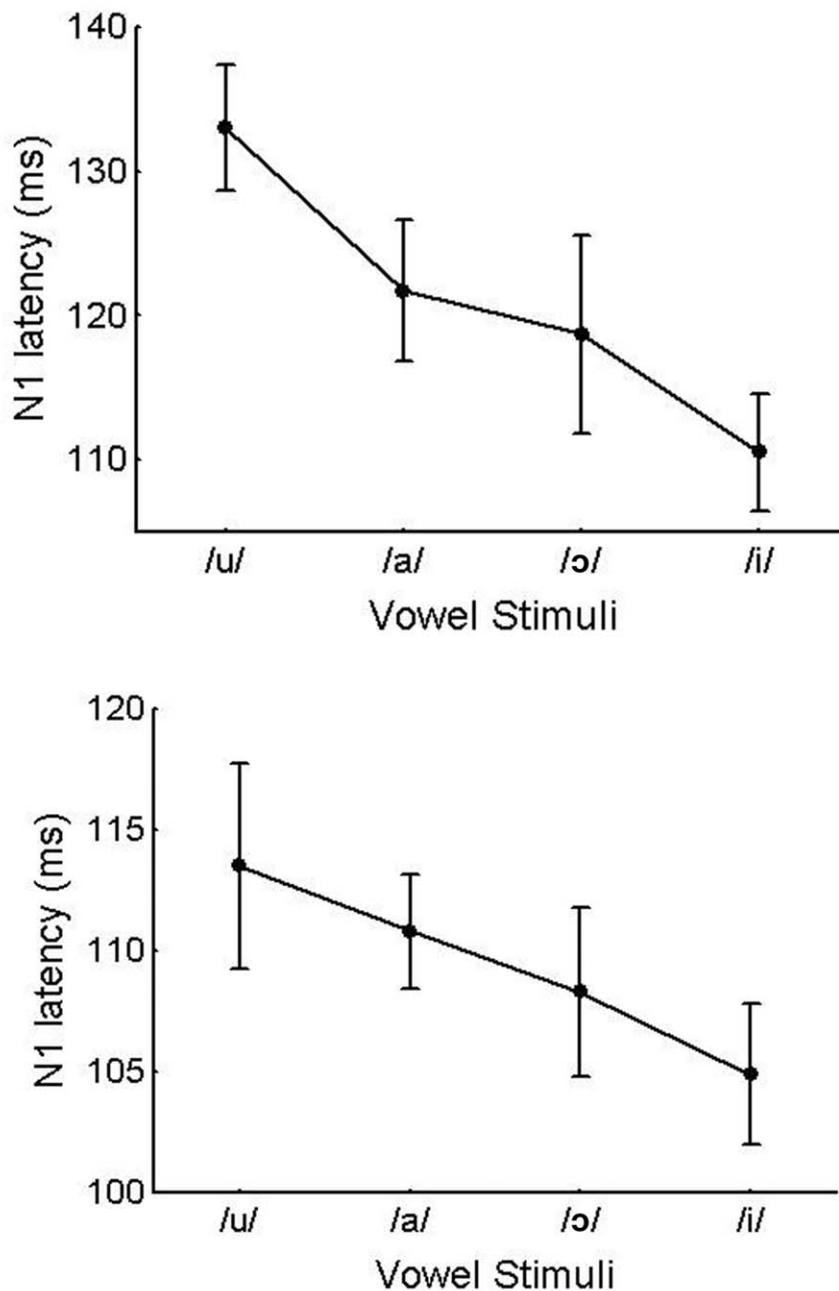
Figure 5. Top: Average N1 latencies for the vowel stimuli for the 500 msec duration stimuli. Error bars show standard deviations. Bottom: Average N1 latencies for the vowel stimuli for the 100 msec duration stimuli. Error bars show standard deviations

There were differences in CAEP amplitudes and latencies across stimuli. These differences were statistically significant for speech sounds with greater spectral and temporal differences; however, the differences were not significantly different for all of the seven speech sounds.

A second aim of the study was to determine stimulus duration effects on the speech-evoked CAEP. The most obvious effect of stimulus duration was the reduction in overall N1 and P2 amplitude for the 500 msec speech sounds compared to the shorter duration 100 msec stimuli. This is evident in the grand average waveforms (see Figure 2) for all speech sounds.

Table 3. Results of the Newman-Keuls Test

P1	/m/	/u/	/a/	/i/	/sh/	/s/	/ɔ/
/m/		0.0037**	0.0540	0.5548	0.0002**	0.0055**	0.8097
/u/	0.6333		0.2230	0.0010**	0.1704	0.9826	0.0040**
/a/	0.4770	0.8822		0.0223*	0.0177*	0.4267	0.0373*
/i/	0.9044	0.5061	0.6329		0.0001**	0.0013**	0.6821
/sh/	0.6197	0.7981	0.8265	0.5916		0.0767	0.0002**
/s/	0.7841	0.5864	0.5852	0.9141	0.6219		0.0071
/ɔ/	0.9285	0.6269	0.7009	0.7562	0.6898	0.9273	

N1	/m/	/u/	/a/	/i/	/sh/	/s/	/ɔ/
/m/		0.0002**	0.4096	0.0119*	0.0001**	0.0003**	0.8089
/u/	0.8899		0.0008**	0.0001**	0.6288	0.5345	0.0002**
/a/	0.2703	0.4276		0.0016**	0.0004**	0.0018**	0.3021
/i/	0.0041**	0.0037**	0.0613		0.0001**	0.0001**	0.0169*
/sh/	0.4276	0.5095	0.8899	0.0566		0.5112	0.0001**
/s/	0.0388*	0.0380*	0.2798	0.3575	0.2257		0.0003**
/ɔ/	0.1169	0.1273	0.4818	0.2638	0.3118	0.5187	

P2	/m/	/u/	/a/	/i/	/sh/	/s/	/ɔ/
/m/		0.0062**	0.6253	0.1500	0.0014**	0.0269*	0.7623
/u/	0.9628		0.0130*	0.0002**	0.5153	0.4436	0.0040**
/a/	0.8024	0.9995		0.0920	0.0038**	0.0341*	0.7075
/i/	0.5966	0.3503	0.7642		0.0001**	0.0004**	0.1176
/sh/	0.4779	0.7380	0.5367	0.2520		0.3344	0.0008
/s/	0.8716	0.9579	0.6395	0.6012	0.6568		0.0224*
/ɔ/	0.9094	1.0000	0.9743	0.6162	0.6422	0.8699	

Note: Each table shows latency differences for the 500 msec stimuli in the top, right-hand half of the table, while latency differences evoked by the 100 msec stimuli are shown in the bottom, left-hand half of the table. Upper, middle, and lower tables show latency differences for the 500 and 100 msec stimuli for P1, N1, and P2, respectively.

*A difference at the 0.05 level.

**A difference at the 0.007 level (Bonferroni adjustment to control for type I errors).

Effects of Stimulus Characteristics

In the current study, speech sounds dominated by higher-frequency spectral energy, such as /s/ and /ʃ/, elicited CAEPs with smaller N1 and P2 amplitudes than speech sounds with dominant spectral energy in the lower frequencies. These findings are consistent with Shestakova et al's (2004) results for the magnetic equivalent of N1, N1m, in response to multiple exemplars in the /a/, /i/, and /u/ vowel categories. Shestakova et al found that N1m amplitude and source locations differed between vowel categories, and vowels with similar spectral envelopes had closer cortical representations than those where spectral differences were greatest.

Similar findings have been reported for tonal stimuli, with low-frequency tones eliciting significantly larger cortical response amplitudes than higher-frequency tones (Jacobson et al, 1992). The current study's findings suggest that N1 and P2 evoked by speech sounds may reflect, at least in part, the tonotopic organization of auditory cortical areas. Cortical areas that respond to low-frequency auditory information are located more superficially (i.e., closer to the surface of the scalp) than cortical regions for higher frequencies (Yetkin et al, 2004). Low-frequency speech sounds may therefore activate more superficial cortical regions and produce larger amplitude cortical responses than higher-frequency speech sounds, when surface scalp recording electrodes are used. It is important to note, however, that there was not just a simple frequency effect on CAEP amplitude. For example, the vowel dominated by the lowest-frequency spectral energy /ɔ/ produced N1 and P2 amplitudes that were not significantly smaller than those for /s/ and /ʃ/. Thus, there was no simple effect of frequency, which is not surprising given the spectral complexity of the stimuli and the evidence for multiple tonotopically organized areas in the human auditory cortex

(Engelien et al, 2002; Talavage et al, 2004).

In contrast to N1 and P2, P1 amplitude did not differ significantly between speech sounds. One reason for the differing effects of spectral changes on P1 versus N1 and P2 amplitude may be because it is difficult to measure stimulus effects on P1 amplitude due to its small size in adults. This may be a result of phase cancellation of P1 by the large N1 in the adult CAEP waveform (Ponton et al, 2002). Another reason for the differing effects of spectral changes on P1 versus N1 and P2 amplitude may be because a major source of activity for P1 is thought to be the lateral portion of Heschl's gyrus (Liegeois-Chauvel et al, 1994). The lateral portion of Heschl's gyrus appears to be sensitive to changes in *temporal* complexity, whereas the processing of *spectral* cues for simple and complex stimuli is thought to occur *between* the lateral portion of Heschl's gyrus and the supratemporal plane (Hall et al, 2002). Since the supratemporal plane is an important site for N1 generation (Liegeois-Chauvel, 1994; Virtanen, 1998; Godey et al, 2001), this could account for the differing effects of changes in the speech stimulus on P1 and N1.

Vowel Contrasts

This study demonstrated that high front vowels such as /i/ evoke CAEPs that have earlier latencies than CAEPs for low mid-back vowels /u/. The CAEP for the midvowel /a/ occurred between these two latencies. Similar finding have been reported by Obleser and colleagues (2004), who found that N1m peaked 5 msec later for the back vowel /o/ compared to the front vowel [ø]. These differences in latencies may be due to changes in cortical activation in response to the phonological features of the vowels, whereby front vowels were found to activate a more inferior and anterior source compared to back vowels (Obleser et al, 2004). Makela et al (2005) found that N1m latencies over the right hemisphere occurred significantly earlier for the mid vowel /a/ compared to the back vowel /u/. Thus, it appears that cortical representations of vowels reflect the phonological features of speech.

However, phonological features alone do not account for N1 and P2 latency differences observed in the current study. In Australian English, /ɔ/ is further retracted than /u/ (see Table 2); however, we found that the

low-back vowel /ɔ/ did not have the latest latency, but rather, it occurred at a similar latency to the midvowel /a/, in between the latencies for high-front /i/ and low mid-back vowel /u/ (see Figure 5). One explanation is that vowels with large F2-F1 differences such as /i/ (~2300 Hz) and /u/ (~1700 Hz) have larger areas of activation and therefore elicit a response that occurs at a different time compared to a vowel with a small F2-F1 distance such as /a/ and /ɔ/. This is supported by Makela et al (2003), who found that different vowels with equal F2-F1 differences produced N1m peaks that did not differ in latency, although the vowels were found to activate distinctly separate areas in the left hemisphere of the auditory cortex. Thus, F2-F1 differences may account at least in part for the latency differences observed for different vowels in the current study.

Other Explanations for Stimulus Effects on CAEPs

Studies investigating the effect of tonal stimulus frequency on N1 latency have shown varying results (see Cone-Wesson and Wunderlich, 2003, for a review). Jacobson et al (1992) reported that N1 in response to 4 kHz high-frequency tone bursts occurred earlier than N1 to 500 Hz, low-frequency tone bursts, whereas an earlier study by Zerlin and Naunton (1974) found longer N1 latencies for 4000 Hz than for 1000 Hz stimuli. There was no simple pattern of latency differences across speech stimuli in the current study. This may be because, unlike tonal data where the stimuli differ only in terms of frequency, naturally produced speech sounds are complex and differ in a number of ways other than frequency. Factors other than spectral differences that may influence speech-evoked CAEPs include bandwidth (Seither-Preisler et al, 2003) and temporal differences (Trebuchon-Da Fonseca et al, 2005). For instance, the speech sound /m/ that is dominated by low-frequency spectral energy differs from the mid-high frequency sound /ʃ/ in periodicity and nasality, as well as frequency content.

Duration Effects

P1, which is thought to have a primary generator in the lateral portion of Heschl's

gyrus, was sensitive to durational differences as reflected in P1 latency that occurred significantly earlier when evoked by the shorter stimulus versus the longer stimulus duration. This finding agrees with previous studies that have shown Heschl's gyrus to be sensitive to differences in voice onset time and temporal regularity (Steinschneider et al, 1999; Gutschalk et al, 2004). However, P1 did not differ significantly in amplitude between the two stimulus durations for any stimuli. As noted above, it may be difficult to show stimulus effects on P1 amplitude in adults due to the waveform being dominated by N1 and P2.

On the other hand, there were clear effects of duration on N1 and P2 latency *and amplitude*, with shorter duration stimuli eliciting responses that were significantly larger in amplitude and earlier in latency. Such an effect contrasts with other reports that demonstrate that increasing stimulus duration up to approximately 30–40 msec typically leads to *increased* N1 amplitudes with longer durations producing minimal change in CAEP amplitude (Davis and Zerlin, 1966; Onishi and Davis, 1968; Skinner and Jones, 1968; Picton et al, 1977; Alain et al, 1997; Gage and Roberts, 2000). The current study compares changes in CAEP amplitudes produced by longer stimulus durations than those investigated in these earlier studies, however.

Increased latencies and reduced amplitudes produced by repeated stimulus presentation have been previously observed for N1 and have been assumed to result from either neural refractoriness (Ritter et al, 1968) or habituation (Budd et al, 1998). Neural refractoriness refers to the time course of neural recovery after a period of excitation and is clearly demonstrated with decreasing interstimulus interval (Naatanen and Picton, 1987), whereas habituation occurs with repeated stimulus presentation (Ritter et al, 1968). In a review of early CAEP research, Picton (1990) noted that the relative refractory period of the P1-N1-P2-N2 response is very slow and lasts more than 10 sec. Picton et al (1977) showed enhancement of CAEP amplitudes with reductions in interstimulus interval from 0.5/sec to 0.1/sec (once every 10 sec). Similarly, Tremblay et al (2004) found small P1 and N1 latency increases and amplitude reductions as ISI was reduced from 1510 to 510 to msec for a

speech and tonal stimulus (180 msec duration) in younger adults. While the ISI and stimulus durations used in the current study are similar to those used by Tremblay et al (2004), the ISI remained the same for the different stimulus durations. Furthermore, the number of samples acquired per average was also identical for the 100 msec and 500 msec stimuli. If we assume that the neural population responding to each of these stimuli was the same, then there should be no change in the amplitude or latency of the evoked waveform.

The existence of duration-sensitive neurons in auditory cortex may explain the reduced amplitudes and increased latencies of N1 and P2 when stimulus duration was increased. Based on their study of stimulus duration effects on tone-evoked N1 in humans, Alain et al (1997) concluded that there are differences in duration sensitivity between different areas of auditory cortex. This is consistent with extracellular recordings in the inferior colliculus of the big brown bat, *Eptesicus fuscus* (Faure et al, 2003) and chinchilla (Chen, 1998) and the auditory cortex of the cat (He et al, 1997) that have revealed the presence of duration-sensitive neurons. These neurons must have different response properties to encode different sound durations; therefore, it is possible that differences also exist in their recovery properties after the offset of the sound stimulus, resulting in different durations of neural refractoriness.

Another possible explanation is that the stimuli used were naturally produced, and hence there were some differences in the temporal and spectral characteristics of the short versus long duration phonemes. The smaller amplitudes and later latencies for the longer duration stimuli may relate to the longer steady-state portion in the 500 msec stimuli. Prolonged auditory stimuli can evoke a *negative* baseline shift in the CAEP waveform through the duration of the stimulus (Picton, 1990). This "sustained potential" was seen by Picton et al (1978a; 1978b) in response to long-duration tonal stimuli (700–1000 msec) and had an onset that occurred between N1 and P2. Thus, the reduction in P2 amplitude seen for the longer duration stimuli in the current study could be due to a sustained potential that overlaps with P2. This does not explain the reduction in N1 amplitudes for the longer duration

stimuli, however. Since the stimuli were naturally produced, the N1 differences may relate to differences in the first 100 msec of the stimuli. N1 is largely determined by stimulus onset characteristics (e.g., Skinner and Jones, 1968; Elfner et al, 1976). There were minimal differences in the onset portion of the acoustic waveforms of the speech stimuli, and hence, other explanations for the effect of stimulus duration on N1, such as differences in response properties of neurons with differing duration sensitivity, seem more likely.

CONCLUSIONS

Cortical auditory evoked responses can be elicited by speech sounds that encompass the entire speech frequency range, with shorter duration sounds producing CAEPs with larger amplitudes and shorter latencies than longer duration sounds. Speech sounds that were dominated by high-frequency energy produced CAEPs that are significantly different in terms of N1 and P2 amplitude when compared to sounds dominated by lower-frequency energy. While CAEPs may be used to objectively measure differences in neural encoding and perception of spectrally different speech sounds, the differences between CAEPs are not sufficient to use as a measure of discrimination for each of the Ling seven sounds. Future research should be directed toward exploring how other features of speech, such as temporal characteristics, also influence the cortical response.

NOTES

1. Most studies of speech-evoked CAEPs have used CV syllables or words, rather than isolated phonemes.

REFERENCES

Agung KB, Purdy SC, Kitamura C. (2005) The Ling sound test revisited. *Aust N Z J Audiol* 27(1):33–41.

Alain C, Woods D, Covarrubias D. (1997) Activation of duration-sensitive auditory cortical fields in humans. *Electroencephalogr Clin Neurophysiol* 104(6):531–539.

Budd TW, Barry RJ, Gordon E, Rennie C, Michie PT. (1998) Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *Int J Psychophysiol* 31:51–68.

Chen GD. (1998) Effects of stimulus duration on responses of neurones in the chinchilla inferior colliculus. *Hear Res* 122:142–150.

Cone-Wesson B, Wunderlich J. (2003) Auditory evoked potentials from the cortex: audiology applications. *Curr Opin Otolaryngol Head Neck Surg* 11(5):372–377.

Cowan RS, Blamey PJ, Galvin KL, Sarant JZ, Alcantara JI, Clark GM. (1990) Perception of sentences, words, and speech features by profoundly hearing-impaired children using a multichannel electrotactile speech processor. *J Acoust Soc Am* 88:1374–1384.

Cox FM. (1996) An acoustic analysis of vowel variation in Australian English. PhD diss., Macquarie University, Sydney.

Crottaz-Herbette S, Ragot R. (2000) Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clin Neurophysiol* 111(10):1759–1766.

Davis H, Zerlin S. (1966) Acoustic relations of the human vertex potential. *J Acoust Soc Am* 39:109–116.

Eggermont JJ, Ponton CW. (2003) Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: correlations with changes in structure and speech perception. *Acta Otolaryngol* 123:249–252.

Elfner LF, Gustafson DJ, Williams KN. (1976) Signal onset and task variables in auditory evoked potentials. *Biol Psychol* 4:197–206.

Engelien A, Yang Y, Engelien W, Zonana J, Stern E, Silbersweig DA. (2002) Physiological mapping of human auditory cortices with a silent event-related fMRI technique. *Neuroimage* 16(4):944–953.

Ertmer DJ, Leonard JS, Pachuilo ML. (2002) Communication intervention for children with cochlear implants: two case studies. *Lang Speech Hear Serv Sch* 33:205–217.

Faure PA, Fremouw T, Casseday JH, Covey E. (2003) Temporal masking reveals properties of sound-evoked inhibition in duration-tuned neurons of the inferior colliculus. *J Neurosci* 23(7):3052–3065.

Gage N, Roberts T. (2000) Temporal integration: reflections in the M100 of the auditory evoked field. *Neuroreport* 11(12):2723–2726.

Godey B, Schwartz D, de Graaf JB, Chauvel P, Liegeois-Chauvel C. (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin Neurophysiol* 112:1850–1859.

Gutschalk A, Patterson RD, Scherg M, Uppenkamp S, Rupp A. (2004) Temporal dynamics of pitch in human auditory cortex. *Neuroimage* 22(2):755–766.

Hall D, Johnsrude I, Haggard M, Palmer A, Akeroyd M, Summerfield A. (2002) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 12(2):140–149.

He JF, Hashikawa T, Ojima H, Kinouchi Y. (1997) Temporal integration and duration tuning in the

dorsal zone of the cat auditory cortex. *J Neurosci* 17:2615–2625.

Hornsby BW, Ricketts TA. (2003) The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding. *J Acoust Soc Am* 113:1706–1717.

Jacobson GP, Lombardi DM, Gibbens ND, Ahmad BK, Newman CW. (1992) The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear Hear* 13:300–306.

Kraus N, McGee T, Carrell TD, Sharma A. (1995) Neurophysiologic bases of speech discrimination. *Ear Hear* 16:19–37.

Kraus N, McGee T, Micco A, Sharma A, Carrell T, Nicol T. (1993) Mismatch negativity in school-age children to speech stimuli that are just perceptibly different. *Electroencephalogr Clin Neurophysiol* 88:123–130.

Ladefoged P. (1993) *A Course in Phonetics.* Orlando: Harcourt Brace College Publishers.

Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P. (2004) Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14(7):731–740.

Ling D. (2002) The Ling six-sound test. Paper presented at the Alexander Graham Bell Convention, St Louis, June 29–July 2.

Mäkelä A, Alku P, May P, Mäkinen V, Tiitinen H. (2005) Left-hemispheric brain activity reflects formant transitions in speech sounds. *Neuroreport* 16(6):549–553.

Mäkelä A, Alku P, Tiitinen H. (2003) The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neurosci Lett* 353(2):111–114.

Martin BA, Boothroyd A. (1999) Cortical auditory evoked potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear* 20:33–44.

Martin BA, Boothroyd A. (2000) Cortical auditory evoked potentials in response to changes of spectrum and amplitude. *J Acoust Soc Am* 107:2155–2161.

Näätänen R, Picton T. (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24(4):375–425.

Novak GP, Kurtzberg D, Kreuzer JA, Vaughan HG Jr. (1989) Cortical responses to speech sounds and their formants in normal infants: maturational sequence and spatiotemporal analysis. *Electroencephalogr Clin Neurophysiol* 73:295–305.

Obleser J, Elbert T, Lahiri A, Eulitz C. (2003) Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cogn Brain Res* 15:207–213.

Obleser J, Eulitz C, Lahiri A. (2004) Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J Cogn Neurosci* 16(1):31–39.

Obleser J, Eulitz C, Lahiri A, Elbert T. (2001) Gender differences in functional hemispheric asymmetry during processing of vowels as reflected by the human brain magnetic response. *Neurosci Lett* 314:131–134.

Onishi S, Davis H. (1968) Effects of duration and rise time of tone bursts on evoked V potentials. *J Acoust Soc Am* 44(2):582–591.

Ostroff JM, Martin BA, Boothroyd A. (1998) Cortical evoked response to acoustic change within a syllable. *Ear Hear* 19:290–297.

Palethorpe S, Watson C, Barker R. (2003) Acoustic analysis of monophthong and diphthong production in acquired severe to profound hearing loss. *J Acoust Soc Am* 114(2):1055–1068.

Picton T. (1990) Auditory evoked potentials. In: Daly DD, Redley TA, eds. *Current Practice in Clinical Electroencephalography.* 2nd ed. New York: Raven Press Ltd.

Picton TW, Alain C, Otten L, Ritter W, Achim A. (2000) Mismatch negativity: different water in the same river. *Audiol Neurootol* 5:111–139.

Picton T, Woods DL, Baribeau-Braun J, Haeley TMG. (1977) Evoked potential audiometry. *J Otolaryngol* 6:90–119.

Picton T, Woods D, Proulx G. (1978a) Human auditory sustained potentials: I. The nature of the response. *Electroencephalogr Clin Neurophysiol* 45(2):186–197.

Picton T, Woods D, Proulx G. (1978b) Human auditory sustained potentials: II. Stimulus relationships. *Electroencephalogr Clin Neurophysiol* 45(2):198–210.

Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M. (2002) Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clin Neurophysiol* 113:407–420.

Purdy SC, Katsch R, Sharma M, Dillon H, Storey L, Ching TY. (2001) Hearing aid evaluation in infants and young children using cortical auditory evoked potentials. *Aust N Z J Audiol* 23:138.

Purdy SC, Katsch RK, Storey LM, Dillon H, Ching TY. (2001) Slow cortical auditory evoked potentials to tonal and speech stimuli in infants and adults. Paper presented at the XVIIth International Evoked Response Audiometry Study Group (IERASG), Biennial Symposium, Vancouver, Canada, July 22–27.

Rance G, Cone-Wesson B, Wunderlich J, Dowell R. (2002) Speech perception and cortical event related potentials in children with auditory neuropathy. *Ear Hear* 23(3):239–253.

Ritter W, Vaughan HG Jr, Costa LD. (1968) Orienting and habituation to auditory stimuli: a study of short-term changes in averaged evoked responses. *Electroencephalogr Clin Neurophysiol* 25:550–556.

Rupp A, Gutschalk A, Uppenkamp S, Scherg M. (2004) Middle latency auditory-evoked fields reflect psychoacoustic gap detection thresholds in human listeners. *J Neurophysiol* 92(4):2239–2247.

Seither-Preisler A, Krumbholz K, Lütkenhöner B. (2003) Sensitivity of the neuromagnetic N100m

deflection to spectral bandwidth: a function of the auditory periphery? *Audiol Neurotol* 8(6):322–337.

Sharma A, Dorman M. (1999) Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am* 106:1078–1083.

Sharma A, Kraus N, McGee T, Nicol T. (1997) Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalogr Clin Neurophysiol* 104(6):540–545.

Sharma A, Marsh C, Dorman M. (2000) Relationship between N1 evoked potential morphology and the perception of voicing. *J Acoust Soc Am* 108(6):3030–3035.

Shestakova A, Brattico E, Soloviev A, Klucharev V, Huotilainen M. (2004) Orderly cortical representation of vowel categories presented by multiple exemplars. *Cogn Brain Res* 21(3):342–350.

Skinner P, Jones H. (1968) Effects of signal duration and rise time on the auditory evoked potential. *J Speech Hear Res* 11(2):301–306.

Stapells DR. (2002) Cortical event-related potentials to auditory stimuli. In: Katz J, ed. *Handbook of Clinical Audiology*. 5th ed. Baltimore: Lippincott Williams and Williams.

Starr A, Picton TW, Sininger Y, Hood LJ, Berlin CI. (1996) Auditory neuropathy. *Brain* 119:741–753.

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA III. (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 82(5):2346–2357.

Talavage T, Edmister W. (2004) Nonlinearity of fMRI responses in human auditory cortex. *Hum Brain Mapp* 22(3):216–228.

Trebuchon-Da Fonseca A, Giraud K, Badier JM, Chauvel P, Liegeois-Chauvel C. (2005) Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. *Neuroimage* 27(1):1–14

Tremblay KL, Billings C, Rohila N. (2004) Speech evoked cortical potentials: effects of age and stimulus presentation rate. *J Am Acad Audiol* 15(3):226–237.

Tremblay KL, Friesen L, Martin BA, Wright R. (2003) Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear Hear* 24(3):225–232.

Virtanen J, Ahveninen J, Ilmoniemi RJ, Naatanen R, Pekkonen E. (1998) Replicability of MEG and EEG measures of the auditory N1/Nm-responses. *Electroencephalogr Clin Neurophysiol* 108:70–379.

Wei WI, Wong R, Hui Y, Au DK, Wong BY, Ho WK, Tsang A, Kung P, Chung E. (2000) Chinese tonal language rehabilitation following cochlear implantation in children. *Acta Otolaryngol* 120:218–221.

Yetkin FZ, Roland PS, Christensen WF, Purdy PD. (2004) Silent functional magnetic resonance imaging (FMRI) of tonotopicity and stimulus intensity coding in human primary auditory cortex. *Laryngoscope* 114:512–518.

Zerlin S, Naunton RF. (1974) Early and late averaged electroencephalic responses at low sensation levels. *Audiology* 13:366–378.