Changes in Sensory Evoked Responses Coincide with Rapid Improvement in Speech Identification Performance

Claude Alain[1,2], Sandra Campeanu[1,2], and Kelly Tremblay[3]

Abstract

■ Perceptual learning is sometimes characterized by rapid improvements in performance within the first hour of training (fast perceptual learning), which may be accompanied by changes in sensory and/or response pathways. Here, we report rapid physiological changes in the human auditory system that coincide with learning during a 1-hour test session in which participants learned to identify two consonant vowel syllables that differed in voice onset time. Within each block of trials, listeners were also presented with a broadband noise control stimulus to determine whether changes in auditory evoked potentials were specific to the trained speech cue. The ability to identify the speech sounds improved from the first to the fourth block of trials and remained relatively constant thereafter. This behavioral improvement coincided with a decrease in N1 and P2 amplitude, and these learning-related changes differed from those observed for the noise stimulus. These training-induced changes in sensory evoked responses were followed by an increased negative peak (between 275 and 330 msec) over fronto-central sites and by an increase in sustained activity over the parietal regions. Although the former was also observed for the noise stimulus, the latter was specific to the speech sounds. The results are consistent with a top–down nonspecific attention effect on neural activity during learning as well as a more learning-specific modulation, which is coincident with behavioral improvements in speech identification. ■

INTRODUCTION

Perceptual learning refers to improvements in sensory identification or discrimination abilities, which can take place over a period of minutes, days, weeks, or years (Watson, 1980). In some instances, learning can be characterized by rapid improvements in performance within the first hour of training (fast perceptual learning) followed by more gradual improvements that take place over several daily practice sessions (slow perceptual learning; Karni & Bertini, 1997). Fast perceptual learning was originally thought to reflect procedural learning (i.e., learning how to do the task; Wright & Fitzgerald, 2001). However, recent findings from animal research suggest that sensory neuron receptor fields can change quickly, within minutes of training (Fritz, Shamma, Elhilali, & Klein, 2003; Ohl, Scheich, & Freeman, 2001; Edeline, Pham, & Weinberger, 1993; Bakin & Weinberger, 1990), suggesting that rapid learning might not only implicate higher cognitive (strategic) functions but also involve experience-related changes in sensory cortices.

The hypothesis that rapid learning may be associated with changes in sensory cortex was recently tested using an auditory evoked potential (AEP) technique (Alain, Snyder, He, & Reinke, 2007). Participants were presented with two vowels simultaneously and asked to identify both of them by successively pressing two buttons. The 1-hour practice session was divided into five blocks of trials. Participants' performance improved gradually from the first to the last block of trials. More importantly, this rapid improvement in performance was paralleled by enhancements in an early evoked response (~130 msec) localized in the right auditory cortex and in a late evoked response (~340 msec) localized in the right anterior superior temporal gyrus and/or inferior prefrontal cortex. No such changes in AEPs were found when a different group of participants listened passively (no responses required) to the stimuli, indicating that these experience-related changes depended on listeners' attention and/or other task-related processes. Moreover, these learning-related changes in AEPs were preserved only if practice was continued; familiarity with the task structure ("procedural learning") was not sufficient.

One important issue that remains unanswered and deserves further empirical work is whether rapid changes in AEP amplitude are stimulus-specific or whether they reflect a more general process related to stimulus and/or task repetition independent of learning. Therefore, we applied a training model previously used to examine the impact of multiple daily training sessions on AEPs recorded during passive listening (i.e., no response required; Tremblay, Shahin, Picton, & Ross, in press; Tremblay, 2007; Tremblay,

[1]Rotman Research Institute, Baycrest Centre, Toronto, Canada, [2]University of Toronto, St. George Campus, Canada, [3]University of Washington, Seattle, WA

Kraus, McGee, Ponton, & Otis, 2001). In these studies, participants were trained to identify two consonant–vowel syllables differing in voice onset time (VOT), and AEPs were recorded before and following training. To determine if the training-related changes found were specific to the VOT cue being trained, Tremblay et al. (in press) examined if post-training physiological enhancements were seen in response to the vowel "a," the portion of the stimulus that did not contain the trained VOT cue. The effects of training were expressed differently for the trained "ba" and control "a" stimuli, but there were also overlapping similarities that might have been related to the common acoustic contents across the inherent stimulus exposure or due to the linguistic nature of the stimuli. For this reason, the present study included a noise stimulus that allowed us to dissociate changes in AEPs that index learning from those related to stimulus exposure or linguistic similarities. We recorded AEPs during a short 1-hour training session, and each practice block included the control stimulus (i.e., broadband noise) for which the participants simply pressed a button to advance to the next trial. Our questions were as follows: (1) Are there rapid physiological changes that take place during a 1-hour focused listening task? If so, are these changes stimulus specific or are the physiological changes also seen in response to an easily identifiable noise stimulus that is acoustically dissimilar from the experimental stimuli? (2) More importantly, is there a brain–behavior relationship? Did the physiological changes coincide with perceptual changes for the experimental but not for the control stimuli? We hypothesized that the ability to identify consonant vowels would improve from the first to the last block of trials and that this improvement in performance would coincide with changes in auditory evoked responses (e.g., N1 and P2 waves). More specifically, we hypothesized that practice effects would be different for the control stimulus (noise), which could easily be identified without training.

METHODS

Participants

Twenty-two right-handed young adults provided written informed consent to participate in the study. Data from two participants were excluded due to poor performance or excessive muscle and ocular artifacts during the AEP recording. The final sample was comprised of 9 men and 11 women aged 20 to 35 years ($M = 24$ years, $SD = 3.7$ years). All participants had pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000 Hz in both ears. For all participants, English was either the first language or was learned at a young age (i.e., between 4 and 7 years, for five participants). Ethical approval and informed consent were obtained according to the guidelines set out by the Toronto Academic Health Sciences Network for Baycrest and the University of Toronto Research Ethics Board.

Stimuli and Task

Stimuli consisted of two synthetic consonant vowels and one broadband noise. The two speech sounds were computer-generated "ba" sounds extracted from a continuum of stimuli used in previous training experiments (Tremblay et al., 2001; Tremblay, Kraus, Carrell, & McGee, 1997). A description of the specific formant frequencies and other stimulus details can be found in these publications; briefly, the two speech sounds were identical to each other spectrally but they differed in terms of VOT. One sound had 30 msec of prevoicing whereas the second stimulus had 20 msec of prevoicing, meaning the vocal-fold vibration simulations occurred 30 and 20 msec before the consonantal release. These particular stimulus contrasts were chosen because prior studies showed that people could learn to identify them as being different from each other quite quickly (McClaskey, Pisoni, & Carrell, 1983), making them easier to identify than the 20- and the 10-msec prevoicing tokens used during long-term training studies (Tremblay et al., 1997, 2001). Because these tokens were extracted from a continuum of stimuli that differed in VOT, a period of silence existed before the onset of the syllable. In the present study, this silent period was removed to ease the synchronization between the onset of the speech sound and the latency of AEPs. The noise sound, which was generated by multiplying white Gaussian noise with the smoothed envelope of the "ba" sound used in the earlier long-term training experiments (Ross & Tremblay, 2009; Tremblay et al., 2001), served as a control stimulus to test whether changes in brain activity reflect learning or whether they index stimulus and task repetition. Sound waveforms (sampling rate = 12,207 Hz) with durations of 196.6 msec were converted to analog signals using the Tucker Davis Technologies RP-2 real-time processor (24-bit, 90-kHz bandwidth) under the control of a custom-made Matlab program. The analog outputs were fed into a headphone driver (HB7, Tucker Davis Technologies, Alachua, FL) and then transduced through a GSI 61 audiometer. Stimuli were low-pass filtered at 6000 Hz through a Krohn–Hite filter and presented binaurally at 80-dB sound pressure level (SPL) through insert earphones (EAR-TONE ER-3a).

Before beginning the experimental task, participants were provided with five examples of each speech stimulus to familiarize themselves with the material. Participants were told in advance which stimulus would be presented for the practice session. For example, when the −20-msec VOT stimulus was presented, participants were told "these are examples of the 'ba' sound." There was no feedback because no judgment was required during this familiarization phase. During the training phase, each participant completed 10 blocks of 90 trials each (30 trials of each stimulus type), which included both speech sounds as well as the noise stimulus. All stimuli were equiprobable and presented in random order with an interstimulus interval of 1900 msec. On each trial, participants were asked to

identify the sound presented by pressing the corresponding key marked with a label "ba," "mba," or "noise" on the response box. No feedback was provided during the experiment. Participants were provided with an optional break after the fifth block or as needed. Upon completion of the experiment, the −30-msec VOT stimulus was scored correct if it had been identified as "mba," the −20-msec VOT stimulus as "ba" and the noise stimulus as "noise."

Electrophysiological Recording and Analysis

Electrophysiological recordings were made during perceptual testing. The electroencephalogram was digitized continuously (sampling rate 500 Hz) from an array of 64 electrodes with a band-pass of 0.05–100 Hz using the NeuroScan Synamps2 (Compumedics, El Paso, TX). We recorded eye movements with electrodes placed at the outer canthi and at the inferior orbits. During recording, all electrodes were referenced to the Cz electrode; for off-line data analysis, they were re-referenced to an average reference. The analysis epoch consisted of 200 msec of pre-stimulus activity and 1000 msec of post-stimulus activity. Epochs contaminated by excessive peak-to-peak deflection ($\pm$100 μV) at the channels not adjacent to the eyes were excluded from the averages. For each participant, the remaining epochs were averaged according to electrode position, stimulus type, and block using BESA (5.1.8). In each block and for each stimulus type, the number of trials included in the average varied from 16 to 30. AEPs were digitally filtered to attenuate frequencies above 20 Hz. AEP amplitudes were measured relative to the mean amplitude over the pre-stimulus interval.

For each participant, a set of ocular movements was obtained before and after the experiment (Picton et al., 2000). From this set, averaged eye movements were calculated for both lateral and vertical eye movements as well as for eye blinks. A principal component analysis of these averaged recordings provided a set of components that best explained the eye movements. The scalp projections of these components were then removed from the experimental AEPs to minimize ocular contamination using BESA (5.1.8).

The effects of rapid learning were examined on N1 and P2 waves. The N1 wave was defined as the largest negative peak between 80 and 160 msec after sound onset. The P2 wave was defined as the largest positive peak between 170 and 280 msec after sound onset. The effects of rapid learning through repeated blocks were examined at 9 fronto-central sites (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, and C2) as well as over the left (T3/T7) and the right (T4/T8) temporal sites. These electrodes were chosen because the sensory evoked responses N1 and P2 are largest at these sites and to ease comparison of our findings with those of prior studies that have examined rapid and/or slow auditory perceptual learning. Two other modulations of interest were also measured, namely, the N2b and the late positive complex, which are maximal over the fronto-

central and parieto-occipital scalp regions, respectively. The N2b was quantified over the 275- to the 330-msec interval using the same electrode array as for the N1 and the P2 wave. The LPC was quantified over the 550–650 interval over the left (P1, PO3) and the right (P2, PO4) parietal sites to explore potential hemispheric differences during learning and processing consonant–vowel syllables. The peak amplitude, the latency, and the mean amplitude measurements were analyzed using repeated measure ANOVAs with block of trials (i.e., first, second, third, fourth, etc.) as the within-subjects factor.

RESULTS

Behavioral Data

Preliminary analysis did not reveal significant differences in accuracy for the "ba" and "mba" speech token nor was the interaction between speech token and block significant; therefore, the behavioral data were collapsed over the two speech tokens. For all participants, we calculated the proportion of trials in which they correctly identified the consonant–vowel "mba" and "ba" stimuli and then divided by the total number of speech stimuli in the respective block of trials. Figure 1 shows the group mean accuracy for the speech and noise sounds as a function of block. As expected, performance for the noise stimulus was maximal for all blocks, whereas for the speech sounds, performance improved with practice. The ANOVA yielded a main effect of Stimulus Type, $F(1,19) = 50.67$, $p < .001$; a main effect of Block, $F(9,171) = 14.42$, $p < .001$, linear trend, $F(1,19) = 56.91$, $p < .001$; and a significant Stimulus $\times$ Block interaction, $F(9,171) = 14.74$, $p < .001$, linear trend, $F(1,19) = 46.08$, $p < .001$). Separate ANOVAs revealed an effect of block only for the speech sounds, $F(9,171) = 15.21$, $p < .001$, linear trend, $F(1,19) = 51.97$, $p < .001$, which indicates that performance gradually improved from the first to the fourth block of trials and then remained stable thereafter (from the 4th to the 10th block of trials, all pairwise comparisons $p > .05$).

Not surprisingly, participants were faster in responding to the noise stimulus than to the speech stimuli, $F(1,19) = 76.90$, $p < .001$. Overall, RT decreased from the first to the last block of trials, main effect of Block, $F(9,171) = 3.97$, $p < .01$, linear trend, $F(1,19) = 11.09$, $p < .005$, but this effect differed as a function of stimulus type, as suggested by the Stimulus $\times$ Block interaction, $F(9,171) = 3.24$, $p < .05$, linear trend, $F(1,19) = 4.68$, $p < .05$. Separate ANOVAs revealed an effect of block for both the speech, $F(9,171) = 4.43$, $p < .01$, linear trend, $F(1,19) = 12.32$, $p < .005$, and the noise stimuli, $F(9,171) = 2.58$, $p < .05$, linear trend, $F(1,19) = 3.58$, $p = .07$.

Electrophysiological Data

To identify neural correlates that are specific to learning, we first compared the effects of repeating the task
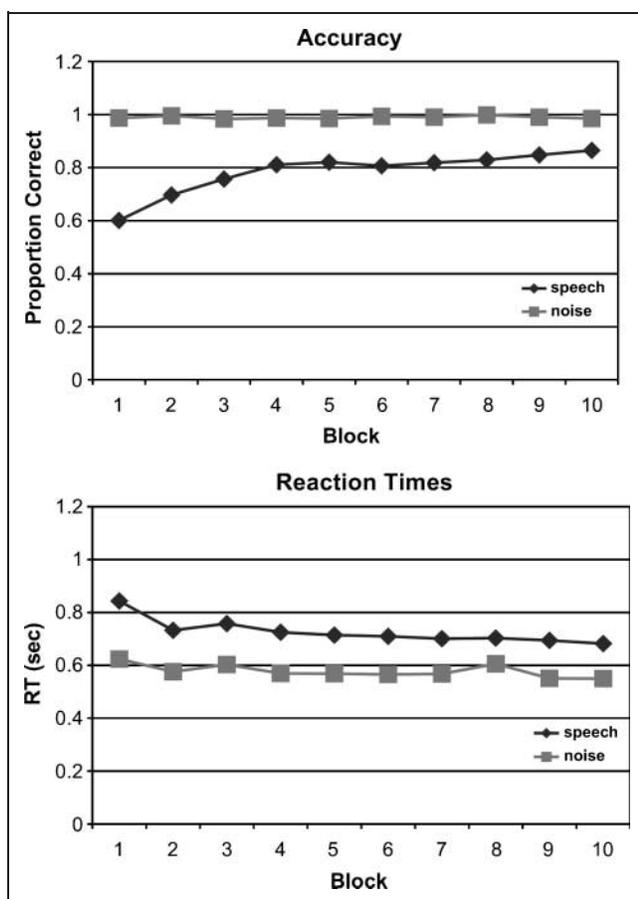
Figure 1. Group mean accuracy for the noise and speech sounds as a function of block.

over several blocks on AEPs from the two speech sounds collapsed together with those recorded for the noise stimulus. Figures 2 and 3 show the group mean AEPs elicited by the consonant–vowel and noise stimuli, respectively, during the first, second, fourth, and eighth blocks. Both stimulus types generated large N1 and P2 deflections, which were followed by an N2b and sustained potential over the fronto-central regions. As expected, the latter reversed in polarity over the parieto-occipital regions.

The N1 Wave

Overall, the N1 wave measured at fronto-central sites peaked earlier for the consonant–vowels (101 msec, $SE = $ 2.26 msec) than for the noise stimulus (105 msec, $SE = $ 2.57 msec), $F(1,19) = 5.60, p = .03$. However, the effects of practice on the N1 latency were not significant, $F < 1$, nor was the Stimulus $\times$ Block interaction, $F = 1.04$. For the N1 peak amplitude, the ANOVA yielded a main effect of Stimulus Type, $F(1,19) = 11.84, p < .001$, and a main effect of Block, $F(9,171) = 6.42, p < .001$, linear trend, $F(1,19) = 23.61, p < .005$. The Stimulus $\times$ Block interaction was not significant ($F < 1$), although the quadratic trend approached significance, $F(1,19) = 3.47, p = .078$,

suggesting that decreases in N1 amplitude as a function of practice may differ between the two stimulus types. Given that learning was greater over the first half of the test session, we further examined the impact of task repetition on the N1 amplitude recorded during the first four blocks. This analysis revealed a significant Stimulus Type $\times$ Block interaction, linear trend, $F(1,19) = 5.39, p < .05$, revealing greater practice-related decreases in N1 amplitude for the noise than for the consonant–vowel stimuli. A separate ANOVA on the N1 amplitude elicited by the consonant–vowel revealed a small, albeit significant decrease in N1 amplitude during the first four blocks, significant linear trend, $F(1,19) = 4.53, p = .047$. For the noise stimuli, the ANOVA revealed a significant practice related decrease in N1 amplitude, main effect of Block, $F(3,57) = 5.20, p < .005$, linear trend, $F(1,19) = 14.26, p < .001$.

The P2 Wave

The P2 wave peaked earlier for the consonant–vowel (198 msec, $SE = 3.56$ msec) than for the noise stimuli (205 msec, $SE = 3.66$ msec), although the difference in latency failed to reach significance, $F(1,19) = 3.82, p < .066$. However, the P2 latency decreased with task repetition, $F(9,171) = 3.10, p < .005$, linear and quadratic trend, $F(1,19) = 3.36$ and $13.35, p = .08$ and $.002$, respectively, and this effect was similar for both stimulus types (Stimulus $\times$ Block interaction, *ns*). A closer examination revealed a progressive decrement in P2 latency from the first to the fourth block of trials for both stimulus types, linear trend, $F(1,19) = 16.87, p < .001$, with no reliable changes in P2 latency from the fifth to the tenth block of trials, $F(1,19) < 1$. Again, there was no Block $\times$ Stimulus Type interaction for the P2 latency during the first four blocks. The ANOVA on P2 mean amplitude (180–220 msec) measured at nine fronto-central electrodes showed larger P2 amplitudes for the noise than for the speech sounds, $F(1,19) = 43.36, p < .001$, and a main effect of Block, linear trend, $F(1,19) = 4.79, p < .05$. More importantly, there was a significant Stimulus $\times$ Block interaction, $F(9,171) = 2.23$, $p < .05$, linear trend, $F(1,19) = 5.14, p < .05$. Separate ANOVAs for each stimulus type revealed a significant decrease in P2 amplitude as a function of block for speech sounds, $F(9,171) = 3.39, p < .05$, linear trend, $F(1,19) = 7.40, p < .02$, whereas the slight increase in P2 amplitude as a function of block for the noise stimulus appeared random and was not statistically reliable, $F(9,171) = 1.71$, $p = .14$, linear trend, $F(1,19) = 1.70, p = .20$.

The N2b Wave

The third modulation showing an effect of practice peaked at about 300 msec after sound onset and was referred to as the N2b wave. The mean amplitude of this N2b wave increased with task repetition, $F(9,171) = 8.27, p < .001$, linear trend, $F(1,19) = 20.82, p < .001$. Although the overall N2b amplitude was more negative for speech than for
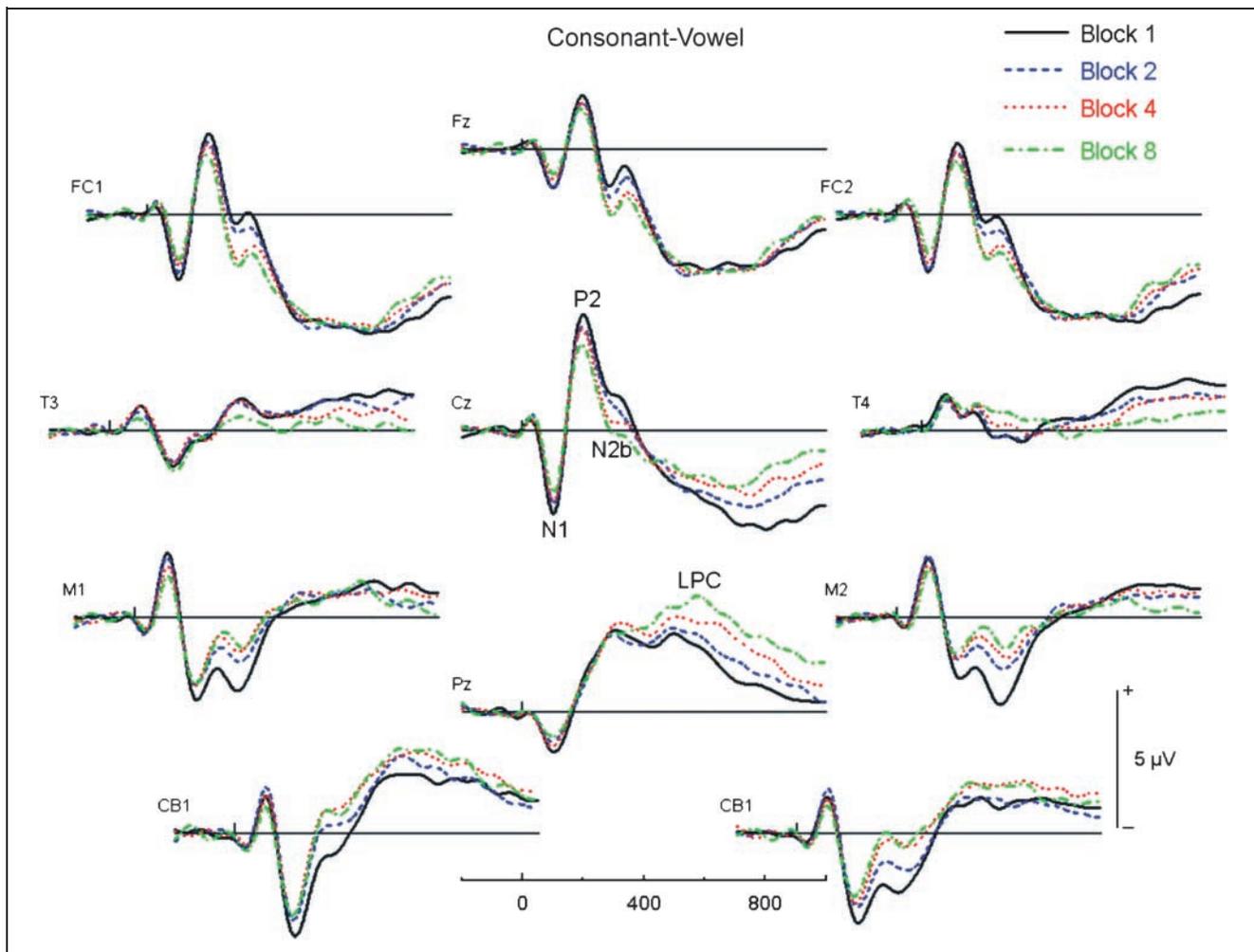
Figure 2. Group mean AEPs elicited by the consonant–vowel stimuli as a function of block from a subset of electrodes over the midline frontal (Fz), the central (Cz), and the parietal (Pz) regions. FC1 = left fronto-central; FC2 = right fronto-central; T3 = left temporal; T4 = right temporal; M1 = left mastoid; M2 = right mastoid; CB1 = left cerebellar; CB2 = right cerebellar.

noise, main effect of Stimulus Type, $F(1,19) = 42.83$, $p < .001$, the Stimulus × Block interaction was not significant ($F < 1$), suggesting that this change in ERP amplitude may index general (i.e., non-stimulus specific) top–down modulations (Figure 4).

The Late Sustained Potential

Processing speech and noise stimuli in this speech learning identification task was also associated with a late sustained potential over the parieto-occipital regions. The effect of practice on this late response mean amplitude was quantified over the left (P1, PO3) and the right (P2, PO4) hemispheres during the 550- to the 650-msec interval. There was a main effect of Stimulus Type, $F(1,19) = 5.50$, $p < .05$, with the speech sounds generating a larger amplitude than the noise stimulus. The main effect of Block was significant, $F(9,171) = 2.78$, $p < .05$, linear trend, $F(1,19) = 7.72$, $p < .05$. The Stimulus × Block interaction was also significant, linear trend, $F(1,19) = 4.76$, $p < .05$. Although the amplitude of the late sustained potential

elicited by the noise stimulus was little affected by practice, $F(9,171) = 1.60$, $p = .12$, linear trend, $F(1,19) = 0.06$, the late positive complex amplitude for the speech sounds increased with it, $F(9,171) = 3.31$, $p < .005$, linear trend, $F(1,19) = 12.06$, $p < .005$.

In a previous study, rapid improvement in identifying two vowels presented concurrently was associated with enhanced positivity in the right temporal scalp region (i.e., T4). We tested whether improvement in consonant–vowel identification would also yield rapid neuroplastic changes over the right auditory cortex. The ANOVA of the mean amplitude for the 110- to the 150-msec interval recorded over the left (T3/T7) and the right (T4/T8) temporal electrodes revealed a main effect of Stimulus Type, $F(1,19) = 17.35$, $p < .001$, with more positive voltages for the speech than the noise stimuli. No other main effect or interaction reached significance ($p > .20$ in all cases). For the 200- to the 300-msec interval, the main effect of Stimulus Type was significant, $F(1,19) = 5.66$, $p < .05$, with the speech sounds producing more positive AEPs than the noise stimuli. The main effect of Block was not significant
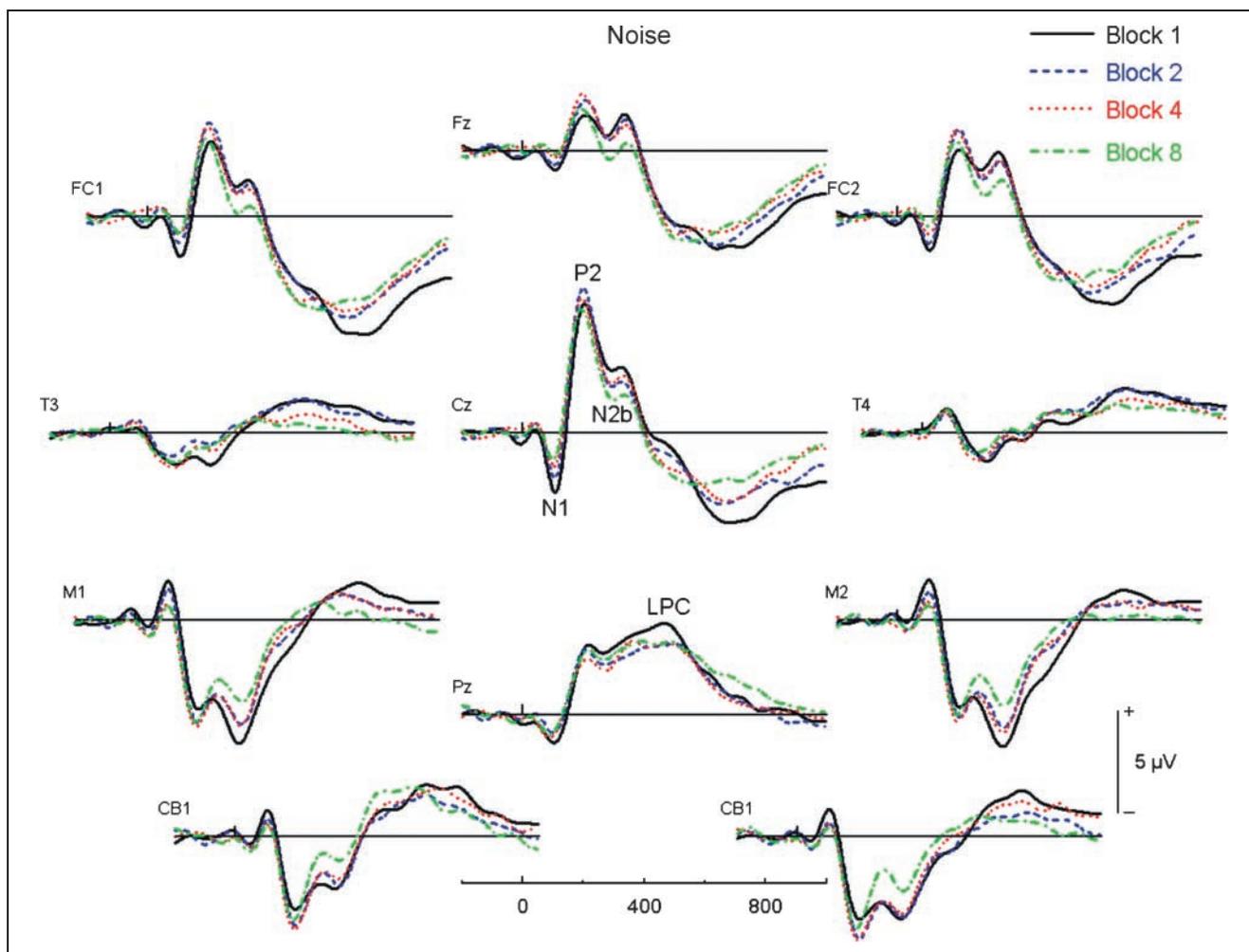
Figure 3. Group mean AEPs elicited by the noise stimuli as a function of block from a subset of electrodes over the midline frontal (Fz), the central (Cz), and the parietal (Pz) regions. FC1 = left fronto-central; FC2 = right fronto-central; T3 = left temporal; T4 = right temporal; M1 = left mastoid; M2 = right mastoid; CB1 = left cerebellar; CB2 = right cerebellar.

($F = 1.71$) nor were the interactions between block and any other factors.

Brain–Behavior Relationship

Previous studies suggest a link between changes in neural activity and behavioral improvement (Tremblay et al., in press; Alain et al., 2007). To further investigate this relationship in the present study, we computed Pearson correlation coefficients between performance and change in AEP amplitude for each of the 20 participants and tested, using one sample t tests, whether the group mean correlations differed from 0 (after Alain, Arnott, & Picton, 2001). For the accuracy data, the relationship between learning rate in identifying the consonant–vowel stimuli and the N1 amplitude for the consonant vowels (averaged over the nine fronto-central electrodes) varied substantially among the participants (Figure 5) but nonetheless reached significance (N1: group mean $r = .19$, $p < .05$; range $= -.66$ to .65). Six participants showed a negative

correlation (i.e., increase in N1 amplitude with increased accuracy). The reasons underlying this opposite pattern in this subset of participants are unclear. However, when these participants were excluded from the analysis, the group mean r increased to .39 ($p < .001$). For the P2 wave, the group mean correlation also differed from 0 (P2: group mean $r = -.27$, $p < .005$; range $= -.83$ to .31) with P2 amplitude decreasing with increasing performance accuracy. However, five participants[1] showed positive correlations (i.e., enhanced P2 amplitude with learning) between the P2 amplitude and the accuracy, and when these participants were excluded, the group mean r increased to $-.40$. A similar analysis for the N2b revealed a consistent increase in N2b amplitude with learning (N2b: group mean $r = -.29$, $p < .005$; range $= -.86$ to .44). Only four participants showed a positive correlation, and when these were omitted from the analysis, the group mean r was $-.43$. Lastly, for the LPC, the group mean correlation also differed from 0 (LPC: group mean $r = .30$, $p < .005$; range $= -.54$ to .79). Five participants showed negative correlations and when these were excluded, the group

r was .49. That is, larger LPC was associated with greater accuracy. These individual correlations suggest a more direct link between the behavioral changes and the AEP measures and provide further evidence supporting the claim that changes in AEP amplitude underlie behavioral improvement.[2]

"ba" Versus "mba"

One implication of the current findings is that the AEP signatures for the "ba" and the "mba" stimuli would likely be more similar at the beginning of the session when the two stimuli were difficult to distinguish than at the end of the session when participants could accurately identify the two speech tokens. To test this hypothesis, we compared the N1 and the P2 peak amplitude and latency for the "ba" and the "mba" as a function of block. For the N1 peak latency, we found a significant Stimulus Type × Block interaction, $F(9,171) = 2.28, p < .05$, linear trend, $F(1,19) = 6.10, p < .05$, reflecting a slight increase in N1 latency for the "ba" stimulus accompanied by a small decrease in N1 latency for the "mba" stimulus as a function of block. This resulted in an increase in the absolute latency differences between the N1 evoked by the "ba" and the N1 evoked by the "mba" stimuli as shown in Figure 6B. For the P2 wave,

the ANOVA revealed only a main effect of stimulus type on the P2 peak latency, $F(1,19) = 19.20., p < .001$, with the P2 peaking later for the "ba" (223 msec, $SE = 5.6$ msec) than for the "mba" (213 msec, $SE = 6.39$ msec) stimulus (Figure 6). For the N1 and the P2 peak amplitudes, the main effect of stimulus type was not significant nor was the Stimulus Type × Block interaction.

DISCUSSION

Adult humans can learn to identify subtle acoustic differences contained in speech sounds in a short period. Our behavioral results are consistent with previous work using similar stimuli (Tremblay et al., 1997, 2001, in press; Tremblay, Kraus, & McGee, 1998; McClaskey et al., 1983) in that individuals can improve their ability to identify subtle differences in VOT. Our study extends previous research by showing that rapid improvements in a speech identification task coincide with changes in exogenous (i.e., N1 and P2 waves) and endogenous (N2b and LPC) evoked responses. These training-related changes in AEP amplitude occur at various latencies, including early sensory registration of acoustic features as reflected by N1 as well as subsequent stages likely involved in other task-related processes (including stimulus classification, response
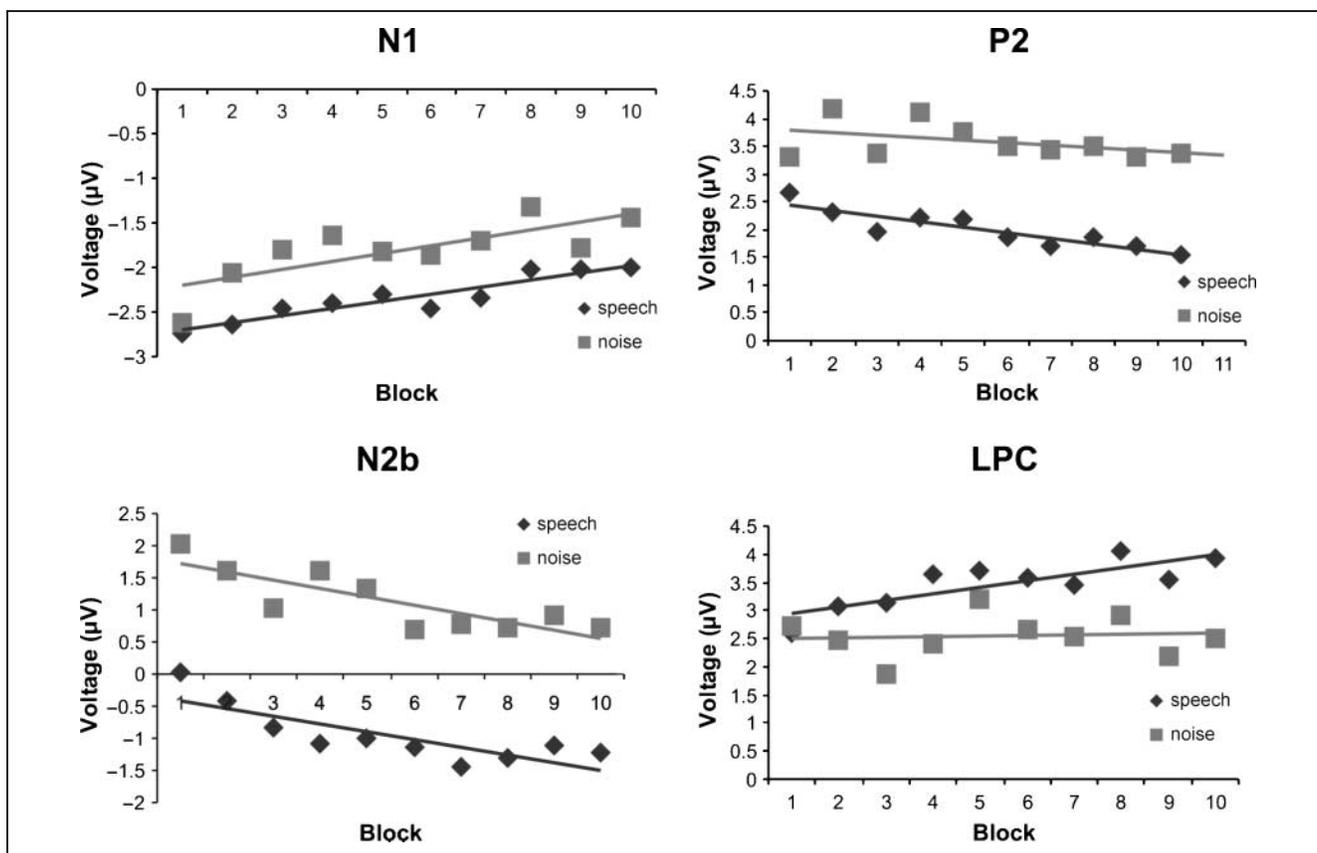

Figure 4. Group mean ERP amplitude as a function of block for the N1, the P2, the N2b, and the LPC. The N1, the P2, and the N2b measurements reflect the mean amplitude over nine fronto-central electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, and C2), whereas the measurement for the LPC reflects mean amplitude over four electrodes over the parietal scalp region (P1, P2, P03, and PO4).
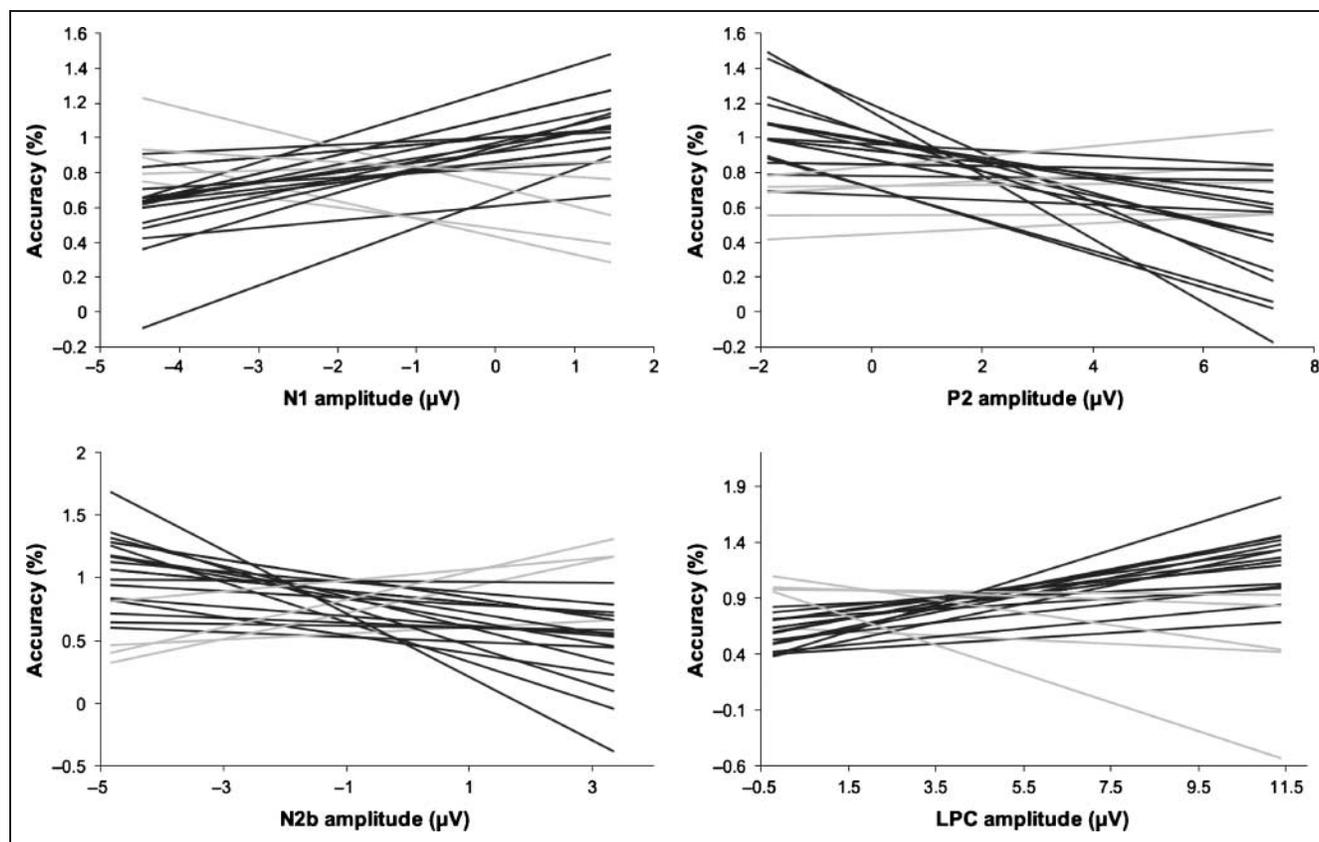
Figure 5. Individual trend lines that best fit the relation between the mean accuracy and AEP amplitude. The gray lines indicate correlations that are opposite from the group mean. The N1 and the P2 mean amplitudes over nine fronto-central scalp sites (see Methods section) were obtained for the 80- to the 120-msec interval and for the 180- to the 220-msec interval, respectively. The N2b and the LPC also reflect mean amplitude over several electrodes (see Methods section).

selection, preparation, and execution). For example, N1 latencies (and similar evoked activity in animals) reflect 10-msec increments in VOT, both in animals (Eggermont, 1995; Steinschneider, Schroeder, Arezzo, & Vaughan, 1995) and in humans (Tremblay, Piskosz, & Souza, 2003; Sharma, Marsh, & Dorman, 2000). During the first block when the two stimuli were difficult to distinguish, N1 latencies did not reflect the 10-msec VOT difference that separated the two stimuli. However, coinciding with improved perfor-

mance, the absolute distance between the two N1 latencies increased and soon approached 10 msec. This effect was most noticeable during the first four blocks.

The fact that changes in AEPs were seen within the first few blocks of trials is consistent with animal studies showing neuroplastic changes in sensory cortex within minutes during classical conditioning (Edeline et al., 1993; Bakin & Weinberger, 1990), instrumental avoidance conditioning (Bakin, South, & Weinberger, 1996), and auditory
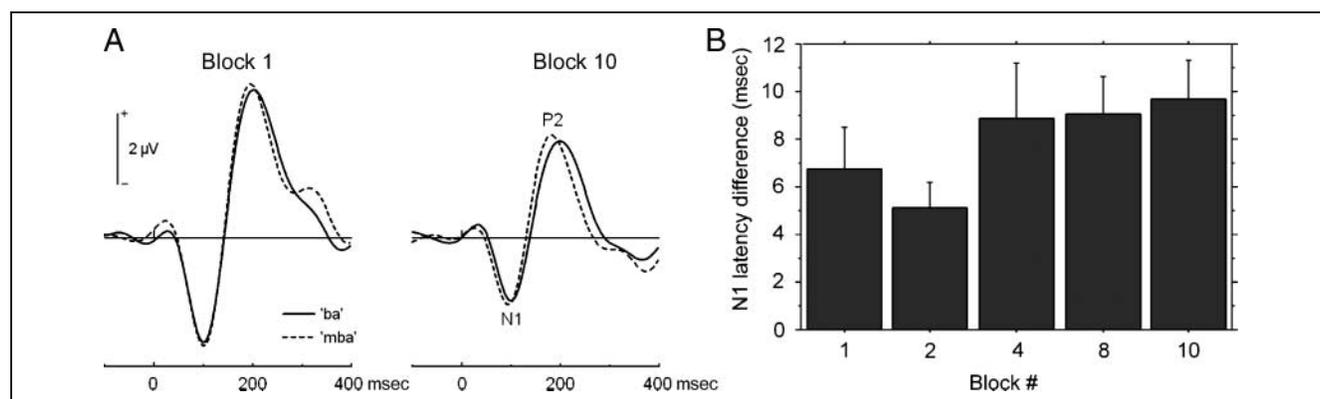

Figure 6. (A) Group mean AEPs elicited by the two consonant–vowel stimuli during the first and the last block of trials at the midline central (Cz) electrode. (B) Absolute peak latency differences separating "ba" and "mba" increase sharply between Blocks 1 and 4 and then stabilize by Block 10.

discrimination learning (Fritz et al., 2003). Rapid changes have also been observed in our previous studies (Alain et al., 2007), where AEPs were measured while listeners learned to segregate concurrent vowels. In that study, learning was associated with enhanced AEP amplitude occurring as early as 130 msec over the right temporal cortex, whereas the N1 wave recorded over the fronto-central scalp sites was little affected by practice. In the present study, the N1 wave recorded over the left or the right temporal cortex did not change significantly as a function of practice. There are several factors that could account for these differences between the two studies. First, the previous study involved vowel stimuli that differed spectrally whereas the current experiment involved the identification of a timing cue. It might be that neural mechanisms and attentional resources vary depending on the acoustic information being trained. Second, the tasks used in the two studies differed. The previous study used a difficult vowel segregation task that yielded slight but significant gains in performance, whereas the current experiment involved an easier identification task and showed sizeable rapid improvements. Despite some difference with previous research, our results reinforce the notion that rapid physiological changes can be seen in humans.

The N1 amplitude decreased with training, and this effect, over the course of the experiment, was similar for both types of stimuli despite the fact that changes in performance were seen only for the speech stimuli. This effect might be explained by a general process related to stimulus repetition and may be partly linked to habituation or sensitization as it was comparable for both speech and noise stimuli. However, during the first four blocks, there were indications of specificity with a more rapid decline in amplitude for the noise stimulus than for the speech sounds. This could perhaps reflect greater attentional resources allocated to the speech sounds at the beginning of the practice session, which could have masked the decrease in amplitude associated with long-term habituation. As the task becomes more automatic, less attention is required resulting in a decrease in N1 amplitude. This is consistent with findings showing that N1 amplitude is modulated by focused attention (Alain & Arnott, 2000; Hillyard, Hink, Schwent, & Picton, 1973).

The first evidence of learning-specific changes over the course of the experiment occurred during the N1 and the P2 intervals. There was a remarkable difference in the N1 latency between the two speech sounds, which emerged with learning to distinguish between them. Moreover, the P2 wave elicited by speech stimuli showed a very different neuroplastic change than that evoked by the noise stimulus. For the speech stimuli, there were consistent decreases in P2 amplitude as a function of block, whereas the noise stimulus fluctuated in amplitude between Blocks 1 and 4 and then decreased in amplitude from Blocks 4 to 10. The practice-related decrease in the P2 elicited by the speech stimuli may reflect changes in the allocation of attentional resources as the discrimination between

the two becomes easier, whereas no such change occurs for the noise stimulus because no discrimination learning was required. The changes in P2 amplitude contrast with our prior study, which showed no significant difference in P2 amplitude during the first hour of training for a concurrent vowel segregation task (Alain et al., 2007). These discrepancies may be related to the task used, the nature of the stimuli, or the rate of learning which differed substantially between the two studies. However, increases in P2 amplitude have also been reported when similar stimuli and tasks were used to examine the effects of long-term training (Tremblay et al., 2001, in press; Tremblay, 2007). In these studies, increases in P2 amplitude were observed following training, when AEPs were collected using a passive listening paradigm on days when training did not occur. Patterns of brain activity across hemispheres were different for the trained and control stimuli, again suggesting some degree of stimulus specific changes in the neural representation following training.

Not all types of training or acoustic learning affect evoked brain activity in the same way, as can be seen when comparing musicians and nonmusicians (Shahin, Bosnyak, Trainor, & Roberts, 2003) or assessing the impact of extended auditory training occurring over multiple sessions (Bosnyak, Eaton, & Roberts, 2004; Reinke, He, Wang, & Alain, 2003). For instance, Shahin et al. (2003) found larger N1c (~140 msec) and P2 amplitudes in musicians relative to nonmusicians. A larger N1m, the magnetic counterpart of the electric N1, has also been found for piano tones compared with pure tones in musicians, whereas no such difference in brain response was found in nonmusicians (Pantev et al., 1998). Neuromagnetic recordings have also revealed an enhanced N1m that is specific to the principal instrument played by the musician (Pantev, Roberts, Schulz, Engelien, & Ross, 2001). Similarly, practice over several daily training sessions decreased the N1 latency (Bosnyak et al., 2004; Reinke et al., 2003) and yielded an augmentation of N1m amplitude (Menning, Roberts, & Pantev, 2000), which may indicate that either more neurons were activated or neurons representing the stimulus were firing more synchronously. In addition, the N1c component showed an increase in amplitude over 15 training sessions (Bosnyak et al., 2004), whereas a training-related increase in P2 amplitude appears after only two (Atienza, Cantero, & Dominguez-Marin, 2002) or three (Bosnyak et al., 2004) daily test sessions. In the present study, the absence of significant P2 increases during the first hour of testing suggests that the P2 amplitude indexes a relatively slow learning process that may depend on consolidation over several days. It is well accepted that sleep plays an important role for the consolidation of newly acquired skills (Atienza, Cantero, & Stickgold, 2004; Gaab, Paetzold, Becker, Walker, & Schlaug, 2004; Mednick et al., 2002; Bonnet & Arand, 1995), and it is possible that P2 enhancement may index processes associated with consolidation.

Collectively, these results reveal progressive changes in the auditory system that vary as a function of task and

length of the training regimen. The current and dominant account is that training-related alterations in N1 and P2 waves are related to changes in the underlying neural generators. From this perspective, practice-related increases in AEPs could be expressed as (1) an increase in the size of the cortical areas representing the trained attribute, (2) a higher degree of synchronization within a particular neural ensemble, (3) a sharpened tuning of cells for the task-relevant (trained) attributes, and/or (4) changes in cortical maps (locus) representing the trained attribute. In the case of rapid learning, the decrease in the AEP amplitudes may reflect sensitization and habituation processes. An alternative possibility to structural and functional changes involving N1 and P2 waves is that training affects another process that overlaps these waves. That is, the N1 and the P2 waves may be superimposed by a slow negative wave associated with the goal-directed task. This may be similar to the processing negativity component used to account for the effects of attention on AEPs (Alain & Arnott, 2000; Alho, Tottola, Reinikainen, Sams, & Näätänen, 1987; Hansen & Hillyard, 1980). The assumption is that all stimuli would elicit this processing negativity whose onset and/or duration would change with learning. For instance, this processing negativity may increase in amplitude during the first hour of learning. Thus, the lack of change in P2 or the decrease in amplitude observed during rapid perceptual learning may be related to increased processing efficiency and increased attention-related processing negativity. This would be consistent with prior research showing that the selective attention effects on AEPs do take time to build up (Hansen & Hillyard, 1988; Donald & Young, 1982). However, as the listeners become more familiar with stimuli, the time or the effort allocated to perceptual decision processing decreases. This would be reflected in the AEPs by a reduced superimposition of a goal-directed component on the N1 and P2 waves. From this perspective, the P2 amplitude increase following several daily practice sessions would occur because the P2 wave is less "contaminated" by other negative components. Such a model could account for the increased negativity that encompasses the P2 and the N2b responses during the first hour of training, which subsequently get smaller as the participants become proficient in performing the perceptual learning task.

Improvement in performance was also associated with enhanced amplitude of the LPC over the parieto-occipital region. This practice-related change was specific to the speech sounds as no reliable difference in LPC amplitude was found for the noise stimulus (for which accuracy did not change throughout the 1-hour session). The practice-related enhancement in LPC amplitude is consistent with a recent study showing larger P3-like responses over the parietal scalp region in musicians who also showed superiority in pitch discrimination accuracy compared with nonmusicians (Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005). Moreover, the P3b has been shown to increase in amplitude with increased stimulus discriminability (Mazaheri & Picton, 2005; Salisbury et al., 1994;

Picton, 1992). In the present study, the enhanced LPC may reflect improvement in stimulus categorization and cannot easily account for the differences in response-related processes because response time to both speech and noise stimuli similarly improved with practice. One possibility is that the LPC indexes a consolidation process and/or memory updating based on new information. The changes in LPC may also reflect a monitoring process that varies with the level of confidence during responding. As the test session progresses and the participants learn about the material, they are likely to feel more confident about their responses. Thus, an important question for future studies would be, to what extent do changes in slow sustained activity over the parietal region reflect memory processes and/or confidence in the observer's perceptual decision.

Concluding Remarks

In the present study, we examined physiological changes that coincide with changes in perception in an attempt to draw links between theories of animal and human learning. We found that observers can quickly learn to identity speech tokens that differ in VOT and that these perceptual changes coincide with changes in exogenous evoked potentials. By varying the stimuli and the tasks, one can identify neuroplastic changes that underlie perceptual learning and these appear to be stimulus and/or task specific. The rapid learning-related changes in AEPs differ from those observed in studies examining the impact of daily training sessions, and further work is needed to assess the extent to which these early changes in performance and neural activity predict subsequent improvement in performance.

Notes

1. The subset of participants that showed opposite correlation with respect to the groups varied from one AEP deflection to the other.
2. Similar analyses on the first four block of trials yield comparable results. The correlation between accuracy and N1 amplitude varied substantially among the participants and did not reach significance (N1: group mean $r = .15$, ns; range $= -.76$ to $.82$) because six participants showed a negative correlation (i.e., increase in N1 amplitude with increased accuracy). When these participants

were excluded from the analysis, the group mean r increased to .45 ($p < .001$). For the P2, the group mean r was −.16 (*ns*; range = −.86 to .79). When seven participants showing a positive correlation (i.e., enhanced P2 amplitude with learning) were excluded, the group mean r increased to −.53 ($p < .001$). For the N2b, the group mean r was −.30 ($p < .05$; range = −.94 to .97) and increased to −.58 ($p < .001$) when six participants with positive correlation were omitted from the analysis. Lastly, for the LPC, the group mean r was .26 ($p = .06$; range = −.90 to .99). When six participants with negative correlations were excluded, the group r was .61 ($p < .001$).

REFERENCES

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience, 5,* D202–D212.

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1072–1089.

Alain, C., Snyder, J. S., He, Y., & Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex, 17,* 1074–1084.

Alho, K., Tottola, K., Reinikainen, K., Sams, M., & Näätänen, R. (1987). Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalography and Clinical Neurophysiology, 68,* 458–470.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning and Memory, 9,* 138–150.

Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience, 16,* 53–64.

Bakin, J. S., South, D. A., & Weinberger, N. M. (1996). Induction of receptive field plasticity in the auditory cortex of the guinea pig during instrumental avoidance conditioning. *Behavioral Neuroscience, 110,* 905–913.

Bakin, J. S., & Weinberger, N. M. (1990). Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Research, 536,* 271–286.

Bonnet, M. H., & Arand, D. L. (1995). Consolidated and distributed nap schedules and performance. *Journal of Sleep Research, 4,* 71–77.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex, 14,* 1088–1099.

Donald, M. W., & Young, M. J. (1982). A time-course analysis of attentional tuning of the auditory evoked response. *Experimental Brain Research, 46,* 357–367.

Edeline, J. M., Pham, P., & Weinberger, N. M. (1993). Rapid development of learning-induced receptive field plasticity in the auditory cortex. *Behavioral Neuroscience, 107,* 539–551.

Eggermont, J. 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.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience, 6,* 1216–1223.

Gaab, N., Paetzold, M., Becker, M., Walker, M. P., & Schlaug, G. (2004). The influence of sleep on auditory learning: A behavioral study. *NeuroReport, 15,* 731–734.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology, 49,* 277–290.

Hansen, J. C., & Hillyard, S. A. (1988). Temporal dynamics of human auditory selective attention. *Psychophysiology, 25,* 316–329.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182,* 177–180.

Karni, A., & Bertini, G. (1997). Learning perceptual skills: Behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology, 7,* 530–535.

Mazaheri, A., & Picton, T. W. (2005). EEG spectral dynamics during discrimination of auditory and visual targets. *Brain Research, Cognitive Brain Research, 24,* 81–96.

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

Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A. A., Pathak, N., et al. (2002). The restorative effect of naps on perceptual deterioration. *Nature Neuroscience, 5,* 677–681.

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport, 11,* 817–822.

Ohl, F. W., Scheich, H., & Freeman, W. J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. *Nature, 412,* 733–736.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392,* 811–814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport, 12,* 169–174.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9,* 456–479.

Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology, 111,* 53–65.

Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Brain Research, Cognitive Brain Research, 17,* 781–791.

Ross, B., & Tremblay, K. L. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research, 248,* 48–59.

Salisbury, D. F., O'Donnell, B. F., McCarley, R. W., Nestor, P. G., Faux, S. F., & Smith, R. S. (1994). Parametric manipulations of auditory stimuli differentially affect P3 amplitude in schizophrenics and controls. *Psychophysiology, 31,* 29–36.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience, 23,* 5545–5552.

Sharma, A., Marsh, C. M., & Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *Journal of the Acoustical Society of America, 108,* 3030–3035.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G., Jr. (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.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research, 161,* 1–10.

Tremblay, K., Kraus, N., Carrell, T. D., & 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: Neurophysiological changes during speech-sound training. *NeuroReport, 9,* 3557–3560.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1–P2 complex after speech-sound training. *Ear and Hearing, 22,* 79–90.

Tremblay, K. L. (2007). Training-related changes in the brain: Evidence from human auditory evoked potentials. *Seminars in Hearing, 28,* 120–132.

Tremblay, K. L., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology, 114,* 1332–1343.

Tremblay, K. L., Shahin, A. J., Picton, T. W., & Ross, B. (in press). Auditory training: Are the physiological effects of phonemic training specific to the trained cue? *Clinical Neurophysiology*.

Tremblay, K. T., Ross, B., & Picton, T. W. (2007). *Auditory training and P1–N1–P2: Age and stimulus effects*. Paper presented at the International Evoked Response Audiometry Group, Bled, Slovenia.

Watson, C. S. (1980). Time course of auditory perceptual learning. *Annals of Otology, Rhinology, & Laryngology. Supplement, 89,* 96–102.

Wright, B. A., & Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 12307–12312.