HHS Public Access

Author manuscript

J Exp Psychol Hum Percept Perform. Author manuscript; available in PMC 2015 August 01.

Published in final edited form as:

J Exp Psychol Hum Percept Perform. 2015 August ; 41(4): 1124–1138. doi:10.1037/xhp0000073.

Incidental Auditory Category Learning

Yafit Gabay[1], **Frederic K. Dick**[2], **Jason D. Zevin**[3], and **Lori L. Holt**[1]

[1]Carnegie Mellon University, Department of Psychology and the Center for the Neural Basis of Cognition, USA

[2]Birkbeck College, University of London, Department of Psychological Sciences, UK

[3]University of Southern California, Departments of Psychology and Linguistics, USA

Abstract

Very little is known about how auditory categories are learned incidentally, without instructions to search for category-diagnostic dimensions, overt category decisions, or experimenter-provided feedback. This is an important gap because learning in the natural environment does not arise from explicit feedback and there is evidence that the learning systems engaged by traditional tasks are distinct from those recruited by incidental category learning. We examined incidental auditory category learning with a novel paradigm, the Systematic Multimodal Associations Reaction Time (SMART) task, in which participants rapidly detect and report the appearance of a visual target in one of four possible screen locations. Although the overt task is rapid visual detection, a brief sequence of sounds precedes each visual target. These sounds are drawn from one of four distinct sound categories that predict the location of the upcoming visual target. These many-to-one auditory-to-visuomotor correspondences support incidental auditory category learning. Participants incidentally learn categories of complex acoustic exemplars and generalize this learning to novel exemplars and tasks. Further, learning is facilitated when category exemplar variability is more tightly coupled to the visuomotor associations than when the same stimulus variability is experienced across trials. We relate these findings to phonetic category learning.

Keywords

Auditory category learning; category training; incidental learning; speech; statistical learning

When we recognize red wines as Barbera, mushrooms as edible, and children's cries as joyful, we rely on categorization. Our ability to treat distinct perceptual experiences as functionally equivalent is vital for perception, action, language and thought. There is a rich literature on category learning (Ashby & Maddox, 2005; Cohen & Lefebvre, 2005; Seger & Miller, 2010), with the vast majority of research conducted using visual objects and training paradigms that capitalize on overt category decisions and explicit feedback. Although we have learned much from this traditional approach, the results of such overt category training

Correspondence concerning this article should be addressed to: Prof. Lori L. Holt, Department of Psychology, The Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, Phone (412)268-4964, loriholt@cmu.edu.

tasks with visual objects may not generalize to category learning in all modalities or all natural environments.

Speech highlights this issue. The acoustic complexity of speech presents an auditory category-learning challenge for learners. Complex multidimensional acoustic attributes define speech categories; as many as 16 different acoustic dimensions co-vary with the consonants /b/ and /p/, for example (Lisker, 1986). Further, the significance of various acoustic dimensions is language-community dependent. For instance, among American English listeners, spectral quality is a strong cue to vowel categories, as in *heel* versus *hill* (Hillenbrand, Getty, Clark, & Wheeler, 1995). By contrast, British English listeners from the South of England rely much more on vowel duration than spectral quality to distinguish these categories (Escudero, 2001). Further complicating the demands on the listener, there is also concurrent acoustical variability unrelated to consonant or vowel category identity, which is associated instead with the talker's voice, emotion, and even with room acoustics. The mapping from acoustics to phonemes can be understood as a process of auditory perceptual categorization (see Holt & Lotto, 2010), whereby listeners must learn to discriminate and perceptually-weight linguistically significant acoustic dimensions and to generalize across within-category acoustic variability in speech.

Although perceptual categorization has long been studied in the cognitive sciences (for a review see Cohen & Lefebvre, 2005), the challenges presented by speech signals are somewhat different from those that have motivated most research on categorization. Speech category exemplars are inherently temporal in nature, with the information signaling categories spread across time. Moreover, unlike typical 'stimulus-response-feedback' laboratory tasks, speech category acquisition 'in the wild' occurs under more incidental conditions, without instructions to search for category-diagnostic dimensions, overt category decisions, or experimenter-provided feedback.

Beyond ecological validity, this is an important issue because there is growing evidence that overt and incidental learning paradigms draw upon neural substrates with distinctive computational specialties (e.g. Doya, 1999; Lim, Fiez, Wheeler, & Holt, 2013; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006). Indeed, research across multiple fields has shown that stimulus structure (Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005; Maddox, Ing, & Lauritzen, 2006), feedback (Maddox & David, 2005), and task timing (Ashby, Maddox, & Bohil, 2002; Maddox, Ashby, Ing, & Pickering, 2004) can have a considerable influence on the category learning mechanisms that are recruited (in the auditory domain see Chandrasekaran, Yi, & Maddox, 2014). To fully understand the general principles underlying category learning, it is vital to understand incidental category acquisition.

In the auditory domain, there has been some recent progress in developing approaches to studying incidental learning (Seitz et al., 2010; Vlahou, Protopapas, & Seitz, 2012; Wade & Holt, 2005). Seitz et al. (2010) report that participants' discrimination of sub-threshold nonspeech sounds improves under task-irrelevant perceptual learning paradigms (Seitz & Watanabe, 2009) whereby sub-threshold sounds are presented in a manner that is temporally correlated with other, supra-threshold task-relevant sound stimuli. Even though participants

do not attend to the sub-threshold sounds, these sounds' alignment with task-relevant goals leads participants to learn about them. Quite surprisingly, the magnitude of this incidental learning is comparable to that achieved through explicit training with direct attention to the sounds, overt decisions, and trial-by-trial performance feedback.

Vlahou et al. (2012) have extended this auditory task-irrelevant perceptual learning approach (Seitz & Watanabe, 2009) to a difficult non-native speech contrast. These studies are innovative in that they examine incidental auditory perceptual learning. However, they do not specifically address auditory *category* learning. Instead, in their approach, learning is measured through improved discriminability thresholds for trained sounds, which may lay a sensory foundation from which to build new auditory categories. However, the relationship of learning in this paradigm to category acquisition remains to be determined. Highlighting the difference, task-irrelevant perceptual learning tends to be limited to stimuli experienced in training, whereas generalization to novel exemplars is a hallmark of categorization.

Wade and Holt (2005) provide more direct evidence that implicit task relevance can result in auditory *category* learning. In their task, participants' objective is to earn points by executing actions to shoot and capture aliens that emerge at specific locations within a space-themed videogame. The task is largely visuomotor, but it is structured such that sound can support success in the game. Most significantly, each alien is associated with multiple, acoustically-variable sounds drawn from an artificial nonspeech auditory category. Upon each appearance of an alien, sounds from its corresponding auditory category are played repeatedly. As the game progresses to more challenging levels the pace becomes faster and generalizing across the acoustic variability that characterizes within-category sound exemplars facilitates game performance. Players can *hear* an approaching alien before *seeing* it appear. Thus, if players have learned the sound categories' relationship with the aliens, they can get a head start on executing the appropriate action. Players may capitalize on the predictive relationship between sound category and game action although they receive no explicit instruction about the relationship's existence or utility. Wade and Holt argue that this predictive relationship encourages participants to learn to treat acoustically variable within-category sounds as functionally equivalent, i.e., to categorize the sounds. However, the learning is incidental, in that it involves no instructions to search for category-diagnostic dimensions, no overt category decisions, and no explicit categorization-performance feedback. Learners' goals and attention are not directed to sound categorization. Yet, participants quickly learn the sound categories and generalize to novel exemplars (Leech, Holt, Devlin, & Dick, 2009; Lim et al., 2013; Lim & Holt, 2011; Liu & Holt, 2011; Wade & Holt, 2005).

Successful auditory category learning within this videogame engages putatively speech-selective left posterior superior temporal cortex for processing the newly-acquired nonspeech categories (Leech et al., 2009; Lim et al., 2013) and warps perceptual space in a manner like that observed in speech category acquisition (Liu & Holt, 2011). The learning evoked in this incidental training task is also effective in speech category learning. Adult native-Japanese second-language learners of English significantly improve in categorizing English /r/-/l/ (a notoriously difficult second-language phonetic learning challnage, Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Ingvalson, Holt, & McClelland, 2012;

Ingvalson, McClelland, & Holt, 2011; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994) with just 2.5 hours of incidental training within the videogame (Lim & Holt, 2011).

Studies like these move us closer to understanding the nature of learning under naturalistic task demands, in that they do not involve overt instructions to search for category structure, explicit categorization decisions, or trial-by-trial experimenter-provided feedback. However, many important questions remain regarding the character of incidental auditory category learning and the processes underlying it.

In the present research, we focus on two of these issues. The first is related to the incidental task demands that are theorized to promote learning in the Wade and Holt (2005) videogame. By design, the videogame models a complex array of factors to simulate the functional use of sound categories in a naturalistic environment. Participants actively navigate the videogame environment and encounter rich multimodal associations and predictive relationships between sound categories and game events. They also experience distributional variability in category exemplars, and a strong relationship between sound category learning and videogame success. Any of these factors might contribute to category learning.

Wade and Holt (2005, see also Lim & Holt, 2011; Lim et al., 2013) argue that the consistent temporal correlation of the visual (alien) and motor (response to the alien) dimensions with the auditory categories may serve as the 'representational glue' that binds together acoustically-distinct category exemplars in the incidental training. This is an interesting possibility because it treats co-occurring stimulus and response dimensions as teaching signals for learning. However, the richness of the cues available in the videogame makes it impossible to test this hypothesis directly within the videogame paradigm. In the present studies, we develop and use a simplified incidental training task -- the Systematic Multimodal Associations Reaction Time (SMART) task -- to assess the influence of visuomotor associations in binding acoustically-variable exemplars together in incidental category learning. We hypothesize that these associations support incidental auditory category learning.

The second issue concerns variability. Research in speech category learning has emphasized the importance of experiencing high acoustic-phonetic variability in training. Experience with multiple speakers, phonetic contexts, and exemplars seems to promote non-native speech category learning and generalization among adult learners (Bradlow et al., 1997; Iverson, Hazan, & Bannister, 2005; Jamieson & Morosan, 1989; Wang, Spence, Jongman, & Sereno, 1999). This notion has been highly influential in empirical and theoretical approaches to speech category learning. However, it has arisen from studies of extensive training across multiple training sessions spanning days or weeks that have examined learning via explicit, feedback-driven tasks in which listeners actively search for category-diagnostic information. How variability impacts incidental auditory learning remains an open question.

In the present studies, we address the issue of variability in incidental auditory category learning in a way that differs from prior research. Thus far, studies examining the impact of

variability have typically compared category learning across stimulus sets characterized by high versus low acoustic variability. In the present studies, we hold acoustic variability constant across experiments and manipulate the relationship of within-category exemplar variability to the visuomotor associations that we predict will serve as "glue" that binds exemplars into categories. We predict more robust auditory category learning under conditions whereby within-category acoustic variability is experienced in association with the visuomotor dimensions compared to the same variability experienced across trials.

Across five experiments, we investigate incidental auditory category learning for the same artificial, nonlinguistic auditory categories studied by Wade and Holt (2005). Experiment 1 tests the main hypothesis that auditory categories can be learned incidentally as participants engage in a seemingly unrelated visual detection task. Experiments 2a and 2b examine whether the learning observed in Experiment 1 depends upon the visuomotor associations we hypothesize to be significant in driving learning. Experiment 3 tests the influence of exemplar variability on incidental learning and Experiment 4 doubles the length of incidental training to compare the outcome to the impact of variability on learning.

The SMART Task

The present experiments examine these questions in the context of a novel incidental training task – the Systematic Multimodal Associations Reaction Time (SMART) task. This task builds from the visuomotor associations we hypothesize to be significant in driving learning in the Wade and Holt (2005) videogame task, but strips away the complexity of the videogame. It thus allows direct assessment of the influence of visuomotor associations in binding acoustically-variable exemplars together in incidental category learning.

In the SMART task, participants must rapidly detect the appearance of a visual target in one of four possible screen locations and report its position by pressing a key corresponding to the visual location. The primary task is thus visual detection. However, a brief sequence of sounds precedes each visual target. Unknown to participants, the sounds are drawn from one of four distinct sound categories. This basic version of the paradigm mimics some of the aspects of incidental training paradigms thought to be important in learning (Lim & Holt, 2011). There is a multimodal (auditory category to visual location) correspondence that relates variable sound category exemplars to a consistent visual object, as in the Wade and Holt (2005) videogame. This mapping is many-to-one, such that multiple, acoustically-variable sound category exemplars are associated with a single visual location (akin to the single alien in the videogame). Likewise, sound categories are predictive of the action required to complete the task; in the case of the SMART task, they perfectly predict the location of the upcoming visual detection target and corresponding response button to be pressed. As with the Wade and Holt (2005) task, the SMART task makes it possible to investigate whether participants incidentally learn auditory categories during a largely visuomotor task. However, the SMART task characteristics are straightforward by comparison to the Wade & Holt (2005) first-person interactive videogame, thereby allowing task manipulations to test the factors necessary and sufficient to produce robust incidental auditory category learning and generalization.

We assess category learning with two measures. The first is more covert and implicit, using changes in visual target detection time as a metric. In the first three blocks of the experiment, there is a perfect correlation between the sound categories and the location of the upcoming visual target. In the terms used above, the visuomotor demands of the task provide a strong signal to bind within-sound-category variability. In a fourth test block, scrambling the mapping between location and sound category destroys this relationship. If participants incidentally learn about the sound categories in the first three blocks then we expect visual detection times to be slower in the random (fourth) test block relative to the (third) block that preceded it. We refer to this implicit measure of auditory category learning as the *RT Cost*. It can be observed without overt auditory categorization decisions or responses. Participants are not alerted to the relationship of the sound to the task and the acoustic variability among within-category sound exemplars assures that there is no simple sound-location association.

We also measure category acquisition via an overt sound categorization task that follows the SMART task. In this task, participants hear novel sound exemplars drawn from the sound categories experienced during the SMART task and guess the location where the visual target would be most likely to appear. However, no visual targets appear in this task and there is no feedback about the correctness of responses. This is thus a strong assessment of generalization of incidental category learning to novel, category-consistent stimuli. It also requires that participants apply the newly learned auditory categories in an explicit task that differs from the learning context. This task presents the opportunity to examine correlations of overt category labeling to the more implicit RT Cost measure collected in the SMART task.

Experiment 1

In Experiment 1 and the experiments that follow, we adopt the same artificial nonspeech auditory categories studied by Wade and Holt (2005; see also Emberson, Liu, & Zevin, 2013; Leech et al., 2009; Lim et al., 2013; Liu & Holt, 2011). The purpose of Experiment 1 is to test whether a predictive relationship between sound categories and the visuomotor aspects of the task (location, response) is sufficient to result in learning the complex auditory categories, and to generalize learning to novel exemplars. Our overarching goal for the entire set of experiments is to understand the factors that drive incidental auditory category learning.

Methods

Participants—In this and all experiments, participants were recruited from the Carnegie Mellon University community. They received payment or course credit, had normal or corrected-to-normal vision, and reported normal hearing. Twenty-five participants were tested in Experiment 1.

Stimuli—The artificial, complex nonspeech sound categories of Wade and Holt (2005; see also Emberson et al., 2013; Leech et al., 2009; Liu & Holt, 2011) were used in Experiment 1, and all experiments that follow. Each auditory category experienced in the SMART task was composed of six sound exemplars. Two of the categories were ‘unidimensional’, and

were differentiated by a single, perceptually salient acoustic dimension. The other two categories were 'multidimensional' and were defined such that no single acoustic dimension determined category membership (see Figure 1 for schematized versions of the six exemplars for each category). Across categories, each sound exemplar was 250 ms in duration and was created by combining a sound made of a lower-frequency spectral peak with another sound made of a higher-frequency spectral peak. Sounds drawn from the unidimensional categories shared the same lower-frequency spectral peak, namely a 100 ms-long 600 Hz square wave carrier that linearly transitioned to a 300-Hz offset frequency across the last 150 ms of the stimulus. Similarly, sounds from each multidimensional category had identical lower-frequency spectral peak characteristics; for these sounds, the 143-Hz square wave carrier transitioned linearly from a 300-Hz starting frequency across 150 ms to 600 Hz, where it was steady-state for the remaining 100 ms of the stimulus.

Uni- and multi-dimensional exemplars were differentiated by the dynamics of the higher spectral peak. The unidimensional category sounds' high spectral peak started and remained at a given steady-state frequency for 100 ms, and then transitioned to an offset frequency across 150 ms. By contrast, the multidimensional exemplars' higher peak immediately transitioned across 150 ms from an onset frequency, and then remained at a given steady-state frequency for the following 100 ms. For multidimensional categories, the high spectral peak was derived from a sawtooth wave of periodicity 150 Hz; for unidimensional stimuli, it was derived from bandpass-filtered uniform random noise[1]. Across all categories, the steady-state portion of the high-frequency peak varied across exemplars in center frequency from 950 to 2950 Hz in 400-Hz steps, thereby carrying no first-order information to category membership.

The linear transitions from the high peak steady-state frequencies were determined by the steady-state frequency and a category-specific offset frequency to which the high peak transitioned. But, to prevent listeners from using the onset/offset frequency alone to determine category membership, the high peak transitioned only about 83% of the distance to the (canonical) onset/offset frequency. As a result, the high peak onset/offset frequencies varied somewhat across exemplars within a category.

The unidimensional category offset frequencies were chosen such that the categories (UD1/UD2) were defined by an upward or downward high-peak frequency trajectory, as shown in Figure 1. Since the offset loci were substantially higher (UD1, 3950 Hz) or lower (UD2, 350 Hz) than the steady-state frequencies (varying between 1000 Hz to 3000 Hz, depending on exemplar), each exemplar within a category possessed a falling or rising high-peak offset transition, with somewhat different slopes and offset-frequencies. This created a perceptually salient cue to category membership that listeners are able to use fairly well to group stimuli (Emberson et al., 2013; Wade & Holt, 2005).

[1]White noise sound sources were generated at 22050 Hz and filtered with an eighth-order elliptical bandpass filter with 2-dB peak-to-peak ripple, 50-dB minimum attenuation, and 500-Hz bandwidth using Matlab (Mathworks, Inc.). After filtering, all spectral peaks (square/sawtooth wave and filtered white noise) were equalized for RMS amplitude within and across categories, and 25-ms linear onset and offset amplitude ramps were applied.

Unlike the higher spectral peak transitions present in the unidimensional categories, the onset frequencies for the high peak in the multidimensional categories (MD1/MD2) were chosen so that the direction of the high peak transition provided no first-order acoustic information with which to differentiate the two categories. Here, onset frequencies (2550 Hz for MD1, 1350 Hz for MD2) fell within the range (1000 Hz to 3000 Hz) of potential steady-state frequencies, which were identical over both multidimensional categories. Hence, high spectral peak onset transitions in both MD1 and MD2 varied from steeply increasing in frequency, to flat to slightly decreasing in frequency (see Figure 1). The multidimensional categories thus lacked consistent necessary and sufficient single cues to category membership, a characteristic intended to model the sound categorization challenge presented by the notoriously non-invariant nature of acoustic dimensions to phonetic categories. Nonetheless, consistent with the characteristics of many phonetic categories (Lindblom, 1996; Lindblom, Brownlee, Davis, & Moon, 1992), the multidimensional categories are linearly separable in higher-dimensional acoustic space. Although there is no first-order acoustic cue with which to differentiate these categories, transition slope and steady-state frequency information provide reliable higher-order information.

In addition to the six exemplars defining each of the categories during training, five additional exemplars per category were created and reserved for testing generalization of category learning to novel exemplars. These stimuli had steady-state frequencies intermediate to those of the training stimuli (900 Hz to 2500 Hz, in 400-Hz steps). In other respects their acoustic characteristics matched those of their category, as described above.

Procedure—All testing took place in a sound-attenuated chamber with participants seated directly in front of a computer monitor. Sounds were presented diotically over headphones (Beyer, DT-150).

SMART Visual Detection Task: Participants first performed a visual detection task in the Systematic Multimodal Associations Reaction Time (SMART) paradigm (see Figure 2). Four rectangles organized horizontally across the computer monitor were present throughout the experiment. On each trial, a red X (the visual target) appeared in one of four rectangles. Across trials, assignment of the X to one of the four rectangles was random; unlike traditional serial reaction time tasks (a well-studied incidental learning paradigm; Nissen & Bullemer, 1987), there was no underlying sequence in the appearance of X's across trials. Using the fingers of the dominant hand, participants indicated the position of the X as quickly and accurately as possible by pressing the U, I, O or P key on a standard keyboard; the keys' left-to-right position mapped straightforwardly to the horizontal screen position of the rectangles. Before the appearance of the visual target, participants heard five repetitions of a *single* sound category exemplar (250 ms sounds, 0 ms ISI, 1250 ms total duration followed immediately by the visual target).

Unbeknownst to participants, the sound category from which each exemplar was drawn perfectly predicted of the horizontal position where the visual target would appear (see Figure 2). For a given subject, presentation of five repetitions of a randomly-selected UD1 exemplar might always precede the appearance of the X in the left-most rectangle and thereby be associated with pressing 'U' on the keyboard. (Note that assignment of sound

categories to horizontal position was counterbalanced across participants). Importantly, this was not a simple associative single-sound-to-position mapping. Each sound category was defined by six complex, acoustically-variable exemplars. Associating the visual target position with the preceding sounds required participants to begin to treat the perceptually discriminable sounds defining a category as functionally equivalent in signaling visuospatial position. Note that the task did not require that participants make use of this functional relationship between sound category and visual target location; the task could be completed perfectly based on visual information alone. However, since the sound category predicts the upcoming spatial position of the visual target, visual detection reaction time (RT) can serve as an indirect measure of sound category learning. If participants come to rely on the sound categories to direct responses to the visual targets, then detection responses should be slower (RT Cost) when the relationship is destroyed.

At the beginning of the experiment, participants completed eight practice trials for which there was no correlation between sound category and the position of the visual target. (Practice trials were identical to experimental trials in all other respects). Following practice, there were 3 blocks of trials for which there was a perfect correlation between sound category and visual target location. Each of these blocks had 96 trials (4 sound categories x 6 exemplars x 4 repetitions of each exemplar). After these three blocks, there was a fourth block in which sound category identity was no longer predictive of the position in which the visual target would appear. In this block, assignment of sound to visual position was fully random; any sound exemplar could precede presentation of the visual target in any position. Block 4 was somewhat shorter than the other blocks (48 trials) so that experience with the random mapping would be less likely to erode any category learning achieved across Blocks 1–3. The final, fifth, block restored the relationship between sound category and the location of the upcoming visual target. This served to re-establish category learning prior to the overt categorization task.

Participants were encouraged to rest briefly between blocks. Reaction times (RTs) were measured from the onset of the visual detection target to the press of the response key.

Overt Categorization Task: A ‘surprise’ explicit sound categorization test immediately followed the SMART visual detection task. On each trial, participants heard a sound exemplar presented five times and observed four rectangles arranged horizontally, just as in the SMART task. Using the dominant hand and same keys (U, I, O, P) as used in the SMART task, participants guessed which visual location matched the sound. No visual targets were presented in the overt task and there was no feedback. Therefore participants could not learn about auditory category in the course of the overt task. Sound-category exemplars in the test were the five novel sounds created for each sound category. These sounds were not experienced in the SMART task and thus tested generalization of category learning to novel exemplars, a characteristic element of categorization.

Results

Results for all experiments are shown in Figures 3 and 4; Experiment 1 results are in the top left-hand corner of Figure 3 and the left-most bar of Figure 4.

SMART Visual Detection Task: Trials for which there was a visual detection error (*M*=3%) or response time (RT) longer than 1500 ms or shorter than 100 ms (*M*=2%) were excluded from analyses. A repeated measures analysis of variance (ANOVA) revealed a significant main effect of Block, $F(4,96)=4.25$, $p=.003$, $\eta_p^2=.150$. Central to the hypotheses, there was a significant *RT Cost*, where RT Cost = $RT_{Block4} - RT_{Block3}$, $t(24)=3.69$, $p=.001$. As seen in Figure 3, participants were on average 38 ms faster to detect the visual target in Block 3 (consistent sound category to location mapping) compared to Block 4, when the sound category/location relationship was destroyed. This indicates that participants were sensitive to the relationship between sound category and visual target and suggests that RT Cost can serve as an index of category learning collected online during the incidental SMART training task.

A repeated measures ANOVA revealed no significant differences in RT cost for the two types of categories, $F < 1$.

Overt Categorization Task: As an overt measure of category learning, we used participants' accuracy in explicitly matching novel sound category exemplars with visual locations consistent with the category-location relationship encountered in the SMART task. The sounds tested in the overt categorization task were not heard during the visual detection task and thus generalization -- a hallmark of category learning -- was required for accurate matching. Participants reliably matched the novel sounds to the experienced visual locations at above-chance levels, $t(24)=6.36$, $p<.0001$ (M=49.73, S.E.=3.89). This was true for both unidimensional, $t(24)=5.77$, $p<.0001$, (M=52.62, S.E.=4.79), and multidimensional, $t(24)=5.80$, $p<.0001$, (M=46.84, S.E.=3.76), sound categories.

Relationship Between Implicit and Overt Measures: There was a significant positive relationship between RT Cost and category labeling accuracy in the overt categorization task ($r=0.596$, $p=0.001$). The slower that visual detection RTs were during the random sound-to-location mapping in Block 4 (relative to the average in Block 3), the more accurate labeling was of novel generalization category exemplars. This is evidence that the online measure of category learning collected during incidental learning in the SMART task relates to generalization of category learning assessed with a more traditional overt labeling task.

Experiments 2a and 2b

The results of Experiment 1 are consistent with incidental auditory category learning via the link to the visuomotor aspects of the primary visual detection task. However, it is possible that the learning arose instead from mere exposure to the sound input. Another alternative hypothesis is that participants in Experiment 1 did not learn auditory categories *per se*, but instead learned sound-location associations between individual sound exemplars and their associated visual positions. We address these possibilities in Experiments 2a and 2b.

In Experiment 2a, we test whether incidental category learning generalizes to novel category exemplars within the SMART task. If participants learn sound-location associations, and not auditory categories, then introducing new category exemplars should produce a RT Cost because no sound-location associations will be known for these stimuli. However, if

participants are learning auditory categories, then they may generalize to these new category-consistent exemplars. In this case, we would observe no RT Cost.

In Experiment 2b, we address the concerns above in a different way. Across blocks, participants experience a deterministic mapping between visual location and sound, but the exemplars mapped to a particular visual location are not necessarily drawn from the same sound category. Thus, overall category exemplar exposure is identical to Experiment 1, but within-category distributional acoustic regularity is not associated with the visuomotor mappings inherent in the SMART task. If listeners are learning from mere exposure then we should observe overt category labeling accuracy on par with that in Experiment 1.

Methods

Participants—Twenty-six participants participated in Experiment 2a. Twenty-five participated in Experiment 2b. Participants had the same characteristics as those of Experiment 1.

Stimuli—Stimuli were identical to those of Experiment 1.

Procedure—The procedure was identical to Experiment 1, except as described below.

Experiment 2a: Blocks 1–3 and Block 5 of the SMART task were identical to those of Experiment 1. However, in Block 4 novel, but category-consistent, sounds were presented. This maintained the category-to-location mapping with category exemplars that had not been previously encountered. The generalization stimuli used in the overt categorization test of Experiment 1 served as the novel generalization sounds in Block 4. To the extent that participants learn the auditory categories across Blocks 1–3 and generalize this learning in Block 4, there should be no RT Cost from Block 3 to Block 4. Experiment 2a did not include an overt categorization test because the generalization stimuli used as a test of category generalization in the overt labeling task of Experiment 1 were used instead in Block 4 of the SMART task.

Experiment 2b: In this experiment, six sound exemplars were again deterministically mapped to each visual target location, but unlike Experiment 1 and Experiment 2a, the set of exemplars associated with each location did not come from a single sound category. The exemplar-to-location mapping was maintained across Blocks 1–3, and in Block 5. In Block 4, a new mapping of exemplars to location was introduced. This mapping also did not obey the category structure of the stimuli; sounds from any category could be assigned to any location, so long as it was not the same location experienced across Blocks 1–3 and 5. If listeners learned specific sound-location associations, then disrupting the exemplar-consistent (but not category-specific) associations established in Blocks 1–3 with a new sound-to-location randomization in Block 4 should produce a RT Cost. However, we expect no RT Cost if the learning observed in Experiment 1 was not a simple sound-location association. Such a finding also would rule out exemplar memorization and mere exposure as the drivers of the Experiment 1 findings. Experiment 2b included an overt categorization test identical to that of Experiment 1.

Results

Experiment 2a

SMART Visual Detection Task (see Figure 3, top middle panel): Trials for which there was a visual detection error (*M*=3%) or response time (RT) longer than 1500 ms or shorter than 100 ms (*M*=3%) were excluded from analyses. A repeated measures analysis of variance (ANOVA) revealed no significant main effect of Block, $F(4,100)<1$. A planned t-test showed that introducing novel generalization stimuli in Block 4 was not associated with a significant RT Cost ($RT_{Block4} - RT_{Block3}$), $t(25)=.45$, $p=.66$ (M=2 ms). In other words, although completely new sounds were introduced in Block 4, participants responded just as quickly to the visual target. Thus, any learning that occurred over Blocks 1–3 generalized to novel category exemplars in Block 4. This pattern of generalization is consistent with category learning in Experiment 1, rather than learning item-specific sound-location associations.

Experiment 2b

SMART Visual Detection Task (see Figure 3, top right panel): Trials for which there was a visual detection error (*M*=3%) or response time (RT) longer than 1500 ms or shorter than 100 ms (*M*=1%) were excluded from analyses. A repeated measures analysis of variance (ANOVA) revealed no significant main effect of Block, $F(4, 96)=1.54$, $p=0.19$. There was no significant RT Cost, $t(24)=-0.70$, $p=0.49$. Thus, although there was a consistent sound exemplar to visual location mapping in Blocks 1–3, disruption of this mapping did not affect the speed at which participants detected the visual targets. This is in contrast to the consequences of disrupting the sound category to visual location mapping in Experiment 1. These results suggest that the pattern of responses observed in Experiment 1 was not the result of mere exposure to the sound exemplars, memorization of individual sound-location mappings, or simple sound-location associations.

Overt Categorization Task (see Figure 4, middle bar): Consistent with the lack of an RT cost in the incidental SMART task, participants' accuracy in overtly matching novel sound category exemplars and visual locations was not significantly different from chance for either uni-dimensional ($t(24)=.42$, $p=0.68$) or multi-dimensional ($t(24)=-0.52$, $p=0.61$) categories. Categories composed of arbitrary samplings of exemplars with no coherent distributional structure in perceptual space were not learned, suggesting that structured distributions are an important factor in incidental category learning. We return to this point in the General Discussion.

Relationship Between Implicit and Overt Measures: There was no correlation between RT Cost ($Block4_{RT}$-$Block3_{RT}$) and overt categorization accuracy, $r=.05$, $p=.4$.

Experiment 3

Experiments 2a and 2b confirmed that the results of Experiment 1 were consistent with auditory category learning and did not arise from mere exposure to the stimuli or from learning individual auditory-visual associations. Moreover, Experiment 2b highlighted the importance of category exemplars that sample an orderly distribution in perceptual space in

supporting learning in the incidental task. Whereas participants learned the auditory categories when six acoustically-variable exemplars sampled from a structured distribution in perceptual space were associated with a visual target location (Experiment 1), they did not learn when six exemplars randomly sampled from the entire set of exemplars across the four auditory categories were consistently associated with one of the visual target locations (Experiment 2b). In Experiment 3, we explored this further by examining the impact of category-consistent exemplar variability across the five sounds preceding the visual target. As highlighted in the Introduction, the issue of variability in training is central in studies of speech category learning. However, the influence of variability on incidental auditory learning is unknown.

In Experiment 3, we examine how the learning we observe in Experiment 1 is modulated by acoustic variability. We take a somewhat different approach compared to prior studies. Whereas investigations of the influence of category exemplar variability on auditory category learning have contrasted learning across category exemplars characterized by more or less variability, we hold variability constant across Experiments 1 and 3. This relates to our hypothesis that visuomotor associations support incidental learning.

The learning observed in Experiment 1, as compared to the failure to learn in Experiment 2b, suggests that the visuomotor associations from the primary visual detection task serve as a strong signal to bind together the acoustically variable auditory category exemplars. We hypothesize that experience that more strongly ties acoustic variability to the teaching signal afforded by the visuomotor associations will promote auditory category learning. To test this, we manipulate exemplar variability *within* a trial while holding it constant (and equivalent to Experiment 1) across the experiment. Specifically, in Experiment 1 five repetitions of a single exemplar drawn from a category preceded a visual target on each trial. By contrast, in Experiment 3, five *unique* exemplars drawn from the same category preceded a visual target's appearance in the category-consistent location. Across experiments, the within-category variability experienced by participants was equivalent. However, in Experiment 3 participants experienced within-category variability within a single trial, tightly coupled with the visuomotor associations we hypothesize to promote incidental category learning, whereas in Experiment 1 participants experienced the variability only across trials.

Methods

Participants—Twenty-five participants with the same characteristics as Experiment 1 were tested.

Stimuli—Stimuli were identical to those of Experiment 1.

Procedure—The experiment was conducted like Experiment 1, except for one change. In Experiment 1, a single category exemplar was chosen and presented 5 times preceding the visual target. In Experiment 3, there were also 5 sounds preceding the visual target. However, instead of a single exemplar, 5 unique exemplars were randomly selected (without replacement) from the 6 category exemplars and presented in a random order. In this way, participants experienced the same category input distributions experienced in Experiment 1.

Across the course of the entire experiment, participants' experience with within- and between-category acoustic variability was identical in Experiments 1 and 3. However, participants in Experiment 3 experienced exemplar variability within a single trial, instead of across trials as in Experiment 1.

Results

SMART Visual Detection Task (see Figure 3, bottom left panel): Trials for which there was a visual detection error (M=4.5%) or response time (RT) longer than 1500 ms or shorter than 100 ms (M=7.5%) were excluded from analyses. A repeated measures analysis of variance (ANOVA) revealed a significant main effect of Block ($F(4,96)=25.61$, $p=0.0001$, $\eta_p^2 = .516$). Most relevant to the hypotheses, there was a large and significant RT Cost ($t(24)=7.78$, $p=0.001$), with participants responding an average of 77 ms slower in Block 4 than Block 3. A repeated measures ANOVA revealed no significant differences in RT cost for the two types of categories, $F < 1$.

Overt Categorization Task (see Figure 4, second bar from right): There was also strong evidence of category learning in the overt post-training categorization task. Participants labeled novel generalization stimuli at above-chance levels, $t(24)=11.92$, $p<0.0001$ (M=65.8%, S.E.=3.42). This was true for both uni-dimensional ($t(24)=11.56$, $p<0.0001$ (M=77.5%, S.E.=4.5)), and multi-dimensional ($t(24)=8.9$, $p<0.0001$ (M=54%, S.E.=3.26)) categories.

Relationship Between Implicit and Overt Measures: There was a significant positive relationship between participants' overt categorization task accuracy and the RT cost elicited from disrupting the category-location mapping in Block 4, $r=0.85$, $p <0.0001$.

Comparison of Category Learning to Experiment 1 Category Learning: Experiments 1 and 3 differed in whether participants experienced within-category exemplar variability within a trial (across the 5 sounds preceding a visual target, Experiment 3) or across trials (5 sounds preceding a visual target were identical, Experiment 1). This factor influenced category learning considerably, as observed in both category learning measures. The RT Cost observed in Experiment 3 (M=77 ms, SE=9.97) was significantly greater than that observed in Experiment 1 (M=38 ms, SE=10.3), $t(48)=2.779$, $p=.008$. In addition, participants in Experiment 3 exhibited greater category learning as indicated by accuracy in the overt labeling task (M=65.75, SE=3.42) than participants in Experiment 1 (M=49.73, SE=3.89), $t(48)=3.09$, $p=.003$. We also examined learning across the three first blocks in Experiment 1 compared with Experiment 3. A repeated measures analysis of variance (ANOVA) revealed a significant interaction between block (Blocks 1–3) and experiment (Exp 1 vs. 3), $F(2, 96) = 4.19$, $p = .017$. Further analysis revealed a significant linear trend in decreased RT across blocks for Experiment 3, $F(1, 48) = 23.16$, $p <.001$, but not for Experiment 1, $F(1, 48) = 1.62$, $p =.207$. Learning was more robust when within-category exemplar variability was linked to the visuomotor associations that support incidental auditory category learning.

Experiment 4

The difference in category learning outcomes between Experiments 1 and 3 suggests that task demands encouraging a link from within-category acoustic variability to a consistent signal (like one of the visual locations in the SMART task) facilitate incidental category learning. In Experiment 4, we sought to establish a benchmark against which to compare the degree of this facilitation. Experiment 4 was identical to Experiment 1 in nearly all respects but that we doubled the number of blocks across which participants experienced a consistent mapping between auditory category and visual location.

Methods

Participants—Twenty-five participants with the same characteristics as Experiment 1 were tested.

Stimuli—Stimuli were identical to those of Experiment 1.

Procedure—The experiment was conducted like Experiment 1, except that participants completed 10 blocks instead of 5 blocks in the SMART task. Randomized blocks whereby the relationship between auditory category and visual target location was destroyed were presented at Block 4 and Block 9. This allowed us to compute two RT Cost measures at two points across training.

Results

SMART Visual Detection Task (see Figure 3, middle bottom panel): Trials for which there was a visual detection error (*M*=4%) or response time (RT) longer than 1500 ms or shorter than 100 ms (*M*=3%) were excluded from analyses. A repeated measures analysis of variance (ANOVA) revealed a significant main effect of Block, F(9,216)=5.07, *p=0*.0001, η_p^2 = 0.175. As shown in Figure 3, there was a significant RT Cost (*t*(24)=3.45, *p*<0.0001), with participants responding an average of 36 ms slower in Block 4 than Block 3. There was also a significant RT Cost, Cost *t*(24)=3.47, *p*=.002, later in training with participants averaging 31 ms slower visual detections in Block 9 than Block 8. A repeated measures ANOVA revealed no significant differences between the two types of categories in the first RT cost (Blocks 3 vs. 4), F (1, 24) = 1.38, p = .252, or in the second RT cost (Blocks 8 vs. 9), F<1.

Overt Categorization Task: There was also evidence of category learning in the overt post-training categorization task. Participants labeled novel generalization stimuli at above-chance levels, (*t*(24)=5.94, *p*<0.0001, M=54.8%, S.E.=5.01). This was the case for both uni-dimensional (*t*(24)=5.22, *p*<0.0001, M=56.82%, S.E.=6.1), and multi-dimensional (*t*(24)=5.76, *p*<0.0001, M=52.7%, S.E.=4.81) categories.

Relationship Between Implicit and Overt Measures: There was no significant relationship between participants' overt categorization task accuracy and the RT cost elicited from disrupting the category-location mapping in Block 4, *r*=.237, *p*=.127 or in Block 9, *r*=.145, *p*=.245.

Comparison of Category Learning in Experiment 4 vs. Experiment 3: In doubling the training trials in the incidental SMART task, Experiment 4 provided a benchmark against which to compare the benefit of variability in Experiment 3. Category learning as assessed via the implicit measure of RT Cost was significantly greater in Experiment 3 than Experiment 4. Randomizing the relationship between sound category exemplars and visual detection targets slowed Experiment 3 participants' visual detection significantly more (M=77 ms, SE=9.66) than Experiment 4 participants, as measured at both the first (M=35.63, SE=10.33, $t(48)=2.95$, $p=.005$) and the second (M=31.45, SE=9.07, $t(48)=3.46$, $p=.001$) block of randomization in Experiment 4. In the overt labeling task, there was no significant difference ($t(48)=1.81$, $p=.076$) across experiments. However, the trend was for greater accuracy in overt labeling of generalization exemplars after training with within-trial variability in Experiment 3 (M=65.75, SE=3.42) compared to training with double the training trials in Experiment 4 (M=54.77, SE=5.01). In all, these comparisons underscore the advantageousness of experiencing within-category exemplar variability in the context of the visuomotor aspects of the primary task hypothesized to support category learning. Indeed, it is notable that experiencing category exemplar variability in this way resulted in as much, or better, learning as doubling the training.

General Discussion

Categorization - the ability to treat distinct perceptual experiences as equivalent - is central to cognition. Accordingly, a rich tradition of research has addressed how humans categorize the perceptual world. Most of what we know about perceptual category learning comes from studies of participants who are actively searching for diagnostic cues in the context of stimulus-response-feedback tasks. However, much less is known about how categories are learned incidentally – that is, without instructions to search for category-diagnostic dimensions, overt category decisions, or experimenter-provided feedback. Although incidental learning has not been a central focus of research, there is evidence that the learning systems engaged by traditional tasks may be distinct from those recruited by incidental category learning (Lim et al., 2013; Tricomi et al., 2006). Since much of the learning we do about categories in the natural auditory world is likely to be incidental rather than driven by explicit, feedback-directed learning, it is important to begin to understand how listeners incidentally acquire perceptual categories.

In the present research, we examined factors driving incidental category learning by studying how participants incidentally learn nonspeech auditory categories. To this end, we developed a novel experimental paradigm – the SMART task – in which participants experienced auditory categories incidentally in the course of participating in a visual detection task. Unbeknownst to participants, auditory category membership predicted the upcoming location of the visual detection target. As a result, the degree to which visual target detection slowed when the tight coupling of auditory category and visual target location was destroyed served as an implicit assessment of sound category learning in the incidental training task. After incidental training, we also assessed auditory category learning using a more traditional overt labeling task to test the generalization of incidental category learning across task and novel sound exemplars.

We focused on two significant issues in incidental category learning. The first was the influence of visuomotor associations in binding acoustically-variable exemplars together in incidental category learning. We hypothesized that these associations would support incidental auditory category learning. Specifically, we expected that the consistent correlation of visual location and the appropriate motor response to indicate the location of the visual target would support auditory category learning as participants performed the visual detection task. This approach treats co-occurring stimulus and response dimensions as teaching signals for learning.

Indeed, we find strong evidence for incidental category learning across experiments. Experiment 1 established that participants suffer a reaction time cost to visual detection responses when the relationship between auditory category and visual target location is destroyed by random assignment in Block 4 of the SMART task. Further, the magnitude of this reaction time cost (in all but Experiment 4) was positively correlated with participants' accuracy in the overt labeling task that followed: the greater the indication of incidental category learning via the reaction time cost, the greater participants' categorization accuracy for novel sound exemplars in the subsequent overt labeling task.

Experiments 2a and 2b corroborate the conclusion that Experiment 1 resulted in incidental category learning via the visuomotor coupling, and not via learning from mere exposure or simple auditory-visual associations. When novel category-consistent auditory exemplars were introduced in Block 4 of Experiment 2a, participants experienced no reaction time cost. The fact that visual detection response times were unaffected suggests that participants were already generalizing category learning to novel sound exemplars in Block 4, consistent with categorization.

The learning also was not a result of simple association of sounds to visual locations. When sounds were arbitrarily assigned to visual location without respect to auditory category membership yet consistently paired with the appearance of a visual target in a particular location in Experiment 2b, participants failed to demonstrate learning in either the incidental or overt tests. Thus, the distributional structure and similarity of within-category exemplars appears to have participated in promoting incidental category learning. This is consistent with the results of Wade and Holt (2005), who found that participants who experienced categories without distributional structure also failed to exhibit above-chance labeling following videogame training.

In contrast to Experiments 2a and 2b, the results of Experiments 1, 3, and 4 showed incidental auditory category learning and robust generalization of this category learning to both novel stimuli and also to an overt labeling task. Participants' attention was not directed to the sounds, they were not informed that the sounds formed categories, they did not actively search for category-diagnostic dimensions and make decisions based on them, and they did not receive overt feedback about category decisions. We find consistent evidence that a pairing of visual detection task elements with the auditory categories can serve as the 'representational glue' that binds together acoustically distinct sound exemplars in incidental training, so long as those exemplars have an underlying distributional structure. It will be of

interest to further probe the boundaries and constraints on the kinds of distributional structure that are readily learnable in incidental auditory category learning in future research.

The second central issue of the present work was the role of stimulus variability in learning. Prior research in speech category learning has emphasized the importance of trained sound variability in promoting category acquisition (e.g., Bradlow et al., 1997; Iverson et al., 2005; Jamieson & Morosan, 1989; Lively, Logan, & Pisoni, 1993; Wang et al., 1999). Although several studies have examined incidental learning in the auditory domain (Seitz et al., 2010; Vlahou et al., 2012; Wade & Holt, 2005), the issue of how variability impacts incidental category learning has not been explored. Drawing off of the hypothesis that consistent pairing of the visual detection task elements with the auditory categories could serve to bind together acoustically-variable within-category exemplars, we took an approach somewhat different from previous studies. With learning in Experiment 1 as a baseline, Experiment 3 was constructed to have equivalent category exemplar variability across the experiment. However, whereas within-category variability was experienced *across* trials (and therefore across visuomotor associations) in Experiment 1, it was experienced *within* trials in Experiment 3. Therefore, the coupling of within-category acoustic variability with the binding signals from the primary visual detection task was more robust in Experiment 3. We predicted that learning would be facilitated by within-trial exemplar variability, with stronger learning in Experiment 3 than Experiment 1, even though overall variability was held constant across experiments.

Indeed, our nontraditional manipulation of category variability had a strong effect. The RT Cost observed in the Experiment 3 SMART task was nearly double that observed in Experiment 1; destruction of the relationship between the sound categories and visual locations had a much more damaging effect on participants' visual detection response speed in Experiment 3 than Experiment 1. This suggests that participants in Experiment 3 were more strongly reliant upon the auditory categories to guide visual detection. The larger reduction in RTs across the first three blocks in Experiment 3 compared with Experiment 1 is another indication for better learning for within-trial variability. The results of the overt labeling task suggest that this reliance was due to more robust category learning. Participants in Experiment 3 exhibited significantly greater accuracy in generalization to novel sound category exemplars in the overt labeling task, compared to Experiment 1 participants. In all, these results demonstrate that variability impacts incidental category learning. It seems that people learn more quickly when within-category exemplar variability is experienced within each trial. Studies of supervised learning suggest that training interleaved across categories may promote learning compared with blocked training (Shea & Morgan, 1979). However, under unsupervised learning conditions the advantage of interleaved over blocked training depends in category similarity (Clapper, 2014). The present data extends this research further by demonstrating the advantage of within-trial variability over variability across trials for promoting incidental auditory category learning.

More than this, these results move forward our thinking about the impact variability in training. The issue here is not one of experiencing more versus less variability in training; variability was equated across Experiments 1 and 3. Rather, the significant factor appears to be how variability relates to the associations supporting learning. By this view, variability

that is experienced in a manner that is more tightly coupled to the binding signals that drive learning is expected to promote category learning. We would predict this to be true of learning via feedback in more traditional tasks as well (i.e., when explicit feedback is the binding signal). This remains a question open for future research.

The careful reader may have noted that we have been attentive to describing the present learning and that observed by Wade and Holt (2005) as *incidental*. We do so to emphasize that the sound categories are learned by virtue of their relationship to success in performing a task defined along other dimensions. Although participants are not overtly searching for dimensions diagnostic to category membership and do not receive overt feedback about categorization performance, it is important to highlight that this learning is neither passive, nor entirely unsupervised or feedback-free. Specifically, in the case of the SMART task, supportive cues (the visual referent and associated motor response) linked to the overt task were correlated with auditory category membership. This buttressed learning beyond what has been observed for these same stimuli under passive, unsupervised learning conditions (Wade and Holt, 2005; Emberson et al., 2013). Nevertheless, there is evidence that quite complex perceptual categories can be acquired through unsupervised learning (for examples in the visual domain see Clapper, 2012; Love, 2002). It will be important to unravel the relative influence of stimulus input distributions, categorization training task, the influence of an active task, and the presence of different types of feedback in future work. This is especially important in light of the fact the incidental approach to category learning described here (and in Wade and Holt, 2005) differs from both passive exposure paradigms and learning via explicit experimenter-provided feedback, the two approaches that have been most influential in understanding auditory category learning relevant to speech categorization.

In the domain of speech, learning via passive exposure has been an influential theoretical perspective (Redington, Chater, & Finch, 1998; Saffran, 2001). By this view, the emphasis is on category learning via passive accumulation of distributional regularities (Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996). The present results are in accord with this perspective with respect to the significance of distributional regularities in the input in supporting category learning. Experiment 2b, in particular, emphasizes that the distributional structure of exemplars in input has an important influence on the learnability of perceptual categories. However, the present results bring up an important point for consideration with respect to passive, statistical learning accounts of auditory (and phonetic) category acquisition. Experiments 2a and 2b demonstrate that exposure alone was not sufficient to elicit category learning. In fact, prior research using these same auditory categories supports the conclusion that passive exposure is not always sufficient for category acquisition. The sounds used here are not acquired in unsupervised sorting tasks with no feedback (Wade & Holt, 2005) or across passive exposure to streams of category exemplars in statistical segmentation tasks like those pioneered by Saffran et al. (1996) (Emberson et al., 2013).

Some have voiced concerns about the extent to which passive, distributional statistical learning could scale up to learning categories in more cluttered natural environments, where there is an explosion of potentially relevant distributional regularities, only some of which

are significant (Pierrehumbert, 2003). The present results suggest that understanding incidental category learning may require broader consideration of the regularities available to learners in natural environments. In the present task, categories that fail to be acquired with passive exposure are learned quite quickly and incidentally in the context of correlations with the visuomotor demands of the primary visual detection task. The present paradigm is simplistic compared to the supportive multimodal correlations potentially available in the natural perceptual world. But, the results suggest that the presence of co-occurring visual referents may support category learning in the context of auditory category learning in complex environments by signaling the distinctiveness of acoustically-similar items across references or the similarity of acoustically-distinct exemplars paired with the same referent.

Though the visual “referent” is a very simple difference in visual target location, it seems to have served to signal a common relationship among category exemplars. A recent study using the videogame paradigm of Wade and Holt (2005) provides support for this possibility (Lim, Lacerda & Holt, in press). Participants played the videogame with sound categories linked to the appearance of specific alien creatures. However, instead of encountering isolated category exemplars upon the appearance of an alien, participants heard acoustically-variable category exemplars embedded in highly variable continuous sound streams. Although variable, the category exemplars were the best predictors of specific aliens and the appropriate game action in the sea of highly variable, continuous sound. Participants learned the categories and generalized learning to novel exemplars without knowledge that there were significant units embedded in the continuous sound, information about the category exemplars’ temporal extent, or awareness of the temporal position of the exemplars within the stream. By contrast, naïve participants failed in unsupervised sorting of these same continuous sounds into categories after passive exposure. Lim et al. hypothesize that the consistent appearance of a unique alien creature, a visual referent, supported learners in acquiring the auditory categories in this complex environment. What this suggests is that objects and events in the world consistently paired with, or predictive of, categories can support category acquisition in complex environments. Since there was no temporal synchronization of the visual referent with the acoustically-variable category exemplars embedded in highly variable continuous sound streams, Lim et al. suggest that the alien may provide a visual referent akin to that provided by the coincidence of words and objects in the world. Imagine being a non-English listener hearing “I found my *keys*! The *keys* were under my book all along. I thought I had lost the *keys* for good!” from a talker holding a set of keys. There is high acoustic variability throughout the utterance, including across the individual instances of *keys*. Nevertheless, the visual referent may serve as a correlated signal that supports discovery of the commonalities across the acoustically-variable instances of *keys* peppering the continuous acoustic stream. The Lim et al. data present the possibility that visual referents may support auditory category learning by signaling the distinctiveness of acoustically similar items across referents and/or the similarity of acoustically distinct items paired with the same referent even for highly variable, continuous sensory input that mimics the complexity of real world learning situations. Incidental learning conditions whereby supportive multi-modality information correlates with category membership may boost learning above and beyond passive exposure. This perspective is in

line with theories positing that in the natural environment infant word learners cut through noisy co-occurrence statistics between words and referents by relying on the convergence of multiple, statistically-sensitive processes (see Smith, Suanda, & Yu, 2014). Challenging learning domains, like the categories of the present studies that are not learned under passive exposure conditions, may be supported by seemingly task-irrelevant "noise" that nonetheless possesses regularity.

This is an important theoretical issue for speech categorization, where it is clear that learning in the natural environment does not arise from explicit feedback of the sort typical of laboratory training tasks. The present results caution that it need not be necessary to posit entirely passive "statistical" distributional learning to accommodate this fact. Statistical input distributions do matter for learning. Nevertheless, although the visuomotor task characteristics in the SMART task are very simple, they support incidental category learning beyond what could be evoked by passive exposure to the sounds. The natural learning environment could be expected to provide even richer supportive regularities and opportunities for learning.

This brings up another theoretical issue of relevance to the present results. Although participants do not engage in explicit categorization and there is no feedback in the traditional sense of the experimenter providing "correct" versus "incorrect" feedback, it would be an error to suggest that there is no feedback in incidental tasks like SMART, or the Wade and Holt (2005) videogame. In the videogame, feedback quite clearly arrives in the success or failure of shooting actions. To the extent that sound categories predict appropriate actions, the outcomes of behavior provide an internal feedback signal that may be influential in driving category learning. In a recent review, Lim, Fiez, and Holt (2014) make the case that such learning signals may be powerful in hastening the system's sensitivity to distributional regularities that would be more slowly acquired through the Hebbian learning principles associated with learning through passive exposure. Indeed, in a recent neuroimaging study of participants as they played the Wade and Holt (2005) videogame, Lim et al. (2013) find evidence for posterior striatal involvement in incidental category learning consistent with this possibility.

Considering this in the context of the present paradigm, it is important to note that participants were nearly uniformly successful in the simple visual detection task. Nonetheless, there was a relationship between sound category and location such that successful sound categorization could facilitate successfully detecting and quickly responding to the visual target. Therefore, predictions about target location made based on sound category are followed by "feedback" about the accuracy of the prediction via the actual appearance of the visual object at a specific location. The present studies do not differentiate the extent to which prediction and auditory-visual association drive category learning, but the SMART paradigm is amenable for discovering this in future research.

This aspect of the task bears some resemblance to the task-irrelevant perceptual learning tasks that produce incidental auditory learning, as reviewed in the Introduction. In the task-irrelevant perceptual learning paradigm, learning may take place for stimulus features, whether or not they are relevant to the task, so long as they are systematically paired with

successfully processed task targets, or rewards, within a critical time window (Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006; Seitz, Nanez, Holloway, Koyama, & Watanabe, 2005; Seitz & Watanabe, 2009; Watanabe, Náñez, & Sasaki, 2001). Applying this approach in the auditory domain, Seitz et al. (2010) demonstrated that when nonspeech sounds modeling aspects of speech formant transitions were paired with targets in an unrelated behavioral task, the discrimination thresholds for detecting whether the formant transition changed in frequency across time decreased. Neither attention nor even awareness of the subthreshold sounds was necessary to evoke this learning. These studies provide evidence of auditory learning from training that is neither passive, nor requiring overt attention nor response to the learned stimulus dimensions. Interestingly, these results also suggest that explicit feedback can sometimes be counterproductive to learning (Vlahou et al., 2012). However, as noted, this incidental learning is not necessarily *category* learning.

Nonetheless, the task-irrelevant learning paradigm task bears some resemblance to the SMART task in that participants' attention is directed away from the learning domain (auditory categories in the present case) and toward another task (here, visual detection). Moreover, learning appears to be closely related to the coordinated timing of task-relevant events and stimuli in the to-be-learned domain. Seitz and Watanabe (2009) argue that task-irrelevant perceptual learning occurs due to diffuse reinforcement signals driven by the primary task and signals driven by the presentation of the task-irrelevant stimuli. To the extent that task-relevant and task-irrelevant stimulus features temporally coincide, then task-irrelevant learning occurs. Both task-irrelevant perceptual learning and the incidental category learning observed in the present studies challenge the prevailing notion that directed attention to the to-be-learned input is a prerequisite for learning.

An interesting aspect of the present data is the generalization of incidental learning to the overt labeling task and to novel category exemplars. Although participants acquired the artificial auditory categories incidentally, they appear to have been able to immediately apply this knowledge to a new task requiring conscious decision making about novel sounds. Moreover, assessments of learning in the incidental and overt tasks were well-correlated. This is an important aspect of the learning we observe. In the present experiments, the superficial relationship of visual location was maintained across the incidental and overt tasks. Future research will need to determine whether this is important. Nonetheless, the degree of across-task generalization we observe is notable. Many "gamified" tasks that attempt to train individuals incidentally have been criticized for "training to the test" with quite poor generalization of new representations to untrained tasks. Ultimately, if approaches to incidental category learning are to have real-world impact such as in training adults to better categorize second language phonetic categories, then generalization of learning to new tasks is essential. It will be informative for future research to establish the extent to which incidental speech category training, for example, generalizes to benefit other language-learning tasks. On a practical level, the present incidental approach to training auditory categories is simple, fast and effective. Moreover, the relationship of visuomotor task elements to auditory categories appears to be the significant factor in driving learning and so the SMART task is quite amenable to embedding in other, perhaps

more engaging, primary tasks (as, for example, the Wade and Holt, 2005 videogame). Our finding that learning generalizes to overt labeling is a necessary first step in this regard.

Incidental category learning draws upon different learning systems than traditional overt categorization training task (Lim et al., 2013; Tricomi et al., 2006). In light of the fact that incidental learning is likely to be typical of category acquisition in the natural world, it is important to begin to understand its basis. To this end, we introduced a novel paradigm for studying the learning mechanisms involved in incidental category learning. With it, we discovered that many-to-one auditory-to-visuomotor correspondences are powerful in supporting incidental auditory category learning. These correspondences serve as a 'representational glue' that binds together acoustically distinct sound exemplars in incidental training, so long as the exemplars have an underlying distributional structure. Moreover, incidental category learning is facilitated when category exemplar variability is more tightly coupled to these visuomotor correspondences than when the same exemplar variability is experienced across trials. These results advance our understanding of incidental auditory category learning and inform the incidental learning mechanisms available to phonetic category learning and category acquisition across modalities.

Acknowledgments

The work was supported by a grant from the National Institutes of Health to LLH, R01DC004674. The authors thank Christi Gomez for her assistance with the experiments.

References

Ashby FG, Maddox WT. Human category learning. Annu Rev Psychol. 2005; 56:149–178. [PubMed: 15709932]

Ashby FG, Maddox WT, Bohil CJ. Observational versus feedback training in rule-based and information-integration category learning. Memory & cognition. 2002; 30(5):666–677. [PubMed: 12219884]

Bradlow AR, Pisoni DB, Akahane-Yamada R, Tohkura Yi. Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. The Journal of the Acoustical Society of America. 1997; 101(4):2299. [PubMed: 9104031]

Chandrasekaran B, Yi HG, Maddox WT. Dual-learning systems during speech category learning. Psychonomic bulletin & review. 2014; 21(2):488–495. [PubMed: 24002965]

Clapper JP. The effects of prior knowledge on incidental category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012; 38(6):1558.

Clapper JP. The impact of training sequence and between-category similarity on unsupervised induction. The Quarterly Journal of Experimental Psychology. 2014:1–21. ahead-of-print.

Cohen, H.; Lefebvre, C. Handbook of categorization in cognitive science. Elsevier; 2005.

Doya K. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural networks. 1999; 12(7):961–974. [PubMed: 12662639]

Emberson LL, Liu R, Zevin JD. Is statistical learning constrained by lower level perceptual organization? Cognition. 2013; 128(1):82–102. [PubMed: 23618755]

Escudero, P. The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. Paper presented at the Proceedings of the 25th annual Boston University conference on language development; 2001.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. The Journal of the acoustical society of America. 1995; 97(5):3099–3111. [PubMed: 7759650]

Holt LL, Lotto AJ. Speech perception as categorization. Attention, Perception, & Psychophysics. 2010; 72(5):1218–1227.

Ingvalson EM, Holt LL, McCLELLAND JL. Can native Japanese listeners learn to differentiate/r–l/on the basis of F3 onset frequency? Bilingualism: Language and Cognition. 2012; 15(02):255–274.

Ingvalson EM, McClelland JL, Holt LL. Predicting native English-like performance by native Japanese speakers. Journal of phonetics. 2011; 39(4):571–584. [PubMed: 22021941]

Iverson P, Hazan V, Bannister K. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. The Journal of the Acoustical Society of America. 2005; 118:3267. [PubMed: 16334698]

Jamieson DG, Morosan DE. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. Canadian Journal of Psychology/Revue canadienne de psychologie. 1989; 43(1):88.

Leech R, Holt LL, Devlin JT, Dick F. Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. The Journal of Neuroscience. 2009; 29(16):5234–5239. [PubMed: 19386919]

Lim, S-J.; Fiez, JA.; Wheeler, ME.; Holt, LL. Investigating the Neural Basis of Video-game-based Category Learning. Paper presented at the Journal of Cognitive Neuroscience; 2013.

Lim S-J, Fiez JA, Holt LL. How may the basal ganglia contribute to auditory categorization and speech perception? Auditory Cognitive Neuroscience. 2014; 8:230.

Lim S-J, Holt LL. Learning Foreign Sounds in an Alien World: Videogame Training Improves Non - Native Speech Categorization. Cognitive science. 2011; 35(7):1390–1405. [PubMed: 21827533]

Lim S-J, Lacerda F, Holt LL. Discovering functional units in continuous speech. Journal of Experimental Psychology: Human Perception & Performance. in press.

Lindblom B. Role of articulation in speech perception: Clues from production. The Journal of the acoustical society of America. 1996; 99(3):1683–1692. [PubMed: 8819859]

Lindblom B, Brownlee S, Davis B, Moon SJ. Speech transforms. Speech communication. 1992; 11(4): 357–368.

Lisker L. “Voicing” in English: a catalogue of acoustic features signaling/b/versus/p/in trochees. Language and speech. 1986; 29(1):3–11. [PubMed: 3657346]

Liu R, Holt LL. Neural changes associated with nonspeech auditory category learning parallel those of speech category acquisition. Journal of Cognitive Neuroscience. 2011; 23(3):683–698. [PubMed: 19929331]

Lively SE, Logan JS, Pisoni DB. Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. The Journal of the acoustical society of America. 1993; 94(3):1242–1255. [PubMed: 8408964]

Lively SE, Pisoni DB, Yamada RA, Tohkura Yi, Yamada T. Training Japanese listeners to identify English/r/and/l/. III. Long - term retention of new phonetic categories. The Journal of the acoustical society of America. 1994; 96(4):2076–2087. [PubMed: 7963022]

Love BC. Comparing supervised and unsupervised category learning. Psychonomic bulletin & review. 2002; 9(4):829–835. [PubMed: 12613690]

Maddox WT, Ashby FG, Ing AD, Pickering AD. Disrupting feedback processing interferes with rule-based but not information-integration category learning. Memory & cognition. 2004; 32(4):582–591. [PubMed: 15478752]

Maddox WT, David A. Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31(1):100.

Maddox WT, Filoteo JV, Lauritzen JS, Connally E, Hejl KD. Discontinuous categories affect information-integration but not rule-based category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31(4):654.

Maddox WT, Ing AD, Lauritzen JS. Stimulus modality interacts with category structure in perceptual category learning. Perception & psychophysics. 2006; 68(7):1176–1190. [PubMed: 17355041]

Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition. 2002; 82(3):B101–B111. [PubMed: 11747867]

Nissen MJ, Bullemer P. Attentional requirements of learning: Evidence from performance measures. Cognitive psychology. 1987; 19(1):1–32.

Pierrehumbert JB. Phonetic diversity, statistical learning, and acquisition of phonology. Language and speech. 2003; 46(2–3):115–154. [PubMed: 14748442]

Redington M, Chater N, Finch S. Distributional information: A powerful cue for acquiring syntactic categories. Cognitive science. 1998; 22(4):425–469.

Saffran JR. The use of predictive dependencies in language learning. Journal of Memory and Language. 2001; 44(4):493–515.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. 1996

Seger CA, Miller EK. Category learning in the brain. Annual review of neuroscience. 2010; 33:203.

Seitz AR, Nanez JE, Holloway S, Tsushima Y, Watanabe T. Two cases requiring external reinforcement in perceptual learning. Journal of vision. 2006; 6(9):9.

Seitz AR, Nanez JE, Holloway SR, Koyama S, Watanabe T. Seeing what is not there shows the costs of perceptual learning. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(25):9080–9085. [PubMed: 15956204]

Seitz AR, Protopapas A, Tsushima Y, Vlahou EL, Gori S, Grossberg S, Watanabe T. Unattended exposure to components of speech sounds yields same benefits as explicit auditory training. Cognition. 2010; 115(3):435–443. [PubMed: 20346448]

Seitz AR, Watanabe T. The phenomenon of task-irrelevant perceptual learning. Vision research. 2009; 49(21):2604–2610. [PubMed: 19665471]

Shea JB, Morgan RL. Contextual interference effects on the acquisition, retention, and transfer of a motor skill. Journal of Experimental Psychology: Human Learning and Memory. 1979; 5(2):179.

Smith LB, Suanda SH, Yu C. The unrealized promise of infant statistical word–referent learning. Trends in cognitive sciences. 2014; 18(5):251–258. [PubMed: 24637154]

Tricomi E, Delgado MR, McCandliss BD, McClelland JL, Fiez JA. Performance feedback drives caudate activation in a phonological learning task. Journal of cognitive neuroscience. 2006; 18(6): 1029–1043. [PubMed: 16839308]

Vlahou EL, Protopapas A, Seitz AR. Implicit Training of Nonnative Speech Stimuli. Journal of Experimental Psychology-General. 2012; 141(2):363. [PubMed: 21910556]

Wade T, Holt LL. Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. The Journal of the Acoustical Society of America. 2005; 118:2618. [PubMed: 16266182]

Wang Y, Spence MM, Jongman A, Sereno JA. Training American listeners to perceive Mandarin tones. The Journal of the acoustical society of America. 1999; 106(6):3649–3658. [PubMed: 10615703]

Watanabe T, Náñez JE, Sasaki Y. Perceptual learning without perception. Nature. 2001; 413(6858): 844–848. [PubMed: 11677607]

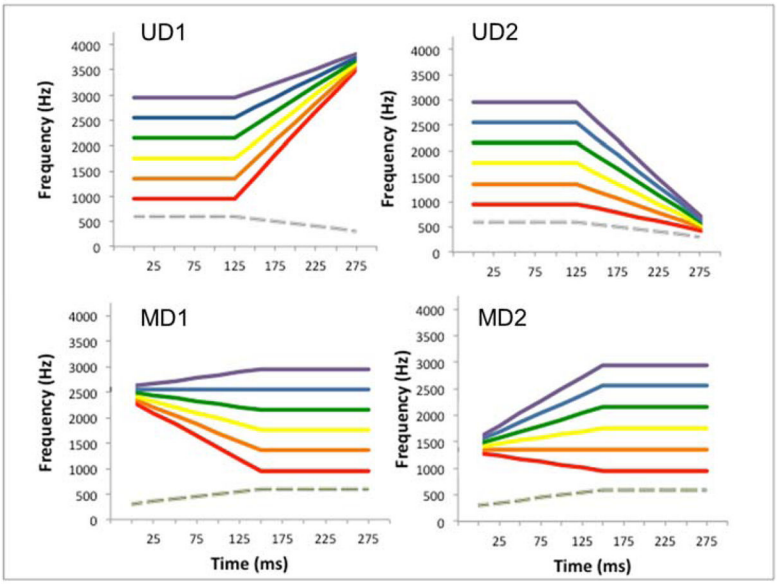

Figure 1.

Schematic spectrograms show the artificial nonspeech auditory category exemplars across time and frequency, for each uni-dimensional (UD1/UD2) and multidimensional (MD1/MD2) category. The dashed grey lines show the lower-frequency spectral peak that is common to all exemplars of a given category. Each colored line shows the higher-frequency spectral peak corresponding to a single category exemplar. See text for further details.

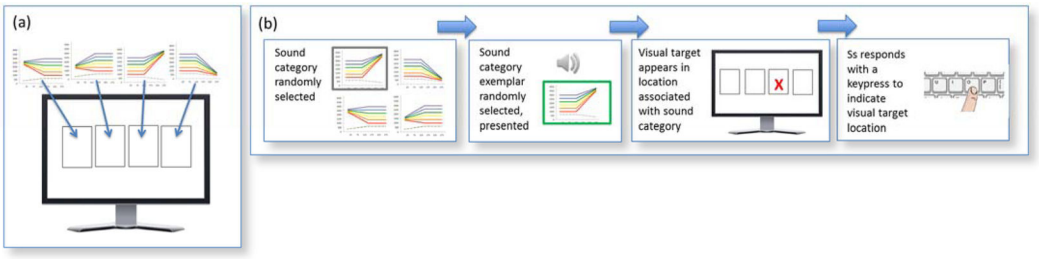

Figure 2.

Overview of the Systematic Multimodal Associations Reaction Time (SMART) task. (a) There is a consistent mapping between auditory categories and screen locations, with acoustically-variable sound exemplars associated with the category-consistent visual location. (b) The order of events in an example trial of the task. A sound category is randomly selected and an exemplar from it is chosen and presented. This is followed the appearance of a red 'X' in the corresponding screen location. Participants then respond by pressing the key corresponding to the position of the 'X'.

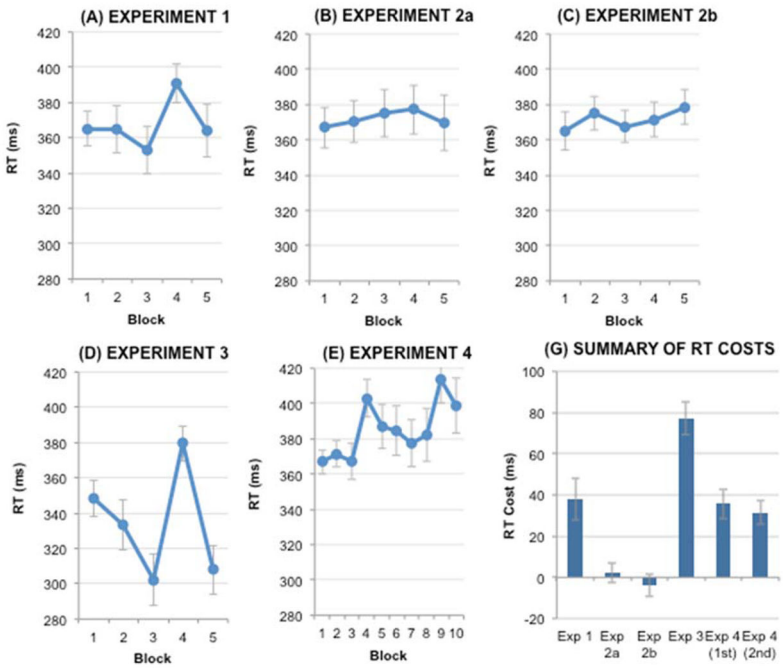

Figure 3.

Reaction time (RT) to detect the visual target as a function of Block, presented across experiments. The RT Cost is the difference in average reaction time across Blocks 3 and 4 (and 8 and 9 in Experiment 4), summarized in the bottom panel.

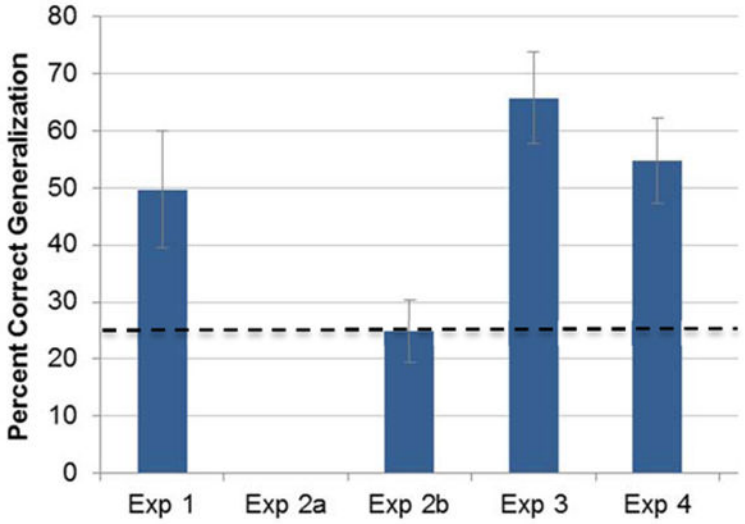

Figure 4.

Average accuracy in the post-training overt categorization task across experiments. Note that there was no overt categorization task conducted in Experiment 2a. All sounds categorized in the overt categorization task were novel category exemplars not experienced in training. The dashed line represents chance-level performance.