Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning

BOAZ M. BEN-DAVID,[a,b,c,d] SANDRA CAMPEANU,[a,e] KELLY L. TREMBLAY,[f]
AND CLAUDE ALAIN[a,e]

[a]Rotman Research Institute, Baycrest Centre, Toronto, Ontario, Canada
[b]Toronto Rehabilitation Institute, Toronto, Ontario, Canada
[c]Oral Dynamics Laboratory, Department of Speech-Language Pathology, University of Toronto, St. George, Ontario, Canada
[d]Centre for Research on Biological Communication Systems, Department of Psychology, University of Toronto, Mississauga, Ontario, Canada
[e]Department of Psychology, University of Toronto, St. George, Ontario, Canada
[f]Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington, USA

Abstract

Performance improvement during an hour of auditory perceptual training is accompanied by rapid physiological changes. These changes may reflect learning or simply task repetition independent of learning. We assessed the contribution of learning and task repetition to changes in auditory evoked potentials during a difficult speech identification task and an easy tone identification task. We posited that only task repetition effects would occur in the tone task but that task repetition and learning would interact in the speech task. Speech identification improved with practice (increased sensitivity d' with a constant response bias $β$). This behavioral improvement coincided with a decrease in the amplitude of sensory evoked responses (N1, P2) and a decrease in the amplitude of a slow wave (peak = 320 ms after onset) over the left frontal and parietal sites. Results show rapid physiological changes associated with learning, distinct from changes related to task repetition.

Descriptors: ERP, AEP, Rapid learning, Perceptual learning, Speech identification

Perceptual learning, an improvement in sensory discrimination after practice (e.g., Watson, 1980), often involves a rapid increase in performance that takes place during the first hour of training. Originally, this rapid improvement in performance was thought to reflect procedural learning, that is, the contingencies between stimulus and response. However, recent studies suggest that increased accuracy during the first hour not only reflects procedural learning but may also involve increased perceptual sensitivity (Hawkey, Amitay, & Moore, 2004). This finding is particularly interesting in light of animal studies showing rapid changes in auditory neuron receptive fields during associative learning (e.g., Bakin, South, & Weinberger, 1996; Edeline, Pham, & Weinberger, 1993). In humans, simple associative learning has also been shown to enhance activity in primary auditory cortex as measured by positron emission tomography (PET; Molchan, Sunderland, McIntosh, Herscovitch, & Schreurs, 1994), providing evidence for rapid learning-related

change in sensory cortices. Evidence from functional magnetic resonance imaging and neuroelectric recording suggests that long-term potentiation may play an important role in these rapid neuroplastic changes in sensory cortices (Clapp, Kirk, Hamm, Shepherd, & Teyler, 2005; Clapp, Zaehle, et al., 2005; Teyler et al., 2005).

More recently, scalp recording of auditory evoked potentials (AEPs) has been used to further investigate the neural underpinnings associated with rapid perceptual learning. For instance, Alain, Snyder, He, and Reinke (2007) showed that the perception of two simultaneously presented vowels could be improved within 1-h of practice and that improvement coincided with enhancements in an early-evoked response (∼130 ms, N1c) localized in the right auditory cortex and a late evoked response (∼340 ms) 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 over several consecutive days; familiarity with the task structure ("procedural learning") was not sufficient. That is, similar rapid learning-related changes in AEPs were observed when the same task was repeated several days later without training.

In a subsequent study, Alain, Campeanu, and Tremblay (2010) sought to distinguish the rapid changes in AEP amplitude that are learning specific from those that reflect more general processes

related to stimulus and/or task repetition independent of learning (practice with task-relevant stimuli involves both processes). During a 1-h test session, participants learned to identify two consonant–vowel syllables that differed in voice onset time (VOT). Within each block of trials, listeners were also presented with a white noise control stimulus to determine if changes in AEPs were specific to the trained speech cue. Behavioral improvement in speech identification was paralleled by a decrease in N1 and P2 amplitudes, which were followed by an increased negative peak (N2, between 275 and 330 ms) over frontocentral sites and an increase in late positive complex (LPC) over the parietal regions. The latter differs from the P3b response, which is usually elicited by infrequent targets embedded in a stream of standard stimuli (Picton, 1992; Polich, 2007). The effects of training on AEPs were expressed differently for the trained VOT cue compared with the control stimulus, something that is also found following long-term training (Tremblay, Shahin, Picton, & Ross, 2009). However, there were also overlapping similarities that might have been related to the learning context, which promoted the processing of both task-relevant (i.e., speech sounds) and task-irrelevant (i.e., noise) stimuli within the same block of trials. To better distinguish learning effects from procedural effects, the present study used a control task in an effort to dissociate neural activity associated with learning from changes in brain activity associated with repeating a task that would not need to be learned. We used the same speech stimuli as in Alain et al. (2010) as well as an easy tone identification task that is similar in procedural demands to the speech task. The control task allowed us to dissociate changes in AEPs that index learning from those related to performing a comparable task without learning (repetition). Moreover, this experimental design also allowed us to use signal detection theory and to assess whether rapid improvement in the speech identification task involves changes in perceptual sensitivity. Alternatively, changes in accuracy may merely reflect adopting a different response criterion or a combination of changes in sensitivity and response bias. If practice improves the discriminability of the speech stimuli, the sensitivity measure (d') would increase. If, on the other hand, practice modifies the criterion espoused by the participants to detect a speech signal, the response bias (β) would be altered. Only changes in the sensitivity measure that are not related to changes in response bias can attest to an improved ability to identify the two speech tokens.

We also hypothesized that the ability to identify speech sounds, as well as perceptual sensitivity, would improve with practice and that this improvement in performance would coincide with changes in auditory evoked responses in components that were shown to be affected by rapid auditory learning (i.e., N1, N1c, P2, N2, and late components). More specifically, we hypothesized that practice effects on AEPs would be different for the speech and the tone identification tasks (e.g., for the speech task only, a decrease in N1 and P2, and an increase in LPC, similar to Alain et al., 2010), providing a dissociation of the effects of learning and task repetition.

Method

Participants

Eighteen right-handed young adults provided written informed consent to participate in the study according to the guidelines set out by the Baycrest Centre for Geriatric Care and the University of Toronto and were paid $10/h. Data from 3 participants were excluded because of excessive muscle and ocular artifacts or equipment failure during the AEP recording. The final sample comprised 5 men and 10 women aged between 19 and 32 years ($M = 23.7$, $SD = 4.1$ years). All participants had pure-tone thresholds within normal limits for octave frequencies ranging from 250 to 8000 Hz (both ears). For all participants, English was either a first language or was learned at a young age (i.e., between 3 and 8 years of age).

Stimuli and Task

Stimuli consisted of two synthetic consonant vowels and two pure tones of 200 ms duration. The two speech sounds were computer-generated /Ba/ sounds extracted from a continuum of stimuli used in previous studies (e.g., Tremblay, Kraus, Carrell, & McGee, 1997; Tremblay, Kraus, McGee, Ponton, & Otis, 2001). 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 differed in terms of VOT. One sound had 30 ms of prevoicing (/mBa/) whereas the other had 20 ms of prevoicing (/Ba/), meaning vocal-fold vibration simulations occurred 30 and 20 ms prior to 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-ms and 10-ms prevoicing tokens used during long-term training (Tremblay et al., 1997, 2001). Because these tokens were extracted from a continuum of stimuli that differed in VOT, a period of silence existed prior to 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 pure tones consisted of low- (350 Hz) and high- (900 Hz) frequency sinusoidal sounds that included a 5-ms rise and fall time.

Pure tones and speech sounds (sampling rate of 12,207 Hz in both cases) were converted to analog signals using Tucker Davis Technologies (TDT, Alachua, FL) 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 (TDT HB7) 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 SPL through insert earphones (EAR-TONE ER-3a).

Each participant took part in two different and alternating tasks: a speech identification task and a tone identification task. Before beginning the first speech or tones experimental block, participants were provided with 12 examples of each of the stimuli (speech or tones) in order to familiarize themselves with the material. During practice trials, participants were told in advance which stimuli would be presented. For example, when the −20-ms VOT stimulus was presented, participants were told "these are examples of the 'Ba' sound." Because no judgment was required during this familiarization phase, there was no feedback.

Participants performed eight blocks of trials (four for each task). Half of the participants began with the speech identification task whereas the other half began with the tone identification task. A total of 120 stimuli were presented in each block of trials (60 trials of each stimulus type). For each block, the two stimuli (e.g., /Ba/ and /mBa/) were presented in random order with an interstimulus interval that varied randomly between 1800 and 2300 ms. During the speech block, participants identified the incoming sound by pressing a key marked with a label 1 (for /mBa/) or 2 (for /Ba/) on the response box. These same labels were used during the tone

blocks with labels 1 (for "low") and 2 (for "high"). No feedback was provided during the experiment. Participants were provided with an optional break after each block, as needed.

Analysis of Behavioral Data

In the speech blocks, *accuracy* was the average of the proportion of correct "Ba" responses (responding "Ba" to − 20-ms VOT signals) and correct "mBa" responses (responding "mBa" to − 30-ms VOT signals). Similarly, in the tone blocks, accuracy was the average of the proportion of correct "low" responses (responding "low" to 350 Hz signals) and correct "high" responses (responding "high" to 900 Hz signals). Sensitivity and response bias measures for the speech blocks were calculated according to the *Signal Detection Theory* (SDT; for a detailed description, see Macmillan and Creelman, 2005). For each block of speech trials, we designated the /Ba/ speech sound as the target and the/mBa/as the nontarget. As shown in Table 1, when a "Ba" signal was present, the rate of correct "Ba" responses was calculated as hits (H) and the rate of incorrect "mBa" responses as misses; when a "mBa" signal was presented, the rate of correct "mBa" responses was calculated as correct rejections and the rate of incorrect "Ba" responses as false alarms (FA). We then calculated the response bias (β, $e^{0.5*((Z(FA))^2 - (Z(H))^2)}$) and sensitivity ($d'$, $Z(H)Z(FA)$, Z stands for Z score[1]). Finally, *response latencies* were measured from the onset of the stimuli to the registration of a button press by the system. Separate analyses of the correct and incorrect trials did not reveal significant interaction (for response latencies) between accuracy and the linear term of the block, $F(1,14) < 1.0$ (for speech signals; the high accuracy level for tones did not permit such an analysis). This lack of interaction with practice indicates no differences in response latencies between correct and incorrect trials as a function of block. As the focus of this article is the effect of practice, we followed the convention in behavioral studies and analyzed correct response times only.

Electrophysiological Recording and Analysis

The electroencephalogram (EEG) was digitized continuously (sampling rate 500 Hz per channel) during perceptual testing from an array of 64 electrodes using the 10–10 system. The EEG was recorded with a bandpass of 0.05–100 Hz using NeuroScan Synamps2 (Compumedics, El Paso, TX). Eye movements were monitored with electrodes placed at the outer canthi and at the inferior orbits. During recording, all electrodes were referenced to the midline central electrode (i.e., Cz); for off-line data analysis, they were re-referenced to an average reference. The analysis epoch consisted of 200 ms of prestimulus activity and 1000 ms of poststimulus activity. For each participant, a set of ocular movements was obtained prior to 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 eyeblinks. 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 in order to minimize ocular contamination, using Brain Electrical Source Analysis (BESA 5.2.0). Epochs contaminated by excessive peak-to-peak deflection (± 100 μV) after correcting for ocular contaminations were excluded from the averages. For each participant, the remaining epochs were averaged according to electrode position, stimulus type, and block, using BESA (5.2.0). In each block and for each stimulus type, the number of

---

[1]The Z function returns the inverse of the standard normal cumulative distribution. The distribution has a mean of 0 and a standard deviation of 1.

Table 1. *A Matrix of Response Designation in a Signal Detection Theory Analysis of Accuracy Data in the Speech Identification Task*

Response	Signal	
	"Ba"	"mBa"
"Ba"	Hit	False alarm
"mBa"	Miss	Correct rejection

trials included in the average varied from 51 to 60 trials. These included correct and incorrect trials. Similar to the analysis of response latencies, separate analyses of the correct and incorrect trials did not reveal significant interaction between accuracy and the linear term of block for any of the AEP deflections. Thus, for AEPs, the effect of practice, the focal point of this study, was not altered by accuracy. This lack of interaction is not surprising when considering that participants received no feedback in this quick, perceptual task. Hence, the AEPs to correct and incorrect trials were averaged together to enhance the signal-to-noise ratio. Lastly, AEPs were digitally filtered to attenuate frequencies above 20 Hz (12 dB/Oct; zero phase). AEP amplitudes were measured relative to the mean amplitude over the prestimulus interval.

The effects of rapid learning were examined on N1 and P2 waves. The N1 wave was defined as the largest negative peak between 70 and 130 ms after sound onset. The P2 wave was defined as the largest positive peak between 150 and 210 ms after sound onset. The effects of rapid learning through repeated blocks were examined at nine frontocentral sites (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) as well as over the left (T3/T7) and 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. Four other modulations of interest were also measured, namely, the N1c, the N2, a sustained negative wave at bifrontal sites, and a late positive complex that was maximal at parietal and parieto-occipital sites. The N1c was quantified over the 115–165-ms interval over the left (T3/T7) and right (T4/T8) temporal sites. The N2 was quantified over the 270–330-ms interval using the same electrode array as for the N1 and P2 wave. The sustained negative potential was quantified over the 300–500-ms interval over left (F7, AF7′, AF3′, and F5) and right (F8, AF8′, AF4′, and F6) frontal sites. The LPC was quantified over the 300–500-ms interval over the left (P1, PO3) and right (P2, PO4) parietal sites to explore potential hemispheric differences during learning and processing consonant–vowel syllables. The peak amplitude, latency, and mean amplitude measurements were analyzed using repeated measure analyses of variance (ANOVAs) with blocks (1, 2, 3, or 4) and task (tones vs. speech) as the within-subjects factors.

Relations between Behavioral and Electrophysiological Measures

In the speech task, we tested linear correlations of Z scores of d' paired with Z scores of AEP amplitudes across the four blocks of practice. In the tone task, we tested linear correlations between Z scores of individual standard deviations of response latencies and Z scores of AEP amplitudes across the four blocks of practice. These two behavioral measures (d' for the identification of speech signals and the standard deviation of response latencies for tones)

were examined, because they were the only behavioral measures that were significantly modulated by block.

Results

Behavioral Data

We first analyzed accuracy of responses, then performed an analysis of accuracy based on Signal Detection Theory and, lastly, we analyzed average response latencies and standard deviation of response latencies for each block.

Accuracy analysis. Figure 1A shows the group mean proportion of stimuli correctly identified for the speech and tone tasks as a function of block. As expected, performance during the tone task was at ceiling for all blocks, whereas for the speech sounds, performance improved with practice. Overall, participants were more accurate in the tone than in the speech identification task, $F(1,14) = 64.3$, $p < .001$, $\eta_p^2 = .82$. The Task $\times$ Block interaction was significant (in this study, henceforth, the effects of practice are measured as the linear term of block effects) $F(1,14) = 6.3$, $p < .05$, $\eta_p^2 = .31$. For the speech task, accuracy increased with practice from 68% in Block 1 to 73% in Block 4 (a linear trend approaching significance[2]) $F(1,14) = 4.1$, $p = .06$, $\eta_p^2 = .22$, whereas for the tone identification task accuracy remained rather stable at ceiling, with a slight decrease from 99.2% in Block 1 to 98.2% in Block 4, paired-sample $t(14) = 2.2$, $p < .05$.

Signal detection theory analysis. To investigate whether practice improved perceptual processing of the speech signals, we performed an SDT analysis of responses in the speech task (it was not meaningful to conduct an SDT analysis for the tone task, because accuracy was at ceiling). Figure 1B shows a steady increase in the d' with practice, from 1.1 in Block 1 to 1.5 in Block 4, a significant linear term, $F(1,14) = 5.6$, $p < .05$, $\eta_p^2 = .28$, with no significant change in β across blocks ($F < 1.0$). Thus, with practice, participants steadily improved their ability to discern between the two speech stimuli without changing their decision criterion.

Response time analysis. Figure 1C shows the average response times during the speech and tone identification tasks as a function of block. Overall, participants were faster in responding during the tone identification task ($M = 473$ ms, $SD = 87$ ms) than during the speech task ($M = 648$ ms, $SD = 162$ ms), a main effect of task, $F(1,14) = 21.9$, $p < .001$, $\eta_p^2 = .61$. In the speech task, response latencies decreased steadily with task repetition, $F(1,14) = 4.3$, $p = .05$, $\eta_p^2 = .24$ for the linear term, but no such differences were observed during the tone task ($F < 1$), resulting in a significant interaction between task and the linear term of block on response times, $F(1,14) = 9.3$, $p < .01$, $\eta_p^2 = .40$.

An analysis of individual standard deviations of response times in each block revealed another interaction between task and the linear term of block, $F(1,14) = 4.7$, $p < .05$, $\eta_p^2 = .25$. In the tone task, standard deviation increased from an average of 130 ms at the first block to 151 ms at the last, paired-sample $t(14) = 2.2$, $p < .05$. Increased variability was not accompanied by an increase in response times with practice; this, along with the slight decrease in accuracy, suggests reduced arousal with task repetition (e.g., Kahneman, 1973). For the speech task, on the

---

[2]Recall that we excluded 3 participants because of technical problems related to eye movements however, their behavioral data were intact. We replicated the analysis with the full set of 18 participants and found the increase in accuracy with practice, from 66% in the first block to 72% in the last, to be significant, $F(1,17) = 6.1$, $p < .05$, $\eta_p^2 = .26$.
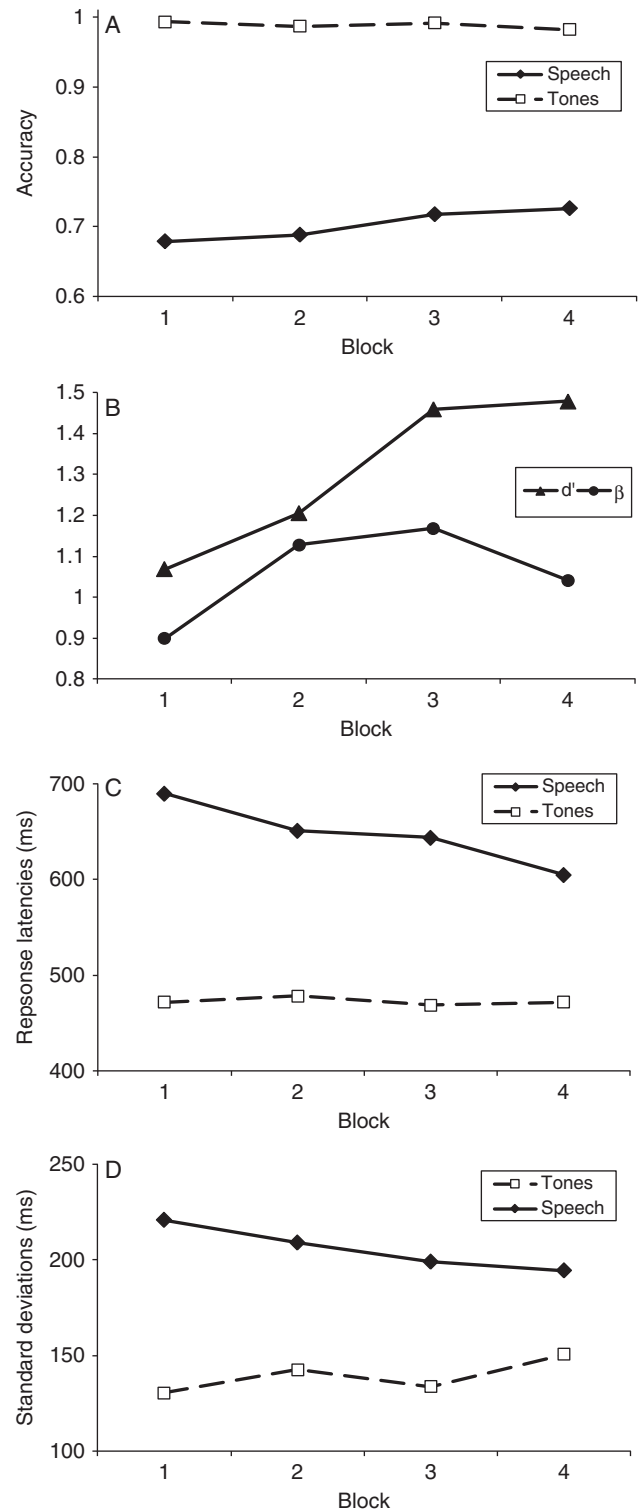

Figure 1. A: Group mean accuracy for the speech (diamonds) and the tone (empty squares) trials as a function of block. B: Sensitivity d' (triangles) and response bias, β (circles) were calculated according to Signal Detection Theory for the speech identification task. C: Group mean response latency for speech (diamonds) and tone (empty squares) identification tasks as a function of block. D: Group mean standard deviation of response latencies for speech (diamonds) and tone (empty squares) identification tasks as a function of block.
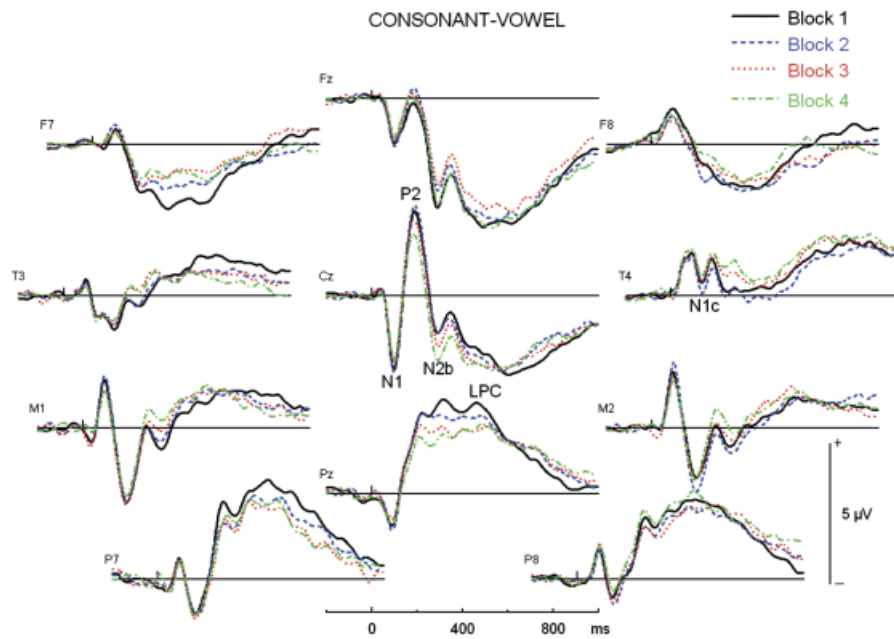
Figure 2. Group mean AEPs elicited by the speech (consonant–vowel) stimuli as a function of block from a subset of electrodes over the midline frontal (Fz), central (Cz), and parietal (Pz) regions. F7 = left frontal; F8 = right frontal; T3 = left temporal; T4 = right temporal; M1 = left mastoid; M2 = right mastoid; P7 = left parietal; P8 = right parietal.

other hand, the variance of latencies remained stable across the four blocks ($F < 1$), and the apparent decrease in standard deviations from an average of 220 ms in the first block to 195 ms at the last block was not statistically significant, paired-sample $t(14) = 1.5$, $p > .1$. Average standard deviation of response times during the speech and tone identification tasks are plotted as a function of practice in Figure 1D.

Electrophysiological Data

Figures 2 and 3 show the group AEPs recorded during the speech and the tone identification tasks, respectively, as a function of block repetition. In both tasks, auditory stimuli generated N1 and P2 waves that were largest at central sites and peaked at about 100 and 180 ms respectively, after sound onset. The N1 and P2 waves inverted in polarity at mastoid sites, consistent with
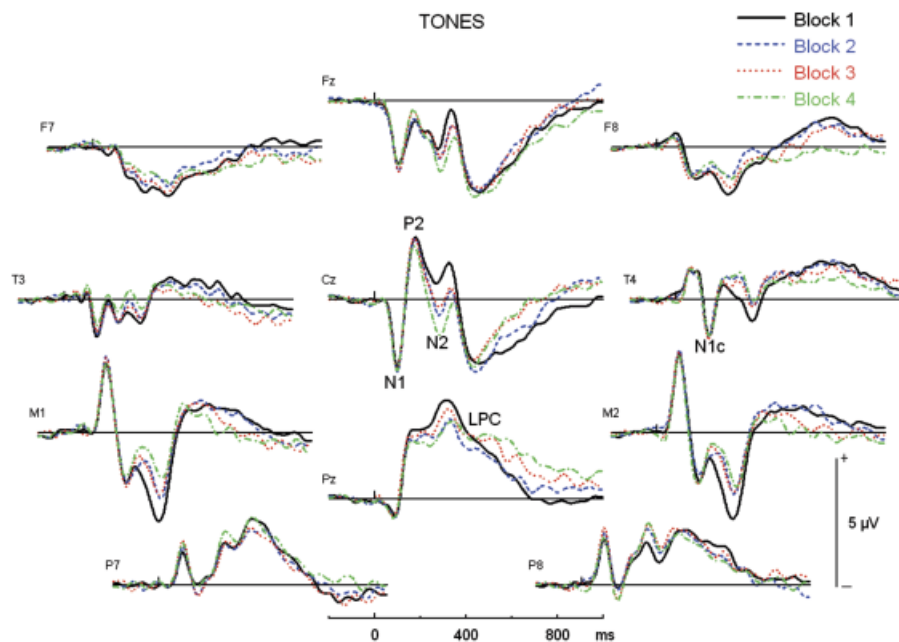

Figure 3. Group mean AEPs elicited by the pure tone stimuli as a function of block from a subset of electrode over the midline frontal (Fz), central (Cz), and parietal (Pz) regions. F7 = left frontal; F8 = right frontal; T3 = left temporal; T4 = right temporal; M1 = left mastoid; M2 = right mastoid; P7 = left parietal; P8 = right parietal.

generators in the superior temporal plane along the Sylvian fissure. In addition to sensory evoked responses, the AEPs comprised an N2 at 280 ms, a sustained frontal negativity at 400 ms, and a LPC between 300 and 600 ms at inferior parietal sites (i.e., P7 and P8 electrodes). In the following paragraphs, we present the effects of practice (linear term of block effects), task (pure tone vs. speech sound identification), and hemisphere (left vs. right) on the amplitude of AEP in components that were shown to be affected by rapid auditory learning (Alain et al., 2007, 2010). An interaction of the effects of the linear term of block and task on AEP amplitude supports our hypothesis of dissociation between the effects of perceptual learning (speech identification task) and mere task repetition (tone identification task) on AEP.

The N1 wave. The N1 peak amplitude measured at the frontocentral sites was larger for pure tones than for speech sounds, a main effect of task, $F(1,14) = 23.15$, $p < .001$, $\eta_p^2 = .62$. The interaction between task and the linear term of block was significant, $F(1,15) = 6.04$, $p < .05$, $\eta_p^2 = .3$. Whereas the N1 amplitude to speech sounds decreased with practice, linear term of block, $F(1,15) = 5.9$, $p < .05$, $\eta_p^2 = .3$, no such change was observed during the tone identification task ($F < 1$).

The N1c component. The N1c component, a negative deflection that peaked at about 140 ms at T3 and T4 electrodes, was larger for pure tones than for speech sounds, a main effect of task, $F(1,14) = 36.9$, $p < .001$, $\eta_p^2 = .72$, and larger over the left than the right hemisphere, a main effect of hemisphere, $F(1,14) = 8.2$, $p < .01$, $\eta_p^2 = .37$. These two effects of task and hemisphere interacted, $F(1,14) = 17.4$, $p < .005$, $\eta_p^2 = .55$: For speech sounds, amplitudes over the left hemisphere were substantially larger than for the right hemisphere, a main effect of hemisphere, $F(1,14) = 18.4$, $p < .05$, $\eta_p^2 = .57$, whereas no reliable hemispheric difference was noted for pure tones. The interaction of task and the linear term of block was not significant, $F(1,14) = .3$, $p > .5$. But, there was a significant three-way interaction of task, the linear term of block, and hemisphere, $F(1,14) = 5.5$, $p < .05$, $\eta_p^2 = .28$, indicating that N1c amplitude *decreased* with practice for tones only over the left hemisphere, an interaction of task and the linear term of block over the left hemisphere, $F(1,14) = 6.1$, $p < .05$, $\eta_p^2 = .30$, and a linear term of block for the tone task, $F(1,14) = 8.0$, $p < .05$, $\eta_p^2 = .37$, with no block effects for the speech task ($F < 1$).

The P2 wave. The P2 peak amplitude at the frontocentral sites was larger for speech sounds than for tones, a main effect of task, $F(1,14) = 10.8$, $p < .01$, $\eta_p^2 = .44$. The interaction between task and the linear term of block was significant, $F(1,14) = 6.8$, $p < .05$, $\eta_p^2 = .33$, with P2 amplitude in the speech identification task decreasing as a function of block, a linear term of block for the speech task, $F(1,14) = 4.0$, $p = .06$, $\eta_p^2 = .22$. Omitting Block 1, given that procedural learning may be dominant during the first block, resulted in a significant effect of the linear term of block, $F(1,14) = 6.6$, $p < .05$, $\eta_p^2 = .32$, but not in the tone identification task ($F < 1$).

The N2 wave. The N2 mean amplitude at the frontocentral sites was larger for speech sounds than for tones, a main effect of task, $F(1,14) = 23.1$, $p < .001$, $\eta_p^2 = .61$, and it increased with practice, linear term of block, $F(1,14) = 33.7$, $p < .001$, $\eta_p^2 = .71$. These two effects interacted, an interaction of task and the linear term of block, $F(1,14) = 5.3$, $p < .05$, $\eta_p^2 = .28$, indicating that the increase in N2 amplitude with practice was only significant in the

tone identification task, a linear term of block, $F(1,14) = 21.4$, $p < .001$, $\eta_p^2 = .6$. For the speech sounds, the N2 amplitude was not significantly affected by practice, a linear term of block, $F(1,14) = 2.5$.

At temporal sites, we can also see a modulation during the same interval as the N2 recorded at central sites (i.e., 260–320 ms; peaking at 280 ms) with an inverted polarity (at mastoid sites). The effect of moving from Block 1 to Block 4 at the temporal sites, however, was comparable for both speech and tone tasks, linear terms of block, $F(1,14) = 15.4$, $p < .005$, $\eta_p^2 = .52$, and $F(1,14) = 13.7$, $p < .005$, $\eta_p^2 = .50$, respectively, with no interaction between task and the linear term of block ($F < 1$).

The LPC wave. The processing of speech and tone signals was associated with a sustained positive wave that peaked at about 450 ms after sound onset at parietal-occipital sites. The effect of task and block on the LPC was quantified over left (P1, PO3) and right (P2, PO4) parietal sites for the 300–500-ms interval. The decrease in amplitude with practice was more substantial in the speech task, an interaction between task and the linear term of block, $F(1,14) = 4.6$, $p < .05$, $\eta_p^2 = .25$, and over the left hemisphere, an interaction between the linear term of block and hemisphere, $F(1,14) = 4.0$, $p = .06$, $\eta_p^2 = .22$. Considering the speech task separately, we can observe an interaction between the linear term of block and hemisphere, $F(1,14) = 6.2$, $p < .05$, $\eta_p^2 = .31$, indicating larger practice effects over the left hemisphere in the speech task. No such effects of block or hemisphere can be found in the tone identification task. The three-way interaction between task, the linear term of block, and hemisphere was not significant, $F(1,14) = 2.9$.

The sustained negative wave. In addition to the sustained positive wave at posterior sites, there was a sustained negative wave at frontal sites that began at about 200 ms after the sound onset and lasted for several hundred milliseconds. The effects of task, block, and hemisphere were quantified for the 300–500-ms interval over the left (F7, AF7, AF3, F5) and right (F8, AF8, AF4, F6) frontal sites. Practice-related amplitude decrement was more robust during the speech than the tone task, Task $\times$ Linear Term of Block interaction, $F(1,14) = 6.2$, $p < .05$, $\eta_p^2 = .31$. In a separate analysis of the speech task, we found that the decrease in amplitude with practice was more substantial over the left hemisphere, a Linear Term of Block $\times$ Hemisphere interaction, $F(1,14) = 6.4$, $p < .05$, $\eta_p^2 = .31$. In the tone identification task, no effects for hemisphere or practice were observed. The three-way interaction between task, the linear term of block, and hemisphere was not significant, $F(1,14) < 1$.

"Ba" versus "mBa". Alain et al. (2010) have shown that the AEP signatures of the "Ba" and "mBa" stimuli were more similar at the beginning of the session when the two stimuli were more difficult to distinguish than at the end of the session when identification was easier. This effect was mainly manifested in latency. In the present study, we did not replicate these effects in any of the sites tested, nor did we find any differences in amplitude between the two speech signals.

Summary of AEP Findings

A summary of the AEP data is presented in Table 2. In most of the investigated components we found a significant interaction of the task and the linear term of block effects. A decrease in amplitude unique for the speech task is found in four components: the N1 and P2 waves at the frontocentral sites, the LPC at the

Table 2. *Summary of the Linear Effects of Practice (from Block 1 to Block 4) on the Amplitude of AEP Recorded for Speech Identification and Tone Identification Tasks*

			Practice effects on amplitude	
	Time (ms)	Site	Speech task	Tone task
N1	70–130	frontocentral	decrease	—
N1c	115–165	left & right temporal	—	decrease (left hemisphere)
P2	150–210	frontocentral	decrease	—
N2	270–330	frontocentral	—	increase
N2 (inverted-polarity)	270–330	left & right temporal	increase	increase
Sustained positivity, LPC	300–500	left & right parietal	decrease (left hemisphere)	—
Sustained negativity	300–500	left & right frontal	decrease (left hemisphere)	—

Note: In all the presented components we found an interaction of the effect of task (tone vs. speech) and the linear effect of practice.

parietal sites, and the sustained negativity at the frontal sites (see Figure 4). In the last two components, the decrease was localized over the left hemisphere. Effects unique to the tone task were found in the N2 wave at the frontocentral sites (an increase in amplitude) and for the N1c wave at the temporal sites (a decrease over the left hemisphere). The amplitude of the inverted-polarity N2 wave (recorded in the temporal sites) was the sole investigated component that showed an equivalent effect (increase with practice) in both the speech and the tone tasks. The analysis of peak latency was found to be of no significance to the study and it will not be further discussed (a summary of peak latency analysis is presented in the Appendix).

Relations between Behavioral and Electrophysiological Measures

Previous studies suggest a link between changes in neural activity and behavioral improvement (Alain et al., 2007; Ross & Tremblay, 2009; Tremblay et al., 2009). Here we correlate behavioral measures with the averaged amplitude recorded across the respective cluster of electrodes for each AEP (one-tailed tests for the averages across participants[3]). In the speech task, we found that the increase in d' with practice (indicating improved perceptual discrimination between the two speech sounds) was negatively correlated with the amplitude of the following waves: the N1 (at the frontocentral sites, $r_p = -.96$, $p < .05$), the P2 (at the frontocentral sites, $r_p = -.61$, n.s.), the LPC (at the left parietal sites, $r_p = -.98$, $p < .01$; averaged across both hemispheres, $r_p = -.97$, $p < .05$), and the sustained negative wave (at the left parietal sites, $r_p = -.88$, $p = .058$; averaged across both hemispheres, $r_p = -.94$, $p < .05$). These correlations indicate that the increase in ability to efficiently identify speech sounds is paralleled by changes in AEP amplitude with practice. We further found that the increase with practice in the amplitude of the inverted-polarity N2 at the temporal sites was paralleled with an increase in d' for speech trials ($r_p = .95$, $p < .05$). In the tone identification task, we found that the increase in standard deviations of response latencies was correlated with an increase in N2 amplitude at frontocentral sites ($r_p = .91$, $p < .05$) and the inverted-polarity N2 amplitude at temporal sites ($r_p = .93$, $p < .05$). All six significant correlations are presented in Figure 5.

---

[3]Because the direction of practice effects on the behavioral measures and on the AEP amplitudes was already established in previous ANOVAs, a one-tail test was indicated.

Discussion

In this study, we aimed to disentangle the contribution of perceptual learning and task-repetition without learning on performance and on the rapid changes observed in auditory evoked potentials. To do this, we compared performance and related AEPs on a demanding speech identification task as well as an easier pure tone identification task. This type of research design enabled us to separately evaluate AEP changes that accompany learning (speech identification task) from task-related brain processes common to both tasks but in the absence of learning (tone identification task). Moreover it permitted us to study brain processes active during the acquisition of VOT contrasts in a way that differed from passive recordings, following long-term training (Tremblay et al., 2009). As was expected, we observed rapid changes in the AEPs that could be attributed to perceptual learning (in the speech identification task) that were different from AEP changes that were most likely related to mere task repetition (in the tone identification task).

Behavioral Findings

Behavioral data replicate and extend those of our prior study (Alain et al., 2010). Listeners' ability to identify subtle acoustic differences contained in speech sounds improved significantly during the first hour of practice. With practice (without feedback), the perceptual discriminability of speech sounds increased (an increase in sensitivity—d'), but the decision criterion adopted by participants did not change (bias—β). The improved accuracy with practice signifies a change in perceptual sensitivity, the ability of listeners to discriminate between the two speech signals. This finding is consistent with a prior behavioral study suggesting that rapid improvement in auditory perception is best accounted for by perceptual learning (i.e., changes in perception) rather than by procedural learning (Hawkey et al., 2004). Response times for the speech stimuli decreased with practice, suggesting that participants were able to improve the efficiency of the processing underlying the speech identification task. Conversely, in the "control" pure tone identification task, we found accuracy to be at ceiling and to decrease slightly with practice from the first to the last block. Response times for this task did not change within an hour of practice, but the variability of response times increased steadily with practice. These behavioral findings show clear effects of task repetition and might be suggestive of a decrease in attention with task repetition to this very easy tone identification task.
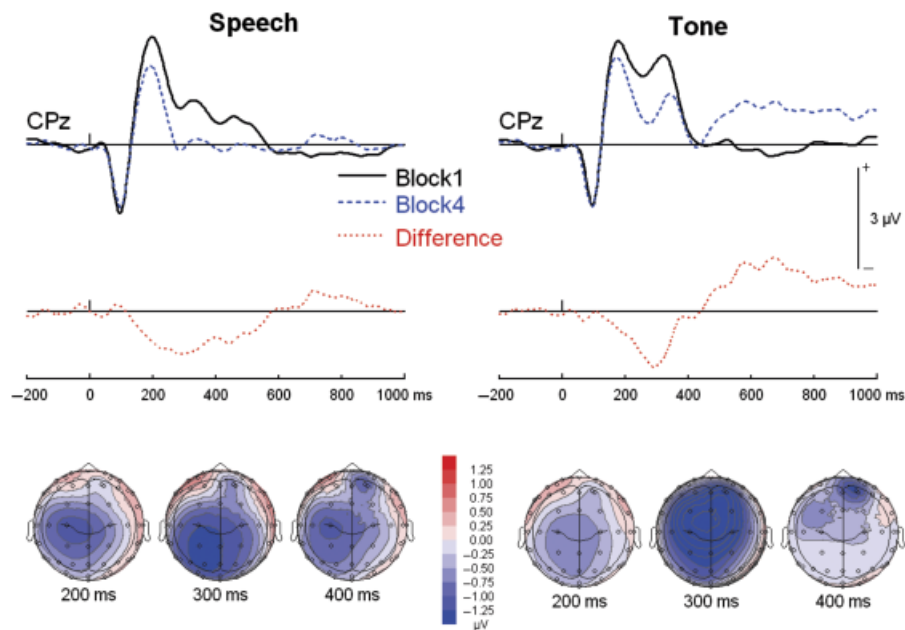
Figure 4. Learning and task-repetition effects. Left side of the panels presents data for the speech identification task and the right side for tone identification task. The top traces show the group mean AEPs recorded at the midline central parietal location (i.e., CPz) for the first and last block of trials. The traces below show the corresponding difference wave between AEPs recorded during Block 1 and Block 4 (i.e., last). The bottom row shows the contour maps from the difference wave at three difference latencies.

AEP Findings

Both the speech and the tone identification tasks involve repeated presentation of the stimuli; however, only the speech task coincides with learning. AEP analysis shows that this behavioral difference is manifested by different rapid neuroplastic changes. Coinciding with a rapid steady improvement in the speech identification task, the electrophysiological data show a steady decrease in N1 and P2 amplitudes recorded at the frontocentral sites, a steady decrease in the mean amplitude for the late positive component over the parieto-occipital region, as well as a steady decrease in the sustained negative wave at frontal sites. These rapid AEP changes differ from those observed during the tone identification (control) task and therefore cannot be attributed solely to task repetition. Instead, our results reveal rapid changes in AEP amplitude that are specific to rapid learning. In the following sections, we consider the consequences of these changes in each component.

N1. In the present study, the N1 amplitude decreased with task repetition, and this change as a function of practice was greater for the speech sounds than for the tones. In addition, the decrease in N1 amplitude was correlated with the increase in d' for speech sounds. The effects of task repetition on the N1 amplitude elicited by the speech sounds replicated findings from a prior study using the same material (Alain et al., 2010). This N1 decrement appears to index rapid learning. The N1 deflection has been shown to reflect the storage of information about auditory stimuli (e.g., Lu, Williamson, & Kaufman, 1992). Consequently, the decrement in N1 amplitude with rapid learning may reflect changes in the representation of the speech signals in auditory memory. Alternatively, it could be attributed to properties akin to adaptation, based on previous patterns reported by Ross and Tremblay (2009) using the same speech sounds (note that their paradigm involved only passive listening).

N1c. The N1c elicited by speech sounds was little affected by practice, in contrast to the parallel effect on the N1 recorded at the frontocentral sites. This finding suggests that different N1 components (i.e., N1a, N1b, and N1c) may have a different sensitivity to task repetition and learning. Although the N1c to speech sounds was little affected by practice, there is evidence that the N1c amplitude increases with multiple daily training sessions (Bosnyak, Eaton, & Roberts, 2004). Our findings, along with those from earlier studies that used multiple daily training sessions, suggest that the N1c may best index slow perceptual learning rather than rapid perceptual learning. In the present study, the N1c elicited by pure tones over the left hemisphere *decreased* slightly with practice. We note that this was found only in a triple interaction, whereas the interaction of the main effects of (the linear term of) block and task was not significant. Nonetheless, this finding was unexpected, and further studies are needed to elucidate the differential pattern of the N1c and N1 for repeated presentations of the speech and tone stimuli recorded at the frontocentral sites.

N2. During the speech identification task, the N2 wave recorded over the frontocentral sites increased with practice, but this effect failed to reach significance, despite improvement in speech identification. In a prior study using more practice blocks (that yielded greater improvement in accuracy), we found a significant increase in N2 amplitude (Alain et al., 2010). The lack of learning-related changes in N2 amplitude for the speech identification task in the present study seems inconsistent with prior research suggesting that the N2 reflects a decision process (e.g., stimulus classification) that controls behavioral responses in identification tasks (Ritter, Simson, & Vaughan, 1983; Ritter, Simson, Vaughan, & Friedman, 1979; Ritter et al., 1984). Given that participants' ability to identify the two speech tokens improved with practice, one would expect corresponding changes in the amplitude and/or latency of the N2 wave. This discrepancy could be due to the small gain in performance following a limited number of practice blocks in our study, relative to Alain et al. (2010).
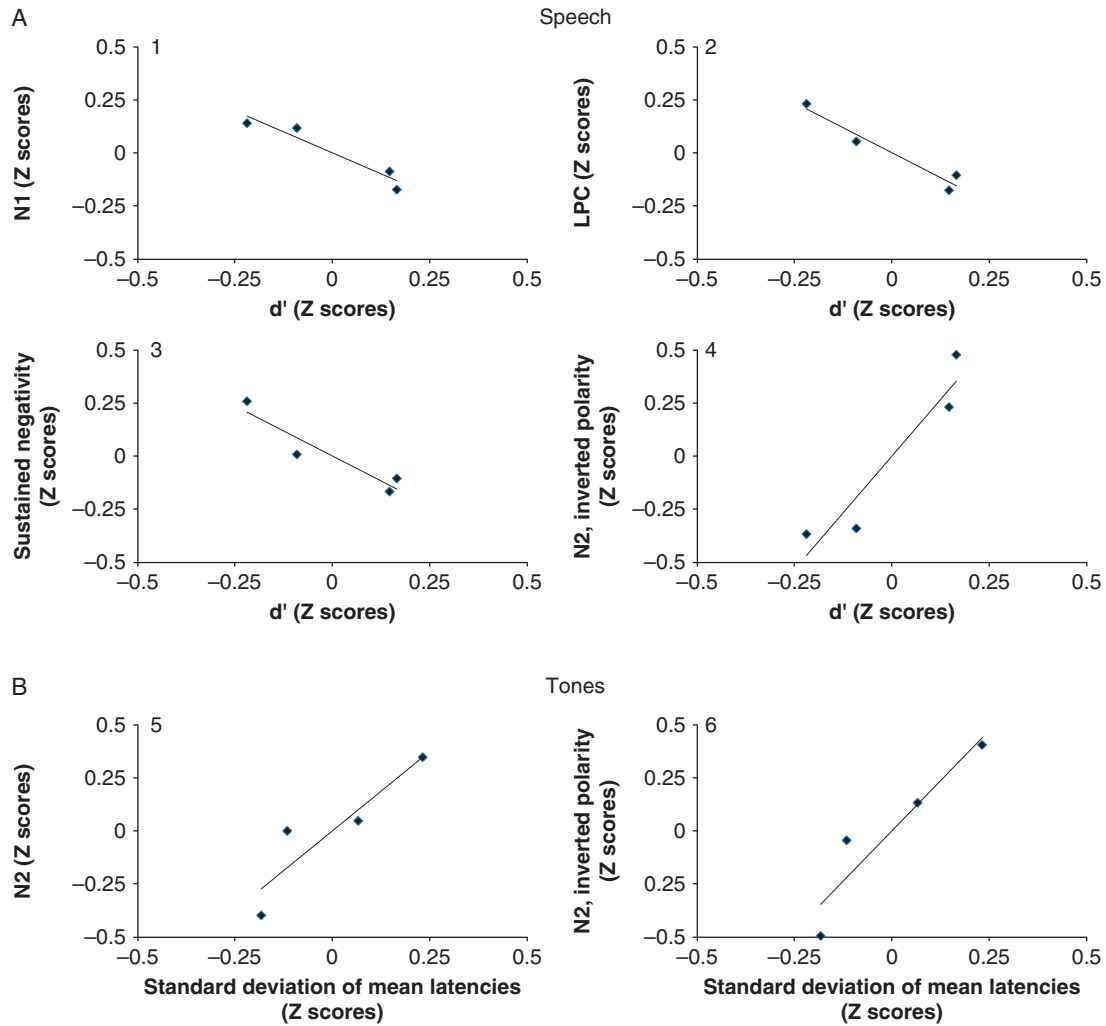
Figure 5. Relations between behavioral and electrophysiological measures. A: Average AEP amplitudes (*Z* scores) per experimental block of (1) N1 (frontocentral sites), (2) LPC (parietal sites), (3) sustained negativity (frontal sites), and (4) N2 (inverted polarity, temporal sites) waves elicited by speech stimuli, as a function of group means (*Z* scores) of *d'* for identification of speech stimuli. B: Average AEP amplitudes per experimental block of (5) N2 (frontocentral sites) and (6) N2 (inverted polarity, temporal sites) elicited by pure tone stimuli as a function of group mean (*Z* scores) of the standard deviation of latencies in pure tone trials. Continuous lines represent the linear correlations of the behavioral and electrophysiological measures.

Interestingly, the N2 amplitude increased during the tone identification task, which did not yield improvement in performance, but rather was accompanied by an increase in variability of response latencies and a slight decrease in accuracy of performance. These findings suggest that processes occurring during the N2 interval are not solely related to stimulus classification and target selection. However, it is important to point out that different processes may underlie the N2 generated during the speech and tone identification tasks. Indeed, the N2 recorded during the tone identification preceded response times by approximately 200 ms, whereas that recorded during the speech identification task preceded the response times by about 300–400 ms. Moreover, the N2 recorded during the speech identification task was largest over the left central-parietal scalp region, whereas the N2 elicited during the tone identification was largest at frontocentral sites. One possibility is that the changes observed during the tone identification task could reflect response-related processes whereas those involved during the speech task may index stimulus categorization.

Late positive and negative components. Highly prominent changes in AEPs associated with improvement in the identification of speech stimuli were the decreased amplitude of the late positive component over the parieto-occipital region and the sustained negative wave over the frontal sites. These neuroplastic changes can be directly related to learning rather than task repetition, because no reliable difference in either the LPC or the sustained negativity amplitude were found for the tone stimulus (for which performance did not improve). The learning-related decrease in LPC and sustained negativity amplitude, which was negatively correlated with an increase in sensitivity, may reflect improvement in stimulus categorization and perceptual processes: With practice, the perceptual discrimination of the two speech stimuli increased, as reflected by an increase in *d'*, and this was negatively correlated to the decrease in these two AEP components. This may also reflect monitoring effects: As the session progresses and identification accuracy increases, participants may feel more confident about their ability to correctly identify the speech sounds. These changes can also index improvement in

memory updating based on new information and/or increased familiarity with the auditory material that modulate the amplitude of evoked responses at frontal sites (e.g., Cycowicz & Friedman, 1998; Lebrun, Clochon, Etevenon, Baron, & Eustache, 1998). In the present study, the effect of practice on the LPC was opposite to that observed in Alain et al. (2010), where the LPC *increased* with practice for the speech sounds. This discrepancy between the two studies may be related to the experimental context. In the Alain et al. (2010) study, the speech and the control (noise) stimuli were presented within the same block of trials, whereas in the present study only the two speech sounds were presented within a block of trials. The number of experimental blocks was largely reduced in the current study. Moreover, in the previous study all speech blocks were consecutive, whereas in the current one the speech identification blocks alternated with the tone identification blocks, incurring a larger time lag and a task switch between speech identification blocks. Future studies should address the meaning of changes in slow sustained activity over the parietal and frontal regions and to what extent they reflect memory, confidence, or categorization processes. Note that learning effects on both of these late components were dominant over the left hemisphere. The left prefrontal lobe (specifically Broca's area) is linked to speech production, articulation, and comprehension related to a motor planning theory of speech comprehension (see a review on the role of the left frontal cortex in language processing in Hillis, 2007).

Concluding Remarks

In the present study, we examined neuroplastic changes that coincide with changes in the perception of speech signals. Listeners' ability to identify two consonant–vowel speech syllables that differ in VOT improved significantly within an hour of practice. An improvement in identification accuracy is related to an increase in sensitivity (d') with practice rather than a change in response bias (β). These perceptual changes coincide with decreases in the amplitude of sensory evoked responses. These AEP changes cannot be explained by task repetition alone, because they did not occur during the repetition of the "control" tone task. The learning-related change in amplitude may indicate improved attentional processes linked to speech categorization or the coding of temporal cues.

REFERENCES

Alain, C., Campeanu, S., & Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *Journal of Cognitive Neuroscience, 22*, 392–403.

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

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.

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.

Clapp, W. C., Kirk, I. J., Hamm, J. P., Shepherd, D., & Teyler, T. J. (2005). Induction of LTP in the human auditory cortex by sensory stimulation. *European Journal of Neuroscience, 22*, 1135–1140.

Clapp, W. C., Zaehle, T., Lutz, K., Marcar, V. L., Kirk, I. J., Hamm, J. P., et al. (2005). Effects of long-term potentiation in the human visual cortex: A functional magnetic resonance imaging study. *NeuroReport, 16*, 1977–1980.

Cycowicz, Y. M., & Friedman, D. (1998). Effect of sound familiarity on the event-related potentials elicited by novel environmental sounds. *Brain and Cognition, 36*, 30–51.

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.

Hawkey, D. J., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7*, 1055–1056.

Hillis, A. F. (2007). Language and the frontal cortex. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 306–316). New York: Guilford Press.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Lebrun, N., Clochon, P., Etevenon, P., Baron, J. C., & Eustache, F. (1998). Effect of environmental sound familiarity on dynamic neural activation/inhibition patterns: An ERD mapping study. *NeuroImage, 8*, 79–92.

Lu, Z.-L., Williamson, S. J., & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science, 258*, 1668–1670.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* Hillsdale, NJ: Erlbaum.

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.

Molchan, S. E., Sunderland, T., McIntosh, A. R., Herscovitch, P., & Schreurs, B. G. (1994). A functional anatomical study of associative learning in humans. *Proceedings of the National Academy of Sciences USA, 91*, 8122–8126.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Cinical 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. *Cinical Neurophysiology, 111*, 53–65.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Cinical Neurophysiology, 118*, 2128–2148.

Ritter, W., Ford, J. M., Gaillard, A. W., Harter, M. R., Kutas, M., Näätänen, R., et al. (1984). Cognition and event-related potentials. I. The relation of negative potentials and cognitive processes. *Annals of the New York Academy of Sciences, 425*, 24–38.

Ritter, W., Simson, R., & Vaughan, H. G., Jr. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology, 20*, 168–179.

Ritter, W., Simson, R., Vaughan, H. G., Jr., & Friedman, D. (1979). A brain event related to the making of a sensory discrimination. *Science, 203*, 1358–1361.

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

Teyler, T. J., Hamm, J. P., Clapp, W. C., Johnson, B. W., Corballis, M. C., & Kirk, I. J. (2005). Long-term potentiation of human visual evoked responses. *European Journal of Neuroscience, 21*, 2045–2050.

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

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., Shahin, A. J., Picton, T., & Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Cinical Neurophysiology, 120*, 128–135.

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

APPENDIX

In this appendix, we present the effects of practice (linear term of block effects), task (pure tones vs. speech sound discrimination), and their interaction on AEP peak latency. The peak latency was defined as the time where the deflection reaches its maximum relative to stimulus onset. The latency for the N1, N1c, P2, and N2 was measured during the 70–130-, 115–165-, 150–210-, and 270–330-ms intervals, respectively. None of the effects were significant for the N1 and N1c waves. The P2 wave at the frontocentral sites peaked earlier for tones than for speech stimuli, $F(1,14) = 12.7$, $p < .005$, $\eta_p^2 = .48$. Moreover, the P2 peaked earlier with practice, $F(1,14) = 12.4$, $p < .005$, $\eta_p^2 = .47$. The interaction between task and practice was not significant ($F < 1$), indicating that this change in P2 latency may index task repetition rather than learning per se. The N2 wave at the frontocentral sites peaked earlier for tones sounds as well, $F(1,15) = 8.2$, $p < .05$, $\eta_p^2 = .37$, but the effect of practice was not significant, nor was the interaction of practice and task.