Dual-learning systems during speech category learning

Bharath Chandrasekaran[1], **Han-Gyol Yi**, and **W. Todd Maddox**
University of Texas, Austin

Abstract

Dual-systems models of visual category learning posit the existence of an explicit, hypothesis-testing 'reflective' system, as well as an implicit, procedural-based 'reflexive' system. The reflective and reflexive learning systems are competitive and neurally dissociable. Relatively little is known about the role of these domain-general learning systems in speech category learning. Given the multidimensional, redundant, and variable nature of acoustic cues in speech categories, our working hypothesis is that speech categories are learned reflexively. To this end, we examined the relative contribution of these learning systems to speech learning in adults. Native English speakers learned to categorize Mandarin tone categories over 480 trials. The training protocol involved trial-by-trial feedback and multiple talkers. Experiment 1 and 2 examined the effect of manipulating the timing (immediate vs. delayed) and information content (full vs. minimal) of feedback. Dual-systems models of visual category learning predict that delayed feedback and providing rich, informational feedback enhance *reflective* learning, while immediate and minimally informative feedback enhance *reflexive* learning. Across the two experiments, our results show feedback manipulations that targeted reflexive learning enhanced category learning success. In Experiment 3, we examined the role of trial-to-trial talker information (mixed vs. blocked presentation) on speech category learning success. We hypothesized that the mixed condition would enhance reflexive learning by not allowing an association between talker-related acoustic cues and speech categories. Our results show that the mixed talker condition led to relatively greater accuracies. Our experiments demonstrate that speech categories are optimally learned by training methods that target the reflexive learning system.

Introduction

A large body of behavioral and neuroscience research suggests that visual category learning is mediated by at least two separate albeit partially overlapping learning systems (Ashby & Maddox, 2005; Knowlton, 1999; Nomura & Reber, 2008; Poldrack & Packard, 2003). The explicit, 'reflective' learning system depends on working memory and executive attention to develop and test hypotheses and rules for explicit classification. Processing in this system is available to conscious awareness and is mediated by a circuit primarily involving dorsolateral prefrontal cortex, anterior cingulate and the anterior caudate nucleus (Ashby & Ell, 2001; Seger & Miller, 2010). The implicit, procedural-based, 'reflexive' learning system, is not consciously penetrable and operates by associating perception with actions that lead to reinforcement via feedback. Dual system models predict that the two systems are complementary in learning various category structures, some of which are reflective-optimal, and others reflexive-optimal. Although more than 20 years of research has motivated the dual-systems framework, this model has not been systematically applied to examine speech category learning.

Previous speech learning studies have examined category learning as an emergent property of unsupervised and/or supervised learning processes (Goudbeek, Cutler, & Smits, 2008; McClelland, Fiez, & McCandliss, 2002; Norris, McQueen, & Cutler, 2003; Toscano & McMurray, 2010; Vallabha, McClelland, Pons, Werker, & Amano, 2007). In unsupervised learning, statistical regularities in the input lead to category representations in sensory regions through a process of implicit Hebbian learning (Goudbeek, et al., 2008; Goudbeek, Swingley, & Smits, 2009; McClelland, et al., 2002). More recent, computationally based unsupervised learning models incorporate competition in addition to statistical learning (McMurray, Aslin, & Toscano, 2009; Toscano & McMurray, 2010). From the neurobiological perspective, unsupervised category learning is driven is instantiated within topographical maps in the primary and secondary auditory regions that are sensitive to input statistics (Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004; Vallabha & McClelland, 2007). In contrast, supervised learning models posit that some form of instructional feedback (lexical or selective attention) to the sensory network is necessary, in addition to Hebbian learning (Norris, et al., 2003). Although significant unsupervised speech learning can occur in adults, category learning with feedback can lead to substantially larger gains (Goudbeek, et al., 2008; McClelland, et al., 2002).

While the role of statistical learning and feedback instruction in the form of lexical influences have been extensively researched in speech learning, there has been less focus on the role of domain-general feedback-based learning processes in mediating category learning success. This is despite the fact that functional neuroimaging studies examining speech category learning in adults implicate reflective and reflexive learning circuitry in addition to the auditory regions (Callan et al., 2003; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006). In dual-systems models, reflective rules are encoded within the sensory areas with bidirectional connections to working memory units within the lateral portion of the prefrontal cortex (PFC). When a new rule is generated, the excitatory input from the PFC to the head of the caudate is strengthened, resulting in the maintenance of a newly established rule. The PFC units that each represents a particular rule are activated by the anterior cingulate to select among various alternative rules. In comparison, during reflexive learning, a single striatal unit (or small group of units) implicitly associates an abstract cortical-motor response with a large group of sensory cells. The critical aspect of learning occurs at cortical-striatal synapses, and synaptic plasticity is facilitated by a dopamine-mediated reinforcement training signal. Despite the different circuitries, both the reflective and reflexive learning systems utilize the sensory component within the primary and association auditory regions. These components are reflectively or reflexively associated with rewards (e.g., instructional feedback).

Our working hypothesis is that speech categories are optimally learned by the *reflexive* learning system. This is because speech categories are often difficult to verbalize, easily learned by infants whose attention and working memory networks are immature (Echols, 1993; Mugitani et al., 2009; Pierrehumbert, 2003), and utilize acoustic cues that are multidimensional, highly redundant, and variable across talkers (Gandour, 1983; Holt & Lotto, 2008, 2010). Creating rules for such a large dimensional space may not be optimal, since generating and testing rules that involve multiple dimensions is resource intensive. In the current paper, we use training manipulations on trial-by-trial feedback (Experiment 1, 2) and talker variability (Experiment 3) to examine the relative contribution of the reflective and reflexive learning systems to speech learning success.

The reflective and reflexive learning systems respond differentially to various training manipulations. For example, delaying the presentation of feedback impairs learning in the reflexive system, but not in the reflective system (Maddox, Ashby, & Bohil, 2003). This is because the reflexive system is critically dependent on dopamine-mediated stimulus-

response implicit reward learning. Delaying feedback interferes with dopamine release, reducing the effectiveness of the association of stimulus-response with reward. Also, 'full' feedback that provides the correctness of the response on each trial as well as information about which category was present speeds learning in the reflective system (Maddox, Love, Glass, & Filoteo, 2008) relative to 'minimal' feedback that provides only the correctness of the response on each trial. Full feedback promotes the generation and testing of rules that are critical to reflective learning but disrupts the transfer of control to the reflexive system (Maddox, et al., 2008). Previous studies have used these timing and feedback manipulations to dissociate the learning systems in artificial category learning, but not in natural speech category learning.

To this end, we conducted three category learning experiments to examine the effect of various training manipulations that target either the reflective or reflexive learning systems on speech category learning success. In each experiment, native English speakers were trained to categorize nonnative Mandarin tone categories produced by multiple talkers (Figure 1) with instructional feedback. This type of training structure (trial-by-trial feedback, high talker variability) is ubiquitous in the speech learning literature. In Mandarin Chinese, tone contours signify differences in word meaning (e.g., /ma/ with a rising tone means "mother", while /ma/ with a falling tone means "to scold," Figure 1). Previous studies show that native English speakers have difficulty in learning tone categories, which is hypothesized to result from inadequate relative weighting of talker-independent pitch direction cues (Chandrasekaran, Sampath, & Wong, 2010; Wang, Jongman, & Sereno, 2003).

Training manipulations used in experiments 1 and 2 were derived from visual category learning studies. Experiment 1 determined the extent to which the *immediacy* of feedback (immediate vs. delayed) impacts tone category learning. Experiment 2 determined the extent to which *information content* of feedback (full versus minimal feedback) impacts tone category learning (Figure 2). Immediate feedback is critical for the reflexive system but not the reflective system (Maddox, et al., 2003), while full feedback selectively speeds reflective learning but impairs reflexive learning (Maddox, et al., 2008). Based on our working hypothesis, we predicted that feedback manipulations that targeted the reflexive learning system (immediate or minimal feedback) will enhance learning relative to those that target the reflective learning system (delayed or full feedback).

While dual systems models of visual category learning make specific predictions about feedback processing, they offer no clear prediction about the impact of talker variability on category learning success, which is argued to be important for generalization to new talkers (Lively, Logan, & Pisoni, 1993) and association of categories with more reliable acoustic cues (Apfelbaum & McMurray, 2011; Rost & McMurray, 2009). While most agree that multi-talker training is advantageous, the role of the order of talker presentation, if any, has not been systematically examined (though see, Perrachione et al., 2011, which we return to in the discussion). Within the framework of the dual learning systems, we predicted that systematically blocked talker presentation will promote reflective learning, whereas a randomly mixed talker presentation will enhance reflexive learning. Our logic here is that blocked talker presentation promotes faster hypothesis testing and validation, and is therefore less resource intensive for the reflective system than is the mixed talker condition. Indeed, previous neuroimaging work show that the mixed talker condition engages the frontal working memory system more extensively than the blocked talker condition (Wong, Nusbaum, & Small, 2004). Further, the mixed talker presentation does not allow learners to predict the next talker in advance, disrupting the reflective generation and testing of talker-specific rules. Therefore, learners are more likely to associate talker-invariant acoustic cues (e.g., pitch direction) with implicit reward than talker-variant cues (e.g., pitch height). Based

on the hypothesis that speech learning is optimally learned by the reflexive learning system, we predicted enhanced learning in the mixed talker condition, relative to the blocked talker condition.

To summarize, enhanced learning in the immediate feedback condition relative to delayed feedback (Experiment 1), minimal feedback relative to full feedback (Experiment 2), and mixed talker condition relative to blocked talker condition (Experiment 3) will be considered as support for the dominance of the reflexive system in speech learning during adulthood.

Methods

Participants

Undergraduate students at the University of Texas were recruited (n = 194; age range: 18-35 years) and monetarily compensated for their participation. Participants reported no history of neurological or hearing deficits and were native speakers of American English with no prior exposure to a tone language. Music history questionnaires were collected to match the groups on musicianship (Wong, Perrachione, & Parrish, 2007). All participants provided informed consent and were debriefed following the experiment. In Experiment 1, participants were divided into immediate (n=25; 15 f) and delayed feedback (n=30; 14 f) groups (equivalent years of musical training: p = .585; immediate: mean = 3.42, SEM = .675; delayed: mean = 2.90, SEM = .664). In Experiment 2, participants were divided into minimal (n=41; 20 f) and full feedback (n=40; 21 f) groups (equivalent years of musical training: p = .979; minimal: mean = 2.20, SEM = .525; full: mean = 2.21, SEM = .393). In Experiment 3, participants were divided into mixed (n=30; 18 f) and blocked talkers (n=28; 15 f) groups (equivalent years of musical training: p = .723; mixed: mean = 3.18, SEM = .787; blocked: mean = 3.57, SEM = .776). Participants did not overlap between groups.

Stimuli for tone category training

Four native Mandarin Chinese speakers (2 f) originally from Beijing produced four Mandarin tones: tone 1 (T1; high-level), tone 2 (T2; low-rising), tone 3 (T3; low-dipping), and tone 4 (T4; high-falling; Figure 1). The tones were produced in citation form in the context of five monosyllabic Mandarin Chinese words (*bu, di, lu, ma,* and *mi*), reflecting variability inherent in natural language. The 80 stimuli were RMS amplitude and duration normalized (70 dB, 0.44 s). Duration normalization was achieved using the PSOLA (Pitch Synchronous Overlap and Add) module incorporated within the Praat software (Boersma & Weenink, 2011). Five independent native Mandarin Chinese speakers rated the stimuli as highly natural and accurately (>95%) identified the tone categories.

Procedure

In all experiments, each auditory stimulus was presented with the written prompt: "Which category? (Press the number key)". Participants generated a response by pressing one of four buttons on a keyboard labeled "1", "2", "3", or "4", corresponding to T1, T2, T3, and T4, respectively. Feedback was displayed for 1000 ms after the response depending on the experimental condition: immediately, or delayed by 500 or 1000 ms. The content of feedback varied depending on the accuracy of the response ("Correct/No") and the experimental condition, in which full feedback informed the participant of the correct answer regardless of the accuracy. The stimulus-response-feedback sequence comprised a single trial. One block consisted of a randomized presentation of all 80 stimuli. Six blocks were presented, yielding in a total of 480 trials per participant. In Experiment 1, response-to-feedback interval was immediate or delayed (1 s), always with full information. In Experiment 2, feedback information content was minimal or full, with a fixed response-to-

feedback interval of 500 ms. In Experiment 3, the order of presentation of talker information was manipulated. In the Mixed Talker condition, the stimuli were presented in a random sequence. In the Blocked Talker condition, the stimuli were only randomized in terms of syllables and tones, but ordered in terms of talker information. The feedback in Experiment 3 was always immediate and minimal (Figure 2). In short, for all experiments, six blocks of 80 trials each were presented to the participants. The order of talkers was completely randomized for Experiments 1 and 2. In Experiment 3, in the blocked condition, the sequence of talkers (n=4) was randomized. Within a talker sequence, participants listened to all stimuli (also randomized) produced by the talker (n=20).

Results

Category learning over training trials

Figure 3 shows the proportion of participants who made a correct category response (sliding window of 80 trials) across conditions. Visually, reflexive and reflective conditions are more equivalent in the beginning than towards the end of the experiment, where the reflexive conditions consistently lead to improved learning progress. This inspection was corroborated with statistical analysis detailed in the next section.

Data Analysis

For each participant, response to each trial was coded as "correct" or "incorrect." A mixed logit analysis was conducted to predict the log odds of producing a correct response, using the lmer program with binomial logit link (Bates, Maechler, & Bolker, 2012). The dependent variable was set as the "correct" or "incorrect" outcome of each response. The fixed effects of interest were the between-subjects condition ("reflexive" vs. "reflective"), trial number (increasing from 1 to 480; mean-centered to 0 and divided by 100), and their interaction term. The model was corrected for by-subject and by-item random intercepts, which was the most complex model as justified by the data, $p < .05$: Outcome ~ Condition * Trial + (1 | Subject) + (1 | Item)

Experiment 1: Immediate versus Delayed Feedback

The trial effect was significant ($p < .0001$; each successive trial increases probability of an accurate response). The trial by condition interaction was significant ($p < .0001$; each successive trial increases probability of an accurate response, more for reflexive-immediate than for reflective-delayed condition). The condition effect was not significant (Table 1).

Experiment 2: Full versus Minimal Feedback

The trial effect was significant ($p < .0001$; each successive trial increases probability of an accurate response). The trial by condition interaction was significant ($p < .0001$; each successive trial increases probability of an accurate response, more for reflexive-minimal than for reflective-full condition). The condition effect was not significant (Table 2).

Experiment 3: Mixed vs. Blocked Talker Presentation

The trial effect was significant ($p < .0001$; each successive trial increases probability of an accurate response). The trial by condition interaction was significant ($p = .021$; each successive trial increases probability of an accurate response, more for reflexive-mixed than for reflective-blocked condition). The condition effect was not significant (Table 3).

Discussion

We hypothesized that the reflexive learning system is optimal for adult speech category learning. The results from Experiments 1-3 strongly support our hypothesis. All three experiments show that training manipulations that targeted the reflexive learning system enhanced learning relative to those that targeted the reflective learning system.

In Experiments 1 and 2, we examined trial-by-trial feedback training manipulations. Both experiments yielded significant interaction between trial and training manipulation, such that accuracy increase over trials was greater for manipulations that targeted the reflexive learning system. In Experiment 1, we found that immediate feedback enhances tone learning towards the end of training, relative to delayed feedback. As per the dual systems models, immediate feedback is a critical requirement of the dopamine-mediated reflexive learning system, but not reflective learning. Therefore, delaying feedback even by just one second disrupts the dopamine-mediated training signal for the reflexive learning system, allowing control to pass to the reflective learning system. In Experiment 2, we found that minimal feedback enhances tone learning towards the end of training, relative to full feedback. Since the full feedback not only provides information about the correctness of the response but also the correct category membership, interpretation of this result is counterintuitive without the consideration of the dual systems perspective; full feedback promotes hypothesis generation and testing, which prevents transfer of control from the reflective system to the reflexive system. Therefore, Experiment 2 is also in support of the hypothesis that the reflexive system is optimal for speech category learning.

In Experiment 3, we manipulated the order of talker presentation: randomly mixed or systematically blocked. This experiment yielded an interaction between trial and training manipulation, such that accuracy increase over trials was greater for the mixed talker condition, relative to the blocked talker condition. This is consistent with our prediction that the mixed condition targets the reflexive learning system by preventing talker-dependent hypothesis generation and testing by the reflective system. As a result, participants in the mixed talker condition are led to rely more on relatively talker-invariant acoustic cues (e.g., pitch direction) than on talker-variant cues (e.g., pitch height). These results are consistent with L1 speech acquisition literature that show that high variability can guide category learning by allowing associations between more invariant cues and the category structure (Apfelbaum & McMurray, 2011; Rost & McMurray, 2009), but not consistent with a previous study that had found greater Mandarin tone learning in the blocked talker condition than in the mixed talker condition (Perrachione, Lee, Ha, & Wong, 2011). These contradictory findings, we believe may be due to methodological differences. The high degree of variability in natural Mandarin tone categories was somewhat reduced in the previous study, where only three of the categories (T1, T2, and T4) were presented and the tone contours interpolated linearly. In contrast, our study utilized the full category structure with the natural tone contours retained (Figure 1). We have argued earlier that the multidimensionality is what makes speech categories hard to verbalize and therefore optimally learned through the reflexive system. In contrast, the reflective system is more likely to suffice as the complexity of the category structure is reduced.

In adult speech learning studies, significant category learning is evidenced without feedback, consistent with unsupervised learning models. The current study did not include a no-feedback condition that could help tease apart the relative contribution of unsupervised implicit learning in mediating category learning success. However, category learning is substantially more efficient with feedback (Goudbeek, et al., 2008; McClelland, et al., 2002). Our theoretical approach does not rule out the statistical learning processes operative within the primary and secondary auditory cortices. Rather, our results demonstrate the

operational specifics of the domain-general feedback-based learning system that is optimal for learning speech categories. Although neuroimaging studies have implicated the neostriatum in speech category learning tasks, the role of the reflexive system, has not been systematically examined. This is despite that fact that animal models clearly demonstrate direct, extensive, and many-to-one connectivity between the primary/secondary auditory cortex and the reflexive systems, suggesting a distinct neurobiological plausibility for a substantial role for this circuitry (Petrides & Pandya, 1988; Yeterian & Pandya, 1998). Future research should systematically examine the relative contribution of various forms of perceptual and learning processes to speech category learning, as well as use other perceptual forms.

In summary, our results demonstrate that speech category learning is optimally learned by the reflexive system. Our results offer practical implications for the development of optimized training approaches that can target the reflexive learning system. Specifically, we hypothesize that for speech categories, learning can be optimized by including minimal and immediate feedback and high trial-by-trial talker variability.

References

Apfelbaum KS, McMurray B. Using variability to guide dimensional weighting: associative mechanisms in early word learning. Cognitive Science. 2011; 35(6):1105–1138. [PubMed: 21609356]

Ashby EG, Maddox WT. Human category learning. Annual Review of Psychology. 2005; 56:149–178.

Ashby FG, Ell SW. The neurobiology of human category learning. Trends in Cognitive Sciences. 2001; 5(5):204–210. [PubMed: 11323265]

Bates, D.; Maechler, M.; Bolker, B. lme4: Linear mixed-effects models using S4 classes. 2012.

Boersma, P.; Weenink, D. [17 September, 2011] Praat: doing phonetics by computer (Version 5.2. 41) [Computer program]. 2011.

Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. Neuroimage. 2003; 19(1):113–124. [PubMed: 12781731]

Chandrasekaran B, Sampath PD, Wong PC. Individual variability in cue-weighting and lexical tone learning. Journal of the Acoustical Society of America. 2010; 128(1):456–465. [PubMed: 20649239]

Echols CH. A perceptually-based model of children's earliest productions. Cognition. 1993; 46(3): 245–296. [PubMed: 8462274]

Gandour J. Tone perception in Far Eastern languages. Journal of Phonetics. 1983; 11:149–175.

Goudbeek M, Cutler A, Smits R. Supervised and unsupervised learning of multidimensionally varying non-native speech categories. Speech Communication. 2008; 50(2):109–125.

Goudbeek M, Swingley D, Smits R. Supervised and unsupervised learning of multidimensional acoustic categories. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35(6):1913–1933. [PubMed: 19968443]

Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA. Representation of sound categories in auditory cortical maps. Journal of Speech Language and Hearing Research. 2004; 47(1):46–57.

Holt LL, Lotto AJ. Speech Perception Within an Auditory Cognitive Science Framework. Current Directions in Psychological Science. 2008; 17(1):42–46. [PubMed: 19060961]

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

Knowlton BJ. What can neuropsychology tell us about category learning? Trends in Cognitive Sciences. 1999; 3(4):123–124. [PubMed: 10322464]

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 Pt 1):1242. [PubMed: 8408964]

Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. Journal of Experimental Psychology. Learning, Memory, and Cognition. 2003; 29(4):650–662.

Maddox WT, Love BC, Glass BD, Filoteo JV. When more is less: feedback effects in perceptual category learning. Cognition. 2008; 108(2):578–589. [PubMed: 18455155]

McClelland JL, Fiez JA, McCandliss BD. Teaching the /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects. Physiology & Behavior. 2002; 77(4-5):657–662. [PubMed: 12527015]

McMurray B, Aslin RN, Toscano JC. Statistical learning of phonetic categories: insights from a computational approach. Developmental Science. 2009; 12(3):369–378. [PubMed: 19371359]

Mugitani R, Pons F, Fais L, Dietrich C, Werker JF, Amano S. Perception of vowel length by Japanese- and English-learning infants. Developmental Psychology. 2009; 45(1):236–247. [PubMed: 19210005]

Nomura EM, Reber PJ. A review of medial temporal lobe and caudate contributions to visual category learning. Neuroscience & Biobehavioral Reviews. 2008; 32(2):279–291. [PubMed: 17868867]

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. Cognitive Psychology. 2003; 47(2):204–238. [PubMed: 12948518]

Perrachione TK, Lee J, Ha LY, Wong PC. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. Journal of the Acoustical Society of America. 2011; 130(1):461–472. [PubMed: 21786912]

Petrides M, Pandya DN. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. Journal of Comparative Neurology. 1988; 273(1):52–66. [PubMed: 2463275]

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

Poldrack RA, Packard MG. Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia. 2003; 41(3):245–251. [PubMed: 12457750]

Rost GC, McMurray B. Speaker variability augments phonological processing in early word learning. Developmental Science. 2009; 12(2):339–349. [PubMed: 19143806]

Seger CA, Miller EK. Category learning in the brain. Annual Review of Neuroscience. 2010; 33:203–219.

Toscano JC, McMurray B. Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics. Cognitive Science. 2010; 34(3):434–464. [PubMed: 21339861]

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

Vallabha GK, McClelland JL. Success and failure of new speech category learning in adulthood: consequences of learned Hebbian attractors in topographic maps. Cognitive, Affective, & Behavioral Neuroscience. 2007; 7(1):53–73.

Vallabha GK, McClelland JL, Pons F, Werker JF, Amano S. Unsupervised learning of vowel categories from infant-directed speech. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(33):13273–13278. [PubMed: 17664424]

Wang Y, Jongman A, Sereno JA. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. Journal of the Acoustical Society of America. 2003; 113(2):1033–1043. [PubMed: 12597196]

Wong PC, Nusbaum HC, Small SL. Neural bases of talker normalization. J Cogn Neurosci. 2004; 16(7):1173–1184. [PubMed: 15453972]

Wong PC, Perrachione TK, Parrish TB. Neural characteristics of successful and less successful speech and word learning in adults. Human Brain Mapping. 2007; 28(10):995–1006. [PubMed: 17133399]

Yeterian EH, Pandya DN. Corticostriatal connections of the superior temporal region in rhesus monkeys. Journal of Comparative Neurology. 1998; 399(3):384–402. [PubMed: 9733085]
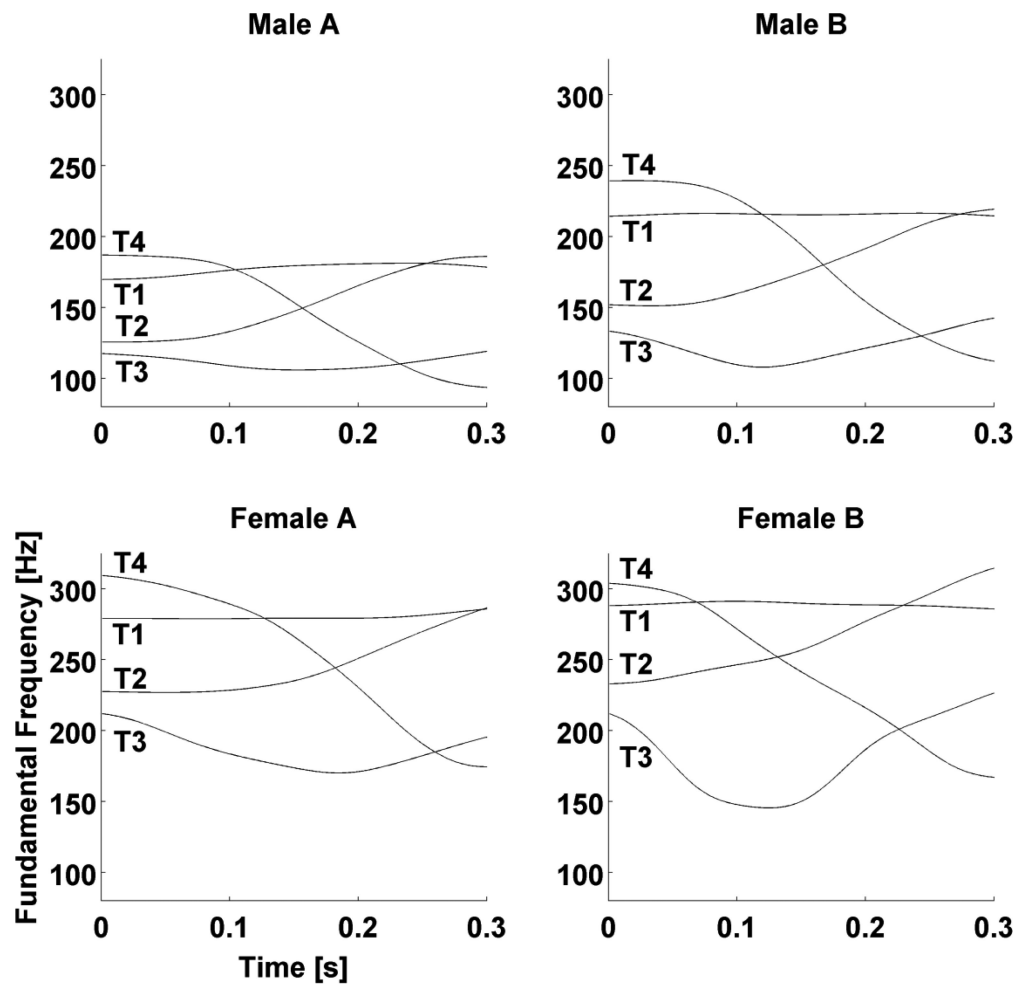
Figure 1.
Example of multiple-talker stimuli used in the category training study. Fundamental frequency contours of the four tones (T1 = high-level; T2 = low-rising; T3 = low-dipping; T4 = falling) produced by four native Mandarin speakers (2 f). Tone contours were obtained using Praat (Boersma & Weenink, 2011).
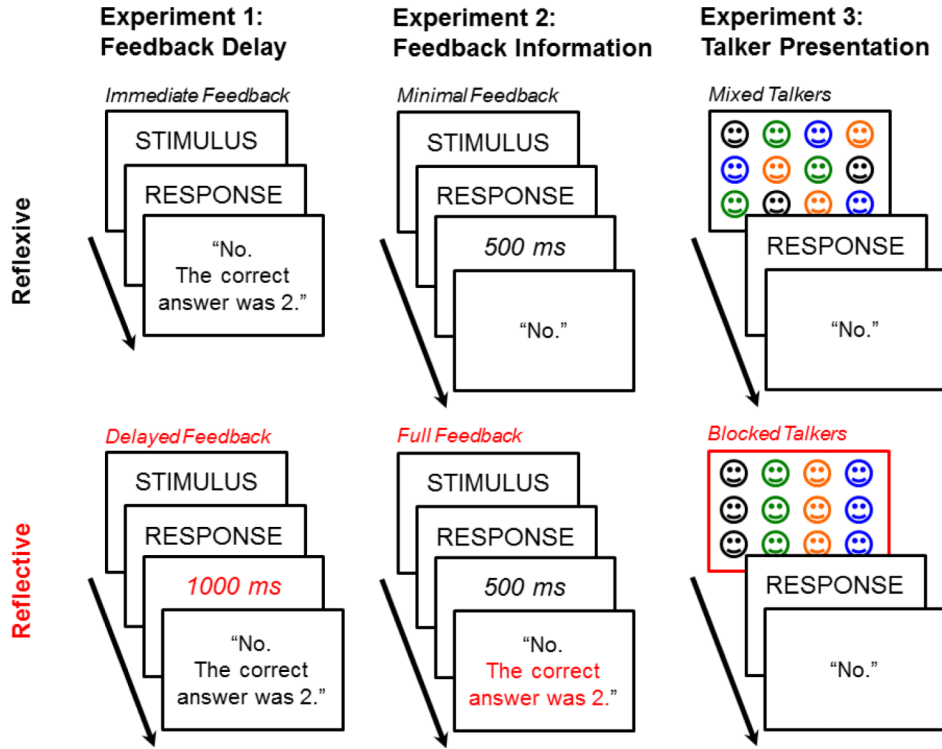
Figure 2.
Experimental procedures. In Experiments 1-3 we examined the effects of reflexive (top) or reflective (bottom) training manipulations on tone category learning success.
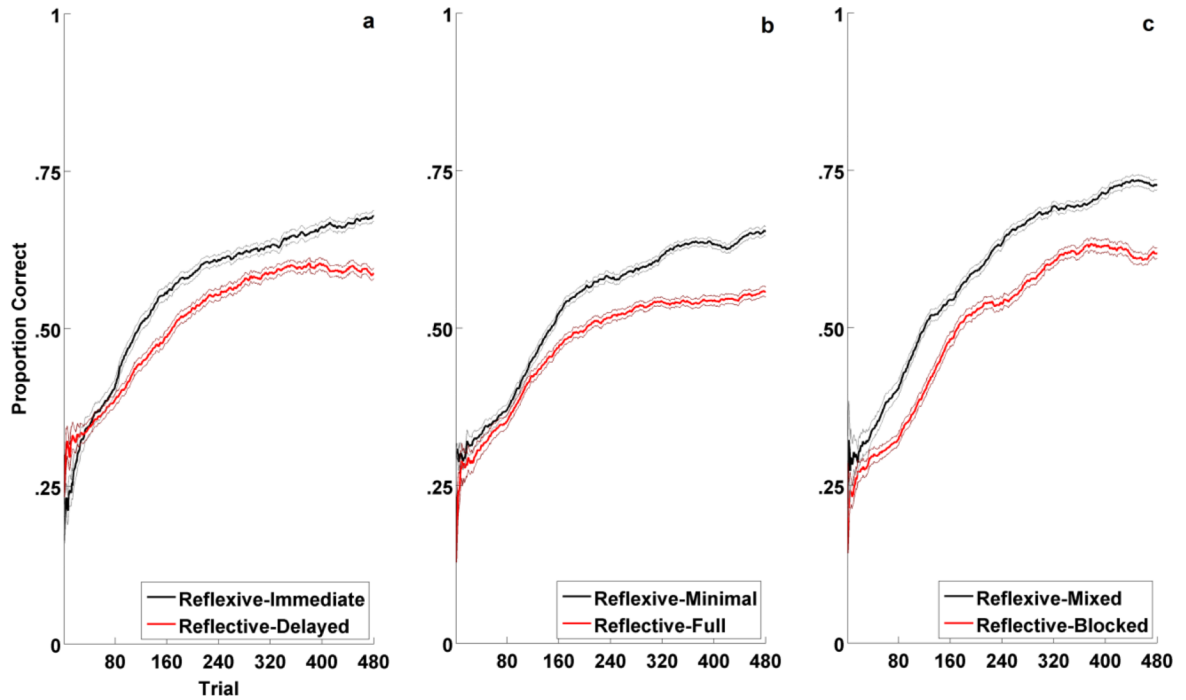
Figure 3.
Category learning curves across reflexive vs. reflective conditions in all three experiments: (a) Experiment 1: feedback delay (immediate vs. delayed); (b) Experiment 2: feedback information (minimal vs. full); (c) Experiment 3: talker variability (mixed vs. blocked). Plotted in solid bold lines are the proportions of correct responses across participants within each condition over the course of learning. The black lines denote the reflexive conditions and the red, the reflective conditions. For purposes of visualization of trial-by-trial data, each point in the line denotes the average number of correct responses in a sliding 80-trial window. For trials preceding the 80th trial, cumulative averages were used. Plotted in thin lines are the ranges of standard error of the averages used in the sliding windows. Visual assessment of the learning curves suggest that both conditions result in equivalent degrees of category learning towards the earlier phase of experiment, but the reflexive condition leads to greater learning than does the reflective condition towards the later phase of the experiment. This pattern is consistent across all three experiments.

Table 1

Effects of Feedback Delay on Category Learning

Fixed Effect	Coefficient	SE	z	P
Intercept	.23888	.21921	1.090	.276
Trial	.25127	.01380	18.211	$< 2e^{-16}$
Condition	.29577	.31519	.0938	.348
Trial*Condition	.08729	.02114	4.129	$3.64e^{-5}$

Note. Coefficients express log odds.

Table 2

Effects of Feedback Information on Category Learning

Fixed Effect	Coefficient	SE	z	P
Intercept	.03810	.18865	.202	.840
Trial	.24091	.01223	19.695	$< 2e^{-16}$
Condition	.34199	.25305	1.351	.177
Trial*Condition	.10610	.01736	6.111	$9.92e^{-10}$

Note. Coefficients express log odds.

Table 3

Effects of Talker Variability on Category Learning

Fixed Effect	Coefficient	SE	z	P
Intercept	.20323	.21452	.947	.343457
Trial	.36940	.01477	25.012	$< 2e^{-16}$
Condition	.45136	.28765	1.569	.116612
Trial*Condition	.07834	.02113	3.708	.000209

Note. Coefficients express log odds.