# Central Auditory Plasticity: Changes in the N1-P2 Complex after Speech-Sound Training

Kelly Tremblay, Nina Kraus, Therese McGee, Curtis Ponton, and Brian Otis

*Objective:* **To determine whether the N1-P2 complex reflects training-induced changes in neural activity associated with improved voice-onset-time (VOT) perception.**

*Design:* **Auditory cortical evoked potentials N1 and P2 were obtained from 10 normal-hearing young adults in response to two synthetic speech variants of the syllable /ba/. Using a repeated measures design, subjects were tested before and after training both behaviorally and neurophysiologically to determine whether there were training-related changes. In between pre- and post-testing sessions, subjects were trained to distinguish the −20 and −10 msec VOT /ba/ syllables as being different from each other. Two stimulus presentation rates were used during electrophysiologic testing (390 msec and 910 msec interstimulus interval).**

*Results:* **Before training, subjects perceived both the −20 msec and −10 msec VOT stimuli as /ba/. Through training, subjects learned to identify the −20 msec VOT stimulus as "mba" and −10 msec VOT stimulus as "ba." As subjects learned to correctly identify the difference between the −20 msec and −10 msec VOT syllabi, an increase in N1-P2 peak-to-peak amplitude was observed. The effects of training were most obvious at the slower stimulus presentation rate.**

*Conclusions:* **As perception improved, N1-P2 amplitude increased. These changes in waveform morphology are thought to reflect increases in neural synchrony as well as strengthened neural connections associated with improved speech perception. These findings suggest that the N1-P2 complex may have clinical applications as an objective physiologic correlate of speech-sound representation associated with speech-sound training.**

(Ear & Hearing 2001;22;79–90)

Department of Speech and Hearing Sciences (K.T., B.O.), University of Washington, Seattle, Washington; Department of Communication Sciences (N.K.), Neurobiology and Physiology, Otolaryngology, Northwestern University, Evanston, Illinois; Department of Communication Sciences (N.K., T.M.), Northwestern University, Evanston, Illinois; and Electrophysiology Department, (C.P.), House Ear Institute, Los Angeles, California.

The present study examines whether the N1-P2 complex event-related potential (ERP) holds promise as a clinical tool for assessing central auditory speech representation as well as changes in neural activity associated with auditory (re)habilitation. The N1-P2 complex is an obligatory ERP that can reflect central auditory speech representation without active patient participation (Martin, Sigal, Kurtzberg, & Stapells, 1997; Ostroff, Martin, & Boothroyd, 1998; Sharma & Dorman, 1999, 2000; Whiting, Martin, & Stapells, 1998). This makes the N1-P2 complex promising for assessing individuals who are affected by communicative or cognitive impairments by providing a window to the brain that is largely free of behavioral confounds such as memory and cognition.

The N1 response reaches maximal amplitudes at fronto-central sites (Vaughan & Ritter, 1970). Depending on the duration of the stimulus, the N1 response manifests itself as a negativity with a peak latency of approximately 100 msec followed by a positivity, labeled (P2), with a peak latency of approximately 175 msec. (Wood & Wolpaw, 1982; Woods, 1995). The N1-P2 complex is thought to reflect synchronous neural activation of structures in the thalamic-cortical segment of the central nervous system in response to auditory stimulation (Naatanen & Picton, 1987; Woods, 1995; Wolpaw & Penry 1975). Several temporally overlapping, spatially distributed neural sources contribute to scalp recorded potentials in the latency region of the N1-P2. However, dipole source models of ERPs and auditory-evoked magnetic fields indicate that the N1-P2 likely reflects neural activity originating from tangentially oriented generator sources located in the supra-temporal plane, in or near primary auditory cortex (Hari, Aittoniemi, Jarvinen, Katila, & Varpula, 1980; Naatanen & Picton, 1987; Scherg & von Cramon 1985; Wood & Wolpaw, 1982).

Several studies have shown that the N1-P2 complex reflects many of the spectral and temporal cues contained in spoken language that are critical for speech perception (Martin et al., 1997; Ostroff et al., 1998; Whiting et al., 1998; Woods & Elmasian, 1986). The temporal cue of interest in this study, voice-onset-time (VOT), is defined as the interval between the release from stop closure and the onset of laryngeal pulsing. For English speakers, increas-

ing VOT changes the perception of /ba/ to /pa/. Pre-voiced sounds are produced when the onset of voicing precedes release of the stop consonants. The English language does not use pre-voicing phonemically. That is, English speakers describe pre-voiced /ba/ sounds to be no different than /ba/ exemplars that are not pre-voiced (Tremblay, Kraus, Carrell, & McGee, 1997; Tremblay, Kraus, & McGee, 1998). However, perception can be altered by experience. With little training, listeners can learn to form arbitrary categories (Carney, Widen, & Viemeister, 1977; McClaskey, Pisoni, & Carrell, 1983). For example, monolingual speakers of English were trained to distinguish between two pre-voiced /ba/ syllables. They learned to label one pre-voiced /ba/ syllable as "mba" and another as "ba" (Tremblay et al., 1997, 1998).

The purpose of this study was to determine whether modifying VOT perception alters neural activity, and whether training-induced changes in neural activity are reflected in the N1-P2 response. This line of investigation is supported by two areas of research reviewed below. First, there is evidence that acoustic components of speech signals such as VOT are reflected both in near and far-field evoked potentials. Second, the structure of the N1-P2 complex is plastic; that is, it reflects changes in neural activity after stimulation and deprivation.

## Neural Encoding of VOT

Acoustic components of speech signals including VOT are reflected in near and far-field evoked potentials in both animals and humans (Eggermont, 1995; Kurtzberg, 1989; Martin et al. 1997; McGee, Kraus, King, & Nicol, 1996; Poeppel et al. 1996; Roberts, Ferrari, & Poeppel, 1998; Roberts & Poeppel, 1996; Sharma & Dorman, 1999, 2000; Sinex & McDonald 1989; Sinex et al., 1991; Steinschneider, Schroeder, Arezzo, & Vaughan, 1994,1995; Steinschneider, Reser, Schroeder, & Arezzo, 1995; Steinschneider, Volkoc, Noh, Garell, & Howard; 1999; Whiting et al. 1998). These studies show that waveform morphology, specifically peak latency, are sensitive indicators of temporal cues. For example, neurons in monkey auditory cortex showed two distinct peaks in response to the consonant-vowel stimulus /ta/, but only a single peak in response to /da/ (Steinschneider et al., 1994). The double-peak response was considered to reflect VOT because the first peak elicited by /ta/ corresponded to the aspiration and the second peak corresponded to the onset of voicing. Because VOT of the /da/ stimulus approximated zero, only one peak was observed.

Similar VOT response patterns have been reported in cats and guinea pigs (Eggermont, 1995; McGee et al., 1996) and most recently in humans (Sharma & Dorman, 1999; Steinschneider et al., 1999). Together these studies demonstrate that near and far-field potentials, both in animals and humans, reflect neurophysiologic coding of temporal cues. Furthermore, patterns of neural activity that are time-locked to the onset of the stimulus, and then again at the onset of voicing, may contribute to our perception of phonemes. Given this relationship, does altering perception through training modify the basic neural representation of the speech-sounds used in training? There are no existing studies on the effects of speech-sound training on the N1-P2 complex. However, previous work has demonstrated that the N1-P2 complex reflects experience-related changes in neural activity.

## The N1-P2 Complex Reflects Central Auditory Plasticity

Latency and amplitude of the N1-P2 complex have been studied in association with auditory system maturation and deprivation, rather than perceptual learning. Studies report significant changes in response latency and amplitude well into adolescence (Courchesne, 1978, 1990; Ponton, Don, Eggermont, Waring, & Masuda, 1996a, 1996b; Ponton, Eggermont, Kwong, & Don, 2000; Sharma, Kraus, McGee, & Nicol, 1997). At slower stimulus presentation rates, as age increases, P1 and N1 responses decrease in latency and P1 amplitude decreases and N1 amplitude increases. Injury-induced changes also affect the N1-P2 complex. Ponton, Vasama, Tremblay, Khosla, Kwong, and Don (2000) showed increases in P1-N1 peak-to-peak amplitude ipsilateral to the intact ear in unilaterally deafened adults. Whereas normal-hearing controls demonstrate asymmetric activity with greater activation over the hemisphere contralateral to the ear of stimulation, after unilateral deprivation, deafened subjects show an increase in neural activity ipsilateral to the ear of stimulation, resulting in increased P1-N1 symmetry. Ponton, Vasama, Tremblay, Khosla, Kwong, and Don (2000) concluded that unilateral hearing loss modifies neural activity in the central auditory pathways and that these changes are reflected in N1-P2 complex. In summary, these studies establish that experience-induced changes in neural activity can be reflected in the N1-P2 complex.

The purpose of this study is to determine whether training novel temporal cues alters the N1-P2 complex. Based on the previously described studies, we hypothesize that experiential learning, resulting in improved perception of VOT, will modify basic neural representation of sound and result in changes in N1 latency and amplitude. If the N1-P2 complex

reflects training-induced changes in perception, there is tremendous potential for clinical application. The N1-P2 complex could be used to monitor neurophysiologic changes during speech-sound acquisition after cochlear implantation, hearing aid use, or any other form of auditory learning. More importantly, physiologic correlates of perception could be particularly valuable when assessing individuals who are difficult to evaluate behaviorally.

However, if the N1-P2 complex is to be of clinical use, it is first necessary to determine whether the N1-P2 response is stable from test to retest, can reflect learning-related changes in individuals, and be recorded using a relatively simple montage. Therefore, our research questions were: 1) is the N1-P2 complex a stable response, showing no significant changes in latency or amplitude from test to retest? 2) does the N1-P2 complex reflect training-induced changes in neural activity, associated with improved VOT perception? and 3) are subtle temporal cues, such as 10 msec of voicing, reflected in the latency or amplitude of the N1-P2 complex?

## MATERIALS AND METHODS

### Participants

Subjects were 10 normal-hearing, right-handed monolingual speakers of English. The age range was 21 to 31 yr and included six female and four male participants.

### Stimuli

The stimuli were generated using a Klatt digital speech synthesizer (Klatt, 1980). Stimuli were synthesized speech tokens modeled after those used in comparable training experiment by McClaskey et al. (1983). The 11-item /ba/-/pa/ VOT continuum varied from −50 to +50 msec in 10-msec steps. A negative VOT indicates pre-voicing when laryngeal pulsing occurs during the stop closure period before the release. Pre-voiced stimuli were intentionally chosen to ensure that subjects were being trained to hear a novel contrast.

The steady-state portion of the stimuli consisted of the vowel /a/, which varied in duration relative to the VOT so that the overall duration for each stimulus remained constant at 180 msec. The formant values for this vowel were: F1 = 700 Hz, BW1 = 90 Hz; F2 = 1200 Hz, BW2 = 90 Hz; F3 = 2600 Hz, BW3 = 130 Hz, F4 = 3300 Hz, BW4 = 400 Hz; F5 = 3700 Hz, BW5 = 500 Hz. The fundamental frequency of the stimuli began at 120 Hz and then fell to 100 Hz during the steady-state portion of the vowel. Formant transitions were 40 msec in duration. To simulate a burst, a turbulent noise source
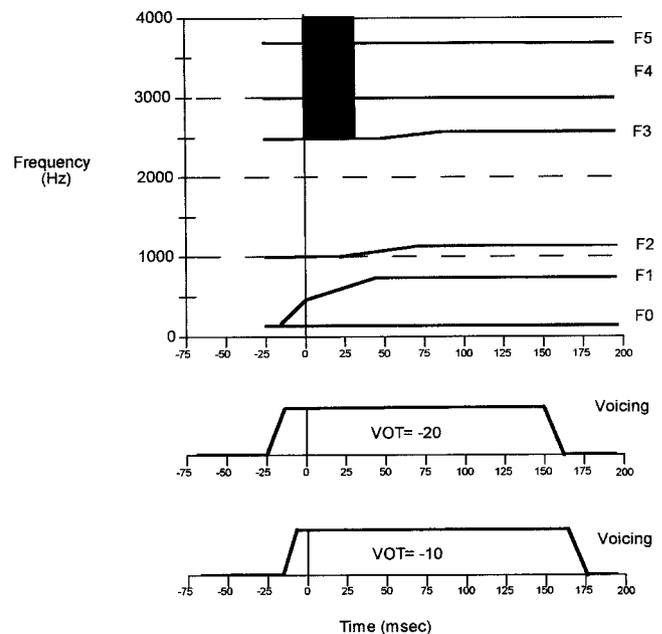


Figure 1. Schematic of formant values for the pre-voiced stimuli. The two pre-voiced stimuli are identical spectrally but differ by 10 msec of pre-voicing. The shaded area represents the spectral component of the burst.

(AF) 10 msec in duration and 60 dB in amplitude was added. The spectrum of the burst was centered around 2500 to 4000 Hz. A schematic representation of the pre-voiced stimuli is shown in Figure 1.

### Procedure

The study took 10 days for each subject to complete. The procedure is described in flowchart form in Figure 2. Behavioral identification tests as well as electrophysiologic measures were conducted on Days 1 and 2. These tests served as test retest control measures to establish the stability of both identification and electrophysiologic measures from day to day. On Days 3, 5, 7, and 9, subjects participated in identification training sessions. Each subjects' ability to identify the stimuli was tested on Days 4, 6, 8, and 10 (1 day after each identification training session). Electrophysiologic testing as well as tests of identification ability were repeated on Day 10 to provide post-test measures. Pre- and post-training measures were made by comparing identification test scores and electrophysiologic responses from Day 1 with Day 10.

### Behavioral Testing and Training

**Identification Testing (Days 1, 2, 4, 6, 8, and 10)**
• During behavioral tests, subjects were seated in a sound treated booth, approximately 1 meter from a 15 inch computer monitor. On Day 1 (pre-test1), to
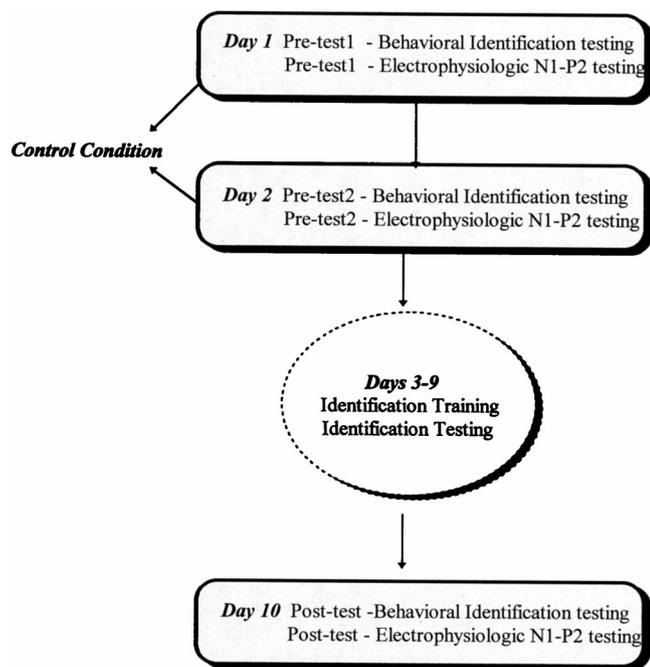
**Figure 2. Procedure flow chart. Subjects were tested behaviorally and electrophysiologically on Days 1 and 2. These tests served as a control condition to determine whether behavioral and electrophysiologic responses were stable from test to retest. After the pretests, identification training took place on Days 3, 5, 7, and 9. Identification testing took place on Days 4, 6, and 8. Behavioral and electrophysiologic tests were repeated on Day 10. Training-induced changes were determined by comparing the results from pre-test1 (Day 1) with post-test (Day 10).**

gain familiarity with the identification task, each subject identified what sound they heard from the −50 to +50 msec VOT continuum. When a single stimulus from the continuum was presented, each subject was asked to label the sound they heard. Three choices were provided on the computer screen: "mba," "ba," and "pa." No feedback was provided and no scores were calculated. Once the subject demonstrated that they understood the task, each subject was presented with either the −20 or −10 msec VOT stimuli and asked to identify the speech token as either "mba" or "ba." Therefore, this paradigm was a two-alternative forced-choice identification task. Fifty tokens were randomly presented. No feedback was provided. The response was scored correct if the subject assigned "mba" to the −20 msec VOT stimulus and "ba" to the −10 msec VOT stimulus. These results served as identification test scores for Day 1. This procedure was repeated on Day 2 and referred to as pre-test2.

Two-alternative forced-choice identification testing also was conducted 1 day after each training session. That is, 50 tokens of either the −20 or −10 msec VOT stimuli were presented, without feed-

back, on Days 4, 6, 8, and 10. Responses were scored correct if the subject assigned "mba" to the −20 msec VOT stimulus and "ba" to the −10 msec VOT stimulus.

**Training Sessions (Days 3, 5, 7, and 9)** • On Day 3, subjects participated in a fading task designed to emphasize the temporal cue (VOT) to be trained. Fading techniques are often used in training studies because they facilitate learning by providing a reference for listeners (Jamieson & Morosan, 1989). In a fading task, subjects are first presented with easy contrasts before being presented with difficult contrasts. Therefore, during the first training session, each subject was instructed to listen to single stimulus presentations, in order, from −50 to +50 msec VOT. Listening to the VOT continuum familiarized the subject with each endpoint as well as the each VOT increment within the continuum. After listening to the VOT continuum, each subject heard randomized presentations of the −10 msec and −30 msec VOT stimuli. Subjects were asked to identify each sound as either a "mba" or "ba." Both choices were presented as text on the computer monitor. Feedback in the form of a green reinforcement light appeared when the subject correctly identified the −30 msec VOT stimuli as "mba" and the −10 msec VOT stimuli as "ba." This task was repeated 50 times to train the listener to label the acoustic pre-voiced component as "mba." The −30 msec and −10 msec VOT stimuli were chosen based on a pilot study in which listeners were able to identify these stimuli well above chance without training. Therefore this session allowed the subjects to listen to the pre-voiced stimuli and orient themselves to the pre-voiced cue using an easy stimulus pair.

After the fading session, each subject began identification training using the −20 msec and −10 msec VOT stimuli. The −20 msec and −10 msec VOT stimuli were chosen because pilot data indicated that these two stimuli were difficult to identify as being different from each other, and the ability to learn to identify this novel distinction according to our previous training studies (Tremblay et al., 1997, 1998).

Identification training sessions consisted of four blocks of 50 trials in which either a −10 msec or −20 msec VOT stimulus was presented. Positive feedback (green reinforcement light) was given when the −20 msec VOT stimuli was labeled as "mba" and the −10 msec VOT labeled as "ba." Each stimulus was presented randomly with an equal probability of occurrence. Percent correct was calculated based on the number of correct responses for each block of 50 trials. Performance on all four blocks of 50 trials was averaged to obtain a final training session score.

## Electrophysiology

**Data Collection (Days 1, 2, and 10)** • During all sessions, subjects were seated in a sound-attenuating booth. Subjects watched closed-captioned videos of their choice while speech-sounds were presented to the right ear using Etymotic Research (ER3) insert earphones. Silver-silver chloride electrodes recorded were placed at Fz, Cz, Pz, over the frontal lobes (Fr and Fl) and temporal lobes (Tr and Tl), as well as the mastoids (A1 and A2). Electrode Tr was situated halfway between T4 and T6. Electrode Tl was situated halfway between T3 and T5 according to the International 10/20 (Jasper, 1958). A nose electrode served as the reference and a forehead electrode as ground. Eye blink activity was monitored using electrodes located on the superior and outer canthus of one eye. Epochs with artifact measuring in excess of 100 microvolts were rejected off-line. The remaining sweeps were prestimulus baselined, then referred to a common reference. Evoked responses were analog bandpass filtered on-line from 0.1 to 100 Hz (12 dB/octave roll off). Using a Neuroscan™ system, EEG channels were amplified with a gain × 500, and converted using an Analog-to-Digital Rate of 1 kHz. Responses were then filtered off-line from 1.0 Hz (high-pass filter, 24 dB/octave) to 40 Hz (low-pass filter, 24 dB/octave).

The recording window included a 100 msec prestimulus period and 500 msec poststimulus time. A PC-based system controlled the timing of stimulus presentation and delivered an external trigger to the evoked potential system. Because the N1 response is rate dependent (e.g., habituating at fast presentation rates), P1, N1 and P2 responses were recorded using two stimulus presentation rates: a slow interstimulus interval (ISI) of 910 msec, as well as a fast ISI of 390 msec (Davis, Mast, Yoshie, & Zerlin, 1966; Naatanen & Picton, 1987; Polich, Aung, & Dalessio, 1987). The 390 msec ISI was included to determine whether training-related changes could be detected at faster stimulus presentation rates, in a paradigm typically used to measure other ERPs such as the mismatch negativity (MMN). If so, training-related changes in the N1-P2 complex could be efficiently recorded simultaneously with discriminative ERPs such as the MMN. Order of presentation (fast versus slow) was randomized across subjects. Seven of the 10 subjects participating in the fast rate experiment also completed the slower rate condition. Each recording session took approximately 15 minutes with approximately 800 sweeps being averaged after artifact rejection in the fast condition, and 200 sweeps in the slow condition.

## RESULTS

Latency and amplitude measures were made from electrode Cz. This measurement location was chosen because the N1-P2 complex was clearer and larger at electrode site Cz in comparison with other electrode sites. Also, it was important to determine whether training effects could be measured from a single electrode.

### Test Retest Stability

To ensure that changes in behavior or waveform morphology resulted from training and not from test retest variability or repeated stimulus exposure, Day 1 (pre-test1) and Day 2 (pre-test2) control conditions were compared for all behavioral and neurophysiologic measures.

**Behavior** • There was no significant change in the subjects' ability to identify the −20 msec VOT stimuli as "mba" and the −10 msec VOT stimulus as "ba" from pre-test1 to pre-test2 according to paired $t$-tests ($t = 0.87, df = 9, p = 0.40$). Mean identification scores were 56% correct for pre-test1 and 59% for pre-test2.

**Neurophysiology** • There were no significant changes in latency or amplitude for either the −20 or −10 msec VOT stimuli, at either presentation rate, fast or slow, from pre-test1 to pre-test2, according to paired $t$-tests. Table 1 lists mean latency and amplitude differences from test to retest as well as paired $t$-test and $p$-values. Figure 3 shows sample pretraining waveforms from an individual subject, as well as group grand averages for pre-test1 and pre-test2, at fast and slow stimulus presentation rates, when recorded from electrode Cz. Visually, the waveforms from pre-test1 and pre-test2 do not appear different. At the slow presentation rate, the first positivity is consistent in latency with a P1 response and is labeled accordingly. The first negativity falls within the latency range of N1, followed by a second positive peak labeled P2. At faster stimulus presentation rates, P1 is followed by a broad bifid negativity with the first negativity approximating 100 msec and the second peaking at 200 msec. The first negativity was labeled N1 and the second negativity was labeled N2. P2 is greatly reduced at this faster stimulus presentation rate. For statistical purposes, P2 was defined as the most positive point in the waveform following N1 and preceding N2.

### Training-Induced Changes

**Behavior** • Before training, individuals had difficulty identifying the difference between the −20 and −10 msec VOT stimuli. Pre-test1 identification

**TABLE 1. Mean latency and amplitude differences from test to retest as well as paired *t*-test and *p*-values.**

| −20 ms VOT | Fast Rate | | | | Slow Rate | | | |
|---|---|---|---|---|---|---|---|---|
| Pre-test1 versus Pre-test2 | Mean Difference | df | *t*-value | *p*-value | Mean Difference | df | *t*-value | *p*-value |
| P1 latency | −1.8 msec | 9 | −0.55 | n/s | −2.63 msec | 6 | −0.43 | n/s |
| N1 latency | 3.5 msec | 9 | 0.78 | n/s | 6.26 msec | 6 | 1.9 | n/s |
| P2 latency | 5.4 msec | 9 | 0.86 | n/s | 5.90 msec | 6 | 1.6 | n/s |
| P1-N1 amp | 0.07 $\mu$v | 9 | 1.1 | n/s | −0.20 $\mu$v | 6 | −0.68 | n/s |
| N1-P2 amp | 0.12 $\mu$v | 9 | 1.3 | n/s | −0.36 $\mu$v | 6 | −0.88 | n/s |
| **−10 ms VOT** | | | | | | | | |
| P1 latency | 2.7 msec | 9 | 1.3 | n/s | 0.29 msec | 6 | 0.03 | n/s |
| N1 latency | −2.3 msec | 9 | −1.3 | n/s | 7.29 msec | 6 | 1.3 | n/s |
| P2 latency | −4.7 msec | 9 | −1.6 | n/s | 7.58 msec | 6 | 0.57 | n/s |
| P1-N1 amp | 0.05 $\mu$v | 9 | 0.76 | n/s | −0.08 $\mu$v | 6 | −0.48 | n/s |
| N1-P2 amp | 0.07 $\mu$v | 9 | 1.3 | n/s | −0.50 $\mu$v | 6 | −1.7 | n/s |

There were no significant changes in latency or amplitude of the neurophysiological responses from pre-test1 to pre-test2 according to paired t-tests. This was true regardless of stimulus (−20 msec or −10 msec VOT) or presentation rate (fast or slow).
n/s = not significant.

scores ranged from 44 to 72% correct with a mean score of 56%. The ability to identify both the −20 msec VOT stimuli as "mba" and the −10 msec VOT stimuli as "ba" improved with training. After training, identification scores ranged from 68 to 100% correct with a mean score of 82%. Figure 4 illustrates the significant improvement from pre-test1 to post-test ($t = 5.24$, $df = 9$, $p < 0.001$).

**Neurophysiology** • Significant changes in waveform morphology were observed after training. Training-induced changes were most apparent at the slower stimulus presentation rate.

**Slow Rate**: As shown in Figures 5 and 6, pretraining ERPs recorded from electrode site Cz were dominated by a large positivity labeled P1 followed by clear N1 and P2 responses. After training, no significant changes in latencies were observed; however, there were significant increases in N1-P2 amplitude.

Latency measures were analyzed using a $2 \times 2 \times 3$ repeated measures analysis of variance (ANOVA), comparing training (pre- versus post-training), stimuli (−20 versus −10 msec VOT), and peak latency (P1, N1, and P2). No significant main effect or interactions for training were obtained [training, main effect, $F = 3.15$, $p = 0.13$; training × stimuli, $F = 1.54$, $p = 0.26$; training × peak latency, $F = 2.09$, $p = 0.17$; training × stimuli × peak latency $F = 0.02$, $p = 0.98$].

Amplitude changes were detected after train-



**Figure 3. Examples of pretraining waveforms from an individual subject, as well as group grand averages for pre-test1 (Day 1) and pre-test2 (Day 2), at fast and slow stimulus presentation rates, when recorded from electrode Cz. Black waveforms were recorded on Day 1. Gray waveforms were recorded on Day 2.**



**Figure 4. Pre-test1 and post-training mean identification scores ±1 standard error of the mean. There was significant improvement in the ability to identify the −20 msec VOT stimulus as /mba/ and −10 msec VOT stimulus as /ba/ after training ($t = 4.53$, df = 9, $p < 0.001$).**
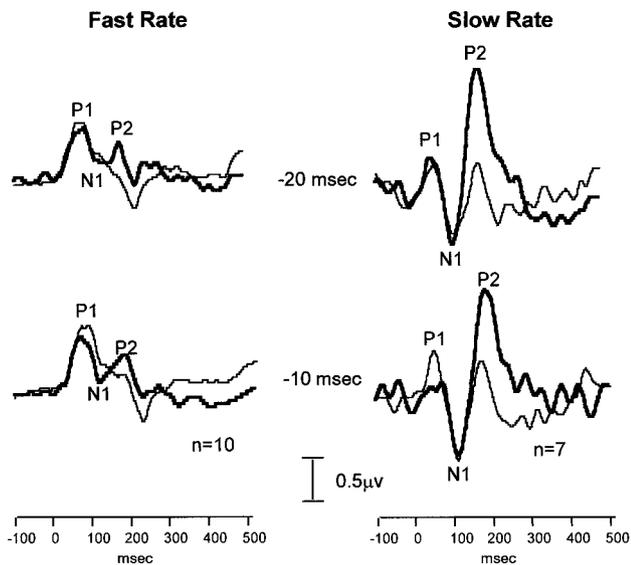
**Figure 5.** Pre- and post-training grand mean waveforms measured from electrode Cz. Pretraining waveforms are thin. Post-training waveforms are thick. As subjects learned to identify the difference between the −20 and −10 msec VOT stimuli, N1-P2 peak-to-peak amplitude increased.



**Figure 6.** Changes in waveform morphology are observed when comparing pretraining responses (thin lines) to post-training responses (thick lines). Grand mean responses measured from electrode Cz are shown on the left. Individual waveforms as well as pre- and post-test identification scores are on the right. Pretest identification scores are shown above the bold post-test scores. Just as individual subjects demonstrate diverse learning patterns, starting and stopping at different identification scores, changes in waveform morphology are also distinct. Yet common to all is an increase in N1-P2 amplitude coinciding with improved perception ability. Of particular interest is Subject 8 who showed little perceptual improvement. Pretest performance was 66% and post-training performance was 68%. There is little change in N1-P2 amplitude for this subject in comparison with the remaining subjects who showed impressive changes in perception and waveform morphology.

ing. A $2 \times 2 \times 2$ repeated measures ANOVA comparing training (pre- versus post-), stimuli (−20 and −10 msec VOT), and peak-to-peak amplitude measures (P1-N1 and N1-P2), revealed significant main effects for training ($F = 7.83, p = 0.03$), stimuli ($F = 10.04, p = 0.02$) as well as a significant interaction for training × peak-to-peak amplitude ($F = 16.24, p = 0.01$). Post hoc Scheffe tests comparing pre- versus post-training measures indicate that amplitude increased significantly for N1-P2 amplitude (mean difference = 1.33, critical difference = 0.61, $p < 0.001$) but not for P1-N1 amplitude (mean difference = 0.19, critical difference = 0.62, $p = 0.53$).

**Fast Rate:** Latency measures were analyzed using the same $2 \times 2 \times 3$ repeated measures ANOVA used to analyze the slow rate data. No significant main effect or interactions for training were obtained [training, main effect, $F = 0.12, p = 0.74$; training × stimuli, $F = 0.01, p = 0.99$; training × peak latency, $F = 0.05, p = 0.95$; training × stimuli × peak latency $F = 0.89, p = 0.43$].

Although N1-P2 amplitude appears to be enhanced with training (Fig. 5), a $2 \times 2 \times 2$ repeated measures ANOVA revealed no significant main effects or interactions for training [main effect, $F = 0.21, p = 0.66$; training × stimuli, $F = 0.08, p = 0.78$; training × peak-to-peak amplitude, $F = 3.26, p = 0.10$; training × stimuli × peak-to-peak amplitude $F = 0.51, p = 0.49$].

## Neural Representation of Temporal Cues

Using the same repeated measures described above, there was a significant main effect for stimulus condition (−20 msec VOT versus −10 msec VOT) for amplitude but not latency measures [amplitude $F = 10.04, p = 0.02$; latency F = 0.26, $p = 0.63$]. As previously described, there were also significant effects for training, as well as a significant interaction for training × amplitude at the slower stimulus presentation rate. Therefore, post hoc Scheffe tests were calculated to determine whether there were significant amplitude differences between the two stimuli before and after training, and whether these differences were seen in the P1-N1 or N1-P2 peak-to peak amplitude.

Post hoc testing indicated that there were no significant differences in P1-N1 amplitude when

## -20 msec VOT versus -10 msec VOT
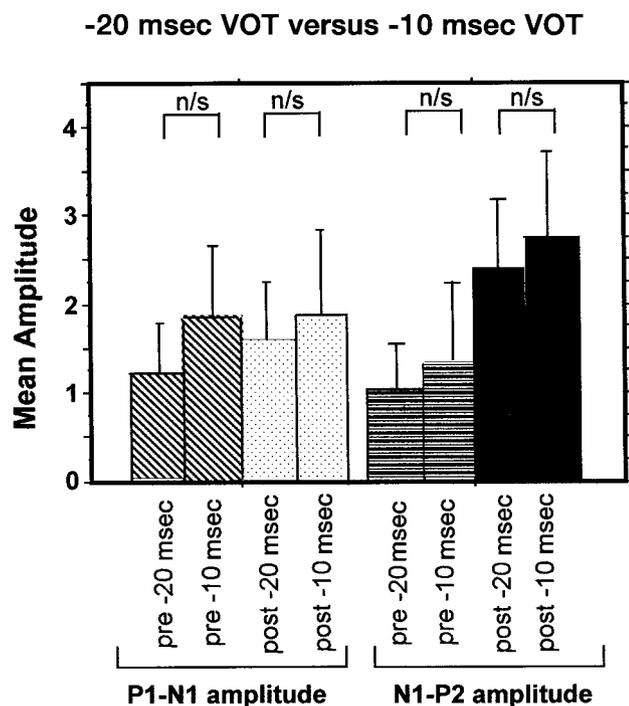


Figure 7. Post hoc Scheffe tests compare P1-N1 and N1-P2 amplitude, for each stimulus, pre- and post-training. No significant amplitude differences were found pre- or post-training. n/s = not significant $p > 0.05$

evoked by the $-20$ msec VOT and the $-10$ msec VOT stimuli before training (mean difference = 0.59, critical difference = 0.78, $p = 0.13$) or after training (mean difference = 0.29, critical difference = 1.04, $p = 0.55$). The same was true for N1-P2 amplitude. There were no significant differences in N1-P2 peak-to-peak amplitude evoked by the $-20$ msec VOT and the $-10$ msec VOT stimuli before training (mean difference = 0.27, critical difference = 0.84, $p = 0.50$) or after training (mean difference = 0.32, critical difference = 1.03, $p = 0.51$) (Fig. 7).

### Summary of Results

In summary, the N1-P2 complex proved to be a stable response showing no significant changes from pre-test1 to pre-test2 especially at slow presentation rates. Through training, subjects learned to identify the $-20$ msec VOT stimulus as "mba" and $-10$ msec VOT stimulus as "ba." After learning to identify the difference between the $-20$ msec and $-10$ msec VOT stimuli, significant increases in N1-P2 peak-to-peak amplitude were observed.

Before training, the N1-P2 response elicited by the $-20$ msec VOT stimuli was not significantly different from the response elicited by the $-10$ msec VOT response. Furthermore, there was no significant difference for N1-P2 amplitude after training

because both the $-10$ and $-20$ msec VOT stimuli elicited larger responses.

## DISCUSSION

### Is the N1-P2 Complex a Stable Measure?

The N1-P2 complex proved to be a stable measure when recorded from electrode Cz, showing no significant changes in latency or amplitude from test to retest, at both slow and fast presentation rates. These results are consistent with others that have demonstrated significant intrasubject reliability (Escera & Grau, 1997; Pekkonen, Rinne, & Naatanen, 1995) and that repeated stimulus exposure does not significantly alter response latency or amplitude.

### Does Speech-Sound Training Alter the N1-P2 Complex?

Auditory training altered the N1-P2 complex. Specifically, as the ability to differentiate the $-20$ and $-10$ msec VOT stimuli improved, N1-P2 amplitude increased. Increases in amplitude are thought to reflect increases in neural synchrony. Changes in neural firing patterns coinciding with learned behaviors are consistent with Hebbian principles of neural plasticity in that behaviorally important inputs that excite neurons, simultaneously in time, are mutually strengthened (Hebb, 1949). In this study, before training, the initial 10 msec portion of the /ba/ syllable was not meaningful because 10 msec differences in pre-voicing are not phonemic in the English language. Subjects perceived both prevoiced sounds as the nearest familiar phonemic category, /ba/. However, when subjects learned that the initial portion of the /ba/ syllable contained an important temporal cue that differentiated the two stimuli, the onset of voicing triggered an increase in neural synchrony. Evidence of this change in neural synchrony, manifested as increased N1-P2 peak-to-peak amplitude, suggests that activation of the neural generators contributing to this ERP are enhanced by training. These findings provide motivation for studying the P2 response and what role it plays in perception.

### Does the N1-P2 Reflect Subtle Differences in VOT?

Previous animal studies have shown monotonic changes in latency corresponding to VOT increments as small as 5 msec. Therefore, one might have expected that the N1 latency in response to the $-20$ msec VOT stimuli might occur 10 msec earlier than the N1 latency in response to the $-10$ msec VOT stimuli, particularly once the two stimuli were per-

ceived to be different from each other. However, no differences in N1 latency were observed when statistically comparing responses elicited by the −20 msec VOT stimuli with responses elicited by the −10 msec VOT stimuli either before or after training.

Studies that report neural representation of fine-grained changes in VOT (e.g., 5- or 10-msec increments) used animal subjects, making it possible to record closer to the source of neural activation in animals (Eggermont, 1995). Therefore, it could be hypothesized that recording closer to the source in humans might show distinct response patterns to VOT increments as small as 10 msec. However, when Steinschneider et al. (1999) recorded directly from human auditory cortex using 20 msec VOT increments ranging from 0 to 80 msec, responses elicited by the 0 msec and 20 msec VOT stimuli were similar in morphology. Distinct waveform morphologies were only seen when VOT exceeded 40 msec. Because perception rapidly changes from a voiced stop consonant /ba/ to an unvoiced consonant /pa/ at an interval of 20 to 40 msec VOT, Steinschneider et al. (1999) suggests that distinct waveform representations may reflect categorical perception of stop consonants.

The present study was not a categorical perception experiment; however, our results are consistent with Steinschneider et al. (1999) in that neural responses evoked by within category variants of the phoneme /ba/ (i.e., −20 msec and −10 msec VOT stimuli) do not appear different. The present study also reinforces that within category distinctions can be enhanced through training (Carney et al., 1977), but that training participants to arbitrarily label two different speech sounds does not result in distinct N1-P2 neural patterns. Increases in N1-P2 peak-to-peak amplitude were seen in response to both stimuli.

## Clinical Application

The N1-P2 complex holds promise for assessing central speech representation and changes in representation after listening training. These results are timely given the recent surge of interest in the neurophysiologic mechanisms underlying central auditory processing disorders as well as learning and reading disabilities (Nagarajan, Mahncke, Salz, Tallal, Roberts, & Merzinech, 1999). The N1-P2 complex may provide a practical tool for clinicians because these responses can be collected using most commercially available systems, require little testing time, and minimal off-line data manipulation compared with other ERPs. For example, the MMN has gained recent attention as a measure of central

speech-sound representation and it too is a preattentive response (Kraus, McGee, Sharma, Carrell, & Nicol, 1992; Kraus, McGee, Micco, Sharma, Carrell, & Nicol, 1993; Kraus, McGee, Carrell, King, Tremblay, & Nicol, 1995; Kraus, McGee, Carrell, Zecker, Nicol, & Koch, 1996; Martin, Kurtzberg, & Stapells, 1999; Sams, Paavilainen, Alho, & Naatanen, 1985; Sandridge & Boothroyd, 1996; Tremblay et al., 1997, 1998). Although the MMN provides insight into physiologic processes underlying speech discrimination and training-related plasticity, this ERP may not be the most efficient response for assessing speech-sound representation in individuals. The MMN is difficult to extract from background electro-encephalic noise and often requires prolonged testing time and off-line analyses (McGee, Kraus, & Nicol, 1997; Ponton, Don, Eggermont, & Kwong, 1997). As a result, clinical use of the MMN response has had limited success.

Although these results encourage the clinical use of the N1-P2 complex, immediate application is premature. Many parametric and nonparametric studies are still needed. Recent work with the MMN has shown that there may be optimal presentation rates, optimal recording periods, and optimal electrode montages to view neural activity (McGee et al., in press; Ponton, Eggermont, Kwong, & Don, 2000). This is probably true for the N1-P2 complex as well (Chermak & Musiek, 1997; Ponton, Eggermont, Kwong, & Don, 2000). For example, in the present study, training-induced N1-P2 amplitude changes at the fast rate were modest suggesting that training-induced changes are best detected at stimulus presentation rates slower than traditionally used to elicit the MMN.

Although this study and others establish that the MMN and N1-P2 reflect changes in neural activity that coincide with improved perception, a one-to-one relationship between changes in perception and changes in physiology, in individuals, has not been found (Kraus et al., 1995; Tremblay et al., 1997, 1998). This may be attributed, in part, to the fact that training-induced changes in physiology and perception do not share the same time course. The MMN and N1-P2 responses reflect preattentive processes that are considered to be precursors to perception. Therefore, training-induced physiologic changes at a preattentive level do not guarantee comparable changes in perception. Perception requires retrieval and integration of neurophysiologic codes into functional behavior. These processes can be affected by cognition, memory and motivation. Therefore, future studies need to be conducted to improve our understanding of the relationship between preattentive ERPs and perception in individ-

uals to explore potential of these measures as objective correlates of perception.

Finally, little is known about the P2 response and why it is enhanced with training. Future studies should be directed at improving our understanding of the relationship between training-induced enhancements in P2 with coincident improvements in perception. Perhaps P2 enhancement is not reflective of increased temporal coding, but rather a by-product of the training task. Because N1 amplitude can be affected by states of arousal and attention, and P2 partly reflects auditory driven output of the mesencephalic reticular activating system, perhaps the training task activates a preattentive alerting mechanism, that contributes to improved perception, but is not necessarily responsible for temporal encoding (Naatanen & Picton, 1987; Rif, Hari, Hamalainen, & Sams, 1991; Wilkinson & Morlock, 1966; Yingling & Skinner, 1977).

Nonetheless, the present study affirms that the central auditory system is capable of change, and that changes in neural activity are reflected in the N1-P2 response. These findings introduce the N1-P2 complex as a potential clinical tool for measuring changes in speech perception. Although the N1-P2 complex is not conventionally thought to correspond directly to perception, the work of Steinschneider et al. (1999) as well as the results of this training study suggest that obligatory responses such as the N1-P2 complex are affected by experience.

## CONCLUSION

1) The N1-P2 is a stable response, with peak latencies showing no significant change from test to retest, at slow and fast stimulus presentation rates. 2) The N1-P2 reflects training-induced changes in neural activity that coincide with improved perception. Training effects were most robust at slow presentation rates. 3) Future studies should be directed at improving our understanding of the relationship between preattentive ERPs and perception in individuals to explore potential application of these measures as objective measures of perception.

Address for correspondence: K. Tremblay, Ph.D., Dept. of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, WA 98105.

## REFERENCES

Carney, A. Widen, G., & Viemeister, N. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *62*, 961–970.

Chermak, G., & Musiek, F. (1997). *Central Auditory Processing Disorders*. San Diego: Singular Publishing.

Courchesne, E. (1978). Neurophysiological correlates of cognitive development: Changes in long-latency event-related potentials from childhood to adulthood. *Electroencephalography and Clinical Neurophysiology*, *45*, 468–482.

Courchesne, E. (1990). Chronology of postnatal human brain development: Event-related potentials, positron emission tomography, myelogenesis, and synaptogenesis studies. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-Related Potentials. Basic Issues and Applications* (pp. 210–241). New York: Oxford University Press.

Davis, H. Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: Recovery process. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Eggermont, J. (1995). Representation of a voice onset time continuum in primary auditory cortex of the cat. *Journal of the Acoustical Society of America*, *98*, 911–920.

Escera, C., & Grau, C. (1997). Short-term replicability of the mismatch negativity. *Electroencapholgraphy and Clinical Neurophysiology*, *100*, 549–554.

Hari, R. Aittoniemi, K., Jarvinen, M. L., Katila, T., & Varpula, T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain. Localization of neural generators. *Experimental Brain Research*, *40*, 237–240.

Hebb, D. O. (1949). *Organization of Behavior*. New York: John Wiley & Son.

Jamieson, D., & Morosan, D. (1989). Training new, non-native speech contrasts in adults: A comparison of the prototype and perceptual fading techniques. *Canadian J of Psychology*, *43*, 88–96.

Jasper, H. H. (1958). The ten-twenty system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Klatt, D. (1980). Software for cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, *7*, 25–32.

Kraus, N., McGee, T., Carrell, T., Zecker, S., Nicol, T., & Koch, D. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science*, *273*, 971–973.

Kraus, N., McGee, T., Micco, A., Sharma, A., Carrell, T., & Nicol, T. (1993). Mismatch negativity in school age-aged children to speech stimuli that are just perceptibly different. *Electroencephalography and Clinical Neurophysiology*, *88*, 123–130.

Kraus, N., McGee, T., Sharma, A., Carrell, T., & Nicol, T. (1992). Mismatch negativity event-related potential to speech stimuli. *Ear and Hearing*, *13*, 158–164.

Kurtzberg, D. (1989). Cortical event-related potentials assessment of auditory system function. *Seminars in Hearing*, *10*, 252–261.

Martin, B. A., Kurtzberg, D., & Stapells, D. (1999). The effects of decreased audibility by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. *Journal of Speech, Language and Hearing Research*, *42*, 271–286.

Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds

/ba/ and /da/. *Journal of the Acoustical Society of America*, *101*, 1585–1599.

McClaskey, C., Pisoni, D., & Carrell, T. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics*, *34*, 323–330.

McGee, T., King, C., Nicol, T., Tremblay, K., Cunningham, J., Allen, J., & Kraus, N. (in press). Long term habituation of the speech-elicited mismatch negativity. *Psychophysiology*. 167.

McGee, T., Kraus, N, King, C., & Nicol, T. (1996). Acoustic elements of speech like stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *Journal of the Acoustical Society of America*, *99*, 3606–3614.

McGee, T., Kraus, N., & Nicol, T. (1997). Is it really a mismatch negativity? An assessment of methods for determining response validity in individual subjects. *Electroencephalography and Clinical Neurophysiology*, *104*, 359–368.

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

Nagarajan, S., Mahncke H., Salz, T., Tallal P, Roberts, T., & Merzinech, M. (1999). Cortical auditory signal processing in poor readers. *Proceedings of the National Academy of Science of the United States of America*, *25*, 6483–6488.

Ostroff, J. M., Martin, B. A., & Boothroyd, A. (1998). Cortical responses to acoustic change within a syllable. *Ear and Hearing*, *19*, 290–297.

Pekkonen, E., Rinne, T., & Naatanen, R. (1995). Variability and replicability of the mismatch negativity. *Electroencephalograph and Clinical Neurophysiology*, *96*, 564–554.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K., & Marantz, A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Cognitive Brain Research*, *4*, 231–242.

Polich, J., Aung, M., & Dalessio, D. J. (1987). Long-latency auditory evoked potentials: Intensity, inter-stimulus interval and habituation. *Pavlovian Journal of Biological Science*, *23*, 35–40.

Ponton, C., Don, M., Eggermont, J. J., & Kwong, B. (1997). Integrated mismatch negativity (MMNi): A noise free representation of evoked responses allowing single point distribution-free statistical tests. *Electroencephalography and Clinical Neurophysiology*, *104*, 381–382.

Ponton, C. W., Don, M., Eggermont, J. J., Waring, M. D., & Masuda, A. (1996a). Auditory system plasticity in children after long periods of complete deafness. *Neuroreport*, *8*, 61–65.

Ponton, C. W., Don, M., Eggermont, J. J., Waring, M. D., & Masuda, A. (1996b). Maturation of human cortical auditory function: Differences between normal hearing and cochlear implant children. *Ear and Hearing*, *17*, 430–437.

Ponton, C. W., Eggermont, J. J, Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. *Electroencephalography and Clinical Neurophysiology*, *111*, 220–236.

Ponton, C. W., Vasama, J. P., Tremblay, K., Khosla, D., Kwong, B., & Don, M. (2001). Experience-related increases in interhemispheric correlations of evoked neurophysiologic activity following profound deafness. *Hearing Research*, *3638*, 1–13.

Riff, J., Hari, R., Hamalainen, M. S., & Sams, M. (1991). Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalography and Clinical Neurophysiology*, *79*, 464–472.

Roberts, T. P., Ferrari, P., & Poeppel, D. (1998). Latency of evoked neuromagnetic M100 reflects perceptual and acoustic stimulus attributes. *NeuroReport*, *9*, 3265–3269.

Roberts, T. P., & Poeppel, D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *NeuroReport*, *7*, 1138–1140.

Sams, M., Paavilainen, P., Alho, K., & Naatanen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *62*, 437–448.

Sandridge, S., & Boothroyd, A. (1996). Using naturally produced speech to elicit the mismatch negativity. *Journal of the American Academy of Audiology*, *7*, 105–112.

Scherg, M., & von Cramon, D. (1986). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalography and Clinical Neurophysiology*, *62*, 32–44.

Sharma, A., & Dorman, M. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset-time. *Journal of the Acoustical Society of America,*, *106*, 1078–1083.

Sharma, A., & Dorman, M. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America*, *107*, 2697–2703.

Sharma, A., Kraus, N., McGee, T., & Nicol, T. (1997). Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalography and Clinical Neurophysiology*, *104*, 540–545.

Sinex, D. G., & McDonald, L. P. (1989). Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America*, *85*, 1995–2004.

Sinex, D. G., McDonald, L. P., & Mott, J. B. (1991). Neural correlates of nonmonotonic temporal acuity for voice onset time. *Journal of the Acoustical Society of America*, *90*, 2441–2449.

Steinschneider, M., Reser, D., Schroeder, C. E., & Arezzo, J. C. (1995). Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Research*, *674*, 147–152.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G. (1994). Speech-evoked activity in primary auditory cortex: Effects of voice onset time. *Electroencephalography and Clinical Neurophysiology*, *92*, 30–43.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G. (1995). Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: Temporal response patterns. *Brain and Language*, *48*, 326–340.

Steinschneider, M., Volkoc, I., Noh, M., Garell, P., & Howard, M. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology*, *82*, 2346–2357.

Tremblay, K., Kraus, N., Carrell, T., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America*, *102*, 3762–3773.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiologic changes during speech-sound training. *Neuroreport*, *9*, 3557–3560.

Vaughan, H. G. Jr., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology*, *28*, 360–367.

Whiting, K. A., Martin, B. A., & Stapells, D. R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Ear and Hearing*, *19*, 218–231.

Wilkinson, R., & Morlock, H. (1966). Auditory evoked response and reaction time. *Electroencephalography and Clinical Neurophysiology*, *23*, 50–56.

Wolpaw, J. R., & Penry, J. K. (1975). A temporal component of the auditory evoked response. *Electroencephalography and Clinical Neurophysiology*, *39*, 609–620.

Wood, C. C., & Wolpaw, J. R. (1982). Scalp distribution of human auditory evoked potentials. II. Evidence for multiple sources and involvement of auditory cortex. *Electroencephalography and Clinical Neurophysiology*, *54*, 25–38.

Woods, D. (1995). The component structure of the N1 wave of the human auditory evoked potential. In G. Karmos, M. Molnar, V. Csepe, I. Czigler, & J. Desmedt (Eds.), *Perspectives of Event-Related Potentials Research* (pp. 102–109). (EEG Suppl. 44).

Woods, D., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology*, *65*, 447–459.

Yingling, C. D., & Skinner, J. E. (1977). Gating of thalamic input to cerebral cortex by nucleus reticularis larminaris. In J. E. Desmedt (Ed.), *Attention, Voluntary Contraction and Event-Related Cerebral Potentials. Progress in Clinical Neurophysiology* (pp. 70–96). Basel, Karger.