Neural Changes Associated with Nonspeech Auditory Category Learning Parallel Those of Speech Category Acquisition

Ran Liu and Lori L. Holt

Abstract

■ Native language experience plays a critical role in shaping speech categorization, but the exact mechanisms by which it does so are not well understood. Investigating category learning of nonspeech sounds with which listeners have no prior experience allows their experience to be systematically controlled in a way that is impossible to achieve by studying natural speech acquisition, and it provides a means of probing the boundaries and constraints that general auditory perception and cognition bring to the task of speech category learning. In this study, we used a multimodal, video-game-based implicit learning paradigm to train participants to categorize acoustically complex, nonlinguistic sounds. Mismatch negativity (MMN) responses to the nonspeech stimuli were collected before and after training, and changes in MMN resulting from the nonspeech category learning closely resemble patterns of change typically observed during speech category learning. This suggests that the often-observed "specialized" neural responses to speech sounds may result, at least in part, from the expertise we develop with speech categories through experience rather than from properties unique to speech (e.g., linguistic or vocal tract gestural information). Furthermore, particular characteristics of the training paradigm may inform our understanding of mechanisms that support natural speech acquisition. ■

INTRODUCTION

In speech perception, the mapping of acoustic variability to meaning is language specific and must be learned through experience with one's native language. It is well established that this learning begins in early infancy and exhibits the hallmarks of categorization, whereby perceptual space is warped to exaggerate differences between native-language speech sound categories and diminish differences within categories (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). Despite there being an appreciation for the profound influence of perceptual categorization on speech processing, we do not yet have a complete understanding of the mechanisms that support speech category acquisition. Part of the challenge lies in the difficulty of experimentally probing the effects of experience; even infant participants have had significant exposure to native-language speech (DeCasper, Lecanuet, Busnel, Granier-Deferre, & Maugeais, 1994; Moon, Cooper, & Fifer, 1993) that is impossible to quantify or systematically control. It is possible, however, to investigate mechanisms supporting general auditory category learning in an experimentally controlled manner by studying the effects of short-term, laboratory-based experience on perception of nonspeech sounds to which listeners have not had previous exposure. Training using nonspeech, as

opposed to foreign language sounds that often experience perceptual interference from native-language speech categories (e.g., Flege & MacKay, 2004; Guion, Flege, Akahane-Yamada, & Pruitt, 2000), provides a means of probing the boundaries and constraints that general auditory perception and cognition bring to the task of speech category acquisition (Holt & Lotto, 2008). Similar lines of research in the visual domain have yielded important insights into the neural mechanisms underlying face perception; for example, the work of Gauthier, Tarr, Anderson, Skudlarski, and Gore (1999) and Gauthier and Tarr (1997) on the perception of nonface objects with which participants develop expertise in laboratory-based training.

The auditory categorization literature is small relative to that of visual categorization, and most studies of auditory category learning have used training methods whereby participants make overt categorization judgments and receive explicit feedback (e.g., Holt & Lotto, 2006; Goudbeek, Smits, Swingley, & Cutler, 2005; Mirman, Holt, & McClelland, 2004; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999). Although social signals (Kuhl, Tsao, & Liu, 2003) and the kinds of feedback they may provide can shape speech acquisition (Goldstein & Schwade, 2008), explicit feedback and category labels do not seem to be necessary or even available for the acquisition of speech categories in infancy (Jusczyk, 1997).

In a recent study, participants received implicit, multimodal auditory training through a video game in which

Carnegie Mellon University

they encountered characters, rapidly identified them, and executed appropriate actions toward them (either "shooting" or "capturing") on the basis of their identities. Throughout the game, characters' visual appearances consistently co-occurred with categories of acoustically varying, auditory nonspeech stimuli (Wade & Holt, 2005). Participants received no explicit directions to associate the sounds they heard to the visual stimuli, but the game was designed such that doing so was beneficial to performance. As an assessment of learning following training, players categorized sounds explicitly by matching familiar and novel sound category exemplars to the visual characters. Participants' performance on this task indicated significant category learning and generalization relative to control participants (who heard the same sounds in the game but did not experience a consistent sound category mapping) after only 30 min of game play (Wade & Holt, 2005). The specific characteristics of the sounds defining the categories required that participants go beyond trivial exemplar memorization or use of unidimensional acoustic cues to learn higher dimensional acoustic relationships typically characteristic of speech categories. Furthermore, participants showed evidence of generalizing their learning to novel acoustic stimuli. These results indicate that even complex nonspeech auditory categories can be learned quickly without explicit feedback if there exists a functional relationship between sound categories and events in other modalities.

Very little is understood, however, about the neural changes that underlie this type of learning and whether they resemble those observed in speech category acquisition. There exists general debate regarding whether general auditory mechanisms, such as those that would be involved in learning and processing nonspeech auditory categories, are involved in learning and processing speech (e.g., Fowler, 2006; Lotto & Holt, 2006). To address these issues, the present study characterized changes in the neural response to complex nonspeech stimuli that result from categorization training via the implicit, multimodal video game paradigm of Wade and Holt (2005).

One marker that can probe training-induced changes in neural response to auditory stimuli is the MMN, a fronto-central component of auditory ERPs that is evoked by any perceptually discriminable change from a repeated train of frequent ("standard") stimuli to an infrequent ("deviant") stimulus (Näätänen, 1995; Näätänen & Alho, 1995; Näätänen, Gaillard, & Mäntysalo, 1978) and is also sensitive to abstract stimulus properties (Paavilainen, Simola, Jaramillo, Näätänen, & Winkler, 2001; Tervaniemi, Maury, & Näätänen, 1994). It is thought to arise from a sensory–memory trace encoding characteristics of the frequently repeated "standard" stimuli (Haenschel, Vernon, Dwivedi, Gruzelier, & Baldeweg, 2005). The amplitude and latency properties of the MMN reflect the magnitude of deviation between the oddball stimulus and the standard, repeated stimulus. Furthermore, it occurs preattentively,

requires no explicit behavioral response, and can be evoked while participants engage in other mental activities that do not obviously interfere with the stimuli being processed (e.g., while watching a silent movie if stimuli are auditory; Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995; Näätänen, 1991). Thus, the MMN can serve as an objective, quantifiable measure of the neural representations of auditory perception that emerge even during simple passive listening. Recent findings also suggest that it can be used not only to probe early perceptual processing but also as a reliable measure of higher level lexical, semantic, and syntactic processing. For example, MMN has been found to reflect distinctions between words versus nonwords in one's native language, the frequency of words as they occur in the native language, and the degree to which a word string violates abstract grammatical rules (for a full review, see Pulvermuller & Shtyrov, 2006). Finally, the MMN is malleable to learning and reflects learned generalizations to novel stimuli (Tremblay, Kraus, Carrell, & McGee, 1997), making it an appropriate measure for investigating the influence of short-term video game training on perceptual organization of nonspeech categories in the current study.

The MMN has been frequently used to index the warping of perceptual space by learned speech categories. Prior experiments with adult listeners have shown that pairs of stimuli drawn from the same speech category evoke either a small or a nonexistent MMN, whereas equally acoustically distinct stimulus pairs drawn from different speech categories elicit a conspicuous MMN (Dehaene-Lambertz, 1997; Näätänen & Alho, 1997; for a review, see Näätänen, 2001). These responses reflect categories that are specific to the native language (Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000; Sharma & Dorman, 2000). Furthermore, studies designed to single out MMN responses to purely phonetic change while controlling for acoustic change (Shestakova et al., 2002; Dehaene-Lambertz & Pena, 2001; for a discussion, see Näätänen, Paavilainen, Rinne, & Alho, 2007) have shown that there is a phonetic-change-specific response that localizes to the left hemisphere in most adults. Not surprisingly, data from infants (Cheour et al., 1998) suggest that such MMN patterns are not present from birth but rather emerge as a result of early native language experience. Specifically, the acquisition of native language speech categories between 6 and 12 months of age is reflected in an enhanced MMN response to pairs of sounds learned to be functionally distinct and thus belonging to different categories. The diminishment of MMN for stimulus pairs drawn from within the same native-language category may take longer to develop; it is not yet observed among 12-month-old infants. In addition, no laterality differences were reported in the MMN responses by 12 months of age, so this phonetic-change marker may also be slower to develop relative to the enhanced across-category amplitude marker.

The present study combined a short-term, 5-day sequence of training on the Wade and Holt (2005) video

game paradigm with two sessions of ERP recordings, one before training and the other after. A comparison of the pretraining and posttraining MMN responses with pairs of novel stimuli belonging to one or more of the trained categories allowed us to probe the effects of short-term categorization training on the early neural response to these sounds. We expected that the pretraining MMN waves would reflect acoustic differences between the stimuli because subjects would be naive to any functional categories before playing the video game. We further hypothesized that if, following training, participants were able to successfully learn the nonspeech sound categories via mechanisms similar to those supporting the natural acquisition of speech categories, the pattern of MMN changes should parallel those observed in prior studies for speech category learning. That is, the main result we expected to observe is a significant increase in the MMN to pairs of nonspeech stimuli drawn from distinct nonspeech categories; this would parallel the changes in MMN observed in infant speech acquisition (Cheour et al., 1998). We also predicted, to a lesser extent, an attenuation of the MMN (relative to pretest baseline) to pairs drawn from the same newly acquired category. Although this change is not observed in early infant speech acquisition, studies on adult listeners frequently report little-to-no MMN response to within-speech-category pairs (Dehaene-Lambertz, 1997; Näätänen & Alho, 1997), and behavioral experiments have suggested that listeners typically exhibit a learned increase in the perceptual similarity of within-category speech sounds as a consequence of native language experience. This is proposed to occur because category centroids behave like perceptual "magnets," drawing other exemplars closer in terms of perceived similarity (Kuhl, 1991; for a review, see Kuhl, 2000). The observation of changes in MMN patterns underlying successful category learning through the game training paradigm would provide evidence that newly learned nonspeech categories tap into some of the same neural mechanisms observed to be malleable for speech categorization (Kraus, McGee, Carrell, King, & Tremblay, 1995).

METHODS

Participants

Sixteen (9 men) adult volunteers aged 18–45 years (median age = 21 years) affiliated with Carnegie Mellon University or the University of Pittsburgh participated and were paid for their time. All were right-handed native English speakers, and none reported any neurological or hearing deficits.

Video Game Training Paradigm

In the video game, players navigate through a pseudo-three-dimensional space and are approached by one of four visually distinct, animated alien characters at a time. To succeed at the game, players must perform one of two distinct sequences of coordinated keystrokes on the basis of the identity of the character that approaches. Players consistently respond to two of the characters, the ones associated with Easy Category 2 and Hard Category 1, with a keystroke sequence labeled "capturing": holding down the "R" key while navigating toward and keeping the character in range using the arrow keys for a fixed time period (after which the character is automatically "captured"). Players respond to the other two characters, the ones associated with Easy Category 1 and Hard Category 2, with a keystroke sequence labeled "shooting": holding down the "Q" key, navigating to the character using arrow keys and pressing the spacebar to shoot the character. Each character's appearance is consistently paired with an experimenter-defined auditory category comprised of a set of six different sounds (see Figure 1). On each trial, the pairing is accomplished by presenting the visual character accompanied by the repeated presentation of one randomly selected exemplar from the corresponding sound category. The number of sound repetitions heard per trial varies on the basis of how long the player takes to execute a response, but because players play for a fixed amount of time per session, there exists an inverse relationship between the average number of sound repetitions per trial and the number of trials in a session. That is, if a player is advancing quickly, he or she will hear fewer sounds repeated per trial but experience more trials overall during the session. Thus, the total number of game-relevant sounds heard in each session is roughly equivalent across participants, regardless of skill or performance in the game.

Players are not explicitly asked to form audio–visual or audio–motor associations at any point in the game, they are not instructed about the significance of the sounds, and there is no overt sound categorization task. During the game, players also hear ambient and game-play-relevant sounds that are unrelated to the visual characters or to the category training; these may further deter players from explicitly recognizing the audio–visual relationships. On each trial, however, because the onset of the repeated sound exemplar slightly precedes the visual appearance of each character, the sound categories are highly predictive of the approaching character. As the game progresses, the speed at which the characters approach increases and they become harder to identify visually, so the rapid anticipation of upcoming characters through consistent multisensory association becomes gradually advantageous for game performance. Thus, in an implicit manner, the relationship of the sound categories to multimodal regularities encourages participants to group acoustically variable sounds into functionally equivalent categories. The effectiveness of this training paradigm in promoting auditory category learning has been verified by prior behavioral results in which, after only 30 min of playing this game, participants showed significant nonspeech

Figure 1. Schematic diagram of stimulus exemplars used in training. The lower frequency component (P1) is constant within a category of sounds and is paired with each of the six higher frequency (P2) components to form the six sounds defining each category. Visual characters and keystroke sequences corresponding to each sound category are shown.
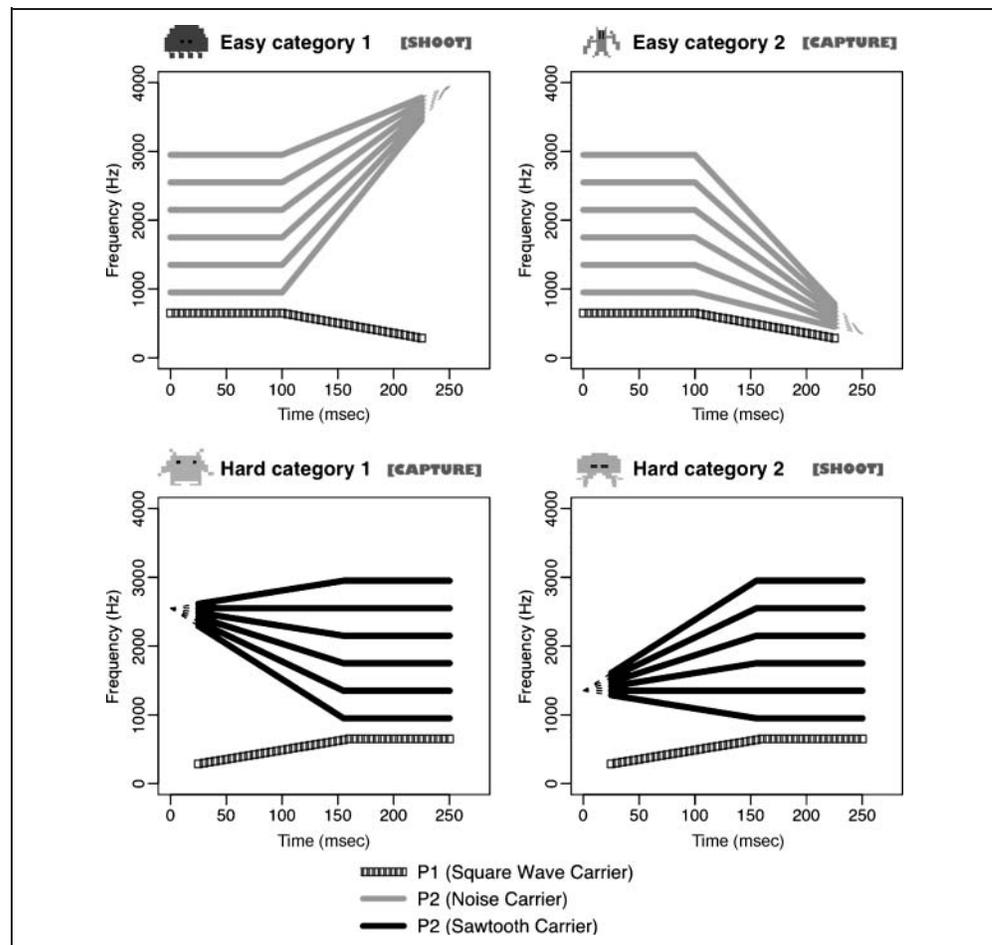
auditory category learning that generalized to novel, non-trained stimuli (for extended discussion, see Wade & Holt, 2005).

For the purposes of the present experiment, we extended the training time to allow participants the opportunity to reach ceiling levels of performance in the game. Results from pilot experiments suggested that five consecutive days of the 30-min training sessions were sufficient for many participants to reach ceiling levels; thus, we adopted this schedule of training in the current investigation.

Stimuli

Training Stimuli

For the training portion of this study, we used the four categories of nonspeech sounds used in the Wade and Holt (2005) behavioral experiments (for schematic spectrograms, see Figure 1; www.psy.cmu.edu/~lholt/php/gallery_irfbats.php for sound stimuli). Two of the categories (Easy Categories 1 and 2) were characterized by a 100-msec steady-state frequency followed by a 150-msec frequency transition. These categories were differentiated by an invariant, unidimensional acoustic cue. For one category, the frequency transition portion of one component

increased in frequency across time; for the other category, it decreased.

Critically, however, no such invariant cue distinguished the other two categories (Hard Categories 1 and 2). Both hard categories had exemplars with both rising and falling frequency patterns, and stimuli across the two hard categories completely overlapped in their steady-state frequencies. Only a higher order, interaction between onset frequency and steady-state frequency created a perceptual space in which the two hard categories were linearly separable (for a comprehensive discussion, see Figure 2; Wade & Holt, 2005).

All stimuli were designed to have two spectral peaks, P1 (carried by a square wave for all categories) and P2 (carried by a noise wave in the easy categories and by a sawtooth wave in the hard categories). In the two easy categories, P2 frequency is steady for the first 100 msec and then rises or falls to an offset frequency during the last 150 msec. In the two hard categories with sawtooth-carried P2 components, P2 frequency either rises or falls linearly from its onset during the first 150 msec and then remains at a steady-state frequency for the final 100 msec. The easy categories differ from the hard categories on several dimensions: The temporal windows during which both the P1 and the P2 frequencies are steady versus

transitioning, the direction of P1 frequency change, and the type of P2 carrier wave.

These stimuli capture many of the spectrotemporal characteristics of speech sounds, including the composition based on combined spectral peaks (modeling speech formants), the syllable-length temporal duration, the combination of acoustic fine-structure modulated by time-varying filters (as for consonants' formant trajectories). The category relationship among stimuli also mirrored speech categories in the use of category exemplars that vary along multiple spectrotemporal dimensions with overlap across categories and lack of a single acoustic dimension that can reliably distinguish the two hard categories. Despite these similarities, the stimuli have a fine temporal structure unlike that of speech, and participants do not report them as sounding "speech-like" at all (see Wade & Holt, 2005).

Test Stimuli

Previous research has demonstrated that with 30 min of video game play, participants successfully learn both the easy and the hard categories above chance level and gen-eralize this category learning to novel sounds drawn from the same categories (Wade & Holt, 2005). For the purposes of the ERP portions of the experiment, we focused on participants' categorization of stimuli drawn from the hard categories because learning categories defined by noninvariant acoustic dimensions is a task more ecologically relevant to speech categorization and is more challenging than the trivial differentiation of rising versus falling frequency transitions that define the easy categories. In addition, we wanted any changes in neural activity to truly reflect changes in general categorization rather than only the perception of specific exemplars that are heard extensively in training. Thus, the ERP stimuli were comprised only of novel, untrained sounds belonging to the hard categories.

Four novel sounds drawn from the hard categories but never heard by listeners during the video-game-based training were selected for the ERP portion of the experiment. The complex nature of these categories required careful selection of test stimuli because there is no simple, single acoustic dimension that distinguishes the two. Thus, the specific characteristics of the test stimuli were selected by plotting the training stimuli of the two hard

Figure 2. Stimuli from the two hard categories (not separable by an invariant cue) plotted in the higher order space that reliably separates the categories. Novel MMN stimuli belonging to the two categories were chosen such that, within this space, Cartesian distances for both within-category and across-category STD/DEV pairs were equated.
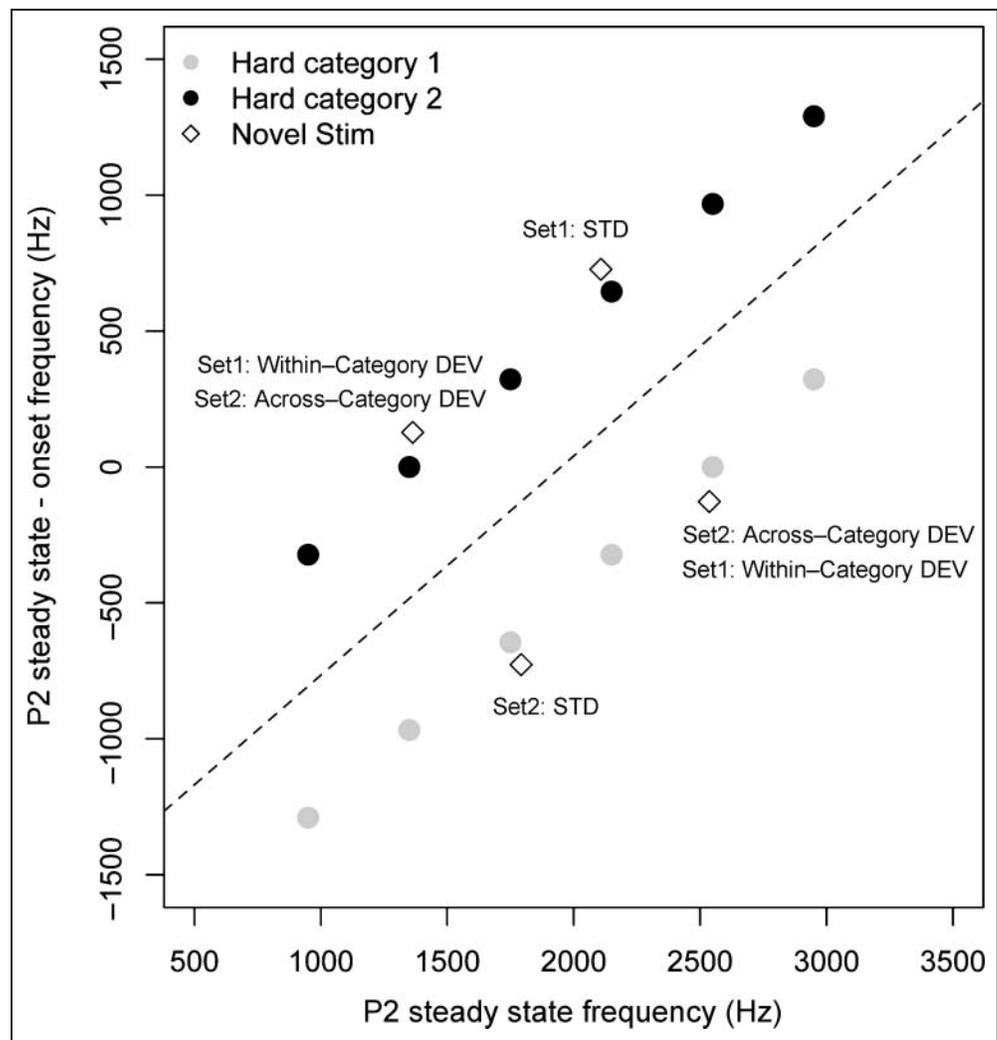
categories in the higher order acoustic space in which they are linearly separable (see Figure 2). Then, points in the space were selected such that (1) there were two points in each category, and (2) each point was equal in geometric distance from its two nondiagonally adjacent points in the higher order acoustic space. One point in each category was selected to define the standard (STD) sound for the MMN stream, and adjacent same-category ("within" category) and opposite-category ("across" category) points were assigned to define deviant (DEV) sounds for each STD sound. Thus, the geometric distance between every STD–DEV pair, regardless of category membership, was equated in the higher dimensional acoustic space that linearly separates exemplars defining the two hard categories. Due to the complex nature of the perceptual space within which categories were defined, we did not expect all stimulus pairs to be equally discriminable to naive listeners, as indexed by the pretraining MMN. However, the use of a pretest versus posttest comparison of neural responses with these stimuli provided individual baselines against which we could assess changes in the MMN that result from category learning.

Design and Procedure

Each participant played the video game for five consecutive days, 30 min per day, and completed an EEG recording session on the first day (before beginning training) and on the fifth day (following the end of training).

Day 1

Participants engaged in a 2.5-hour EEG recording session (see the following section for details). Immediately following this session, each participant was given instructions and asked to play the video game using the procedure described by Wade and Holt (2005). During the instructions, no explicit mention was made of the sound categories or their importance to game performance. Participants were instructed that their objective was to achieve the highest possible score in the game. They wore headphones (Beyer DT-100) calibrated to approximately 70 dB and played the game for 30 min.

Days 2–4

Participants played the video game for 30 min each of these 3 days, without any additional instructions.

Day 5

Participants played the video game for 30 min. Immediately following this session, a 10-min behavioral posttraining cat-

egorization task was administered in which participants were presented, in random order, four repetitions of each of 40 acoustic stimuli for a total of 160 trials. Each of the four sound categories (see Figure 1) contributed 10 distinct sounds: 6 familiar sounds per category were drawn directly from the sounds experienced in the game, and the remaining 4 sounds were the novel sounds that listeners were tested on during ERP sessions. The order of trial presentation was determined by a random in-place shuffle of all 160 trials. On each trial, participants heard the sound repeated through headphones (Beyer DT-100) while they saw the four visual alien characters of the video game and chose the character that best matched the sound they heard. Participants were instructed to respond as quickly as possible. Sounds were repeatedly played, with a 30-msec silent gap between presentations (as in the video game training), and they terminated when a button response was made. If participants failed to select a character within the first 1.5 sec of the trial, they heard a buzz sound and were presented with the next trial. Those trials to which participants did not respond to in time were not repeated later in the experiment and the responses were recorded as incorrect. Following this task, participants engaged in a 2.5-hour posttraining EEG recording session that was procedurally identical to the pretraining session.

Electrophysiological Recording

The EEG waves were recorded, using Ag/AgCl sintered electrodes embedded in an elastic fabric Quik-Cap, from 32 scalp locations arranged according to an extended 10–20 array: FP1, FPz, FP2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, POz, O1, Oz, and O2 (naming system of American Electroencephalographic Society; Sharbrough et al., 1991). Additional electrodes were placed on both the left and the right mastoid positions. Horizontal and vertical EOGs were calculated from recordings of pairs of electrodes placed on the outer canthus of each eye and above and below the right eye, respectively. The ground electrode was placed at location AFz. Signals were recorded in continuous mode using Neuroscan Acquire software (Compumedics NeuroScan, El Paso, TX) on a PC-based system through SynAmp2 amplifiers. All electrode signals were referenced to the left mastoid during recording, and impedances were kept below 10 kΩ throughout each participant's session. The data were sampled at a rate of 1 kHz with a band-pass filter of 0.1–200 Hz.

Stimulus presentation during EEG recording was divided into four blocks, one to test each STD–DEV stimulus pair (Set 1 STD and within-category DEV, Set 1 STD and across-category DEV, Set 2 STD and within-category DEV, and Set 2 STD and across-category DEV; see Figure 2).

Figure 3. Pretest butterfly plots, global field power (GFP), and dissimilarity (Dis) waves from 0 to 600 msec following stimulus onset and topographic maps at the time of peak GFP (indicated on GFP plot by red line) for the MMN responses to all four STD–DEV pairs.
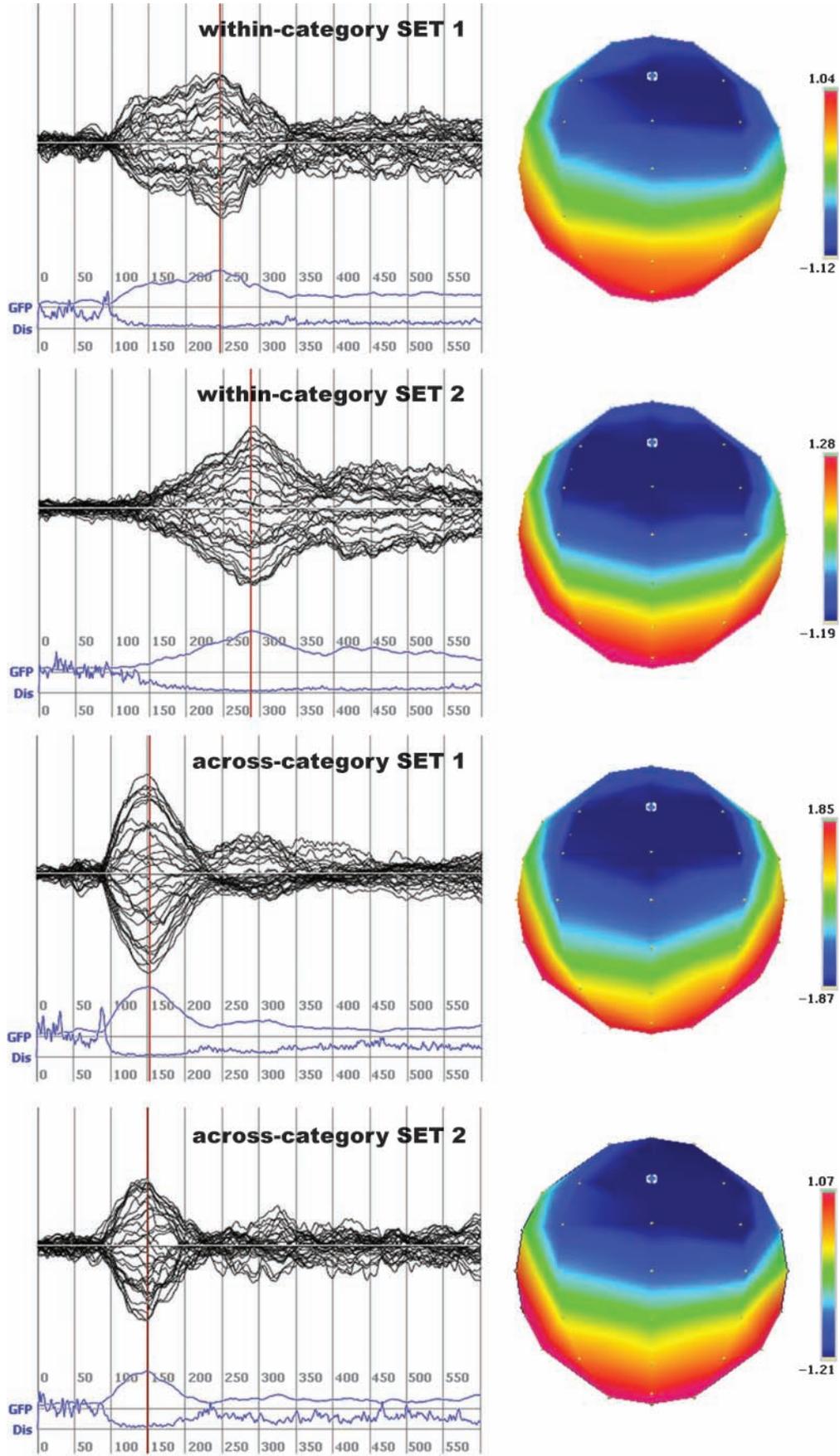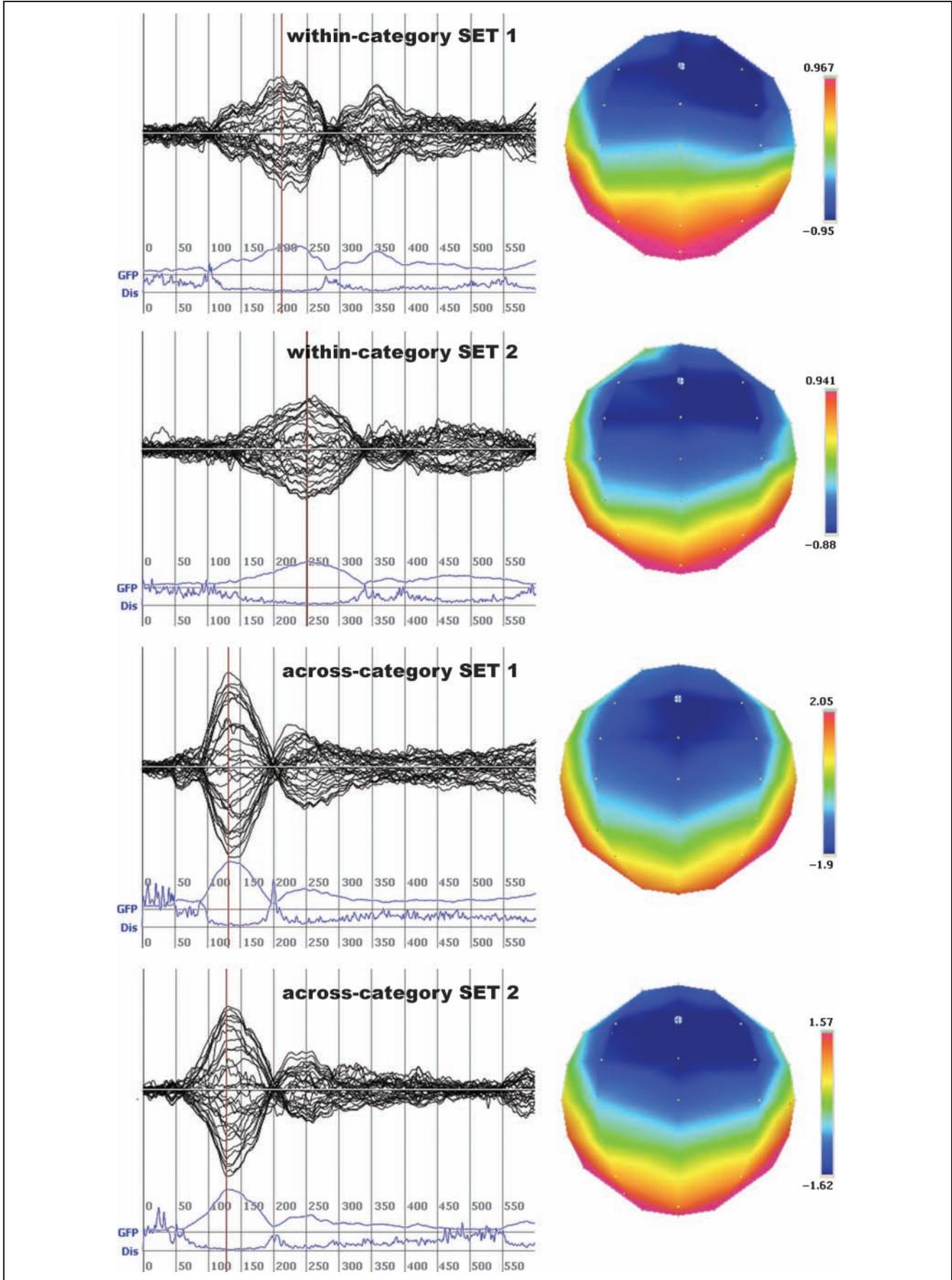
Within each block, the STD and DEV stimuli were presented in a pseudorandom sequence across which at least three consecutive standard stimuli separated any two presentations of deviant stimuli. The sequence was constructed by assigning a trial to be the STD sound with 100% probability if preceded by fewer than three consecutive STD trials or assigning it to be STD with 50% probability and DEV with 50% probability if preceded by three or more consecutive STD trials. This randomization scheme results in the presentation of approximately 1600 standard (80%) and 400 deviant stimuli (20%). Sounds were separated by a fixed stimulus onset-to-onset asynchrony of 500 msec. The presentation order of the four experimental blocks was counterbalanced across participants to eliminate any potential order effects. For each participant, block order was kept constant for the pretraining and posttraining sessions. Each block was 17 min long, and participants took a 3- to 5-min break between blocks, for a total experiment length of approximately 80 min.

Participants were seated comfortably in an electrically shielded experimental chamber and heard the stream of acoustic stimuli presented diotically over headphones (E-A-RTONE Gold 3A 50 ohm tubephone) while they watched a self-selected movie without sound or subtitles. Because the MMN reflects preattentive auditory deviance detection, participants were instructed to ignore the sounds and to focus on watching the movie.

Data Processing and Statistical Analyses

Data preprocessing was done using NeuroScan Edit software. All EEG channels were first rereferenced to the mean of the left and right mastoid recordings. The continuous EEG data were then band-pass filtered from 0 to 30 Hz and corrected for ocular movement artifacts using a regression analysis combined with artifact averaging (Semlitsch, Anderer, Schuster, & Presslich, 1986). Stimulus-locked epochs that ranged from 100 msec before stimulus onset to 500 msec after stimulus onset were extracted, baseline corrected relative to the period from -100 to 0 msec before stimulus onset, and run through an artifact rejection procedure in which epochs with amplifier saturation or voltage changes greater than 100 μV were automatically discarded. The stimulus-locked epochs were averaged for each possible combination of participant, session (pretraining vs. posttraining), stimulus pair (within vs. across category, Sets 1 and 2), and stimulus type (standard vs. deviant) to generate ERP waves. Finally, MMN responses were calculated from the waves of every participant/session/ condition by subtracting the average response to the standard stimulus in each block from the average response to the deviant stimulus with which it was paired.

Topographic and global field analyses were performed using Cartool (http://brainmapping.unige.ch/Cartool.htm). Well-defined MMN responses peaking between 100 and 300 msec were present in all four stimulus pairs at both pretest and posttest, verified by visual inspection of butterfly plots and mean global field power peaks (Figures 3 and 4). Topographic analyses revealed that, for all four pairs at both recording sessions, the MMN response was strongest at electrode Fz. An extensive literature has shown that this finding is typical in MMN studies (e.g., Winkler et al., 1999; Näätänen et al., 1997; Sams, Paavilainen, Alho, & Näätänen, 1985; for a review, see Näätänen, 2001). We thus focused the statistical analyses of averaged MMN waves on the response measured at electrode Fz.

Raw data from the averaged waves were imported into R (http://www.R-project.org) for statistical analyses. Peak amplitude and peak latency values of each averaged wave at electrode Fz were calculated across the range of 100- to 300-msec poststimulus, where MMN is typically observed (Maiste, Wiens, Hunt, Scherg, & Picton, 1995; for a review, see Näätänen, 2001) and within which peak amplitudes for all conditions in the present study were observed.

RESULTS
Behavioral

Categorization accuracies across all participants on the explicit posttraining task are summarized in Figure 5. On the basis of data from the Wade and Holt (2005) behavioral studies, we expected that if participants successfully learned sound-character associations in the video game, they would be able to match characters to sounds significantly above chance (25%) level. Indeed, overall accuracy across participants was well above chance, mean = 64.1%, $t(15) = 9.03$, $p < .001$. Furthermore, participants showed evidence of learning the two difficult sound categories defined in higher dimensional acoustic space and of generalizing the learning to untrained stimuli belonging to the categories. Under the assumption that the task of distinguishing easy and hard categories from each other was easily achieved (for empirical evidence, see Experiment 2 in Wade & Holt, 2005; Emberson, Liu, & Zevin, 2009), chance level for successfully categorizing an exemplar drawn from one of the two hard categories would be 50%. Participants showed performance significantly above this level of categorization accuracy for familiar hard category stimuli experienced in the video game, mean = 64.6%, $t(15) = 4.12$, $p < .001$, as well as for novel sounds drawn from the hard categories but not experienced during training, mean = 62.5%, $t(15) = 2.14$, $p = .049$. Thus, the explicit posttraining categorization task indicates that for

Figure 4. Posttest butterfly plots, global field power (GFP), and dissimilarity (Dis) waves from 0 to 600 msec following stimulus onset and topographic maps at the time of peak GFP (indicated on GFP plot by red line) for the MMN responses to all four STD–DEV pairs.
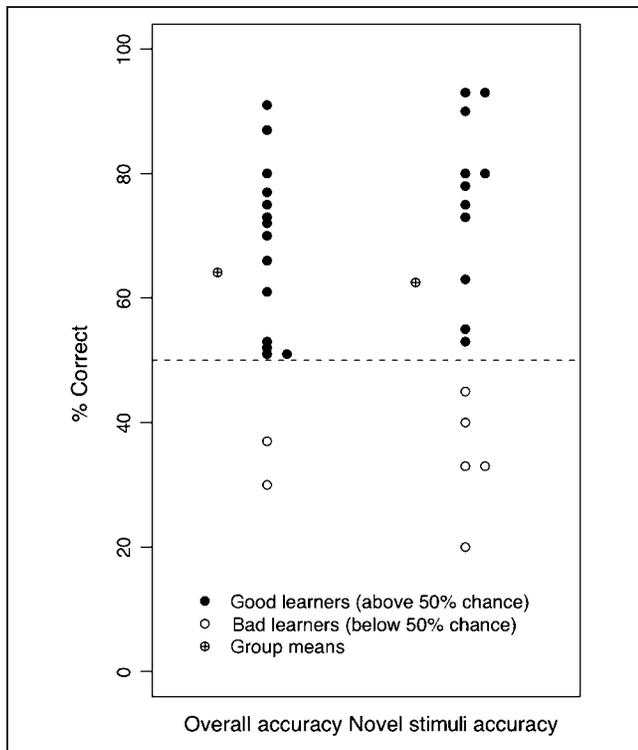
Figure 5. Categorization accuracy levels of all subjects on the posttraining explicit categorization task. Overall accuracy values reflect performance on both trained and novel stimuli. Novel stimuli accuracy values reflect performance on the novel hard category sounds that were heard during ERP recording. The two individuals scoring below the 50% chance level for overall accuracy are a subset of the five individuals scoring below chance level for novel stimuli accuracy; these five individuals were identified as "bad learners" for their failure to learn the categories robustly enough to generalize them to the novel, untrained MMN stimuli.

both the trained stimulus set and for the specific stimuli tested in the MMN paradigm, participants reliably learned the sound categories and generalized the learning to untrained stimuli.

Although the behavioral results show strong evidence of learning across participants, there also was individual variability in learning, as is typical of perceptual categorization tasks (Leech, Holt, Devlin, & Dick, 2009; Tremblay, Kraus, & McGee, 1998). Not all participants learned the Hard sound categories robustly enough to generalize them to the novel stimuli heard during ERP recording, judging by the conservative criterion of above 50% chance level categorization accuracy for the novel stimuli alone. This criterion was chosen to assure a stringent assessment of hard category acquisition and generalization because we expected that changes in MMN activity across training consistent with our hypotheses should emerge only for participants who learned successfully enough to correctly categorize the novel sounds presented during ERP recording. There were five nonlearners as indexed by a novel-stimulus categorization accuracy of less than 50% (see Figure 5), and we expected that these participants would

show little-to-no MMN changes in our hypothesized directions from pretest to posttest.

Posttraining categorization task performance also correlated positively with performance on the video game as indexed by both maximum level ($r = 0.55$, $p = .02$) and maximum score ($r = 0.63$, $p = .009$) achieved in the game, confirming that success in the training paradigm is strongly related to category learning and is a critical factor in explaining the individual differences seen in the posttraining behavioral results (for extended discussion, see Wade & Holt, 2005).

These results replicate the major findings reported by Wade and Holt (2005) but reveal an even higher degree of posttraining categorization performance for both trained, $t(24) = 2.793$, $p = .004$, and novel, $t(24) = 2.31$, $p = .03$, stimuli, measured by percentage of sounds accurately categorized in the explicit categorization task. Thus, extended experience with the video game seems to yield even better category learning. This further supports the efficacy of this particular training paradigm in driving the formation of distinct, spectrotemporally complex, and acoustically noninvariant auditory categories.

Pretraining ERPs

We used novel, nonspeech sounds to acquire experimental control over listener experience while modeling some of the complexities of speech category learning. However, the complex nature of the stimuli introduces some challenges for using MMN as a method because MMN is exquisitely sensitive to fine acoustic differences among stimuli. We made efforts to orient our category distributions in acoustic space such that our STD–DEV stimulus pairs were equivalently acoustically distinct (by equating the geometric distances of all pairs of stimuli in the higher dimensional acoustic space that reliably separate the categories). However, auditory perceptual space is not homogeneous; discontinuities exist and are known to influence the acquisition of auditory categories, including speech categories (Holt, Lotto, & Diehl, 2004).

Acknowledging this, we examined the group-averaged pretraining ERP waves (see Figure 6). Judging by differences in average MMN peak amplitude and peak latency across conditions, it is evident that participants did not naively perceive the pairs of sounds to be equally distinct across conditions.

Recognizing that these differences exist before experience with the nonspeech sounds, the critical measurement in assessing category learning reflected at the neural level was the *relative* change in MMN waves from pretraining to posttraining for each STD–DEV stimulus pair. Participants' naive MMN responses to the stimuli serve as baselines by which the influence of learning are assessed.

Changes in ERPs following Training

The averaged pretraining and posttraining ERP waves at electrode Fz for the entire group of participants are shown

in Figure 6. Paired t tests run for each condition on peak amplitudes and peak latencies of posttraining waves, relative to those for pretraining waves, revealed a significant decrease in within-category peak amplitude, $t(15) = -2.7372, p = .008$, and an earlier across-category peak latency, $t(15) = 2.598, p = .01$, in stimulus Set 1. A significant increase in across-category peak amplitude, $t(15) = 3.0623, p = .004$, was found in stimulus Set 2. No other peak amplitude or latency changes were significantly different across training.

Considering the degree of individual variability that the behavioral posttraining results revealed, we additionally investigated the neural changes of only those participants who exhibited robust learning and generalization for all categories. On the basis of our hypotheses, we were most interested in the neural changes underlying *successful* category learning, so the subset of participants that showed this behaviorally were most relevant for ERP analyses. To this end, we averaged together pretraining and posttraining ERP waves for the "good learners" ($n = 11$), indexed by above 50% posttest categorization accuracy (chance level assuming the easy vs. hard category distinction has

been accomplished) for the novel ERP stimuli. These averaged waves are shown in Figure 7. Paired t tests revealed that, for stimulus Set 1, there was a significant decrease in within-category peak amplitude, $t(10) = -2.392, p = .01$, a significant increase in across-category peak amplitude, $t(10) = 1.753, p = .05$, and earlier across-category peak latency, $t(10) = 2.261, p = .024$. For stimulus Set 2, a significant increase in across-category peak amplitude, $t(10) = 3.4181, p = .003$, was found.

Our primary objective was to investigate learning-driven neural changes, but failures of learning are also informative. Although the video game training was successful in training the majority (69%) of listeners to learn the categories, five participants did not exhibit behavioral evidence of learning sufficient enough to generalize the categories to untrained stimuli (see Figure 5). Because we predicted changes in the MMN driven by category learning, we would not expect significant MMN changes among these noncategorizers. The video game training was effective enough for this sample to be too small for strong statistical tests (which would test a predicted null effect). However, qualitative observation of pretraining versus posttraining MMN

Figure 6. Pretraining and posttraining MMN waves at electrode Fz, averaged across all participants ($n = 16$). *Statistically significant ($p < .05$) changes in either peak amplitude or peak latency from pretraining to posttraining.
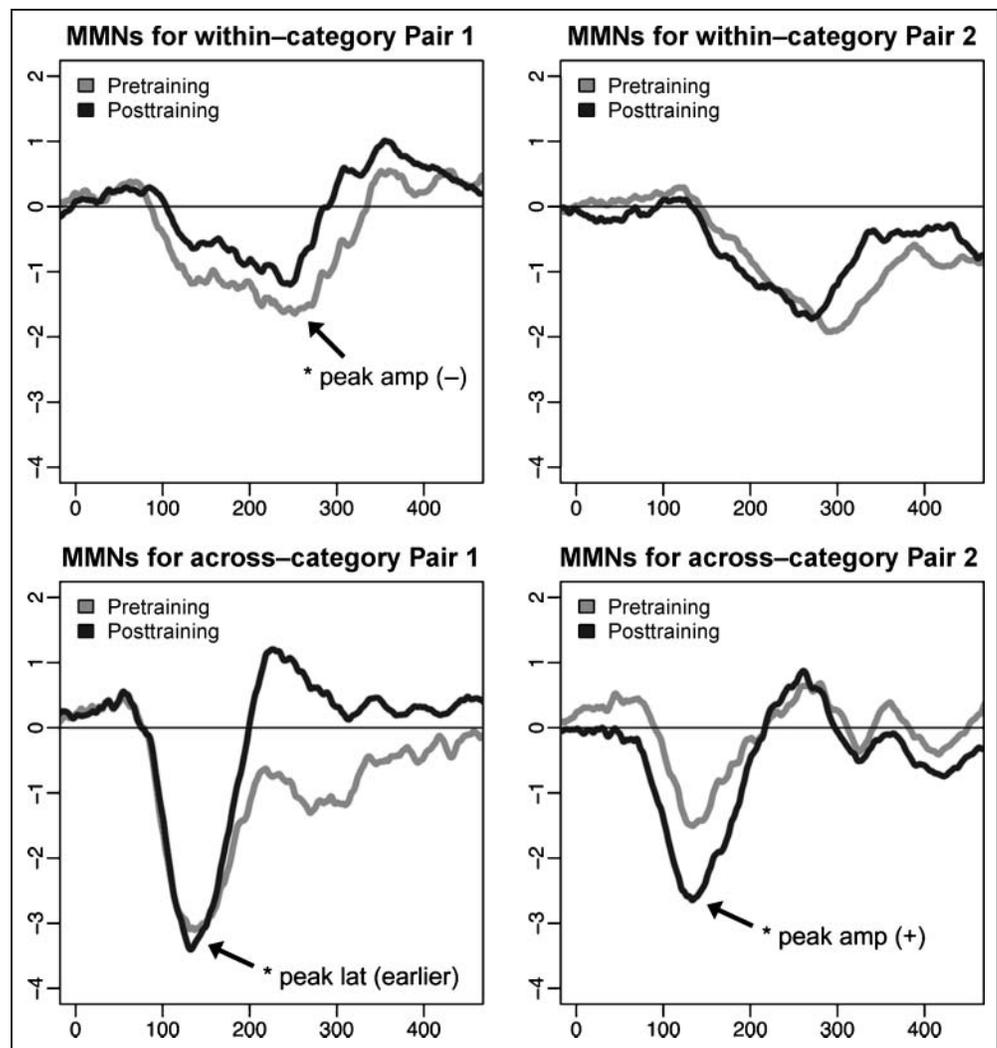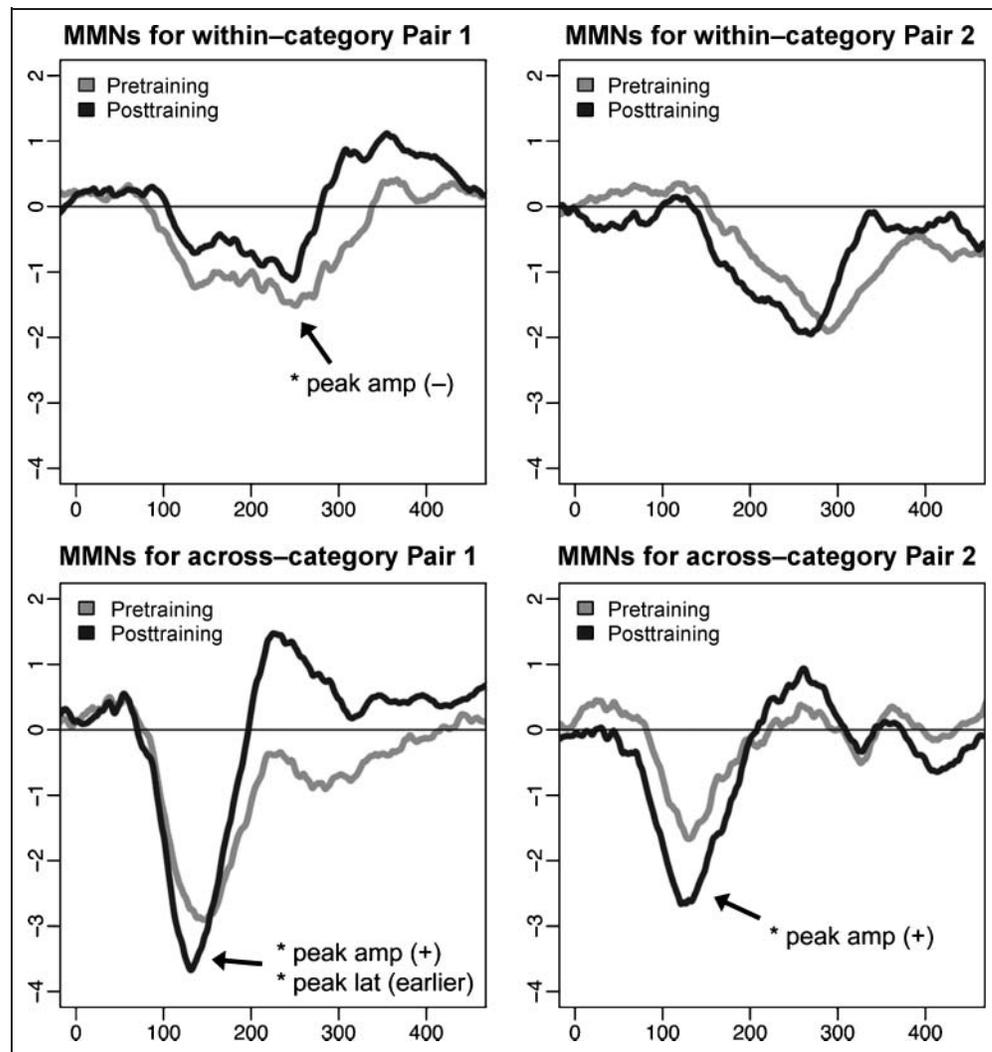
Figure 7. Pretraining and posttraining MMN waves at electrode Fz, averaged across all "good learners," those with above-chance performance on both trained and novel stimuli on all categories (*n* = 11). *Statistically significant (*p* < .05) changes in either peak amplitude or peak latency from pretraining to posttraining.


among these participants indicates very little change compared with those observed for participants who learned the categories (see Figure 8). One exception is the within-category pair in stimulus Set 2, which shows a decrease in peak amplitude following training. Interestingly, this is the only pair for which our hypothesized changes were not statistically significant among the participants who learned the categories well.

MMN was the chosen measure of neural change in this investigation, as it has been extensively used in studies of speech category acquisition, is malleable to training using nonspeech stimuli, and can systematically index degree of perceptual deviance by its magnitude. These characteristics of the MMN allowed us to make specific, theoretically driven directional predictions of expected changes following our training paradigm. The P3a, a frontal aspect of the novelty P3 response that sometimes follows MMN in oddball paradigms in which the deviant stimulus is sufficiently perceptually distinct from the standard stimulus, has been less extensively studied in the speech categorization and auditory training literature. Friedman, Cycowicz, & Gaeta (2001) have argued that the P3a response marks

an involuntary capture of attention by the deviant event as well as an evaluation of the event for subsequent behavioral action. Because our paradigm was passive and participants were specifically instructed not to attend to the sounds, it was unclear whether the perceptual disparity between sound pairs ought to elicit a P3a response at either pretest or posttest. On the basis of visual inspection, our data showed fairly distinct P3a responses in the posttest ERP waves (Figures 4, 6–8). Focusing on the "good learner" group as we did for the MMN results, an exploratory analysis reveals a significant increase (from pretest to posttest) in the P3a peak amplitude for the across-category stimulus Set 1 pair (*p* < .005) as well as a marginally significant increase for the across-category stimulus Set 2 pair (*p* = .07). No other pairs showed significant directional changes in P3a amplitude from pretest to posttest. The posttraining increase in P3a amplitude for across-category stimuli may indicate that the video game training cultivated an "attention-grabbing" orienting response to changes in stimuli spanning across categories that demanded behaviorally distinct responses, but further experimentation focusing specifically on this question and on

Figure 8. Pretraining and posttraining MMN waves at electrode Fz, averaged across all "bad learners," those at chance performance or below on either trained or novel stimuli in the two hard categories (*n* = 5).
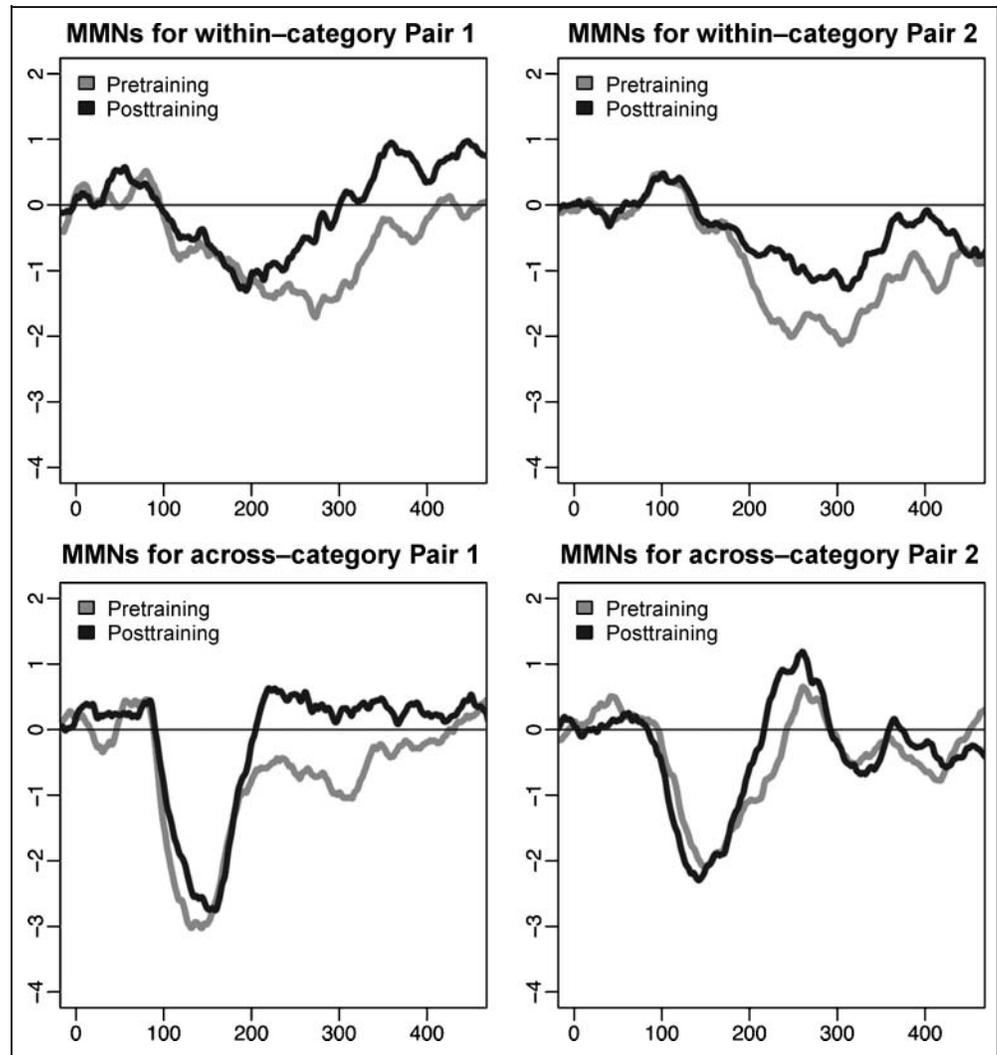
the P3a response as a marker is necessary to draw firm conclusions.

DISCUSSION

Two and a half hours of video game training, spread across five consecutive days, was sufficient for most participants to learn noninvariant auditory categories that were defined in higher dimensional acoustic space and modeled some of the complexity of speech categories. Listeners generalized to novel category exemplars not experienced in training, as evidenced by performance on an explicit behavioral test. These results replicate and extend the behavioral findings of Wade and Holt (2005), demonstrating that playing the game for longer periods translates into better category learning.[1] More generally, these results demonstrate that listeners are able to learn novel auditory categories modeled after speech categories without explicit feedback, suggesting that general mechanisms of auditory category learning may contribute to speech category learning. In the present work, we aimed to assess the degree to which learning mechanisms

engaged by this category learning paradigm and those that support speech category learning elicit common patterns of change in neural response.

It is important to recognize that successful auditory category learning in our experiment is the behavioral outcome that is most analogous to the robust category acquisition that occurs for natural speech. Thus, our conclusions are focused on the training-induced MMN changes that were observed in the subset of participants who were successful category learners in our paradigm (Figure 7). Among this group of learners, we observed a significant increase in peak MMN amplitude for both across-category pairs, suggesting an acquired perceptual distinctiveness between pairs of stimuli trained to belong to different categories (Lawrence, 1950). This finding nicely parallels those from infant MMN data, which suggest that the primary changes associated with natural speech category acquisition are increases in across-category MMN amplitude for learned native-language categories (Cheour et al., 1998). Furthermore, we also observed a significant decrease in within-category MMN amplitude for one of the two stimulus sets. This may reflect an acquired perceptual similarity between

within-category stimuli; both behavioral (Kuhl, 1991; for a review, see Kuhl, 2000) and ERP (Dehaene-Lambertz, 1997; Näätänen, 1997) work have suggested that this is characteristic of naturally acquired speech categories. There are several reasons that may explain why we only observed this acquired similarity for one pair of stimuli as opposed to both. The naive, pretraining MMN to within-category pairs was weak compared with those for across-category pairs, which left less room for becoming weaker as a result of training. In addition, the sound category input distributions were not designed in a manner that emphasized internal category structure (see Kluender, Lotto, Holt, & Bloedel, 1998; Kuhl, 1991), so there was a lack of rich within-category distributional information. This may have promoted mechanisms of category learning that emphasized distinguishing across-category exemplars rather than increasing within-category exemplar similarity (see Guenther et al., 1999). Alternatively, sensitivity to within-category similarities change may develop more slowly. In the speech domain, infants seem to show rapid changes in their neural responses to across-category stimuli but not to within-category stimuli (Cheour et al., 1998), yet adults who have had more experience with their native languages show diminished MMN responses that suggest high level of within-category perceptual similarity.

The MMN changes observed across training appear to result from auditory category learning and not from mere exposure to the sounds during training or ERP recording. The specific directional predictions that were supported by the empirical results rely on the acquisition of category–level relationships among novel sounds. Familiarity with the sounds through mere exposure would predict more uniform changes across all pairs. In addition, although all listeners were exposed to the stimuli across five training sessions, only the participants showing behavioral evidence of category learning and generalization exhibited the predicted MMN changes.

The degree to which the patterns of MMN change underlying acquisition of complex nonspeech categories in the current study parallel those found in studies of speech category learning suggests that similar neural mechanisms support both processes. This conclusion is additionally supported by a recent fMRI study investigating the influence of learning within the Wade and Holt (2005) video game training paradigm on the neural response to passive presentation of the same nonspeech sounds used in the present study. Before training, when participants had not had experience with the sound categories, the nonspeech sounds activated regions of left posterior STS relatively less than they did posttraining when participants had learned the auditory categories (Leech et al., 2009). This is particularly interesting for interpreting the present results because the left posterior STS is typically activated more to speech than to environmental sounds (Cummings et al., 2006). Thus, this putatively speech-specific area may be more generally involved in categorizing complex sounds with which listeners have particular expertise. The present

MMN results indicate that, like speech categories, learning nonspeech auditory categories affects perceptual processing even within the first 300 msec following stimulus presentation. Furthermore, our results suggest that there may be differences in the timing or specific mechanisms by which increased across-category distinction and increased within-category similarity are achieved as part of the category acquisition process.

These results support a growing literature investigating speech category acquisition as a general process rooted in domain general, auditory perceptual, and cognitive mechanisms (Leech et al., 2009; Goudbeek, Cutler, & Smits, 2008; Holt & Lotto, 2006; Wade & Holt, 2005; Holt et al., 2004; Mirman et al., 2004; Lotto, 2000; Guenther et al., 1999). Analogous to findings from the "Greebles" literature that suggest that mechanisms underlying face processing may be based on acquisition of general visual expertise (Gauthier et al., 1999; Gauthier & Tarr, 1997), the current MMN results suggest that part of what makes speech processing seem "unique" compared with nonspeech auditory processing is the acquisition of expertise with speech sounds that comes with linguistic experience and the perceptual changes elicited by such expertise. This is particularly relevant in that most studies comparing speech versus nonspeech neural processing do not control for level of expertise with or degree of category learning of the nonspeech stimuli (e.g., Benson et al., 2001; Vouloumanos, Kiehl, Werker, & Liddle, 2001).

The ability to carefully manipulate listeners' experience with nonspeech categories provides an opportunity to investigate the learning mechanisms available to speech category acquisition in greater depth than is typically possible with speech category learning. The present data demonstrate robust auditory category learning, indexed by both behavioral and neural signatures, with very short-term training. This learning was elicited without an explicit category learning task and without overt feedback, learning characteristics that appear to be important in modeling the natural acquisition of speech categories. The more natural relationship of auditory categories to function within the video game highlights the significance of implicit learning, multisensory association, acquired functional equivalence of stimuli that vary in physical properties, and intrinsic reward in complex auditory category learning. Investigating these processes in greater depth will be essential to understanding auditory category acquisition, in general, and speech category acquisition, in particular. Leveraging the experimental control afforded by nonspeech auditory category learning paradigms makes such direct investigations possible.

Reprint requests should be sent to Ran Liu, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, or via e-mail: ranliu@cmu.edu.

Note

1. The longer training time of the present study (relative to that of Wade & Holt, 2005) likely translated into a greater proportion of successful category learners. However, the present study did not explicitly manipulate overall training duration to determine its potential effect on MMN independently, and it is possible that similar changes in MMN may arise even for shorter training regimens to the extent that they promote successful category learning. Conservatively estimating 4 hours of speech experience a day, a 1-year-old child has heard nearly 1500 hours of speech. If this experience includes just 200 exemplars of a particular speech category each day, the 1-year-old child will have heard over 73,000 category instances. Compared with speech experience, the longer training regimen's 2.5 hours is very brief, and it is notable that such training induces neural changes similar to those observed for speech categories.

REFERENCES

Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language, 78,* 364–396.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1,* 351–353.

Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research, 1115,* 92–107.

DeCasper, A. J., Lecanuet, J.-P., Busnel, M.-C., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior & Development, 17,* 159–164.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8,* 919–924.

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport, 12,* 3155–3158.

Emberson, L. L., Liu, R., & Zevin, J. D. (2009). *Statistics all the way down: How is statistical learning accomplished using novel, complex sound categories?* Proceedings of the Cognitive Science Society 31st Annual Conference, Amsterdam, Netherlands.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1–34.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics, 68,* 161–177.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews, 25,* 355–373.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37,* 1673–1682.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience, 2,* 568–573.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19,* 515–522.

Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication, 50,* 109–125.

Goudbeek, M., Smits, R., Swingley, D., & Cutler, A. (2005). Acquiring auditory and phonetic categories. In H. Cohen & C. Lefebvre (Eds.), *Categorization in cognitive science* (pp. 497–513). Amsterdam: Elsevier.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106,* 2900–2912.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107,* 2711–2724.

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience, 25,* 10494–10501.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119,* 3059–3071.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17,* 42–46.

Holt, L. L., Lotto, A. J., & Diehl, R. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America, 116,* 1763–1773.

Jusczyk, P. W. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. B. (1998). Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models. *Journal of the Acoustical Society of America, 104,* 3568–3582.

Kraus, N., McGee, T., Carrell, T., King, C., & Tremblay, K. (1995). Central auditory system plasticity associated with speech discrimination learning. *Journal of Cognitive Neuroscience, 7,* 27–34.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, U.S.A., 97,* 11850–11857.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 9096–9101.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608.

Lawrence, D. H. (1950). Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology, 40,* 175–188.

Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *Journal of Neuroscience, 29,* 5234–5239.

Lotto, A. J. (2000). Language acquisition as complex category formation. *Phonetica, 57,* 189–196.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics, 68,* 178–183.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing, 16,* 68–90.

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of non-speech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America, 116,* 1198–1207.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development, 16,* 495–500.

Näätänen, R. (1991). Mismatch negativity (MMN) outside strong attentional focus: A commentary on Woldorff et al. *Psychophysiology, 28,* 478–484.

Näätänen, R. (1995). The mismatch negativity—A powerful tool for cognitive neuroscience. *Ear and Hearing, 16,* 6–18.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38,* 1–21.

Näätänen, R., & Alho, K. (1995). Mismatch negativity—A unique measure of sensory processing in audition. *International Journal of Neuroscience, 80,* 317–337.

Näätänen, R., & Alho, K. (1997). Mismatch negativity (MMN)—The measure for central sound representation accuracy. *Audiology and Neuro-Otology, 2,* 341–353.

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42,* 313–329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Houtilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118,* 2544–2590.

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology, 38,* 359–365.

Pulvermuller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology, 79,* 49–71.

Ritter, W., Deacon, D., Gomes, H., Javitt, D. C., & Vaughan, H. G., Jr. (1995). The mismatch negativity of event-related potentials as a probe of transient auditory memory: A review. *Ear and Hearing, 16,* 52–67.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research, 111,* 13–23.

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

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology, 23,* 695–703.

Sharbrough, F., Chatrian, G., Lesser, R., Luders, H., Nuwer, M., & Picton, T. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Electroencephalography and Clinical Neurophysiology, 8,* 200–202.

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

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., et al. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *NeuroReport, 13,* 1813–1816.

Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport, 5,* 844–846.

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.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13,* 994–1005.

Wade, T., & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of the Acoustical Society of America, 118,* 2618–2633.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49–63.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36,* 638–642.