Published in final edited form as:

Biling (Camb Engl). 2014 October 1; 17(4): 709–728. doi:10.1017/S1366728913000783.

Tests of a Dual-systems Model of Speech Category Learning

W. Todd Maddox¹ and Bharath Chandrasekaran

University of Texas, Austin

Abstract

In the visual domain, more than two decades of work posits the existence of dual category learning systems. The *reflective* system uses working memory to develop and test rules for classifying in an explicit fashion. The *reflexive* system operates by implicitly associating perception with actions that lead to reinforcement. Dual-systems models posit that in learning natural categories, learners initially use the reflective system and with practice, transfer control to the reflexive system. The role of reflective and reflexive systems in second language (L2) speech learning has not been systematically examined. Here monolingual, native speakers of American English were trained to categorize Mandarin tones produced by multiple talkers. Our computational modeling approach demonstrates that learners use reflective and reflexive strategies during tone category learning. Successful learners use talker-dependent, reflective analysis early in training and reflexive strategies by the end of training. Our results demonstrate that dual-learning systems are operative in L2 speech learning. Critically, learner strategies directly relate to individual differences in category learning success.

Keywords

L2 acquisition; dual-learning systems; reflexive processing; reflective processing; individual differences

Introduction

A large body of neuropsychological, neuroimaging, and behavioral studies in vision has identified two dissociable systems that are operative during category learning: a *reflective* system, wherein processing is under conscious control, and a *reflexive* system that is not under conscious control (Ashby & Maddox, 2011; Ashby & Spiering, 2004; Maddox & Ashby, 2004; Nomura et al., 2007; Nomura & Reber, 2008; Poldrack et al., 2001; Poldrack & Foerde, 2008; Poldrack & Packard, 2003). The reflective learning system uses working memory and executive attention to develop and test verbalizable rules based on feedback (Ashby, Maddox, & Bohil, 2002; Maddox & Ashby, 2004; Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005; Maddox, Love, Glass, & Filoteo, 2008). In contrast, the reflexive learning system is not consciously penetrable, non-verbalizable, and operates by automatically associating perception with actions that lead to reward (Seger, 2008; Seger & Cincotta, 2005; Seger & Miller, 2010). This system is not dependent on working memory

Address correspondence to: W. Todd Maddox, maddox@psy.utexas.edu or Bharath Chandrasekaran, bchandra@utexas.edu.

¹The contribution of both authors was equal.

and executive attention (Decaro, Thomas, & Beilock, 2008), is implicit and procedural. Although, there is anatomical evidence of extensive connectivity between auditory regions and the reflective and reflexive systems (Petrides & Pandya, 1988; Yeterian & Pandya, 1998), the dual-systems framework has not been systematically applied in the auditory domain.

This paper applies the dual-systems theoretical framework in speech category learning. We use computational modeling approaches within a dual-systems framework to examine learning natural speech (Mandarin tone) categories by adult English speakers. Computational modeling provides insight into specific learning operations (e.g. reflective vs. reflexive) that participants employ (Cleeremans & Dienes, 2008) and can account for individual variability in task performance. Although accuracy rates provide information regarding the level of performance, they provide little information regarding the specific strategy being used by a participant because there exists a number of reflexive and reflective strategies that yield the same accuracy rate. The computational models used in the current study are constrained by our understanding of the neurobiology of the reflective and reflexive learning systems. Therefore, modeling allows a more systematic examination of the computational strategies learners' use while acquiring novel speech categories. We will review the neurobiology underlying the two learning systems next.

Dual-learning systems: Neurobiology

The Competition between Verbal and Implicit Systems (COVIS) model captures the dual-systems framework, and proposes a neural circuitry involved in visual category learning (Ashby & Maddox, 2011; Maddox & Ashby, 2004; Nomura & Reber, 2008). In COVIS, processing in the reflective, hypothesis-testing system is available to conscious awareness and is mediated by a circuit primarily involving dorsolateral prefrontal cortex, anterior caudate nucleus, anterior cingulate, and medial temporal lobe structures. Processing in the reflexive, procedural-based learning system is not consciously penetrable and operates by associating perception with actions that lead to reinforcement. Learning in the procedural system is mediated primarily by the posterior caudate nucleus and putamen (Ashby & Maddox, 2005; Ashby & O'Brien, 2005). COVIS assumes that the two systems compete throughout learning, but that initial learning is dominated by the reflective, hypothesis-testing system. Learners will continue to use the hypothesis-testing system until the output of the procedural system is more accurate when control is passed to the reflexive, procedural system.

Anatomical studies in animal models suggest that the primary and association auditory cortical regions are strongly connected to the reflective and reflexive systems. Retrograde anatomical labeling studies in primates demonstrate that the primary and association cortex are bi-directionally connected to the prefrontal cortex and form many-to-one projections to the caudate (Petrides & Pandya, 1988; Yeterian & Pandya, 1998). Fibers from the superior temporal gyrus run through the extreme capsule to the frontal lobe (ventral route). A second group of fibers curve around the Sylvian fissure to the frontal lobe (dorsal route). With respect to the reflexive system, massive many-to-one projections from secondary auditory areas connect to the tail and body of the caudate, as well as the putamen—which are key

areas in the reflexive system. These studies lend neurobiological plausibility to the application of a dual-systems framework in the auditory domain. Note that extant dual-systems models restrict themselves to the visual domain (Ashby & Ell, 2001; Ashby & Ennis, 2006; Ashby & Maddox, 2005, 2011), although the underlying systems are assumed to be domain general.

Dual-systems framework: category structures

COVIS assumes that the reflective and reflexive systems are complementary and excel in the learning of different types of category structures. Category structures that are learnable by the reflective system, like those shown in Figure 1a, are referred to as rule-based (RB) structures (Ashby & Maddox, 2011). The optimal verbal rule (denoted by the solid horizontal and vertical lines) is to “respond A to short, high frequency sounds, B to short, low frequency sounds, C to long, high frequency sounds, and D to long, low frequency sounds”. In contrast, the structure in Figure 1b is not learnable by the reflective system because the optimal rule (denoted by the solid diagonal lines) is not verbalizable (i.e., frequency and duration involve incommensurable units). However, such structures can still be learned and are referred to as information-integration (II) category structures. Participants who learn the II category structure are unable to articulate verbal rules used to learn these categories, but likely integrate dimensions prior to decision-making.

These artificial category structures have been used to test the extent to which the two learning systems are dissociable. A large number of dissociations have been found between the two learning systems in the visual domain. A critical difference between the systems is the reliance on working memory (WM) (DeCaro, Carlson, Thomas, & Beilock, 2009; Decaro, et al., 2008). The reflective learning system critically depends on WM, while the reflexive learning system does not. Individuals with high working memory capacity (WMC) are faster at learning rule-based categories relative to those with low WMC (DeCaro, et al., 2009; Decaro, et al., 2008). Interestingly, this pattern does not hold true for learning II category structures. While artificial category structures are a good test-bed to examine dissociations between the two systems, most naturally occurring categories cannot be cleanly demarcated as either RB or II categories.

Dual-systems framework in second language acquisition (SLA)

While the dual-systems framework has not been systematically examined in L2 speech category learning, there has been significant work on the role of explicit and implicit learning systems in SLA (see (DeKeyser, 2008), for an extensive review). The mechanistic role of multiple learning systems has generated considerable interest as well as controversy, both in empirical and practical domain (e.g. second language instruction) (Hulstijn, 2005; Krashen, 1982; Schmidt, 1995, 2012), as well as in the theoretical domain (e.g. models of SLA) (Paradis, 1985, 2004; Ullman, 2004, 2006). Early models argue for the existence of conscious and subconscious systems in language acquisition, with the subconscious system posited to be more critical to language learning. For example, Krashen’s Monitor model argues that instruction may not be optimal for all language learners, because complex language rules are mostly acquired subconsciously (Krashen, 1982). In contrast, other accounts (e.g. Noticing hypothesis) suggest the conscious learning processes are critical to

SLA (Schmidt, 1995, 2012). As per this account, conscious awareness and attention are critical for L2 learning. More recent dual-systems models argue for a differential role for explicit and implicit learning systems in mediating different components of language. In the DP (Declarative-Procedural) model (Ullman, 2004, 2006); for a related model and application see (McClelland, McNaughton, & O'Reilly, 1995) a declarative, rule-based system, subserves vocabulary learning, and a more automatic, procedural system, mediated by the basal ganglia, subserves key aspects of grammar processing. Learning by the declarative system involves fast mapping. On the other hand, learning by the procedural system is more gradual, requiring more extensive practice. In SLA, the DP model predicts that the fast-mapping declarative system initially dominates during grammar processing, making L2 processing effortful and less automatic (Hernandez & Li, 2007). With practice and training (e.g. advanced L2 learners) however, the more gradual-learning procedural system becomes more dominant, allowing more automaticity in L2 processing. The neurolinguistic theory of bilingualism also hypothesizes the existence of declarative and procedural learning systems that mediate metalinguistic and implicit linguistic competence, respectively (Paradis, 2004). In this model, implicit linguistic competence is achieved via incidental learning. This model proposes that L2 learners will have maximum difficulty acquiring phonetics and prosody, features that are implicitly acquired. While much of the literature on implicit vs. explicit learning processes has focused on morphology, syntax, or vocabulary learning, much less is known about the role of the dual systems in L2 phonetic acquisition. Here we extend a dual-systems approach to understanding phonetic learning. From a neurobiological perspective, our dual-systems approach differs from extant dual-systems models in the language domain in several important ways. In SLA literature, the declarative system, mediated by the medial temporal lobe, is argued to play a critical role in explicit processing. The neural bases of the implicit system is however, is less clearly delineated. For example, the DP model implicates the basal ganglia as a putative structure for procedural learning. Others suggest a non-specific locus for implicit processes (Reber, 2013), or view implicit processes as a form of computing statistics (Evans, Saffran, & Roberts-Torres, 2009; Hay, Pelucchi, Graf Estes, & Saffran, 2011; Romberg & Saffran, 2010; Saffran, Aslin, & Newport, 1996). In our dual-systems theoretical framework, the medial temporal lobe is not the primary locus of explicit learning; the primary role is performed by the prefrontal cortex (PFC). The PFC, in coordination with the head of the caudate, and the anterior cingulate, is important in generating, testing, and revising hypothesis, based on feedback. In contrast, the MTL-based declarative memory system keeps track of hypothesis that have been tested and those that have been rejected when rules are complex. The implicit 'reflexive' learning system, in our model, is instantiated within anatomic regions in the basal ganglia (body and tail of the caudate, putamen). Importantly, the computational modeling used in the current study is neurobiologically-constrained, based on the knowledge about the reflective/reflexive systems, and their communication with the primary and secondary sensory regions.

Learning L2 speech categories as adults

A significant challenge in speech perception and learning is mapping highly variable signals, produced by multiple talkers to relevant categories, which can be likened to a categorization problem (Holt & Lotto, 2008, 2010). Thus, successful categorization requires resolving the

highly variable signals, produced by multiple talkers into specific categories. Electrophysiological work argues for the existence of stored representations of speech categories that are language-specific (Cheour et al., 1998; Naatanen et al., 1997). Extant neuroscientific models argue that these category representations are instantiated within association auditory regions in the posterior superior temporal gyrus (Hickok & Poeppel, 2004, 2007). In neuroimaging studies examining laboratory-based L2 speech learning, short-term auditory training enhances activation of the auditory association areas (Wang, Sereno, Jongman, & Hirsch, 2003a), and results in enhanced electrophysiological responses to newly learned categories.

The mechanisms underlying speech category learning has been a focus of several studies. Previous research has theorized several reasons for difficulties in L2 speech learning, attributed to various levels of speech processing, including interference caused by existing speech categories, and interference due to a ‘warping’ of auditory-perceptual space by prior experience with native speech categories (Best, 1993, 2006; Best, Morrongiello, & Robson, 1981; Flege, 1999; Francis, Ciocca, Ma, & Fenn, 2008; Francis & Nusbaum, 2002). Laboratory training studies have typically utilized trial-by-trial feedback and high variability (multiple talkers) training to teach L2 speech categories (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Lim & Holt, 2011; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006; Zhang et al., 2009). Feedback enhances learning by reducing errors, and multiple-talker training results in learners refocusing their attention to cues that are relevant for distinguishing speech categories and/or reducing attention to irrelevant cues (Bradlow, 2008). Although unsupervised training results in some amount of speech learning in adults, the addition of feedback results in substantially larger learning gains (Goudbeek, Cutler, & Smits, 2008; McClelland, Fiez, & McCandliss, 2002; Vallabha & McClelland, 2007). Studies have also examined the role of high-variability (using multiple talker) training in speech learning. Training with multiple talkers leads to better real-world generalization, relative to single-talker training (Bradlow, 2008; Lively, Logan, & Pisoni, 1993; Lively, et al., 1994). Learners trained using a multiple-talker paradigm show significant transfer to a novel talker (Lively et al., 1993). On the other hand, learners trained using a single-talker paradigm showed significant talker-dependent learning, but showed poor transfer to a novel talker. This suggests that high variability training may be a beneficial method to create robust representations of categories that are resistant to variation across talkers. However, the ability to successfully learn with high talker variability during training may depend on the learner. For example, Perrachione et al. (2011) showed that high variability training was beneficial to some individuals, but others performed better when exposed to less variability in training.

Several speech perceptual learning studies have examined the role of talker normalization processes in speech category learning (Kraljic & Samuel, 2005, 2006; Samuel & Kraljic, 2009). Perceptual learning studies show that adults constantly adjust their phonemic categories to incorporate phonetic variations produced by novel talkers. Talker-dependent perceptual analyses during categorization is driven by the nature of the speech categories (e.g. fricatives vs. stops). For fricatives, spectral mean is a critical acoustic cue differentiating categories (e.g. [sh] vs. [s], [sh] has a lower spectral mean); this cue also

correlates with gender differences (male talkers have a lower spectral mean than female talkers). Perceptual training resulted in gender-specific adjustment of [s] and [sh] categories (Kraljic & Samuel, 2007). In contrast, perceptual training leading to adjustment of stops (cued by voice-onset time, a timing cue that is not gender-specific) was not gender-specific.

Taken together, the studies reviewed before suggests that feedback and talker variability lead to significant L2 speech learning. While much of this research has focused on the mechanics of the perceptual system in speech learning, much less is known about the role of the dual category learning systems, which previous studies suggest is critical to learning RB and II category structures. This leads us to an important point, are speech categories similar to RB category structures, or II category structures? Speech categories are typically difficult to verbalize, have multiple dimensions, and are highly variable. Generating and testing hypothesis for categories involving multiple dimensions is resource-intensive. Since the reflective system is dependent on working memory and attention, generating rules/hypothesis for multiple dimensions may not be efficient. Further, the redundancy, and variability of cues available during speech perception prevents a simple one-to-one mapping of cues to categories. These suggest that reflexive learning may be most optimal for speech categories. Does this mean there is no role for the reflective system during L2 speech learning? Talker-dependent analyses engage working memory and executive attention resources (Wong, Nusbaum, & Small, 2004a). Therefore, it is possible that resolving talkers in a multi-talker training paradigm may engage the reflective system substantially. Our hypothesis is therefore that speech learning is reflexive-optimal, but may involve some reflective analysis. During natural visual category learning, the dual-systems framework assumes that the reflective and reflexive learning systems compete throughout learning for control (Ashby & Maddox, 2011). Early learning is mostly reflective and involves actively testing hypotheses and using feedback to validate or invalidate rules. With practice, learners switch to the more automatic, reflexive learning if the output of the reflexive system is more accurate than the reflective system. In line with dual-systems prediction, we propose that resolving talker variability is at least partially a reflective process. In contrast, learning natural speech categories, we argue, is reflexive-optimal. We examine these hypotheses across two experiments involving Mandarin tone category learning by native English speakers with no prior experience with tone languages.

The Current Study

In the current study we employ trial-by-trial feedback, and high-talker variability to examine the computational strategies that native English adult speakers employ while learning non-native Mandarin Tone category structures. Mandarin Chinese has four tone categories (ma¹ ‘mother’ [T1], ma² ‘hemp’ [T2], ma³ ‘horse’ [T3], ma⁴ ‘scold’ [T4]), described phonetically as high level, high rising, low dipping, and high falling, respectively (Fig. 2a). Native English speakers find it particularly difficult to learn tone categories (Wang, Jongman, & Sereno, 2003), but training can enhance tone identification and discrimination in native English speakers although such training paradigms have typically resulted in significant inter-individual differences in learning success (Perrachione, Lee, Ha, & Wong, 2011a).

Previous speech learning studies have typically relied on behavioral measures of accuracy to examine category learning. This is problematic since the same accuracy rate can often be achieved by using qualitatively different strategies (e.g., reflective or reflexive). We apply reflective and reflexive computational models to a) determine the extent to which speech category learning is mediated by reflective and reflexive learning, and b) examine the source of individual differences in category learning success. Our working hypothesis, consistent with dual-systems prediction is that successful learners' initially rely on the reflective learning system to perform talker-dependent analyses, and switch to the reflexive learning system by the end of training. We will expand on our predictions in the next section.

Tone category learning: category structure

A number of dimensions (e.g., pitch height, pitch direction) may serve as cues to tone identification. The perceptual saliency of these dimensions may be influenced by the presence of specific types of pitch patterns in a language's tonal inventory (Gandour, 1978; Gandour, 1983) as well as by the occurrence of abstract tonal rules in the listeners' phonological system (Hume & Johnson, 2001). Multidimensional scaling analysis of dissimilarity judgments, reaction time measures, and event-related responses, converges on two primary dimensions that underlie the tone space: pitch height and pitch direction (Fig. 2b).

Native speakers of Mandarin Chinese emphasize pitch direction more than pitch height. In contrast, English listeners place less emphasis on pitch direction in disambiguating tone categories. This is consistent with cue-weighting theories, that suggest that category learning difficulties in adults may be due to reduced emphasis on critical dimensions that are more invariant, and greater reliance on dimensions that are less invariant (Francis, et al., 2008). This is consistent with reasons attributed for learner difficulties with other speech categories. For example, one reason attributed to difficulties in learning the /l/ vs. /r/ category difference in Japanese listeners is reduced emphasis on the third formant, a critical dimension that differentiates the two categories (Hattori & Iverson, 2009; Iverson et al., 2003).

In Figure 3a, we plot the 80 stimuli used in our experiments (five consonant-vowel segments X four talkers X four tones) along two dimensions (pitch height: average fundamental frequency (x-axis) and pitch direction: slope (y-axis)). A visual inspection of this space suggests that this category structure is most likely information-integration (compare with Fig. 1b), and therefore most likely learned by the reflexive learning system. Of the two major acoustic dimensions, pitch height provides information about speaker identity and gender (a simple verbalizable strategy, e.g. male vs. female). Across languages, high pitch is typically aligned with a female talker, and a low pitch is aligned with a male talker. Therefore, in a high-variability training environment involving male and female speakers, the initial reflective strategy that learners may attempt is one that creates gender-dependent perceptual spaces. This is consistent with hypothesis from a previous study that showed gender-specific perceptual learning for category structures that are cued by dimensions that are also involved in processing sex-differences across speakers (Samuel & Kraljic, 2009). In line with this, we posit that separating the perceptual space on the basis of sex of the speaker (Fig. 3b, 3c), is a simple verbalizable strategy ("male" vs. "female") that listeners use. As

seen in Figures 3b and 3c, separating the perceptual space by sex of talker allows for significantly less overlap between tone categories, leading to less category confusion. Also, within the male and female perceptual spaces, there is little to distinguish between talkers (i.e., male talker 1 vs. male talker 2). Our working hypothesis is that learners' use a reflective strategy early in training to resolve talker variability, and a reflexive strategy later in training to differentiate tonal categories based on feedback.

Experiment 1: Introduction

In Experiment 1, we examine the extent to which learners use reflective and reflexive strategies while learning tone categories. Our prediction based on the dual-systems framework is that successful learners use a combination of reflective and reflexive strategies. Second, we predict that reflexive learning may be more optimal for successful categorization of tone categories. Using separate perceptual spaces (male vs. female) may be a good learning strategy to reduce tone confusion, but likely requires more working memory resource than a strategy that does not use separate perceptual space. In Experiment 2, we study the extent to which the two strategies (Talker Separation vs. Non-Separation) are mediated by working memory and executive attention resources. Specifically, we examine WMC differences between talker separators and non-separators. We predict that individuals with lower WMC will be less effective in using separate perceptual spaces, and therefore, less effective in tone category learning. The logic here is that separating perceptual space using a verbalizable strategy (male or female) is a reflective strategy that therefore requires working memory resources. We predict that individual differences in category learning success are related to differential strategy use between learners.

Experiment 1: Material and Methods

Model Details

In the next section, we describe the stimulus space used to examine strategy in this study. We will then describe the model fitting approach in detail.

Stimulus Characteristics—Stimuli consisted of natural native exemplars of the four Mandarin tones, tone 1 (T1), tone 2 (T2), tone 3 (T3), and tone 4 (T4). Monosyllabic Mandarin Chinese words (*bu*, *di*, *lu*, *ma*, and *mi*) that are minimally contrasted by the four tone categories were used in the experiment. Since these syllables exist in the American English inventory, this circumvents the need to learn phonetic structures additional to the tone distinction (Alexander, Wong, & Bradlow, 2005). By using different segments and multiple talkers, our aim is to expose learners to variability inherent in natural language. Each of these syllables was produced in citation form with the four Mandarin tones. Talkers in Experiment 1 consisted of two male and two female native speakers of Mandarin Chinese originally from Beijing. Two of these talkers (one male and one female) were used in Experiment 2. Stimuli were RMS amplitude and duration normalized (70 dB, 0.4 s) using the software Praat. Duration and amplitude envelope are potentially useful cues to disambiguate lexical tones. However, behavioral studies (Howie, 1976) as well as multidimensional scaling (MDS) analyses have shown that dimensions related to pitch (especially pitch height and pitch direction) are used primarily to distinguish tone categories

(Francis et al., 2008). In fact, phonetically, Mandarin tones 1–4 are described using these two dimensions as ‘high-level’, ‘low-rising’, ‘low-dipping’, and ‘high-falling’ respectively. Five native speakers of Mandarin were asked to identify the tone categories (they were given four choices) and rate their quality and naturalness. High identification (> 95%) was achieved across all 5 native speakers. Speakers rated these stimuli as highly natural.

Model Fitting Approach—We fit each model at the individual participant level because of problems with interpreting fits to aggregate data (Estes, 1956, Ashby, Maddox & Lee, 1994, Maddox, 1999). We fit each model to a block of 80 trials assuming a fixed decision strategy on each trial within the block. We fit three classes of models with multiple instantiations possible within a class. The first class is computational models of the reflexive procedural learning system. This is instantiated with the Striatal Pattern Classifier (SPC) (Maddox, Molis, & Diehl, 2002). The SPC is a computational model whose processing is consistent with what is known about the neurobiology of the procedural-based category learning system thought to underlie information-integration classification performance (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Ennis, 2006; Maddox, et al., 2002; Nomura, et al., 2007; Seger & Cincotta, 2005). The second class is reflective rule-based and instantiate hypothesis-testing strategies such as the application of uni-dimensional or conjunctive rules. These are verbalizable strategies. The third model is a random responder model that assumes that the participant guesses on each trial. The model parameters will be estimated using maximum likelihood procedures (Ashby, 1992; Wickens, 1982) and models will be compared using Akaike weights (Wagenmakers & Farrell, 2004) as described in detail in the Results section. We provide the specifics of each model in the next section.

Striatal Pattern Classifier—The SPC assumes that stimuli are represented perceptually in higher level auditory areas, such as the superior temporal gyrus. Because of the massive many-to-one (approximately 10,000-to-1) convergence of afferents from the primary and secondary auditory cortices to the striatum (Ashby & Ennis, 2006; Wilson, 1995), a low-resolution map of perceptual space is represented among the *striatal units*. Within the auditory domain it is well known that there are direct projections from secondary auditory areas such as superior temporal gyrus and supratemporal plane to the caudate (Arnauld, Jeantet, Arsaut, & Desmotes-Mainard, 1996; Hikosaka, Sakamoto, & Usui, 1989; Yeterian & Pandya, 1998). During learning the striatal units become associated with one of the category labels, so that, after learning is complete, a category response label is associated with each of a number of different regions of perceptual space. In effect, the striatum learns to associate a response with clumps of cells in the auditory cortex². The SPC assumes that there is one striatal “unit” in the pitch height-pitch direction space for each category, yielding a total of four striatal units. Because the location of one of the units can be fixed, and since a uniform expansion or contraction of the space will not affect the location of the

²It is important to be clear that the SPC is a computational model that is inspired by what is known about the neurobiology of the striatum. Because of this fact, the striatal “units” are hypothetical and could be interpreted within the language of other computational models (e.g., as “prototypes” in a multiple prototype model like SUSTAIN; Love, Medin, & Gureckis, 2004). In addition, we do not model learning in the SPC in the sense that we do not update association weights between units and category labels. Learning models have been proposed (Ashby & Maddox, 2011), but are not utilized here due to their complexity.

resulting response region partitions, the SPC contains six free parameters--five that determine the location of the units, and one that represents the noise associated with the placement of the striatal units. Figure 4a displays a scatterplot of the responses and response regions for the four tone categories in Figure 3a generated from a version of the Striatal Pattern Classifier. It is worth mentioning that versions of the SPC have been previously applied in the auditory domain. Specifically, Maddox, Ing and Lauritzen (2006) applied the model to data from an artificial auditory category learning task, and Maddox, Molis and Diehl (2002) applied the model to data from an auditory vowel categorization task.

Conjunctive Rule-Based Model—A conjunctive rule-based model that assumes that the participant sets two criteria along the pitch height dimension and one criterion along the pitch direction dimension was also applied to the data. The model assumes that the two criteria along the pitch height dimension are used to separate the stimuli into those that are of low, medium or high pitch height. Low pitch height items are classified into tone category 3 (T3) and high pitch height items are classified into tone category 1 (T1). If an item is classified as of medium pitch height then the pitch direction dimension is examined. The single criterion along the pitch direction dimension is used to separate the stimuli into low and high pitch direction. Stimuli that have medium pitch height and low pitch direction (negative slope) are classified into tone category 4 (T4) and medium pitch height items of high pitch direction are classified into tone category 2 (T2). Figure 4b displays a scatterplot of the responses and response regions for the four tone categories in Figure 3a generated from a version of the Conjunctive model. This model contains four free parameters—three criteria and one noise parameter.

Uni-Dimensional Rule-Based Model—A uni-dimensional_height rule-based model that assumes that the participant sets three criteria along the pitch height dimension was also applied to the data. The model assumes that the three criteria along the pitch height dimension are used to separate the stimuli into those that are of low, medium-low, medium-high or high pitch height with each of these being associated with one of the tone categories. Notice that this model completely ignores the pitch direction dimension. Although 24 versions of the model are possible given four category labels, some are highly unrealistic (e.g., a model that assumes that tone category 1 (T1) was the lowest in pitch height). We examined the eight most reasonable variants of the model.

A uni-dimensional_direction rule-based model that assumes that the participant sets three criteria along the pitch direction dimension was also applied to the data. The model assumes that the three criteria along the pitch direction dimension are used to separate the stimuli into those that are of low, medium-low, medium-high or high pitch direction with each of these being associated with one of the tone categories. Notice that this model completely ignores the pitch height dimension. Although 24 versions of the model are possible given four category labels, many are highly unrealistic. We examined the two most reasonable variants of the model. Figure 4c displays a scatterplot of the responses and response regions for the four tone categories in Figure 3a generated from a version of the Uni-Dimensional_Height model, and Figure 4d displays a scatterplot of the responses and response regions for the four tone categories in Figure 3a generated from a version of the Uni-

Dimensional_Direction model. The uni-dimensional models each contain four free parameters—three criteria and one noise parameter.

Random Responder Model—The random responder model assumes a fixed probability of responding tone 1, tone 2, tone 3, and tone 4 but allows for response biases. The model has three free parameters to denote the predicted probability of responding “1,” “2,” or “3” with the probability of responding “4” equal to one minus the sum for the other three.

Modeling Talker Separation—An a priori prediction was that learners use a verbalizable strategy (talker sex) to separate out male and female perceptual spaces. We predicted that using separate perceptual spaces reduces overlap between categories and increases successful learning of the tone categories. The model procedure assumes that each model applied to a block of 80 trials using the 80 stimuli displayed in Figure 3a is a modeling procedure that assumes no talker separation (referred to as Non-Separation models).

To model the presence of talker separation (referred to as Separation models), we assume that the participant converted the 80 stimuli perceptual space in Figure 3a into two separate perceptual spaces, one that characterizes the 40 stimuli spoken by male talkers and one that characterizes the 40 stimuli spoken by female talkers. A scatterplot of the stimuli associated with the male and female sub-perceptual spaces are displayed in Figures 3b and 3c.

We fit each of the models outlined above (SPC, Conjunctive, Uni-Dimensional_Height, Uni-Dimensional_Direction, and Random Responder) separately to the 40 trials with a female speaker and the 40 trials with the male speaker and estimated separate parameters for each of the relevant perceptual spaces (male or female). For example, whereas the Conjunctive model required four parameters when no talker separation was assumed, it required eight parameters when talker separation was assumed.

Experiment 1: Methods

Participants

Twenty-four participants were recruited from the undergraduate population at the University of Texas at Austin (age range 18–35 years) and were compensated \$10 per hour for their participation. Participants were monolingual, raised in monolingual English households, as reported in detailed background questionnaires. Participants with significant exposure to another language before age 12 were not included. No participant reported prior exposure to a tone language. Participants reported no history of neurological, visual, or hearing deficits. A detailed language history questionnaire was used to ensure that participants had no previous exposure to a tone language. All participants were asked to provide informed consent and were debriefed following the conclusion of the experiment consistent with UT IRB requirements. Participants ($n=2$) with negative learning slopes (i.e., poorer performance in the final block relative to the first block), were excluded leaving 22 participants' data to include in the statistical analyses. Average accuracy in the final block for the excluded participants was 29%, which is close to chance-level performance (25%)

Tone category training procedure

On each of multiple trials, participants were presented with single exemplars from one of four Mandarin tone categories (T1, T2, T3, or T4) and instructed to categorize these stimuli into one of four categories. Participants were given feedback on each trial and exposed to multiple talkers throughout the training program. Participants listened to 80 stimuli per block (four tone categories X five syllables X four talkers). Within a block, the talkers were randomized. Each participant completed six 80-trial blocks of training and was instructed to categorize sounds into four equally likely categories. Further, participants were instructed that high accuracy levels are possible. Participants generated a response by pressing one of four buttons on a response box, labeled “1,” “2,” “3,” or “4.” Corrective feedback was provided for 1 s on the screen immediately following the button press and consisted of the word “Correct” or “Error” followed by the label of the tone that was actually presented. For example, on a correct tone 2 (T2) trial the feedback display was as follows: “Correct that was a 2.” On an incorrect response trial where tone 3 was the correct response the feedback display was as follows: “Error that was a 3”. A 1s ITI followed the feedback.

Experiment 1: Results

Overall Accuracy Results

Figure 5a displays the average accuracy along with standard error bars. We also include average accuracy for participants best fit by a Separation or Non-Separation model (discussed below). We conducted a two-way ANOVA (block X tone category) and found significant learning across blocks [$F(5, 95) = 23.98, p < 0.001, \text{partial } \eta^2 = .56$] with average accuracy increasing from 42% in block 1 (b1) to 71% in block 6 (b6). We also found a main effect of tone category [$F(3, 57) = 4.51, p = 0.007, \text{partial } \eta^2 = .19$]. Average accuracy was greatest for T3 (69%); pair-wise comparison showed significantly better learning for T3 stimuli, relative to T2 ($p = 0.004$) and T4 ($p = 0.007$). No other pair-wise comparison was significant. The interaction between tone category and block was not significant [$F(15, 285) = 1.11, p = 0.35, \text{partial } \eta^2 = .05$], suggesting similar learning patterns across the four tones. A similar pattern held for tones produced by female and male talkers. Specifically, average accuracy increased from 44% in block 1 to 72% in block 6 for tones spoken by female talkers and increased from 41% in block 1 to 69% in block 6 for tones spoken by male talkers.

Modeling Results

Model Fitting and Model Comparison

As outlined above, each model was fit to the data from each participant on a block-by-block basis. The models were fit to the Mandarin tone category learning data from each trial by maximizing negative log-likelihood. We used Akaike weights to compare the relative fit of each model (Akaike, 1974, Wagenmakers & Farrell, 2004). Akaike weights are derived from Akaike’s Information Criterion (AIC), which is used to compare models with different numbers of free parameters. AIC penalizes models with more free parameters. For each model, i , AIC is defined as:

$$AIC_i = -2\log L_i + 2V_i \quad (1)$$

where L_i is the maximum likelihood for model i , and V_i is the number of free parameters in the model. Smaller AIC values indicate a better fit to the data. We first computed AIC values for each model and for each participant's data in each block. Akaike weights were then calculated to obtain a continuous measure of goodness-of-fit. A difference score is computed by subtracting the AIC of the best fitting model for each data set from the AIC of each model for the same data set:

$$\Delta_i(AIC) = AIC_i - \min AIC \quad (2)$$

From the differences in AIC we then computed the relative likelihood, L , of each model, i , with the transform:

$$L(M_i|data) \propto \exp \left\{ -\frac{1}{2} \Delta_i(AIC) \right\} \quad (3)$$

Finally, the relative model likelihoods are normalized by dividing the likelihood for each model by the sum of the likelihoods for all models. This yields Akaike weights:

$$w_i(AIC) = \frac{\exp \left\{ -\frac{1}{2} \Delta_i(AIC) \right\}}{\sum_k \exp \left\{ -\frac{1}{2} \Delta_k(AIC) \right\}} \quad (4)$$

These weights can be interpreted as the probability that the model is the best model given the data set and the set of candidate models (Wagenmakers & Farrell, 2004). Akaike weights range from 0 to 1.0 with an Akaike weight of 0 implying that the given model is the best model with probability 0, and an Akaike weight of 1 implying that the given model is the best model with probability 1.0. Equivocal evidence in support of a given model is associated with an Akaike weight of $1/n$ where n denotes the number of models being compared (e.g., with two models, an Akaike weight of 0.5 implies equivocal support for the given model).

Best Fitting Model vs. Random Responder Model—We began by comparing the Akaike weights from the best fitting uni-dimensional, conjunctive or SPC model that assumed Non-Separation or Separation with the best fitting Random Responder model. This comparison allowed us to determine whether the best fitting model is capturing noise or is capturing meaningful strategic responding. The results were clear. The resulting Akaike weights were .964, .990, .946, .956, .991, and .991 in blocks 1 – 6, respectively. In every case these values were significantly above 0.5 based on a one-sample t-test (all p 's < .0001) which denotes that the best fitting models are effectively fitting the data.

Best Fitting Non-Separation Model vs. Best Fitting Separation Model—Next we compared the Akaike weights from the best fitting Separation model against the best fitting

Non-Separation model. This comparison allows us to determine whether the best fitting model is truly capturing additional strategic responding or just more noise. Again the results were clear. When a Separation model provided the best account of the data, the Akaike weights ranged from .880 – .982 and in every block were significantly above 0.5 based on a one-sample t-test (all p 's < .001). When a Non-Separation model provided the best account of the data, the Akaike weights ranged from .893 – .982 and in every block were significantly above 0.5 based on a one-sample t-test (all p 's < .001). These findings suggest that the best fitting model (separation or non-separation) is capturing meaningful strategic variance in the data and not just random noise.

Distribution of Best Fitting Non-Separation and Separation Model—Because Talker Separation is hypothesized to improve performance, we predicted that the number of participants whose data is best fit by one of the Separation models will increase with experience relative to the number of participants whose data is best fit by one of the Non-Separation models. Figure 6 displays the proportion of participants whose data was best fit by a Separation or Non-Separation model in each block. As a formal test of our hypothesis, we compared the number of Separators and Non-Separators across the first and final block. A χ^2 test suggested that the number of Separators increased while the number of Non-Separators decreased from the first to the final block of trials [χ^2 (1, $N = 22$) = 5.94, $p < .005$].

Separation Strategy Distribution For Final Block Separators and Final Block Non-Separators—Another way to examine changes in the use of Separation strategies across blocks is to compare the number of blocks of trials best fit by the Separation model for participants whose final block of trials is best fit by a Separation (hereafter referred to as final block Separators) or Non-Separation (hereafter referred to as final block Non-Separators) model. We hypothesized that participants whose data is best fit by a Separation model in the final block of trials will be more likely to use Separation strategies earlier in learning as well. The results supported our prediction (see Figure 7a) with significantly more blocks of trials being best fit by a Separation model for final block Separators (4.9 blocks) than for final block Non-Separators (1.0 blocks) [$F(1, 20) = 15.449$, $p < .001$, partial $\eta^2 = .436$].

We also examined the first block of trials for which a Separation model provided the best fit to the data for final block Separators versus final block Non-Separators. We hypothesized that final block Separators will begin to talker-normalize sooner. The results supported our prediction (Figure 7b) with final block Separators talker-normalizing earlier (1.65 blocks) than Non-Separators (4.50 blocks) [$F(1, 20) = 7.20$, $p < .005$, partial $\eta^2 = .265$].

Learning Curves for Final Block Separators and Final Block Non-Separators—Figure 5a displays the learning curves for final block Separators and final block Non-Separators. A 2 model strategy x 6 block mixed ANOVA was conducted on these data. We observed a main effect of model strategy [$F(1, 20) = 6.70$, $p < .05$, partial $\eta^2 = .251$] with performance being significantly better for Separators (.65) than for Non-Separators (.27). We observed a main effect of block [$F(5, 100) = 3.83$, $p < .01$, partial $\eta^2 = .161$] suggesting significant learning. We also observed a significant model strategy by block interaction [$F(5,$

100) = 2.59, $p < .05$, partial $\eta^2 = .151$]. The interaction is characterized by significant learning in the Separator group [$F(5, 95) = 28.77$, $p < .001$, partial $\eta^2 = .602$], and non-significant learning in the Non-Separators [$F(5, 5) < 1.0$].

Reflective and Reflexive Strategies and Accuracy Rates for Final Block

Separators—Here we examine performance for the reflective and reflexive strategies used by final block Separators. Of the 20 final block Separators, 10 were best fit by the SPC, 2 by the Conjunctive rule-based model, and 8 by the Uni-dimensional_Height model. Because the optimal strategy requires application of a reflexive strategy, we predicted that reflexive participants would outperform reflective participants. To test this, we compared overall accuracy across the 10 reflexive separators and the 10 reflective separators. The effect of strategy was significant [$F(1, 18) = 9.44$, $p < .01$, partial $\eta^2 = .344$] with participants using a reflexive strategy (.762) outperforming those using a reflective strategy (.532).

Experiment 1: Discussion

Experiment 1 examined Mandarin tone category learning in native English speakers. A series of computational models were applied that are derived from a dual-systems model of visual category learning. The models capture two aspects of learning that are hypothesized to be critical to a complete understanding of Mandarin tone category learning. First, the models capture the distinction between reflective category learning strategies that are available to conscious awareness and require working memory and executive attention, and reflexive category learning strategies that are not available to conscious awareness and operate without relying on working memory and executive attention (Ashby et al., 1998, Ashby & Maddox, 2011). This distinction is modeled by placing constraints on the nature of the participant's decision process (see Figure 4). Second, the models capture talker-dependent strategies. The lack of talker dependencies is modeled by assuming that the participant generates categorization responses from a perceptual space that makes no distinction between talkers (see Figure 3a). Talker Separation, on the other hand, is modeled by assuming that the participant first determines whether the talker is male or female and then generates categorization responses from the relevant male or female pitch height-pitch direction perceptual space (see Figures 3b and 3c).

Several results emerged from Experiment 1. Behaviorally, significant learning occurred within a session of training. Importantly, learning across blocks was similar for all four tonal categories and did not differ between male and female talkers. We found a significant main effect of tone, driven by the fact that T3 was easier to learn, relative to T2 and T4. This is an interesting finding because T3 is by far the most distinct category for native English speakers. T2 and T4 can be mapped on to existing intonational categories (rising and falling pitch are important cues in intonational processing, for e.g., cueing the difference between a question and a statement). This finding is consistent with predictions made by the Speech Learning Model (SLM) (Flege, 1995; Flege, 1999). Specifically, SLM predicts that a native category that is similar to a non-native, but not identical, may interfere with processing and learning of the non-native category. By this account, T2 and T4 categories receive interference from existing intonational categories; in contrast T3 is learned better because there is no interference from existing category representation.

The basic computational modeling approach received strong validation from the data. Akaike weights that can be interpreted as the probability that a particular model is the best model given the data set and the set of candidate models were very large for the best fitting model. When the best fitting model was compared with a Random Responder model all Akaike weights were larger than .946. In addition, when the best fitting model assumed Talker Separation, the Akaike weights for this model compared with the best fitting Non-Separation model were all larger than .880. When the best fitting model assumed no Talker Separation, the Akaike weights for this model compared with the best fitting Separation model were all larger than .893. Second, we found that the number of participants whose data was best fit by a Separation model increased across blocks suggesting that the prevalence of Talker Separation increased with experience. Third, we found that those participants whose final block of data was best fit by a Separation model were more likely to use talker-dependent strategy in other blocks and showed Talker Separation earlier in learning than those participants whose final block of data was best fit by a Non-Separation model. Finally, amongst final block Separators we found that those who used a reflexive strategy were more accurate than those who used a reflective strategy. Taken together, these results provide strong validation of our modeling approach and demonstrate its usefulness.

Experiment 2: Introduction

In Experiment 2, we conduct a large-scale replication of Experiment 1 and we extend the study by collecting measures of working memory capacity. We hypothesized that individuals using separate perceptual spaces will be more effective in learning tone categories. Second, we predicted that learners who use separate perceptual spaces (Separators) will show higher WMC than individuals who do not use separate perceptual spaces (Non-Separators). This prediction is very similar to one proposed by Tagarelli, Mota & Rebuschat (2001) who showed that working memory capacity correlated with performance on an explicit, but not an implicit artificial language task (Tagarelli, Borges-Mota, & Rebuschat, 2011) and earlier work by Reber and colleagues that demonstrate that aptitude measures influence explicit learning processes, but not implicit learning (Reber, Walkenfeld, & Hernstadt, 1991). Similarly, a recent neuroimaging study showed that in an artificial grammar learning task, individual differences in working memory predicted performance in participants who learned the task explicitly, but not in participants who learned the task implicitly (Yang & Li, 2012).

Experiment 2: Methods

Participants

Ninety-eight monolingual participants were recruited from the undergraduate population at the University of Texas (age range 18–35 years) and were compensated \$10 per hour for their participation. No participant reported prior exposure to a tone language. Participants reported no history of neurological or hearing deficits. All participants provided informed consent and were debriefed following the conclusion of the experiment consistent with UT IRB requirements. Participants ($n=16$) with negative learning slopes (i.e., poorer performance in the final block relative to the first block), were excluded leaving 82

participants' data to include in the statistical analyses. Average accuracy in the final block for these participants was 30%, which is close to chance (chance performance=25%).

Stimuli for tone category training

Stimuli were identical to those from Experiment 1 with the exception that only one male and one female native speaker of Mandarin Chinese were included. This resulted in a shorter training experiment that allowed a working memory measure to be included.

Procedure

The procedure was identical to that from Experiment 1 except that participants listened to 40 stimuli per block (four tone categories X five syllables X two talkers) across five blocks of training. In addition, following completion of the tone category learning task, each participant completed the immediate recall portion of the logical memory test in the Wechsler Memory Scale, 3rd edition (WMS-III) (Wechsler, 1997). In this test, two stories were read at conversational rate. One story was read only once, the other twice. Participants were told to listen to the stories and repeat as much as they can recall immediately after each story was read, for a total of three story recall opportunities. The total number of key phrases or words correctly recalled across stories served as their raw score for the task.

Experiment 2: Results

Overall Accuracy Results

Figure 5b displays the average accuracy along with standard error bars. We also include average accuracy for participants best fit by a Separation or Non-Separation model (discussed below). Participants showed significant learning across blocks [$F(4, 324) = 133.59, p < .0001, \text{partial } \eta^2 = .623$] with accuracy increasing from 40% in block 1 to 74% in block 5. A similar pattern held for tone spoken by female and male talkers. Specifically, average accuracy increased from 40% in block 1 to 74% in block 5 for tones spoken by female talkers and increased from 40% in block 1 to 73% in block 5 for tones spoken by male talkers³.

Modeling Results

The Experiment 1 model fitting and model comparison approach was used.

Best Fitting Model vs. Random Responder Model—First we compared the Akaike weights from the best fitting model and compared that with the fit of the Random Responder model to determine whether the best fitting model is capturing noise or is capturing meaningful strategic responding. The resulting average Akaike weights were .944, .975, .990, .994, and .993 in blocks 1 – 5, respectively. In every case these values were significantly above 0.5 based on a one-sample t-test (all p 's < .0001).

³It is worth noting that performance was very similar across the two experiments despite the fact that four speakers were included (with 80 stimuli per training block) in Experiment 1 and only two speakers were included (with 40 stimuli per training block) in Experiment 2. In Experiment 1, two male and two female speakers were included (see Figure 3) whereas in Experiment 2, one male and one female speaker were included. Importantly, the perceptual features (i.e., pitch height and pitch direction) were very similar across the two male and across the two female speakers. This likely explains why performance was similar across the two studies.

Best Fitting Non-Separation Model vs. Best Fitting Separation Model—Next we compared the Akaike weights from the best fitting Separation model against the best fitting Non-Separation model. When a Separation model provided the best account of the data, the Akaike weights ranged from .908 – .944 and in every block were significantly above 0.5 (all p 's < .001). When a Non-Separation model provided the best account of the data, the Akaike weights ranged from .775 – .853 and in every block were significantly above 0.5 based on a one-sample t-test (all p 's < .01). These findings suggest that the models are capturing meaningful strategic variance in the data.

Distribution of Best Fitting Non-Separation and Separation Model—Here we test the hypothesis that Talker Separation leads to improved performance and thus that the number of participants best fit by one of the Separation models will increase with experience. Figure 6b displays the proportion of participants whose data was best fit by a Separation or Non-Separation model in each block. In support of our hypothesis, a χ^2 test suggested that the number of Separators increased while the number of Non-Separators decreased from the first to the final block of trials [$\chi^2(1, N=82) = 22.61, p < .00001$].

Separation Strategy Distribution For Final Block Separators and Final Block Non-Separators—Here we examine changes in the use of Separation strategies across blocks by comparing the number of blocks of trials best fit by the Separation model for final block Separators and final block Non-Separators. As predicted (see Figure 7c) final block Separators (3.84 blocks) used Separation strategies in significantly more blocks than final block Non-Separators (1.58 blocks) [$F(1, 80) = 56.90, p < .001$, partial $\eta^2 = .416$].

We also examined the first block of trials for which a Separation model provided the best fit to the data for final block Separators and Non-Separators. As predicted (see Figure 7d) final block Separators began using separate perceptual spaces earlier (1.95 blocks) than final block Non-Separators (2.68 blocks) [$F(1, 80) = 5.09, p < .05$, partial $\eta^2 = .060$].

Learning Curves for Final Block Separators and Final Block Non-Separators—Figure 5b displays the learning curves for final block Separators and final block Non-Separators. A 2 model strategy $\times$ 5 block mixed ANOVA was conducted on these data. We observed a main effect of model strategy [$F(1, 80) = 19.34, p < .001$, partial $\eta^2 = .195$] with performance being significantly better for Separators (.67) than for Non-Separators (.44). We observed a main effect of block [$F(4, 320) = 78.28, p < .001$, partial $\eta^2 = .495$] suggesting significant learning. We also observed a significant model strategy by block interaction [$F(4, 320) = 4.33, p < .005$, partial $\eta^2 = .051$]. Post hoc analyses suggested that performance in the Separators was superior to that for the Non-Separators in every block (all p 's < .05). Both groups showed significant learning, but a comparison of performance in block 1 with block 5 suggested that Separators showed greater learning (an increase of .37) than Non-Separators (an increase of .22) [$F(1, 80) = 8.64, p < .005$, partial $\eta^2 = .097$].

Working Memory Capacity and Talker Separation—As outlined above, we expect that individuals who use Talker Separation strategies will be more likely to have high working memory capacity. As a test of this hypothesis we compared the WMS-III scores for final block Separators and final block Non-Separators. As predicted final block Separators

remembered significantly more story items (mean = 43.22; standard error = 1.21) than Non-Separators (mean = 37.95; standard error = 1.79) as measured by the WMS-III [$F(1, 79) = 4.80$, $p < .05$, partial $\eta^2 = .057$].

Experiment 2: Discussion

In Experiment 1, we examined the extent to which reflective and reflexive strategies were utilized in tone category learning across different blocks. We showed that learners who used a reflexive strategy at the end of training were more accurate in categorization relative to those who used reflective strategies. Further, our data demonstrated that a majority of learners use talker-dependent strategies to deal with the multi-talker variability in the training paradigm. The extent to which this strategy is beneficial in category learning success, and whether this strategy is indeed dependent on working memory (and therefore, reflective) was not addressed in Experiment 1. In Experiment 2, we use a larger sample, and examine working memory ability to address both these issues. Our results show that participants who use the Talker Separation strategy showed enhanced learning relative to participants who did not use a Talker Separation strategy. Moreover, participants who did not use a Talker Separation strategy showed lower working memory capacity relative to participants who used a Talker Separation strategy. Experiment 2 thus shows that reflective strategy use is an important determiner of success in the category learning task.

General Discussion

In this study we provide a computational account of strategies that adults use while learning non-native speech sound categories using a dual-systems framework. We examined learning of Mandarin tone categories by native English speakers with no prior experience with tone languages across two experiments. In both experiments, participants showed learning across training blocks. Our results demonstrate significant individual differences in the computational strategies participants use during category learning. Importantly, these differences in strategies have a direct bearing on category learning success.

We hypothesized that tone category learning is reflexive-optimal, but learners may use reflective, speaker-dependent strategies early in training. Consistent with these predictions, we demonstrate in Experiment 1 that learners' who used reflexive strategies by the end of training were more accurate in category learning than those who used reflective learning strategies. We show that successful learners' initially separate perceptual spaces based on a reflective strategy ('male/female') to perform category-related computations. Successful learners then transition to a reflexive computational strategy (SPC) by the end of the training session. As per the dual-systems model, working memory capacity is a critical component of the reflective learning system. Reflective learning is resource intensive, requiring working memory and executive attention to develop, test, and update verbal rules. We predicted that WMC may therefore be an important determiner of individual differences in category learning success. Indeed, in Experiment 2 we demonstrate that individuals who use the verbalizable talker-dependent learning strategy also show enhanced learning and have higher WMC, as measured by standardized tests of working memory.

The extent to which L2 learning in adulthood is explicit versus implicit has been a significant topic of inquiry in SLA literature (DeKeyser, 2008). Some argue that optimal L2 learning is driven by implicit processes (Krashen, 1982), while others argue for a critical role for attentive processes (Robinson, 2008; Schmidt, 1995). Applied in the L2 context, the DP model predicts that L2 learners are more declarative during language learning (even for grammar processing) (Ullman, 2001). With increased competence, they may switch to the more automatic, slowly learning procedural system. Much of this work has focused on grammar, morphology, and vocabulary learning. In contrast, much less is known about the role of the explicit/implicit processes during phonetic learning. Speech sounds are typically difficult to verbalize and involve multiple, redundant dimensions. These features argue for a dominance of the procedural-based reflexive learning system. Indeed, our results show that participants, who use reflexive strategies at the end of training, have higher learning accuracies, relative to those who use reflective strategies. Yet, some reflective strategies are advantageous in certain situations. For example, in multi-talker training paradigms (Lively et al., 1994) talker variability has been shown to be important for generalizing categories to novel talkers. Indeed, for speech learning to be effective in the real world, categorization performance needs to transfer to novel talkers. However, a previous tone category learning study showed that not all learners benefit from high talker variability (Perrachione, Lee, Ha, & Wong, 2011b). This study argued that some learners may find high variability training too resource intensive, and therefore may be less successful. The reasons why some learners have difficulty in such high variability environment is unclear. Our results show large individual differences in strategies that learners use to deal with talker-variability. Tone categories are cued by speaker-dependent and speaker-independent dimensions. We predicted that an initial reflective strategy is to create speaker-dependent perceptual spaces. Our computational modeling results show that reduced effectiveness in utilizing this strategy, results in substantially lower category learning success. While a majority of participants use separate perceptual space for male and female talkers, a smaller group of participants operate on a single perceptual space that encompasses male and female talkers. Across both experiments, our results demonstrate that learning is enhanced when participants use separate perceptual spaces. However, this is a reflective strategy, and therefore requires working memory resources. Thus individuals with low WMC may have difficulty maintaining separate perceptual spaces for talkers. Individuals who show high WMC, on the other hand, have an advantage in learning, a finding that addresses a source of individual differences in category learning success. Our results suggest a trade-off in category learning. Multiple-talker training benefits real-world generalization, but requires working memory resources, and thus may lead to individual differences in learning success.

Previous studies have focused on perceptual differences as a source of individual variability in speech learning success. For example, neuroanatomical and neurophysiological studies show that pre-training differences in auditory regions is a significant predictor of individual differences in successful speech learning (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007; Wong, Chandrasekaran, Garibaldi, & Wong, 2011; Wong, Perrachione, & Parrish, 2007). Individual differences in phonetic acquisition under natural learning situations were found to relate to neural processing of speech sounds, as evidenced by preattentive change-detection electrophysiological responses, but not basic psychoacoustic differences between

learners (Diaz, Baus, Escera, Costa, & Sebastian-Galles, 2008). Impressively, functional connectivity between frontal and parietal brain regions at rest, also relates to individual differences in speech learning success (Ventura-Campos et al., 2013). Our results show that individual differences in computational strategies may further contribute to variability in speech learning success.

From a mechanistic perspective, previous speech learning studies have primarily focused on the perceptual processes mediating speech learning. Typically, these studies have viewed speech learning from the perspective of unsupervised learning (McClelland, et al., 2002; Vallabha & McClelland, 2007) or interactive (lexically-mediated), supervised learning mechanisms (McClelland, Mirman, & Holt, 2006; Mirman, McClelland, & Holt, 2006). The role of feedback-based multiple learning systems, known to play an important role in feedback-based visual category learning are unclear. Our theoretical approach argues for two competing learning systems operative when feedback is provided; a reflective learning system that uses feedback to develop rules; and an implicit, reflexive learning system that unconsciously associates feedback with stimulus and reward. Our computational approach demonstrates that a predictor of learning success is the computational strategies that participants employ while learning category differences. Thus, our framework views L2 speech learning as a category learning problem (Lotto, 2000). That is, with perceptual monitoring and feedback, what strategies do learners employ to make decisions on category membership? Our data show that these strategies are indeed an important contributor to inter-individual differences in learning success.

We are aware that our computational approach is a limited beginning to understanding dual-learning system influences on L2 speech learning. Speech categories are extensively multidimensional (Diehl, Lotto, & Holt, 2004), a factor that distinguishes speech learning from visual category learning (Lotto, 2000). The two-dimensional perceptual space (pitch height/direction) that is derived likely underestimates the inherent complexity in listeners' processing of linguistic tone. However, our model fits suggest that this approach is a good beginning to examine strategy use during speech learning. Future experiments need to build on complexity to allow a better understanding of real-world speech processing. Second, we use Mandarin tone categories to evaluate speech learning. Previous research suggests that several similarities between segmental and suprasegmental speech learning. From a neural perspective, the extent to which lexical tones behave like segmental information (vowels and consonants) has been extensively studied (Gandour & Dardarananda, 1983; Gandour et al., 2003; Gandour et al., 2000; Gandour, Wong, & Hutchins, 1998; Klein, Zatorre, Milner, & Zhao, 2001; Wong, Parsons, Martinez, & Diehl, 2004b; Xu, Gandour, & Francis, 2006). For the purpose of this paper, three main points are particularly relevant. First, native speakers use a left hemisphere dominant network to process linguistic tones that is indistinguishable from the network used to process other speech sound categories (Gandour & Dardarananda, 1983; Gandour, et al., 1998; Wong, et al., 2004b). Second, when non-native participants learn lexical tone categories, there is increased activity in left-hemisphere anterior and posterior language regions (Wang, Sereno, Jongman, & Hirsch, 2003b). Third, native listeners process Mandarin tones categorically, in a manner similar to consonants (Xi, Zhang, Shu, Zhang, & Li, 2010). While this suggests a certain degree of constancy in the processing of linguistically 'tainted' stimuli, the extent to which our dual-systems approach

can be applied to segmental learning is unclear. The SPC model has been successfully applied to examine vowel categorization (Maddox, et al., 2002), suggesting likely generalization to segmental learning. The applicability of our approach to learning segmental information is an important direction for future research. Finally, our data suggests that successful learners use separate perceptual spaces for male and female talkers to deal with high variability across talkers. We argue that this strategy is an explicit process because participants who use this strategy have higher working memory ability, relative to participants who do not use this strategy. However, many studies have shown that normalizing for talker differences is an automatic process (Holt, 2006; Huang & Holt, 2012; Laing, Liu, Lotto, & Holt, 2012), reliant on general auditory processes (e.g. neural adaptation to the long-term average spectrum of speech) rather than cognitive effort. We reconcile these differences with the fact that processing talker information while categorizing L1 speech sounds may be fundamentally different from the role of talker information during the learning of L2 speech categories. Initial learning of categories in a high talker variability environment may be an effortful process requiring talker-dependent analysis. Indeed, a number of studies demonstrate a processing cost in multi-talker paradigms, as well as showing more cognitive effort in mixed-talker presentations relative to blocked-talker presentations (Wong, et al., 2004a). However, with practice and a switch to reflexive analysis within the perceptual space, such effortful analysis may not be necessary. Although the current experiments were not designed to distinguish between various models of talker normalization, we believe that the computational modeling approach developed here could potentially contribute to this topic of research.

In conclusion, we used a computational modeling approach to examine the role of the dual-learning systems during speech category learning. Our results demonstrate that learners use a variety of reflective and reflexive strategies to learn new categories. Importantly, we demonstrate that category learning success is dependent on the computational strategy listener's use. Successful learners use a combination of reflective and reflexive strategies and are likely to be more reflexive with practice. The computational modeling approach developed here provides a foundation for future work on speech learning and perception.

Acknowledgments

This research was supported by NIMH grants MH077708 and DA032457 to WTM. We thank the Maddox Lab RA's for data collection.

References

- Alexander, JA.; Wong, PCM.; Bradlow, AR. Lexical tone perception in musicians and non-musicians. Paper presented at the INTERSPEECH-2005; 2005.
- Arnould E, Jeantet Y, Arsaut J, Desmotes-Mainard J. Involvement of the caudal striatum in auditory processing: c-fos response to cortical application of picrotoxin and to auditory stimulation. *Molecular Brain Research*. 1996; 41:27–35. [PubMed: 8883931]
- Ashby FG. Multivariate probability distributions. 1992
- Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychol Rev*. 1998; 105(3):442–481. [PubMed: 9697427]
- Ashby FG, Ell SW. The neurobiology of human category learning. *Trends Cogn Sci*. 2001; 5(5):204–210. [PubMed: 11323265]

- Ashby FG, Ennis JM. The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*. 2006; 46:1–36.
- Ashby FG, Maddox WT. Human category learning. *Annu Rev Psychol*. 2005; 56:149–178. [PubMed: 15709932]
- Ashby FG, Maddox WT. Human category learning 2.0. *Ann N Y Acad Sci*. 2011; 1224:147–161. [PubMed: 21182535]
- Ashby FG, Maddox WT, Bohil CJ. Observational versus feedback training in rule-based and information-integration category learning. *Mem Cognit*. 2002; 30(5):666–677.
- Ashby FG, O'Brien JB. Category learning and multiple memory systems. *Trends Cogn Sci*. 2005; 9(2):83–89. [PubMed: 15668101]
- Ashby FG, Spiering BJ. The neurobiology of category learning. *Behav Cogn Neurosci Rev*. 2004; 3(2):101–113. [PubMed: 15537987]
- Best CT. Emergence of Language-Specific Constraints in Perception of Nonnative Speech - a Window on Early Phonological Development. *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. 1993; 69:289–304.
- Best CT. Language-specific attunement of speech perception. *Australian Journal of Psychology*. 2006; 58:3–3.
- Best CT, Morrongiello B, Robson R. Perceptual Equivalence of Acoustic Cues in Speech and Nonspeech Perception. *Perception & psychophysics*. 1981; 29(3):191–211. [PubMed: 7267271]
- Bradlow AR. Training non-native language sound patterns. *Phonology and second language acquisition*. 2008:287–308.
- Bradlow AR, Akahane-Yamada R, Pisoni DB, Tohkura Y. Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Perception & psychophysics*. 1999; 61(5):977–985. [PubMed: 10499009]
- Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Naatanen R. Development of language-specific phoneme representations in the infant brain. *Nat Neurosci*. 1998; 1(5):351–353. [PubMed: 10196522]
- Cleeremans A, Dienes Z. Computational models of implicit learning. *Cambridge handbook of computational psychology*. 2008:396–421.
- DeCaro MS, Carlson KD, Thomas RD, Beilock SL. When and how less is more: reply to Tharp and Pickering. *Cognition*. 2009; 111(3):397–403. [PubMed: 19344894]
- Decaro MS, Thomas RD, Beilock SL. Individual differences in category learning: sometimes less working memory capacity is better than more. *Cognition*. 2008; 107(1):284–294. [PubMed: 17707363]
- DeKeyser R. Implicit and Explicit Learning. *The handbook of second language acquisition*. 2008:313.
- Diaz B, Baus C, Escera C, Costa A, Sebastian-Galles N. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proc Natl Acad Sci U S A*. 2008; 105(42):16083–16088. [PubMed: 18852470]
- Diehl RL, Lotto AJ, Holt LL. Speech perception. *Annu Rev Psychol*. 2004; 55:149–179. [PubMed: 14744213]
- Evans JL, Saffran JR, Robe-Torres K. Statistical learning in children with specific language impairment. *J Speech Lang Hear Res*. 2009; 52(2):321–335. [PubMed: 19339700]
- Flege JE. Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*. 1995:233–277.
- Flege JE. Age of learning and second language speech. *Second Language Acquisition and the Critical Period Hypothesis*. 1999:101–131.
- Francis AL, Ciocca V, Ma L, Fenn K. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*. 2008; 36(2):268–294.
- Francis AL, Nusbaum HC. Selective attention and the acquisition of new phonetic categories. *Journal of experimental psychology Human perception and performance*. 2002; 28(2):349–366. [PubMed: 11999859]
- Gandour J. Perceived dimensions of thirteen tones: a multidimensional scaling investigation. *Phonetica*. 1978; 35:169–179. [PubMed: 674389]

- Gandour J. Tone perception in Far Eastern languages. *Journal of Phonetics*. 1983; 11:149–175.
- Gandour J, Dardarananda R. Identification of tonal contrasts in Thai aphasic patients. *Brain Lang*. 1983; 18(1):98–114. [PubMed: 6220759]
- Gandour J, Dziedzic M, Wong D, Lowe M, Tong Y, Hsieh L, Lurito J. Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang*. 2003; 84(3):318–336. [PubMed: 12662974]
- Gandour J, Wong D, Hsieh L, Weinzapfel B, Van Lancker D, Hutchins GD. A crosslinguistic PET study of tone perception. *J Cogn Neurosci*. 2000; 12(1):207–222. [PubMed: 10769317]
- Gandour J, Wong D, Hutchins G. Pitch processing in the human brain is influenced by language experience. *Neuroreport*. 1998; 9(9):2115–2119. [PubMed: 9674604]
- Golestani N, Molko N, Dehaene S, LeBihan D, Pallier C. Brain structure predicts the learning of foreign speech sounds. *Cereb Cortex*. 2007; 17(3):575–582. [PubMed: 16603709]
- Goudbeek M, Cutler A, Smits R. Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*. 2008; 50(2):109–125.
- Hattori K, Iverson P. English /r/-/l/ category assimilation by Japanese adults: individual differences and the link to identification accuracy. *J Acoust Soc Am*. 2009; 125(1):469–479. [PubMed: 19173432]
- Hay JF, Pelucchi B, Graf Estes K, Saffran JR. Linking sounds to meanings: infant statistical learning in a natural language. *Cogn Psychol*. 2011; 63(2):93–106. [PubMed: 21762650]
- Hernandez AE, Li P. Age of acquisition: its neural and computational mechanisms. *Psychological bulletin*. 2007; 133(4):638–650. [PubMed: 17592959]
- Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*. 2004; 92(1–2):67–99. [PubMed: 15037127]
- Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci*. 2007; 8(5):393–402. [PubMed: 17431404]
- Hikosaka O, Sakamoto Y, Usui S. Functional properties of monkey caudate neurons: III. Activities related to expectation of target and reward. *Journal of Neurophysiology*. 1989; 61:814–832. [PubMed: 2723722]
- Holt LL. The mean matters: effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America*. 2006; 120(5 Pt 1):2801–2817. [PubMed: 17091133]
- Holt LL, Lotto AJ. Speech Perception Within an Auditory Cognitive Science Framework. *Curr Dir Psychol Sci*. 2008; 17(1):42–46. [PubMed: 19060961]
- Holt LL, Lotto AJ. Speech perception as categorization. *Atten Percept Psychophys*. 2010; 72(5):1218–1227. [PubMed: 20601702]
- Huang J, Holt LL. Listening for the norm: adaptive coding in speech categorization. *Frontiers in psychology*. 2012; 3:10. [PubMed: 22347198]
- Hulstijn JH. Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in second language acquisition*. 2005; 27:129–140.
- Hume E, Johnson K. 2001A model of the interplay of speech perception and phonology. *The Role of Speech Perception in Phonology*. :3–26.
- Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, Siebert C. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. 2003; 87(1):B47–57. [PubMed: 12499111]
- Klein D, Zatorre RJ, Milner B, Zhao V. A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*. 2001; 13(4):646–653. [PubMed: 11305893]
- Kraljic T, Samuel AG. Perceptual learning for speech: Is there a return to normal? *Cogn Psychol*. 2005; 51(2):141–178. [PubMed: 16095588]
- Kraljic T, Samuel AG. Generalization in perceptual learning for speech. *Psychon Bull Rev*. 2006; 13(2):262–268. [PubMed: 16892992]
- Kraljic T, Samuel AG. Perceptual adjustments to multiple speakers. *Journal of Memory and Language*. 2007; 56(1):1–15.
- Krashen, S. *Principles and practice in second language acquisition*. Oxford Pergamon; 1982.

- Laing EJ, Liu R, Lotto AJ, Holt LL. Tuned with a Tune: Talker Normalization via General Auditory Processes. *Frontiers in psychology*. 2012; 3:203. [PubMed: 22737140]
- Lim SJ, 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]
- 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. *J Acoust Soc Am*. 1993; 94(3 Pt 1):1242–1255. [PubMed: 8408964]
- Lively SE, Pisoni DB, Yamada RA, Tohkura Y, Yamada T. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *J Acoust Soc Am*. 1994; 96(4):2076–2087. [PubMed: 7963022]
- Lotto AJ. Language acquisition as complex category formation. *Phonetica*. 2000; 57(2–4):189–196. [PubMed: 10992139]
- Love BC, Medin DL, Gureckis TM. SUSTAIN: A Network Model of Category Learning. *Psychological review*. 2004; 111(2):309–332. [PubMed: 15065912]
- Maddox WT, Ashby FG. Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behav Processes*. 2004; 66(3):309–332. [PubMed: 15157979]
- Maddox WT, Filoteo JV, Lauritzen JS, Connally E, Hejl KD. Discontinuous categories affect information-integration but not rule-based category learning. *J Exp Psychol Learn Mem Cogn*. 2005; 31(4):654–669. [PubMed: 16060771]
- 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]
- Maddox WT, Molis MR, Diehl RL. Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels. *Perception & psychophysics*. 2002; 64(4):584–597. [PubMed: 12132760]
- McClelland JL, Fiez JA, McCandliss BD. Teaching the /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects. *Physiol Behav*. 2002; 77(4–5):657–662. [PubMed: 12527015]
- McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*. 1995; 102(3):419. [PubMed: 7624455]
- McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? *Trends in cognitive sciences*. 2006; 10(8):363–369. [PubMed: 16843037]
- Mirman D, McClelland JL, Holt LL. An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic bulletin & review*. 2006; 13(6):958–965. [PubMed: 17484419]
- Naatanen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*. 1997; 385(6615):432–434. [PubMed: 9009189]
- Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, Parrish TB, Reber PJ. Neural correlates of rule-based and information-integration visual category learning. *Cereb Cortex*. 2007; 17(1):37–43. [PubMed: 16436685]
- Nomura EM, Reber PJ. A review of medial temporal lobe and caudate contributions to visual category learning. *Neurosci Biobehav Rev*. 2008; 32(2):279–291. [PubMed: 17868867]
- Paradis M. On the representation of two languages in one brain. *Language Sciences*. 1985; 7(1):1–39.
- Paradis, M. A neurolinguistic theory of bilingualism. Vol. 18. John Benjamins Publishing; 2004.
- Perrachione TK, Lee J, Ha LY, Wong PC. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *J Acoust Soc Am*. 2011a; 130(1):461–472. [PubMed: 21786912]
- Perrachione TK, Lee J, Ha LY, Wong PC. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*. 2011b; 130:461. [PubMed: 21786912]
- Petrides M, Pandya DN. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J Comp Neurol*. 1988; 273(1):52–66. [PubMed: 2463275]

- Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA. Interactive memory systems in the human brain. *Nature*. 2001; 414(6863):546–550. [PubMed: 11734855]
- Poldrack RA, Foerde K. Category learning and the memory systems debate. *Neurosci Biobehav Rev*. 2008; 32(2):197–205. [PubMed: 17869339]
- 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]
- Reber AS, Walkenfeld FF, Hernstadt R. Implicit and explicit learning: individual differences and IQ. *J Exp Psychol Learn Mem Cogn*. 1991; 17(5):888–896. [PubMed: 1834770]
- Reber PJ. The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*. 2013
- Robinson, P. The handbook of second language acquisition. Blackwell Publishing Ltd; 2008. Attention and Memory during SLA; p. 631–678.
- Romberg AR, Saffran JR. Statistical learning and language acquisition. *Wiley Interdiscip Rev Cogn Sci*. 2010; 1(6):906–914. [PubMed: 21666883]
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274(5294):1926–1928. [PubMed: 8943209]
- Samuel AG, Kraljic T. Perceptual learning for speech. *Atten Percept Psychophys*. 2009; 71(6):1207–1218. [PubMed: 19633336]
- Schmidt R. Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning*. 1995:1–63.
- Schmidt R. Attention, awareness, and individual differences in language learning. *Perspectives on Individual Characteristics and Foreign Language Education*. 2012; 6:27.
- Seger CA. How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neurosci Biobehav Rev*. 2008; 32(2):265–278. [PubMed: 17919725]
- Seger CA, Cincotta CM. The roles of the caudate nucleus in human classification learning. *J Neurosci*. 2005; 25(11):2941–2951. [PubMed: 15772354]
- Seger CA, Miller EK. Category learning in the brain. *Annu Rev Neurosci*. 2010; 33:203–219. [PubMed: 20572771]
- Tagarelli, K.; Borges-Mota, M.; Rebuschat, P. The role of working memory in implicit and explicit language learning. Paper presented at the Proceedings of the 33rd Annual Conference of the Cognitive Science Society; Austin, TX: Cognitive Science Society; 2011.
- 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]
- Ullman MT. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition*. 2001; 4(02):105–122.
- Ullman MT. Contributions of memory circuits to language: The declarative/procedural model. *Cognition*. 2004; 92(1):231–270. [PubMed: 15037131]
- Ullman MT. The declarative/procedural model and the shallow structure hypothesis. *Applied Psycholinguistics*. 2006; 27(01):97–105.
- Vallabha GK, McClelland JL. Success and failure of new speech category learning in adulthood: consequences of learned Hebbian attractors in topographic maps. *Cogn Affect Behav Neurosci*. 2007; 7(1):53–73. [PubMed: 17598735]
- Ventura-Campos N, Sanjuán A, González J, Palomar-García M-Á, Rodríguez-Pujadas A, Sebastián-Gallés N, Ávila C. Spontaneous Brain Activity Predicts Learning Ability of Foreign Sounds. *The Journal of Neuroscience*. 2013; 33(22):9295–9305. [PubMed: 23719798]
- Wagenmakers EJ, Farrell S. AIC model selection using Akaike weights. *Psychon Bull Rev*. 2004; 11(1):192–196. [PubMed: 15117008]
- Wang Y, Jongman A, Sereno JA. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J Acoust Soc Am*. 2003; 113(2):1033–1043. [PubMed: 12597196]

- Wang Y, Sereno JA, Jongman A, Hirsch J. fMRI evidence for cortical modification during learning of Mandarin lexical tone. *J Cogn Neurosci*. 2003a; 15(7):1019–1027. [PubMed: 14614812]
- Wang Y, Sereno JA, Jongman A, Hirsch J. fMRI evidence for cortical modification during learning of Mandarin lexical tone. *J Cogn Neurosci*. 2003b; 15(7):1019–1027. [PubMed: 14614812]
- Wickens, TD. *Models for behavior: Stochastic processes in psychology*. WH Freeman; San Francisco: 1982.
- Wilson CJ. The contribution of cortical neurons to the firing pattern of striatal spiny neurons. *Models of information processing in the basal ganglia*. 1995:29–50.
- Wong FC, Chandrasekaran B, Garibaldi K, Wong PC. White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *J Neurosci*. 2011; 31(24):8780–8785. [PubMed: 21677162]
- Wong PC, Nusbaum HC, Small SL. Neural bases of talker normalization. *J Cogn Neurosci*. 2004a; 16(7):1173–1184. [PubMed: 15453972]
- Wong PC, Parsons LM, Martinez M, Diehl RL. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *J Neurosci*. 2004b; 24(41):9153–9160. [PubMed: 15483134]
- Wong PC, Perrachione TK, Parrish TB. Neural characteristics of successful and less successful speech and word learning in adults. *Hum Brain Mapp*. 2007; 28(10):995–1006. [PubMed: 17133399]
- Xi J, Zhang L, Shu H, Zhang Y, Li P. Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*. 2010; 170(1):223–231. [PubMed: 20633613]
- Xu Y, Gandour JT, Francis AL. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J Acoust Soc Am*. 2006; 120(2):1063–1074. [PubMed: 16938992]
- Yang J, Li P. Brain Networks of Explicit and Implicit Learning. *PloS one*. 2012; 7(8):e42993. [PubMed: 22952624]
- Yeterian EH, Pandya DN. Corticostriatal connections of the superior temporal region in rhesus monkeys. *J Comp Neurol*. 1998; 399(3):384–402. [PubMed: 9733085]
- Zhang Y, Kuhl PK, Imada T, Iverson P, Pruitt J, Stevens EB, Nemoto I. Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *Neuroimage*. 2009; 46(1):226–240. [PubMed: 19457395]

Figure 1. Artificial category structures (left panel(a) Rule Based (RB), right panel (b) Information-integration (II)) used to study dissociations between reflective and reflexive learning systems.

Figure 2.

(Left panel (a)) Sample fundamental frequency contours of four Mandarin tone (T1: high-level; T2: rising; T3: dipping; T4: falling) produced by a male native Mandarin speaker used in the experiment. (Right panel (b)) The four tones plotted in a two-dimensional perceptual space (x-axis: pitch height, y-axis: pitch direction). Pitch height (dim. 1) and pitch direction (dim. 2) are major cues used to distinguish the tone categories.

Figure 3.

(Top panel (a)) In the tone category training paradigm, we use 80 stimuli (5 segments X 4 talkers X 4 tones) that are plotted on the two-dimensional space (pitch height, direction). In the middle (b) and lower (c) panel, the stimuli are separated by male and female talkers. Within the male (b) and female (c) perceptual spaces, category separation is clearer than the perceptual space that uses all talkers (a).

Figure 4.

Scatterplot of the responses along with the decision boundaries that separate response regions from versions of the (a) Striatal Pattern Classifier, (b) Conjunctive rule-based, (c) Uni-Dimensional_Height, and (d) Uni-Dimensional_Direction models as applied to the stimuli from Figure 3a.

Figure 5.
Overall proportion correct for final blocks Separators versus Non-Separators in Experiment 1 (A) and Experiment 2 (B).

Figure 6. Proportion of participants whose data was best fit by a Non-Separation or Separation model as a function of block in Experiment 1 (A) and Experiment 2 (B).

Figure 7.

A. Average number of blocks best fit by a Separation model for final block Separators and final block Non-Separators in Experiment 1. B. Average block first best fit by a Separation model for final block Separators and final block Non-Separators in Experiment 1. C. Average number of blocks best fit by a Separation model for final block Separators and final block Non-Separators in Experiment 2. D. Average block first best fit by a Separation model for final block Separators and final block Non-Separators in Experiment 2.