Published in final edited form as:
Psychol Aging. 2013 December ; 28(4): . doi:10.1037/a0034969.

Dual systems of speech category learning across the lifespan

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

Abstract

Although categorization is fundamental to speech processing, little is known about the learning systems that mediate auditory categorization and even less is known about changes across the lifespan. Vision research supports dual-learning systems that are grounded in neuroscience and are partially-dissociable. The *reflective, rule-based* system is prefrontally mediated and uses working memory and executive attention to develop and test rules for classifying in an explicit fashion. The *reflexive, information-integration* system is striatally mediated and operates by implicitly associating perception with actions that lead to reinforcement. We examine the extent to which dual-learning systems mediate auditory and speech learning in younger and older adults. We examined auditory category learning when a rule-based strategy (Experiment 1) or information-integration strategy (Experiment 2) was optimal, and found an age-related rule-based deficit, but intact information-integration learning. Experiment 3 examined natural auditory category learning, and found an age-related performance deficit. Computational modeling suggested that this was due to older adults' persistent reliance on sub-optimal, uni-dimensional strategies when two-dimensional strategies were optimal. Working memory capacity was also found to be associated with improved rule-based and natural auditory category learning, but not information-integration category learning. These results suggest that dual-learning systems are operative in speech category learning across the lifespan, and that performance deficits, when present are due to deficiencies in frontally-mediated, rule-based processes.

Keywords

Aging; auditory category learning; rule-based; information-integration; working memory; reflective system; reflexive system

Introduction

Category learning is a fundamental aspect of human cognition that is critically important during all stages of life. When we judge whether a talker or a musical piece is familiar or not, an environmental sound indicates danger or not, or a speaker sounds angry or happy, we are categorizing. Quick and efficient categorization is as important later in life, as it is early, and thus an understanding of age-related changes in categorization is an important area of scientific inquiry.

The neurobiological underpinnings of visually-mediated category learning suggest that the learning of different types of category structures is mediated by different systems that have unique, but interacting neural substrates (for reviews see, Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005, 2010; Filoteo & Maddox, 2007; Keri, 2003;

[1]This research was supported by NIMH grants MH077708 and DA032457 to WTM. We thank Michelle Lee, Kathryn Curry, Neda Abdul-Razzak, Jacqueline Gonzalez, Taylor Denny, Kadee Bludau and the other MaddoxLab RAs for help with data collection.

Address correspondence to W. Todd Maddox (maddox@psy.utexas.edu) or Bharath Chandrasekaran (bchandra@austin.utexas.edu).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Nomura & Reber, 2008; Poldrack & Foerde, 2008; Price, Filoteo, & Maddox, 2009; Seger, 2008). The COmpetition between Verbal and Implicit Systems (COVIS) model captures this dual-system framework (Ashby, et al., 1998; Ashby, Paul, & Maddox, 2011). COVIS postulates that optimal rule-based category learning involves the application of verbalizable strategies (Bruner, Goodnow, & Austin, 1956; Shepard, Hovland, & Jenkins, 1961) that are mediated by an explicit, reflective, hypothesis-testing system that relies on working memory and executive attention, and is highly dependent on the prefrontal cortex (Filoteo et al., 2005; Lombardi et al., 1999; Monchi, Petrides, Petre, Worsley, & Dagher, 2001; Nomura et al., 2007; Schnyer et al., 2009; Seger & Cincotta, 2006). Optimal information-integration category learning, on the other hand, involves strategies that maximize accuracy that cannot be described verbally and instead involve integrating information from two or more stimulus dimensions at some pre-decisional stage (Ashby & Waldron, 1999). Information-integration category learning does not rely on working memory and executive attention and is mediated by an implicit, reflexive, procedural-based learning system that is highly dependent upon the striatum (Aron et al., 2004; Filoteo, Maddox, Salmon, & Song, 2005; Maddox & Filoteo, 2001, 2005; Nomura, et al., 2007; Poldrack et al., 2001; Seger, 2008; Seger & Cincotta, 2005).

COVIS posits that in information-integration tasks, learners initially use the reflective (rule-based) system, but switch to the reflexive (information-integration) system with practice. Given the extensive literature suggesting frontal and striatal declines with normal aging (Greenwood, 2000, 2007; Grieve, Williams, Paul, Clark, & Gordon, 2007; Gunning-Dixon et al., 2003; Park & Reuter-Lorenz, 2009; Raz, 2000; Raz et al., 2003; Reuter-Lorenz & Park, 2010), it is not surprising that visually-mediated rule-based and information-integration category learning deficits are associated with normal aging (Filoteo & Maddox, 2004; Filoteo, Maddox, Ing, Zizak, & Song, 2005; Gorlick et al., 2012; Maddox, Pacheco, Reeves, Zhu, & Schnyer, 2010; Racine, Barch, Braver, & Noelle, 2006; Ridderinkhof, Span, & van der Molen, 2002).

In contrast to the visual domain, much less is known about the role of the dual-learning systems in auditory category learning. To our knowledge, no studies have examined auditory category learning in healthy aging, despite its importance in everyday life. 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 show that the primary and association cortices are connected to the prefrontal cortex via dorsal and ventral routes, and form many-to-one projections to the striatum (Petrides & Pandya, 1988; Yeterian & Pandya, 1998). The convergent 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 (procedural-based) learning system (Ashby & Ennis, 2006; Waldschmidt & Ashby, 2011). The primary auditory cortex, in contrast, is less densely connected to the tail of the caudate. While current knowledge about the role of dual-learning systems in category learning has been derived from studies in the visual domain, the anatomical labeling studies lend neurobiological plausibility to the application of a dual-systems framework in the auditory domain.

The auditory category learning literature has primarily focused on perceptual processes involved in categorization, but a few studies have applied the dual-systems framework in the auditory domain. Maddox, Ing and Lauritzen (2006) examined rule-based and information-integration category learning with artificial auditory stimuli, and Maddox, Molis and Diehl (2002) applied the dual system approach to data from an auditory vowel categorization task. More recently, the dual systems framework has been applied to learning of second language (L2) speech categories {Maddox, 2013 #2870;Chandrasekaran, in press #2903}. Adult, native speakers of American English were trained to learn Mandarin tone category structures with feedback. Mandarin Chinese has four linguistically-relevant tone categories that differ primarily on the basis of pitch pattern (ma^1 ‘mother’ [T1], ma^2 ‘hemp’ [T2], ma^3 ‘horse’ [T3], ma^4 ‘scold’ [T4]), described phonetically as high level, high rising, low falling rising, and high falling pitch patterns, respectively. Two dimensions (pitch height and pitch direction) serve as primary cues in categorizing tone patterns, and these cues are differentially weighted across languages. Successful learners tended to use multidimensional strategies, whereas less successful learners tended to use simple uni-dimensional strategies.

While these studies demonstrate the applicability of the dual-systems framework to study auditory category learning, the population examined has been young adults (age 18–35). The current goal is to extend this approach to study auditory category learning in healthy aging. Experiment 1 examines rule-based category learning, and Experiment 2 examines information-integration category learning. Experiment 3 uses a combination of behavioral analysis and computational modeling to examine the effect of normal aging on natural speech category learning (Mandarin tone categories). In Experiments 1 and 2, based on the visual category learning literature, we would predict deficits in rule-based as well as information-integration learning in older adults. However, several studies have demonstrated modality-specific deficits in sensory processes during aging, and fundamental differences in the role of inhibitory processes across domains (Ceponiene, Westerfield, Torki, & Townsend, 2008; Guerreiro, Murphy, & Van Gerven, 2010, 2013). This warrants a systematic examination of auditory category learning across the lifespan. In Experiment 3, we examine the extent to which aging impacts learning novel speech categories. Using computational models, we are able to evaluate the extent to which category-learning success relates to strategy use in younger and older adults. Together, these experiments will systematically evaluate lifespan changes in auditory category learning across the dual category learning systems.

Experiment 1

Experiment 1 examines age-related changes in rule-based category learning.

Method

Participants—Seventeen older adults (average age 67.59) from the greater Austin, Texas community and 21 younger adults from the University of Texas community were paid $10 per hour for their participation. Informed consent was obtained from all participants and the experiment was approved for ethics procedures using human participants. All participants passed a hearing-screening test (thresholds of <40 dB HL at frequencies of 500, 1,000, 2,000, and 4,000 Hz) and reported no significant issues related to hearing. Older and younger adult groups did not significantly differ in musicianship: the age at which the participant began music practicing, years of practice, and hours practiced per week. Stimuli were presented at comfortable supra-threshold listening levels, as judged by the participants.

Neuropsychological Testing Procedures—Older adults were given a series of standardized neuropsychological tests designed to assess general intellectual ability across *attention* (WAIS-III Digit Span; WAIS-III Vocabulary; Wechsler, 1997), *executive functioning* [Trail Making Test A&B (TMT), (Lezak, 1995); FAS, Wisconsin Card Sorting Task (WCST), (Heaton, Chelune, Talley, Kay, & Curtiss, 1993)] and *memory* (California Verbal Learning Test; CVLT; Delis, Kramer, Kaplan, & Ober, 1987). The tests were administered over the course of two two-hour sessions spaced approximately a week apart.

Normative scores for each subject were calculated for each neuropsychological test using the standard age-appropriate published norms. Table 1a shows the means, standard deviations, and ranges of standardized z-scores on each test for older adults. All WAIS subtest percentiles were calculated according to the testing instructions and then converted to standardized z-scores. The CVLT and WCST standardized t-scores were calculated according to testing directions then converted to standardized z-scores, and the TMT standard z-scores were calculated according to the testing instructions. Participants were excluded from participation if they scored more than two standard deviations below the standardized mean on more than one neuropsychological test in the same area (memory, executive functioning, or attention). Only subjects who were within normal ranges were asked to participate in the experiment.

Stimuli—Stimuli consisted of auditory tones presented via headphones. Each stimulus was four dimensional with one of two possible values for each dimension being presented (16 stimuli total). The stimuli varied along the four auditory dimensions of pitch (high vs. low; 180 Hz vs. 80 Hz), duration (long vs. short; 500 ms vs. 250 ms), number (1 vs. 2 non-overlapping tones), and vowel (/a/ vs. /i/). For the rule-based task, categories were defined by arbitrarily making two stimulus dimensions relevant (e.g., pitch and duration), and two stimulus dimensions irrelevant (e.g., number and vowel). For the two relevant dimensions the binary properties of each dimension were arbitrarily given the values 1 or −1 (e.g., high pitch = 1 and low pitch = −1; long duration = 1 and short duration = −1). Stimuli in category A were those with values of 1 on both relevant dimensions (high pitch with long duration) or values of −1 on both relevant dimensions (low pitch with short duration). Stimuli in category B were those with a value of 1 on one relevant dimension and a value of −1 on the other relevant dimension (high pitch with short duration or low pitch with long duration). A schematic of one possible conjunctive rule-based problem is displayed in Figure 1a.

Procedure—Participants performed the experiment on a personal computer in a well-controlled testing room. Participants wore Sennheiser HD 280 Pro headphones to listen to the stimuli presented. Participants were informed that they would be listening to sounds that vary across trials in pitch, duration, vowel, and number of tones. They were informed that each sound was a member of one of two categories: A or B, and that their task was to determine the category membership for each sound by using the computer key and pressing either the "z" button which corresponded to category A or the "m" button which corresponded to category B. Participants were informed that they would receive feedback following each response that would state whether their response was "correct" or "incorrect". Finally, they were informed that their goal was to generate 10 correct responses in a row. Once they achieved 10 correct responses in a row, or after 200 trials, whichever came first, the task would end.

Results

To determine whether there were rule-based performance differences between healthy older and younger adults we conducted two analyses. First, we compared the number of trials needed to reach criterion. If an individual did not reach criterion after 200 trials, we assumed a trials-to-criterion of 200. Figure 2a displays the average trials-to-criterion for the older and younger adults. A t-test confirmed that older adults (trials-to-criterion = 186) took significantly longer to learn the task than younger adults (trials-to-criterion = 129) [$t(36) = 2.72$, $p = .01$, partial $\eta^2 = .171$]. Second, we compared the number of older adults who reached or did not reach criterion with the number of younger adults who reached or did not reach criterion. Figure 2b displays the proportion of older (.18) and younger (.48) adults who reached criterion. A $\chi2$ test confirmed that significantly fewer older than younger adults reached criterion [$\chi2(1) = 3.75$, $p = .05$].

Summary

These findings suggest that older adults do show a significant rule-based category learning deficit relative to younger adults when the stimulus dimensions are presented within the auditory domain.

Experiment 2

Experiment 2 examines age-related changes in information-integration category learning.

Method

Participants—Seventeen older adults (average age 68.41) from the greater Austin, Texas community and 19 younger adults from the University of Texas community were paid $10 per hour for their participation. Procedures for informed consent and the hearing screening were identical to those from Experiment 1.

Neuropsychological Testing Procedures—The neuropsychological testing procedures were identical to those from Experiment 1. Table 1b shows the means, standard deviations, and ranges of standardized z-scores on each test for older adults.

Stimuli—The stimuli were identical to those used in Experiment 1 however Experiment 2 used an information-integration category structure. To create the information-integration category structures we first made one stimulus dimension irrelevant (e.g. pitch). Then for the three remaining relevant stimulus dimension, the possible properties of each stimulus were given a value of 1 or −1 (e.g. for duration, long = 1 and short = −1). Then, each category structure was created by the following mathematical formula (where the three relevant stimulus dimensions are X, Y, and Z):

If $X + Y + Z > 0$, then “A,” else “B.”

A schematic of one possible information-integration problem is displayed in Figure 1b.

Procedure—The procedures were identical to those used in Experiment 1.

Results

We used the same data analytic approach used in Experiment 1. Figure 3a displays the average trials-to-criterion for the older and younger adults. Older adults (trials-to-criterion = 141) took slightly longer to learn the task than younger adults (trials-to-criterion = 129), but this difference was non-significant [$t(34) = .52$, $p = .61$, partial $\eta^2 = .008$]. Figure 3b displays the proportion of older (.53) and younger (.58) adults who reached criterion. Although slightly fewer older adults than younger adults reached criterion, a $\chi 2$ test confirmed that this difference was non-significant [$\chi 2(1) = .09$, $p = .77$].

As with any information-integration category learning task, rule-based strategies exist that can yield good, albeit non-optimal, performance. In the current task, a uni-dimensional rule can accurately classify 12 of the 16 stimuli. Thus, it is possible that learners (i.e., those who generated 10 correct responses in a row) are not utilizing the optimal information-integration strategy but rather are using a uni-dimensional strategy on a run of 10 trials for which the uni-dimensional rule yields perfect performance. As a test of this hypothesis we computed the accuracy rate predicted by the optimal information-integration rule during the final 10 trials for learners, which by definition is 100%, and compared that with accuracy rate predicted by the most accurate uni-dimensional rule. If the accuracy rate predicted by the most accurate uni-dimensional rule is 100% for a given participant, then it is equivocal whether that participant used the optimal information-integration rule or a uni-dimensional rule. The results were clear. For 14 of the 20 learners, the responses during the final 10 trials were inconsistent with a uni-dimensional rule (7 of 9 older adult learners and 7 of 11 younger adult learners) suggesting that learners were using information-integration strategies at a much higher rate than uni-dimensional strategies.

Interestingly though, there was evidence that some non-learners were using uni-dimensional strategies. Specifically, 10 of the 16 non-learners' data was better accounted for by a uni-dimensional rule than by the optimal information-integration rule. This finding is expected since sub-optimal information-integration performance often results when individuals fail to transition from rules to information-integration strategies.[2]

Summary

The results from Experiments 1 and 2 suggest an age-related rule-based, but not information-integration category learning deficit. This suggests that within the auditory category learning domain, age-related performance deficits emerge with respect to frontal, rule-based processes but not striatal, procedural learning processes. Experiments 1 and 2 use artificially constructed category structures. In Experiment 3, we examine natural speech category learning in older and younger adults.

Experiment 3

Learning new speech categories in adulthood is known to be a difficult task. Many theories have been proposed to account for such difficulty. In general, the difficulty may arise because learners tend to attend to dimensions that are relevant in their native language (L1) and are less focused on dimensions that are more relevant in the second language (L2). Previous studies suggest that speech categories are optimally learned implicitly, although listeners do tend to use a variety of strategies when learning speech categories (Chandrasekaran, Yi, & Maddox, in press; Lim & Holt, 2011; Seitz et al., 2010). However, this work has entirely focused on young adults. The strategies used by older adults in learning L2 speech categories has not been systematically examined.

In the current experiment we examine novel speech category learning in younger and older adults using exemplars from Mandarin Chinese, a tone language. Two dimensions, pitch height, and pitch direction are important for discerning tone categories across languages. For example, on the pitch height-pitch direction continuum, the four Mandarin tone categories can be differentiated as “high-level”, “low-rising”, “low-dipping”, and “high-falling”. The pitch height dimension (average pitch across the syllable) is important for distinguishing low tones from high tones; the pitch direction dimension is important in distinguishing rising tones from falling tones, as well as documenting changes within the syllable. We use Mandarin tonal categories as a test-bed to examine strategy differences across the lifespan. Based on Experiments 1 and 2, one possibility is that speech category learning success does not differ between older and younger adults, given that speech categories are reflexive-

[2]Another possibility is that learners are using a rule-plus exception strategy in which they classify 12 of the 16 stimuli using a uni-dimensional rule and then memorize the remaining 4 stimuli. Importantly, because the rule-plus-exception and optimal information-integration strategies yields identical responses, they are mathematically equivalent and thus cannot be teased apart based on the behavioral responses. Even so, there are two reasons to believe that the rule-plus-exception hypothesis in not viable in the present case. First, in a recent paper, Davis, Love and Maddox (2012) found that older adults show performance deficits in category learning when the optimal strategy is a rule-plus-exception strategy. In particular, older adults struggle to learn the exceptions. Second, the pattern of correlations between information-integration performance and working memory capacity argue against the use of a rule-based strategy in the information-integration task. A rule-plus-exception strategy would put a heavy demand upon working memory and thus good performance should be associated with high working memory capacity. Instead, we find that good performance is associated with low working memory capacity. Thus, although one must always be aware of the possibility that participants might use rule-based strategies when solving an information-integration task, learners in the present study appear to be relying upon information-integration strategies.

optimal. However, given the rule-based category learning deficit seen in Experiment 1, one possibility is that older adults may perseverate with simple uni-dimensional rules (rules based on pitch height or pitch direction), and unable to use verbal rules with multiple dimensions (pitch height and pitch direction), or implicitly integrate across dimensions. Note that COVIS posits that natural category learning is initially dominated by the reflective system, irrespective of whether category structure is reflective or reflexive-optimal. Therefore, if simple uni-dimensional rules persist, the learner may not be able to transition to more optimal strategies.

To test these possibilities, we not only evaluate accuracy measures, which will discern performance difference between younger and older adults, but also evaluate strategy differences between the two groups. We will evaluate strategy differences by using neurobiologically inspired computational models of reflective and reflexive learning.

Method

Participants—Thirty-five older adults (average age 68.14) from the greater Austin, Texas community and 38 younger adults from the University of Texas community were paid $10 per hour for their participation. Exact hearing thresholds were recorded in approximately half of the participants (17 older adults and 21 younger adults). We used this subset of the data to conduct additional analyses to evaluate the influence of hearing status on category learning accuracy. Procedures for informed consent and the hearing screen were identical to those from Experiments 1 and 2. Groups did not significantly differ with respect to the age at which the participant began music practicing, years of practice, and hours practiced per week.

Neuropsychological Testing Procedures—The neuropsychological testing procedures were identical to those from Experiments 1 and 2. Table 1c shows the means, standard deviations, and ranges of standardized z-scores on each test for older adults.

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, the use of these stimuli 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 consisted of native speakers (N = 2; 1 f) of Mandarin Chinese originally from Beijing. Stimuli were RMS amplitude and duration normalized (70 dB, 0.4 s) using the software Praat (Alexander, et al., 2005; Perrachione, Lee, Ha, & Wong, 2011; Wong, Perrachione, Gunasekera, & Chandrasekaran, 2009). 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 height and direction, are used primarily to distinguish tone categories (Francis, Ciocca, Ma, & Fenn, 2008). In fact, phonetically, Mandarin tones 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. A scatter-plot of the 40 stimuli in the pitch height-pitch direction space is displayed in Figure 4a.

Procedure—On each trial, participants were presented with a single exemplar from one of four Mandarin tone categories (T1, T2, T3, or T4) and instructed to categorize the stimulus into one of four categories. Participants were given feedback on each trial and exposed to multiple talkers throughout the training program. Participants listened to 40 stimuli per block (4 tone categories × 5 syllables × 2 talkers). The talkers were randomized within a block. Each participant completed five 40-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 number button keys on the left side of the computer keyboard, 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 “No.” followed by the label of the tone that was actually presented. For example, on a correct T2 trial the feedback display was as follows: “Correct, that was a category 2”. On an incorrect response trial where T3 was the correct response the feedback display was as follows: “No, that was a category 3”. A 1-s ITI followed the feedback.

Accuracy Results

We first present accuracy analyses comparing block-by-block performance across older and younger adults, and then we present model-based analyses to explore the types of strategies that participants use to solve the task.

Learning curves for the younger and older adults are presented in Figure 5a. We begin with a 2 participant group (younger vs. older adults) × 5 block mixed design ANOVA on the accuracy data. The main effect of participant group was significant [$F(1, 71) = 24.45$, $p < .001$, partial $\eta^2 = .256$] and suggested an age-related deficit in performance. The main effect of block was also significant [$F(4, 284) = 46.06$, $p < .001$, partial $\eta^2 = .393$]. Finally, the interaction between participant group and block was significant [$F(4, 284) = 5.59$, $p < .001$, partial $\eta^2 = .073$]. Post hoc analyses suggested that the older adult performance deficit emerged in all blocks of trials (all p’s < .005), but that the learning rate was lower in older than in younger adults.

To rule out a confounding role of hearing status, we ran a 2 participant group (younger vs. older adults) × 5 block ANOVA on the accuracy data with left and right pure tone average (PTA; average of thresholds 500, 1000, and 2000 Hz) as covariates. Note that we used a smaller subset for whom exact thresholds were collected (17 older adults and 21 younger adults). Adding these covariates did not significantly alter the pattern of results. That is, the main effect of participant group was significant [$F(1, 33) = 5.27$, $p =0.028$, partial $\eta^2 = .138$] and suggested an age-related deficit in performance. The main effect of block was also significant [$F(4, 284) = 46.06$, $p < .001$, partial $\eta^2 = .393$]. Finally, the interaction between participant group and block was significant [$F(4, 132) = 3.706$, $p = .007$, partial η^2]. Post hoc analyses suggested that the older adult performance deficit emerged in all blocks of trials (all p’s < .005), but that the learning rate was lower in older than in younger adults. The covariates were not significant ($p>0.6$) and did not significantly interact with block. Further, correlations between PTA (left and right) and final block accuracy was not significant for older adults ($p>0.5$). Taken together, these data suggest that differences in hearing status between older and younger adults do not confound task performance.

Modeling Results

The accuracy based analyses suggest an age-related performance deficit in Mandarin tone category learning, but tell us nothing about the nature of this age-related performance deficit; in particular, whether older and younger adults use similar strategies for solving the task, but with older adults using a more sub-optimal version of that strategy, or whether older and younger adults use qualitatively different strategies to solve the task. Model-based analyses provide this window onto cognitive processing.

Computational Modeling—We fit a series of decision-bound models on a block-by-block basis at the individual participant level because of problems with interpreting fits to aggregate data (Ashby, Maddox, & Lee, 1994; Estes, 1956; Maddox, 1999). We assume that the two-dimensional space (pitch height vs. pitch direction) displayed in Figure 4 accurately describes the perceptual representation of the stimuli. We acknowledge that this is a reductionist approach, given the number of cues that differentiate tone categories (as with any speech category). However, our previous work (Maddox & Chandrasekaran, 2013) revealed significant insights (over traditional accuracy measures) that validate this innovative approach. Note that as long as the major dimensions are known, these modeling procedures can be applied to any type of speech category structure. For example, Maddox et al. (2002) found that the Striatal Pattern Classifier (SPC; Ashby & Waldron, 1999), a computational model of processing in the procedural based learning system that will be described next provided good fits for vowel category structures.

The Appendix provides details of each model, as well as the model fitting and model comparison procedure. Here we provide a brief description of each model, as well as an interpretation of the model results. Each model assumes that decision-bounds were used to classify stimuli into each of the four Mandarin tone categories (T1, T2, T3, or T4). We applied three classes of models. The first class is computational models of the implicit, procedural based learning category learning system. This is instantiated with the Striatal Pattern Classifier (SPC). The SPC is a computational model whose processing is consistent with the neurobiology of the procedural based category learning system and is thought to underlie information-integration (II) classification performance (Nomura, et al., 2007; Seger & Cincotta, 2005). The second class is models of the explicit, hypothesis-testing system (Maddox, Ashby, & Bohil, 2003; Maddox, Filoteo, Hejl, & Ing, 2004). A number of conjunctive and uni-dimensional rule-based (RB) models were examined. Conjunctive RB models assume that the participant sets criteria along the pitch height and pitch direction dimensions that are then combined to determine category membership. Uni-dimensional RB models assume that the participant sets criteria along the pitch height or pitch direction dimension that are then used to determine category membership. The third model is a random responder (RR) model that assumes that the participant guesses on each trial.

In a previous study we found that many learners use separate male and female perceptual spaces during category learning (Maddox & Chandrasekaran, 2013). We therefore also examined talker separation models. The model procedure described before assumes that each model is applied to a block of 40-trials using the 40 stimuli displayed in Figure 4a is effectively a modeling procedure that assumes no Talker Separation (hereafter referred to as Non-Separation models). To model the presence of Talker Separation (hereafter referred to as Separation models), we assumed that the participant converted the 40 stimulus perceptual space in Figure 4a into two separate perceptual spaces, one that characterizes the 20 stimuli spoken by the male talker and one that characterizes the 20 stimuli spoken by the female talker. A scatterplot of the stimuli associated with the male and female sub-perceptual spaces are displayed in Figures 4b and 4c, respectively. We fit each of the models outlined above separately to the 20-trials with a female speaker and the 20-trials with the male speaker.

Distribution of the 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-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Separation models. Table 2 displays the number of younger and older 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 separately for younger and older adults. A χ^2 test suggested that the number of Separators did increase while the number of Non-Separators did decrease from the first to the final block of trials for younger [$\chi^2(1) = 12.05$, $p < .001$] and older adults [$\chi^2(1) = 16.57$, $p < .001$]. To determine whether the rate of change in the number of Separators and Non-Separators differed as a function of age, we compared the number of younger and older Non-Separators across the first and last block. A χ^2 test suggested that age did not affect the distribution of Non-Separators across the first to the final block of trials [$\chi^2(1) = .001$, $p = .97$], nor did it affect the distribution of Separators across the first to the final block of trials [$\chi^2(1) = .63$, $p = .43$]. Thus, older and younger adults did not differ in their ability to use talker-dependent strategies that more effectively parse out the four categories.

Rule-Based and Information-Integration Strategies and Accuracy Rates for Final Block Separator—Although we did not find differences in the proportion of Separators and Non-Separators across age groups there is reason to believe that we will see differences in the types of strategies used. In this section we examine performance for the reflective, rule-based and reflexive, information-integration strategies used by older and younger adults whose final block of data was best fit by a Separation model. Of the 25 older adult final block Separators, 7 were best fit by the SPC, 1 by the Conjunctive rule-based model, 13 by the Uni-dimensional_Height model, and 4 by the random responder model. Of the 29 younger adult final block Separators, 11 were best fit by the SPC, 9 by the Conjunctive rule-based model, and 7 by the Uni-dimensional_Height model, and 2 by the random responder model. Because the SPC and Conjunctive rule-based strategies often yield similar accuracy rates, they can be difficult to tease apart. However, both involve processing of both stimulus dimensions, in contrast to the uni-dimensional model that involves the processing of only a single dimension. In light of these facts, we combined the SPC and Conjunctive rule-based model frequencies and compare those with the combined uni-dimensional and random model frequencies. We found a significant age group difference in the distribution of model strategies [$\chi^2(1) = 7.35$, $p < .01$]. This difference suggests that younger adults are more likely to use two-dimensional strategies (n=20) than one-dimensional or random strategies (n=9), whereas older adults are more likely to use one-dimensional or random strategies (n=17) than two-dimensional strategies (n=8).

Because the optimal strategy clearly requires a two-dimensional strategy, we also predicted that participants using a two-dimensional strategy would outperform those using a one-dimensional or random strategy. As a test of this hypothesis, and to determine whether this might account for the age-based performance difference, we conducted a 2 age group × 2 strategy group (two-dimensional vs. one-dimensional) ANOVA. These data are presented in Figure 5b. As expected we observed a main effect of age group [$F(1, 50) = 18.69$, $p < .001$, partial $\eta^2 = .272$] with younger adults outperforming older adults. We also observed a main effect of strategy group [$F(1, 50) = 39.02$, $p < .001$, partial $\eta^2 = .438$] with participants using a two-dimensional strategy outperforming those using a one-dimensional or random model strategy. These were qualified by a significant interaction [$F(1, 50) = 7.96$, $p < .01$, partial $\eta^2 = .137$]. Post hoc analyses suggested that there was a large and significant age-based performance difference for participants using the two-dimensional strategy [$t(26) = 5.36$, $p < .001$], but not for the participants using the one-dimensional or random strategies [$t(24) = 1.00$, $= .33$].

Summary

Experiment 3 revealed an age-based Mandarin tone category learning deficit across 200 trials of training. The model-based analyses suggested that older adults were able to separate by talker at the same rate as younger adults, but were much less likely to learn and apply a two-dimensional strategy to solve the task, instead relying heavily on sub-optimal uni-dimensional strategies (for a similar finding in the visual domain see Maddox, et al., 2010).

Working Memory Correlates of OA Performance

In this section we explore the working memory correlates of performance in each of the three experiments using the classic digit span task as our measure of capacity. The predictions for the rule-based (Experiment 1) and information-integration (Experiment 2) tasks are straightforward. Rule-based category learning is heavily dependent on working memory, whereas information-integration category learning is not. Thus, we predict that working memory capacity will correlate with rule-based category learning performance but not with information-integration category learning performance. In a recent study using younger adults and visually presented stimuli, DeCaro et al. (2008; however see Lewandowsky, Yang, Newell, & Kalish, 2012; Tharp & Pickering, 2008) found that working memory capacity was positively correlated with rule-based performance but was negatively correlated with information-integration performance.

We also correlated category learning performance with a number of other measures from our neuropsychological battery. These include: Trails A, Trails B, total number of errors in the WCST, and Stroop interference. Although no strong a priori predictions are offered for these measures, these are included for completeness. The correlations for all three Experiments are presented in Table 3 with correlations that are significant at the .05 level in bold type.

The results are clear. In the Experiment 1 rule-based task, high working memory capacity was associated with fewer trials needed to learn the task. In the Experiment 2 information-integration task, on the other hand, high working memory was associated with slower learning, although this result did not reach significant. In the Experiment 3, speech category learning task, high working memory was associated with better final block accuracy. Finally, in all studies, none of the other measures correlated significantly with category learning performance. Taken together, these results support the previous work suggesting working memory correlates of rule-based but not information-integration category learning, and extend it to the auditory realm and to healthy aging.

General Discussion

To our knowledge, this study represents the first attempt to comprehensively examine age-related changes in auditory category learning. The aim was to examine the extent to which the dual-systems model accounts for auditory category learning across the lifespan. We examined age-related changes in learning of simple auditory sounds that were grouped into rule-based (Experiment 1) or information-integration (Experiment 2) category structures. In Experiment 3, we examined age-related changes in learning of non-native tone categories. Finally, we examined working memory correlates of performance in the older adults across all three experiments.

Our results demonstrate a clear age-related rule-based category learning deficit with more young adults learning the rule-based task, relative to older adults, and young adults learning the task faster than older adults. However, we found that older and younger adult participants did not differ significantly in the speed or proportion learning the information-integration task. This suggests that older adults have may preserved reflexive, procedural based learning ability, although clearly more work is needed. Interestingly, this is inconsistent with a previous study by Ashby and colleagues that used the same information-integration category structure but with visual stimuli (Ashby, Noble, Filoteo, Waldron, & Ell, 2003). We hypothesize that these may reflect differences in the neurobiology of the two sensory domains. Connectivity to the primary auditory cortex, a region that shows large age-related changes, is sparse. In contrast, connectivity to the secondary auditory cortex is large but diffuse. The diffuse nature of the connectivity patterns may potentially preserve learning across the lifespan. However, such a proposal needs to be tested with greater rigor. Indeed, although the difference in information-integration learning ability did not significantly differ between the two groups, younger adults were numerically faster than older adults to reach the learning criteria. Future studies are needed to establish the nature of this preserved ability, in the context of deficits in visual information-integration learning.

While Experiments 1 and 2 examined artificially created categories controlled to be rule-based or information-integration optimal, in Experiment 3, we examined age-related effects on natural auditory category learning. In Experiment 3, monolingual older and younger participants learned to distinguish Mandarin tone categories with feedback. Accuracy data revealed a clear age-related accuracy deficit across all blocks of learning. We further examined the nature of this deficit using computational models. We found that older adults were able to separate by talker at the same rate as younger adults but older adults tended to use more sub-optimal, uni-dimensional strategies. That is, older adults showed less effectiveness in utilizing both cues (pitch height and pitch direction), and instead predominantly used the pitch height cue (for a related finding in vision see Maddox, et al., 2010). Pitch height is an important cue in English, that distinguishes talker sex (e.g., low pitch='male'; high pitch='female'). However, the persistent use of this cue is sub-optimal. As seen in Figures 4 and A1, the use of rules based on pitch height would lead to confusions between tone 2 and tone 4, which are distinguished based on pitch direction (tone 2 is 'rising'; tone 4 is 'falling'), but have similar pitch height. Younger adults, on the other hand, use pitch height and pitch direction either in conjunction (reflective strategy) or integrate these two dimensions (reflexive strategy). Thus, they predominantly use both dimensions, which likely requires a refocus (either attentionally in the form of a conjunctive rule, or predecisionally in the form of reflexive SPC strategy) from language-specific strategies.

Taken together, the results from the three experiments suggest that the reflective system, which uses verbal rules to develop hypothesis about category structure is more affected by systemic aging than the 'procedural-based' reflexive system. These findings parallel results from studies examining language processing in older adults. Dual-system models in the language domain have posited that grammar learning is predominantly procedural-based, while vocabulary learning uses the declarative system (Morgan-Short, Finger, Grey, & Ullman, 2012; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010). Extrapolating from our results, an interesting question is the extent to which aging affects vocabulary learning relative to grammar learning. Indeed, consistent with our findings, a recent study showed preserved implicit grammar learning ability in older adults relative to younger adults (Kurten, De Vries, Kowal, Zwitserlood, & Floel, 2012). In contrast, vocabulary learning is impaired in older adults, relative to younger adults (Service & Craik, 1993).

The computational modeling results can inform the development of more optimal training strategies for older adults. The use of a large number of talkers and high talker variability, may lead to more reliance on pitch direction as a dimension. On the other hand, the use of high variability ensures that a uni-dimensional rule is less rewarded and allows older adults to switch to a more reflexive, procedural strategy. Since reflexive learning appears to be better preserved, this may effectively enhance category learning success. Future work should explore this possibility. Another fruitful approach would be to include test items (not presented during training) that can be used to rigorously test between reflective rule-based and reflexive information-integration strategies, as in work by Mata and colleagues (Mata, von Helversen, Karlsson, & Cupper, 2012). As outlined above, information-integration and conjunctive strategies yield very similar accuracy rates in the Experiment 3 stimulus set. The inclusion of additional test trials could be used to more accurate tease apart these two strategies.

We also explored the relationship between working memory capacity and performance in older adults by correlating the backward span score with trials-to-criterion in Experiments 1 and 2, and final block accuracy in Experiment 3. As predicted from recent work in the visual domain (DeCaro, et al., 2008; however see Lewandowsky, et al., 2012; Tharp & Pickering, 2008) we found that increased working memory capacity was significantly predictive of good rule-based category learning and was predictive of poor information-integration category learning, although the latter effect was not statistically significant. In the tone category learning experiment we found that increased working memory capacity was significantly correlated with good performance, which is in line with the modeling results that suggest the persistent use of simple one-dimensional rules (mostly pitch height) in older adults (Maddox, et al., 2010).

Conclusions

This represents the first study to comprehensively examine age-related changes in auditory category learning. We found an age-related deficit in rule-based category learning, but not in information-integration category learning. We found an age-related deficit in natural auditory category learning that was due to older adults' persistent reliance on sub-optimal, uni-dimensional strategies when two-dimensional strategies were optimal. Working memory capacity was found to correlate with perform in the rule-based and natural auditory category learning tasks, but not in the information-integration task. The implications of this work for second language learning in older adults were discussed.

References

Akaike H. A new look at the statistical model identification. Transactions on Automatic Control. 1974; 19:716–723.

Alexander JA, Wong PCM, Bradlow AR. Lexical tone perception in musicians and non-musicians. Interspeech. 2005

Aron AR, Shohamy D, Clark J, Myers C, Gluck MA, Poldrack RA. Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. J Neurophysiol. 2004; 92(2): 1144–1152. [PubMed: 15014103]

Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. Psychological Review. 1998; 105:442–481. [PubMed: 9697427]

Ashby FG, Ennis JM. The role of the basal ganglia in category learning. The Psychology of Learning and Motivation. 2006; 47(1–36)

Ashby FG, Maddox WT. Human Category Learning. Annual Review of Psychology. 2005; 56:149–178.

Ashby FG, Maddox WT. Human category learning 2.0. Annals of the New York Academy of Sciences. 2010

Ashby FG, Maddox WT, Lee WW. On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. Psychological Science. 1994; 5(3):144–151.

Ashby FG, Noble S, Filoteo JV, Waldron EM, Ell SW. Category learning deficits in Parkinson's disease. Neuropsychology. 2003; 17(1):115–124. [PubMed: 12597080]

Ashby FG, Paul E, Maddox WT. Wills EPA. COVIS 2.0. Formal Approaches in Categorization. 2011

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Ashby FG, Waldron EM. On the nature of implicit categorization. Psychonomic Bulletin & Review. 1999; 6(3):363–378. [PubMed: 12198775]

Bruner, JS.; Goodnow, J.; Austin, G. A study of thinking. New York: Wiley; 1956.

Ceponiene R, Westerfield M, Torki M, Townsend J. Modality-specificity of sensory aging in vision and audition: evidence from event-related potentials. Brain research. 2008; 1215:53–68. [PubMed: 18482717]

Chandrasekaran B, Yi H-G, Maddox WT. Dual-learning systems during speech categroy learning. Psychon Bull Rev. (in press).

Davis T, Love BC, Maddox WT. Age-related declines in the fidelity of newly acquired category representations. Learn Mem. 2012; 19(8):325–329. [PubMed: 22815536]

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]

Delis, DC.; Kramer, JW.; Kaplan, E.; Ober, BA. The California verbal learning test: Adult version manual. San Antonia, Texas: The Psychological Corporation; 1987.

Estes WK. The problem of inference from curves based on group data. Psychological Bulletin. 1956; 53:134–140. [PubMed: 13297917]

Filoteo JV, Maddox WT. A Quantitative Model-Based Approach to Examining Aging Effects on Information-Integration Category Learning. Psychology & Aging. 2004; 19(1):171–182. [PubMed: 15065940]

Filoteo, JV.; Maddox, WT. Category learning in Parkinson's Disease. In: Sun, MK., editor. Research Progress in Alzheimer's Disease and Dementia. Nova Sciences Publishers; 2007.

Filoteo JV, Maddox WT, Ing AD, Zizak V, Song DD. The impact of irrelevant dimensional variation on rule-based category learning in patients with Parkinson's disease. J Int Neuropsychol Soc. 2005; 11(5):503–513. [PubMed: 16212677]

Filoteo JV, Maddox WT, Salmon DP, Song DD. Information-Integration Category Learning in Patients With Striatal Dysfunction. Neuropsychology. 2005; 19(2):212–222. [PubMed: 15769205]

Filoteo JV, Maddox WT, Simmons AN, Ing AD, Cagigas XE, Matthews S, Paulus MP. Cortical and subcortical brain regions involved in rule-based category learning. Neuroreport. 2005; 16(2):111–115. [PubMed: 15671857]

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.

Gorlick MA, Giguere G, Glass BD, Nix BN, Mather M, Maddox WT. Attenuating Age-Related Learning Deficits: Emotional Valenced Feedback Interacts With Task Complexity. Emotion. 2012

Greenwood PM. The frontal aging hypothesis evaluated. J Int Neuropsychol Soc. 2000; 6(6):705–726. [PubMed: 11011517]

Greenwood PM. Functional plasticity in cognitive aging: review and hypothesis. Neuropsychology. 2007; 21(6):657–673. [PubMed: 17983277]

Grieve SM, Williams LM, Paul RH, Clark CR, Gordon E. Cognitive aging, executive function, and fractional anisotropy: a diffusion tensor MR imaging study. AJNR Am J Neuroradiol. 2007; 28(2):226–235. [PubMed: 17296985]

Guerreiro MJ, Murphy DR, Van Gerven PW. The role of sensory modality in age-related distraction: a critical review and a renewed view. Psychological bulletin. 2010; 136(6):975–1022. [PubMed: 21038938]

Guerreiro MJ, Murphy DR, Van Gerven PW. Making sense of age-related distractibility: the critical role of sensory modality. Acta psychologica. 2013; 142(2):184–194. [PubMed: 23337081]

Gunning-Dixon FM, Gur RC, Perkins AC, Schroeder L, Turner T, Turetsky BI, Gur RE. Age-related differences in brain activation during emotional face processing. Neurobiol Aging. 2003; 24(2):285–295. [PubMed: 12498962]

Heaton, RK.; Chelune, GJ.; Talley, JL.; Kay, GG.; Curtiss, G. Wisconsin Card Sorting Test manual. Odessa: Psychological Assessment Resources, Inc.; 1993.

Howie, J. Acoustical studies of Mandarin vowels and tones. Cambridge, UK: Cambridge University Press; 1976.

Keri S. The cognitive neuroscience of category learning. Brain Research Reviews. 2003; 43(1):85–109. [PubMed: 14499464]

Kurten J, De Vries MH, Kowal K, Zwitserlood P, Floel A. Age affects chunk-based, but not rule-based learning in artificial grammar acquisition. Neurobiol Aging. 2012; 33(7):1311–1317. [PubMed: 21093109]

Lewandowsky S, Yang LX, Newell BR, Kalish ML. Working memory does not dissociate between different perceptual categorization tasks. J Exp Psychol Learn Mem Cogn. 2012; 38(4):881–904. [PubMed: 22746954]

Lezak, MD. Neurpsychological assessment. 3rd edition. New York: Oxford University Press; 1995.

Lim SJ, Holt LL. Learning foreign sounds in an alien world: videogame training improves non-native speech categorization. Cogn Sci. 2011; 35(7):1390–1405. [PubMed: 21827533]

Lombardi WJ, Andreason PJ, Sirocco KY, Rio DE, Gross RE, Umhau JC, Hommer DW. Wisconsin Card Sorting Test performance following head injury: dorsolateral fronto-striatal circuit activity predicts perseveration. J Clin Exp Neuropsychol. 1999; 21(1):2–16. [PubMed: 10420997]

Love BC, Medin DL, Gureckis TM. SUSTAIN: A Network Model of Category Learning. Psychological Review. 2004; 111(2):309–332. [PubMed: 15065912]

Maddox WT. On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. Perception & Psychophysics. 1999; 61(2):354–375. [PubMed: 10089766]

Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. J Exp Psychol Learn Mem Cogn. 2003; 29(4):650–662. [PubMed: 12924865]

Maddox WT, Chandrasekaran B. Tests of a dual system model of speech category learning. under review. 2013

Maddox WT, Filoteo JV. Striatal contributions to category learning: Quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. Journal of the International Neuropsychological Society. 2001; 7(6):710–727. [PubMed: 11575593]

Maddox, WT.; Filoteo, JV. The Neuropsychology of Perceptual Category Learning. In: Cohen, H.; Lefebvre, C., editors. Handbook of Categorization in Cognitive Science. Elsevier, Ltd.; 2005. p. 573-599.

Maddox WT, Filoteo JV, Hejl KD, Ing AD. Category number impacts rule-based but not information-integration category learning: further evidence for dissociable category-learning systems. Journal of experimental psychology. Learning, memory, and cognition. 2004; 30(1):227–245.

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

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]

Maddox WT, Pacheco J, Reeves M, Zhu B, Schnyer DM. Rule-based and information-integration category learning in normal aging. Neuropsychologia. 2010; 48(10):2998–3008. [PubMed: 20547171]

Mata R, von Helversen B, Karlsson L, Cupper L. Adult age differences in categorization and multiple-cue judgment. Dev Psychol. 2012; 48(4):1188–1201. [PubMed: 22059450]

Monchi O, Petrides M, Petre V, Worsley K, Dagher A. Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. J Neurosci. 2001; 21(19):7733–7741. [PubMed: 11567063]

Morgan-Short K, Finger I, Grey S, Ullman MT. Second language processing shows increased native-like neural responses after months of no exposure. PLoS One. 2012; 7(3):e32974. [PubMed: 22470434]

Morgan-Short K, Sanz C, Steinhauer K, Ullman MT. Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. Lang Learn. 2010; 60(1):154–193. [PubMed: 21359123]

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]

Park DC, Reuter-Lorenz P. The adaptive brain: aging and neurocognitive scaffolding. Annu Rev Psychol. 2009; 60:173–196. [PubMed: 19035823]

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. 2011; 130(1):461–472. [PubMed: 21786912]

Petrides M, Pandya DN. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. The Journal of comparative neurology. 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]

Price A, Filoteo JV, Maddox WT. Rule-based category learning in patients with Parkinson's disease. Neuropsychologia. 2009; 47(5):1213–1226. [PubMed: 19428385]

Racine CA, Barch DM, Braver TS, Noelle DC. The effect of age on rule-based category learning. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. 2006; 13(3–4):411–434. [PubMed: 16887781]

Raz, N. Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In: Craik, FIM.; Salthouse, TA., editors. Handbook of Aging and Cognition. Nahwah: Lawrence Erlbaum Associates; 2000. p. 1-90.

Raz N, Rodrigue KM, Kennedy KM, Head D, Gunning-Dixon F, Acker JD. Differential aging of the human striatum: longitudinal evidence. AJNR Am J Neuroradiol. 2003; 24(9):1849–1856. [PubMed: 14561615]

Reuter-Lorenz PA, Park DC. Human neuroscience and the aging mind: a new look at old problems. J Gerontol B Psychol Sci Soc Sci. 2010; 65(4):405–415. [PubMed: 20478901]

Ridderinkhof KR, Span MM, van der Molen MW. Perseverative behavior and adaptive control in older adults: performance monitoring, rule induction, and set shifting. Brain Cogn. 2002; 49(3):382–401. [PubMed: 12139960]

Schnyer DM, Maddox WT, Ell S, Davis S, Pacheco J, Verfaellie M. Prefrontal contributions to rule-based and information-integration category learning. Neuropsychologia. 2009; 47(13):2995–3006. [PubMed: 19643119]

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. Journal of Neuroscience. 2005; 25(11):2941–2951. [PubMed: 15772354]

Seger CA, Cincotta CM. Dynamics of frontal, striatal, and hippocampal systems during rule learning. Cereb Cortex. 2006; 16(11):1546–1555. [PubMed: 16373455]

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

Service E, Craik FIM. Differences between Young and Older Adults in Learning a Foreign Vocabulary. Journal of Memory and Language. 1993; 32(5):608–623.

Shepard RN, Hovland CI, Jenkins HM. Learning and memorization of classifications. Psychological Monographs. 1961; 75(13)

Tharp IJ, Pickering AD. A note on DeCaro, Thomas, and Beilock (2008): Further data demonstrate complexities in the assessment of information-integration category learning. Cognition. 2008

Wagenmakers EJ, Farrell S. AIC model selection using Akaike weights. Psychon Bull Rev. 2004; 11(1):192–196. [PubMed: 15117008]

Waldschmidt GA, Ashby FG. Cortical and striatal contributions to automaticity in information-integration categorization. under review. 2011

Wechsler, D. Wechsler Adult Intelligence Scale - Third Edition. San Antonio: Harcourt Brace & Company; 1997.

Wilson, CJ. The contribution of cortical neurons to the firing pattern of striatal spiny neurons. Cambridge: MIT Press; 1995.

Wong PC, Perrachione TK, Gunasekera G, Chandrasekaran B. Communication disorders in speakers of tone languages: etiological bases and clinical considerations. Semin Speech Lang. 2009; 30(3): 162–173. [PubMed: 19711234]

Yeterian EH, Pandya DN. Corticostriatal connections of the superior temporal region in rhesus monkeys. J Comp Neurol. 1998; 399(3):384–402. [PubMed: 9733085]

Appendix

Model Details

Three classes of models were applied to the data and were used to classify stimuli into each of the four Mandarin tone categories (T1, T2, T3, or T4). The first provides a model of the implicit, procedural based learning category learning system and is instantiated with the Striatal Pattern Classifier (SPC). The SPC is a computational model whose processing is consistent with the neurobiology of the procedural based category learning system and is thought to underlie optimal information-integration (II) classification performance (Nomura, et al., 2007; Seger & Cincotta, 2005). 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 cortex to the striatum (Ashby & Ennis, 2006; Wilson, 1995), a low-resolution map of perceptual space is represented among the *striatal units*. During feedback-based 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 cortex. 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 resulting response region partitions, the SPC contains six free parameters--5 that determine the location of the units, and one that represents the noise associated with the placement of the striatal units. Fig. A1a displays a scatterplot of the responses and response regions for the four tone categories in Fig. 4a generated from a hypothetical participant using one version of the Striatal Pattern Classifier. It 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).

The second class of models instantiate explicit, hypothesis-testing strategies (Maddox, et al., 2003; Maddox, et al., 2004). A number of conjunctive and uni-dimensional rule-based (RB) models were examined. A conjunctive RB model assumes that the participant sets two criteria along the pitch height dimension and one criterion along the pitch direction dimension. 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. Stimuli with low pitch are classified as T3 and high pitch height items are classified as T1. If an item is classified as of medium pitch height then the pitch direction is examined to discern between T2 and T4 (making this a conjunction of the two dimensions). Stimuli with medium pitch height and negative slopes are classified as T4 and those with medium pitch height and positive slopes are classified as T2. Fig. 6b displays a scatterplot of the responses and response regions for the four tone categories in Fig. 4a generated from a hypothetical participant using one version of the Conjunctive model. This model contains four free parameters—three criteria and one noise parameter. 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. <u>This model ignores the pitch direction dimension</u>. We will examine the 8 most reasonable variants of the model that differ only in the assignment of categories to response regions (low, medium-low, medium-high and high). Fig. A1c displays a scatterplot of the responses and response regions for the four tone categories generated from a hypothetical participant using one version of the Uni-Dimensional_Height model. We also examined a Uni-Dimensional_Direction model that separated the direction dimension into four response regions while ignoring pitch height. Figure A1d displays a scatterplot of the responses and response regions for the four tone categories from a hypothetical participant using one version of the Uni-Dimensional_Direction model. The Uni-Dimensional models each contain four free parameters—three criteria and one noise parameter and ignore the second dimension.

The third model is a random responder (RR) model that assumes that the participant guesses on each trial.

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:

$$AIC_i = -2 log L_i + 2V_i \quad \text{(A-1)}$$

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:

$$\Delta_i(AIC) = AIC_i - min AIC \quad \text{(A-2)}$$

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

$$L(M_i|data) \propto \exp\left\{-\frac{1}{2}\Delta_{\mathrm{i}}(\mathrm{AIC})\right\} \quad \text{(A-3)}$$

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:

$$w_i(AIC) = \frac{\exp\left\{-\frac{1}{2}\Delta_{\mathrm{i}}(\mathrm{AIC})\right\}}{\exp\left\{-\frac{1}{2}\Delta_{\mathrm{k}}(\mathrm{AIC})\right\}} \quad \text{(A-4)}$$

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. For example, 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 for older adults were .781, .759, .820, .827, and .865 in blocks 1 – 6, respectively. The resulting Akaike weights for younger adults were .868, .893, .917, .945, and .935 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 < .01) 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 .806 – .899 for older adults and from .905 – .958 for younger adults and in every block were significantly above 0.5 based on a one-sample t-test (all p’s < .01). When a Non-Separation model provided the best account of the data, the Akaike weights ranged from .758 – .908 for older adults and from .806 – .879 for younger adults 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 best fitting model (separation or non-separation) is capturing meaningful strategic variance in the data and not just random noise.

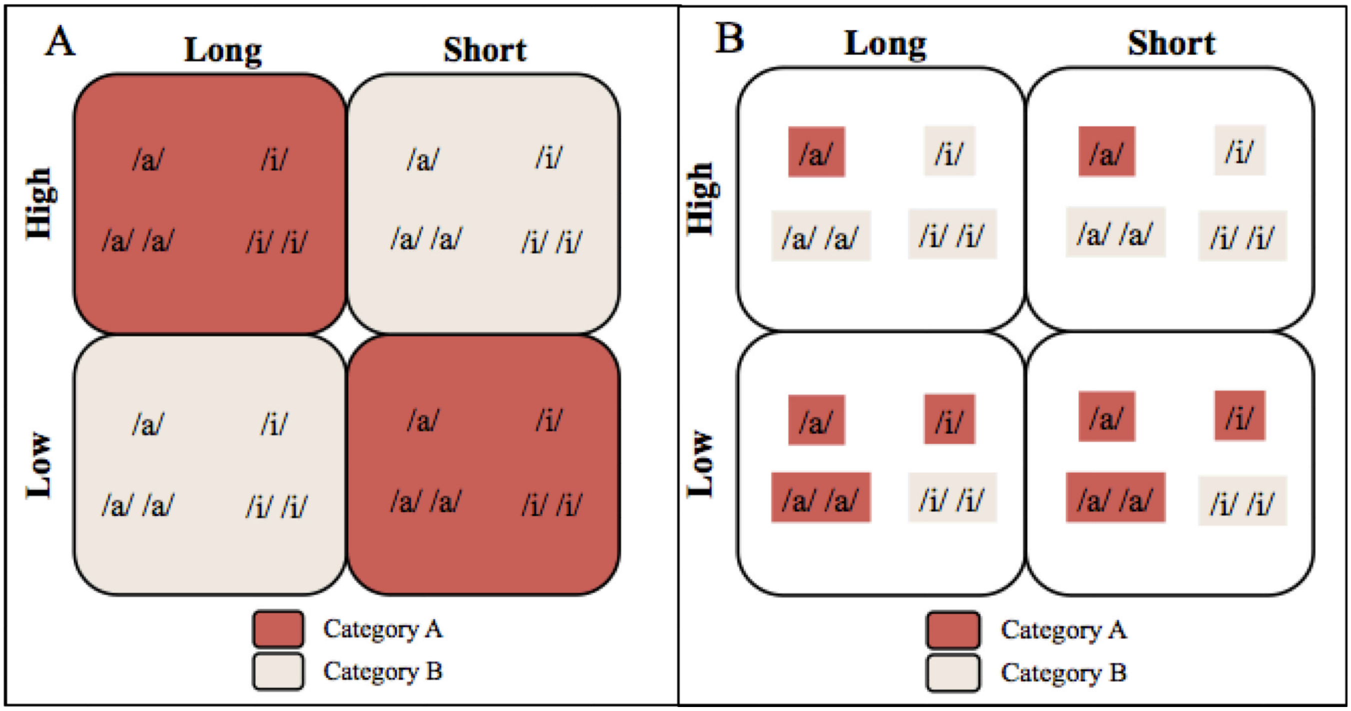

Figure 1.
(A) A schematic of one possible conjunctive rule-based problem in which the pitch (High or Low) and duration (Long or Short) dimensions are relevant. (B) A schematic of one possible information-integration problem in which the duration dimension (Long or Short) is irrelevant.

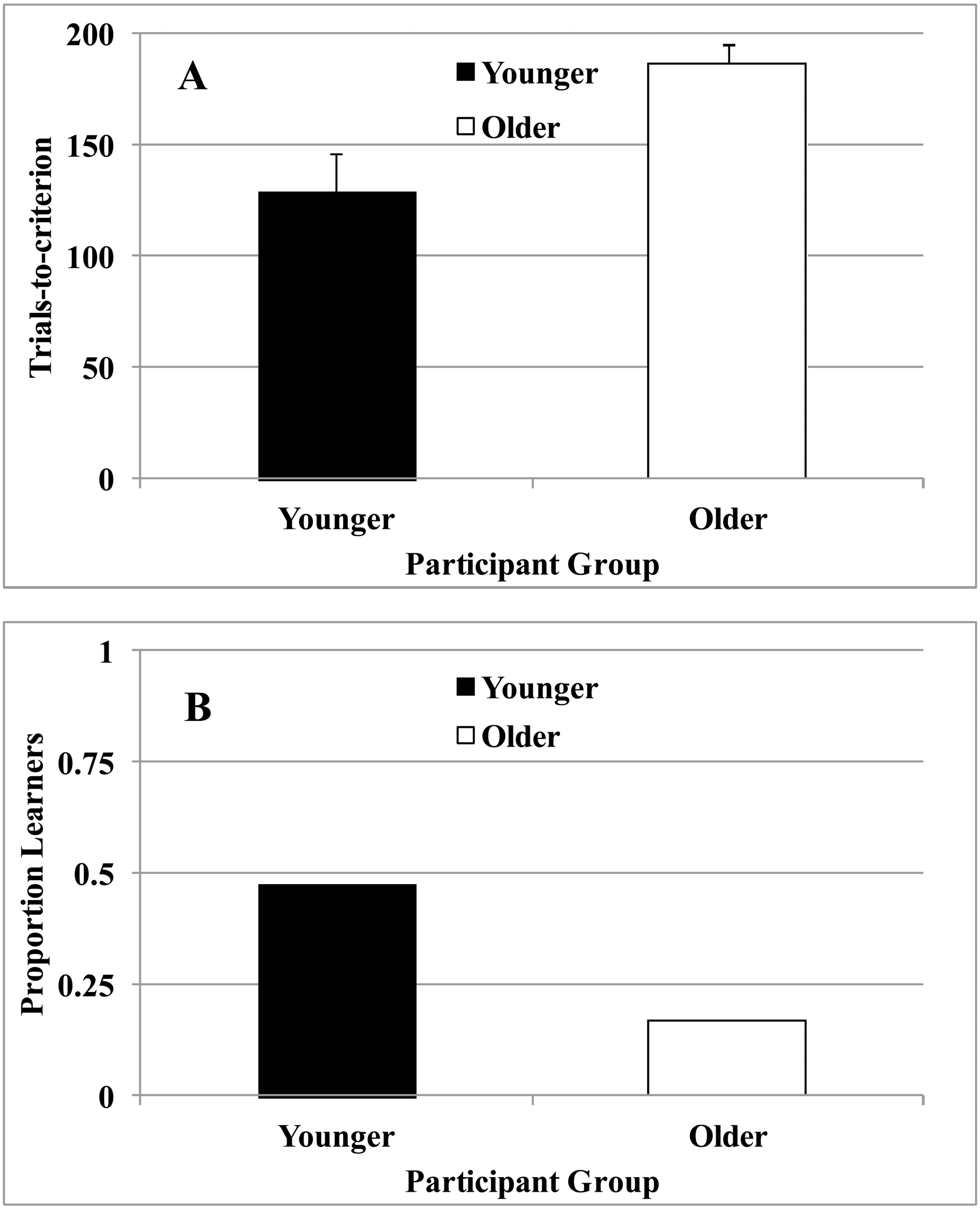

Figure 2.
(A) Trials-to-criterion for younger and older adults in the rule-based category learning condition. Standard error bars included. (B) Proportion of younger and older adult learners in the rule-based category learning condition.

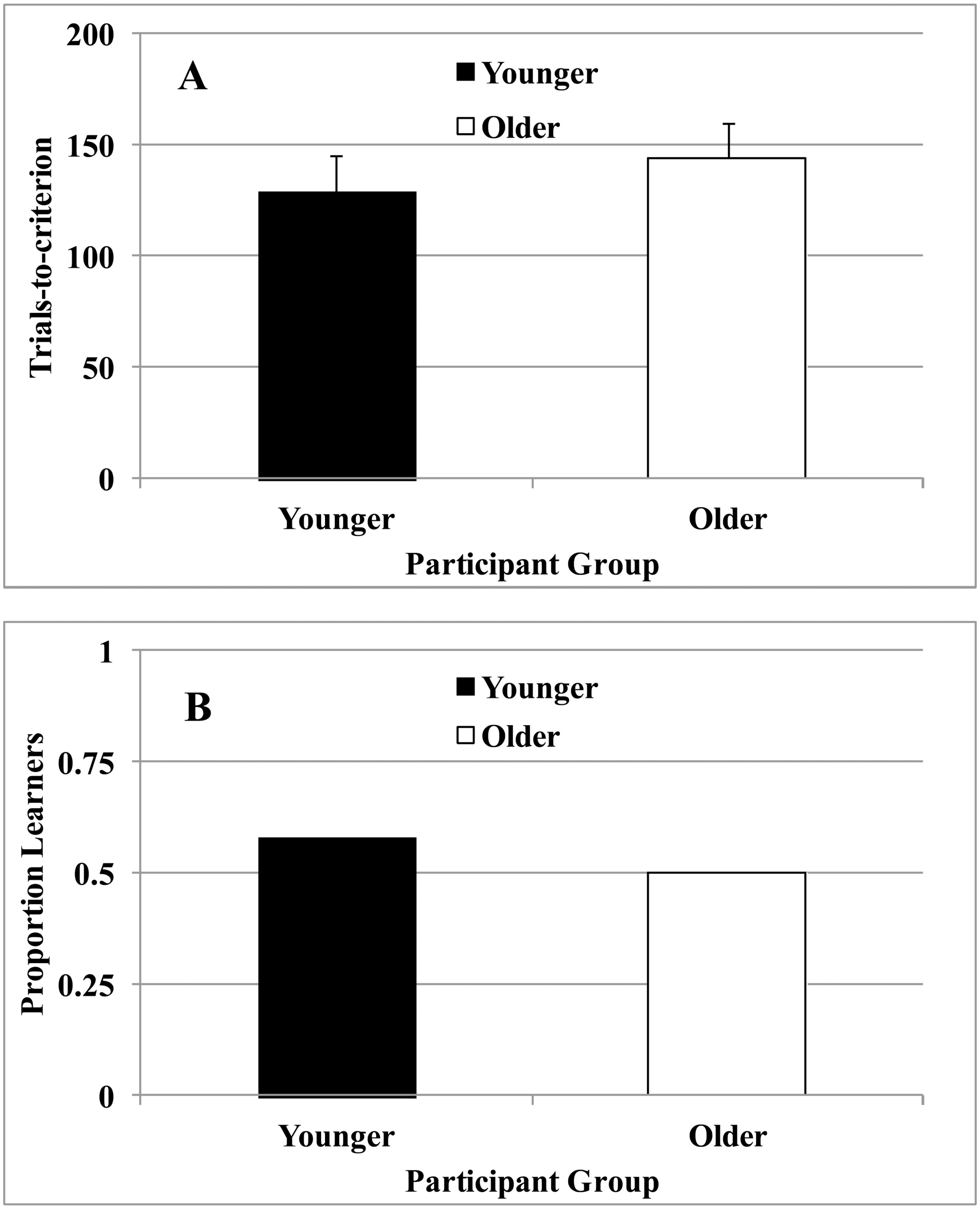

Figure 3.
(A) Trials-to-criterion for younger and older adults in the information-integration category learning condition. Standard error bars included. (B) Proportion of younger and older adult learners in the information-integration category learning condition.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

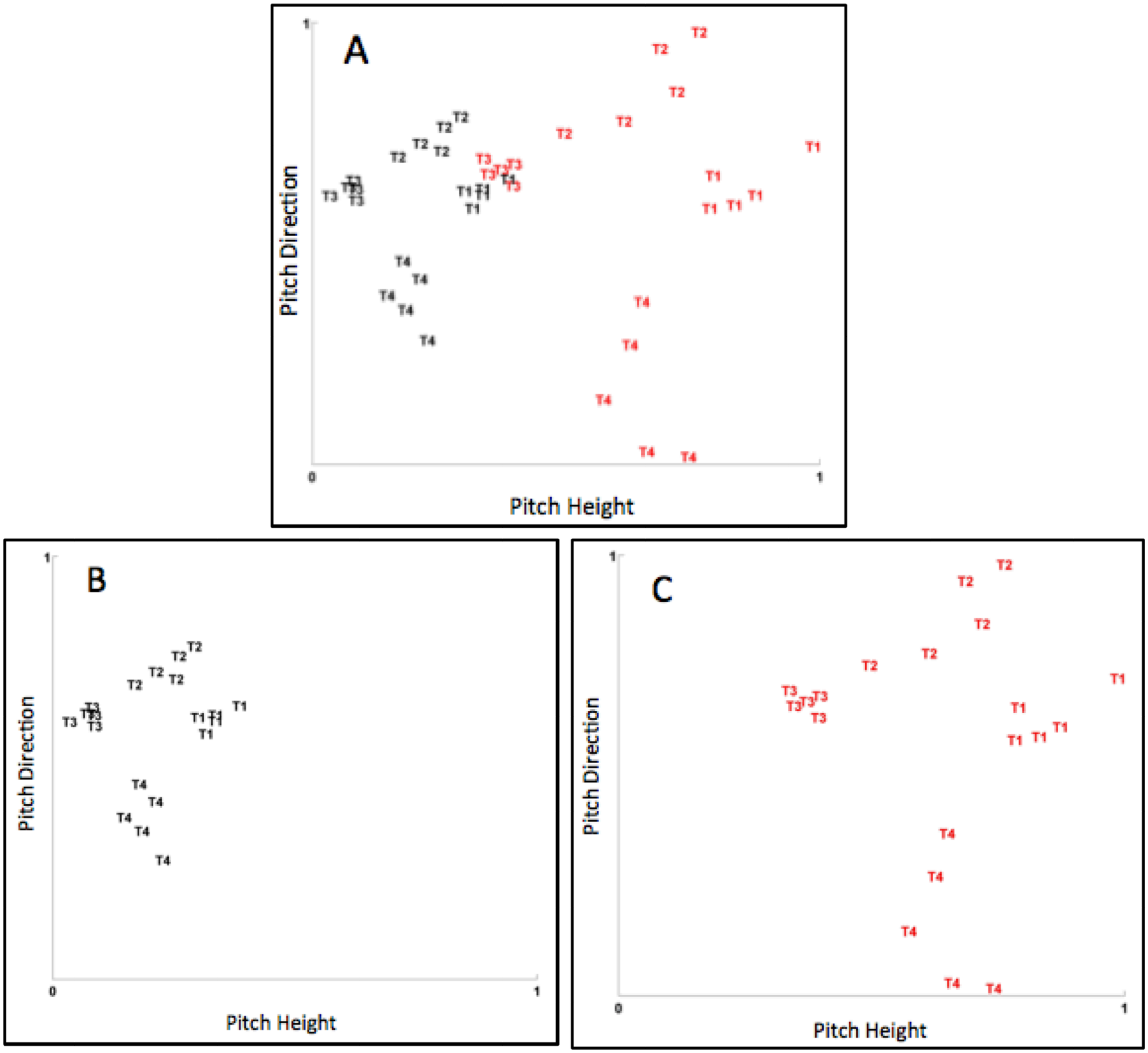

Figure 4.
(A) Scatterplot of all stimuli from Experiment 3. (B) Scatterplot of male-talker stimuli from Experiment 3. (C) Scatterplot of female-talker stimuli from Experiment 3. Stimuli dimensions (Pitch Height and Pitch Direction) were normalized between 0 and 1.

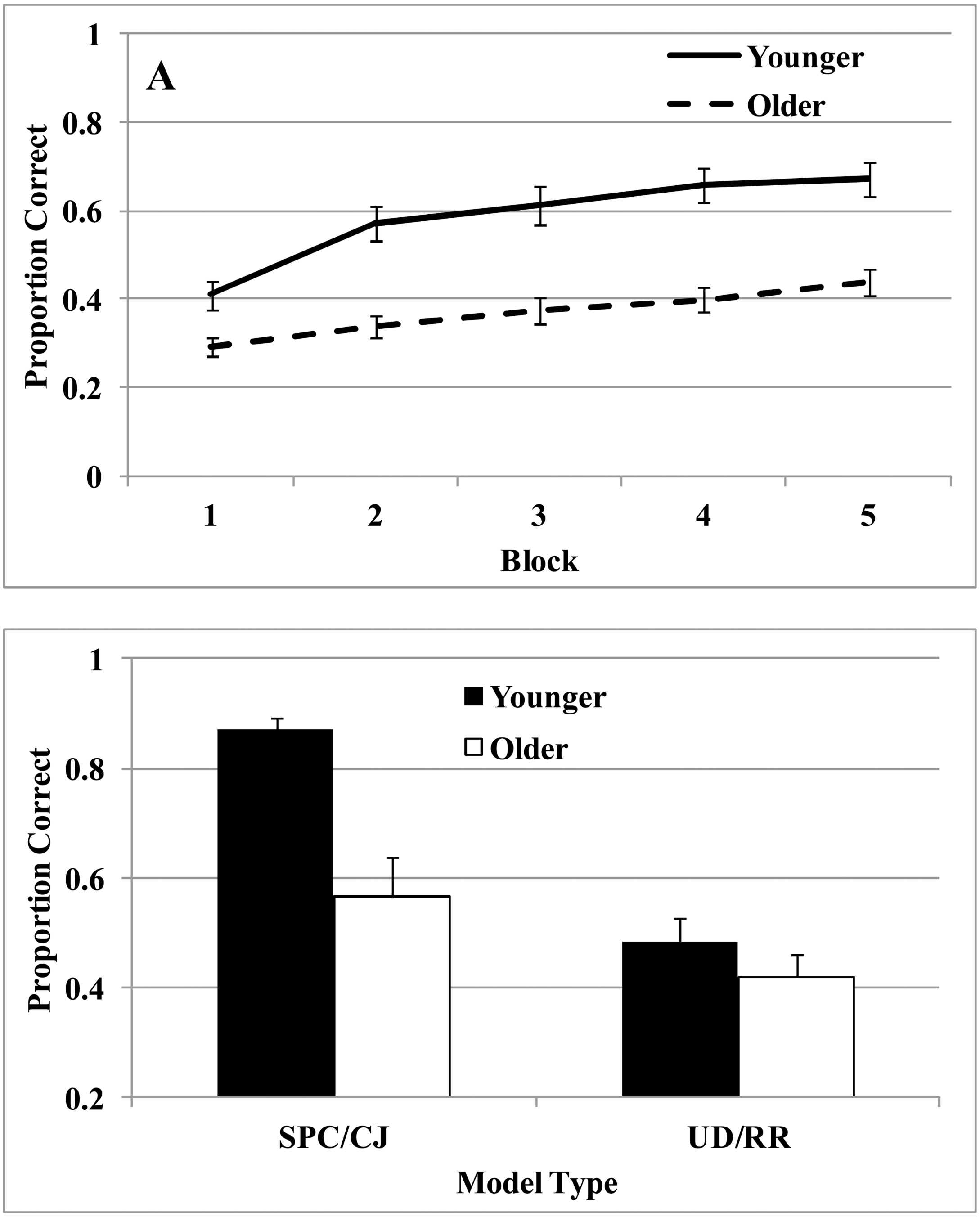

Figure 5.
(A) Proportion correct for younger and older adults in the Mandarin tone category learning task. (B) Final block proportion correct for younger and older adult Separators as a function of model type. Standard error bars included.

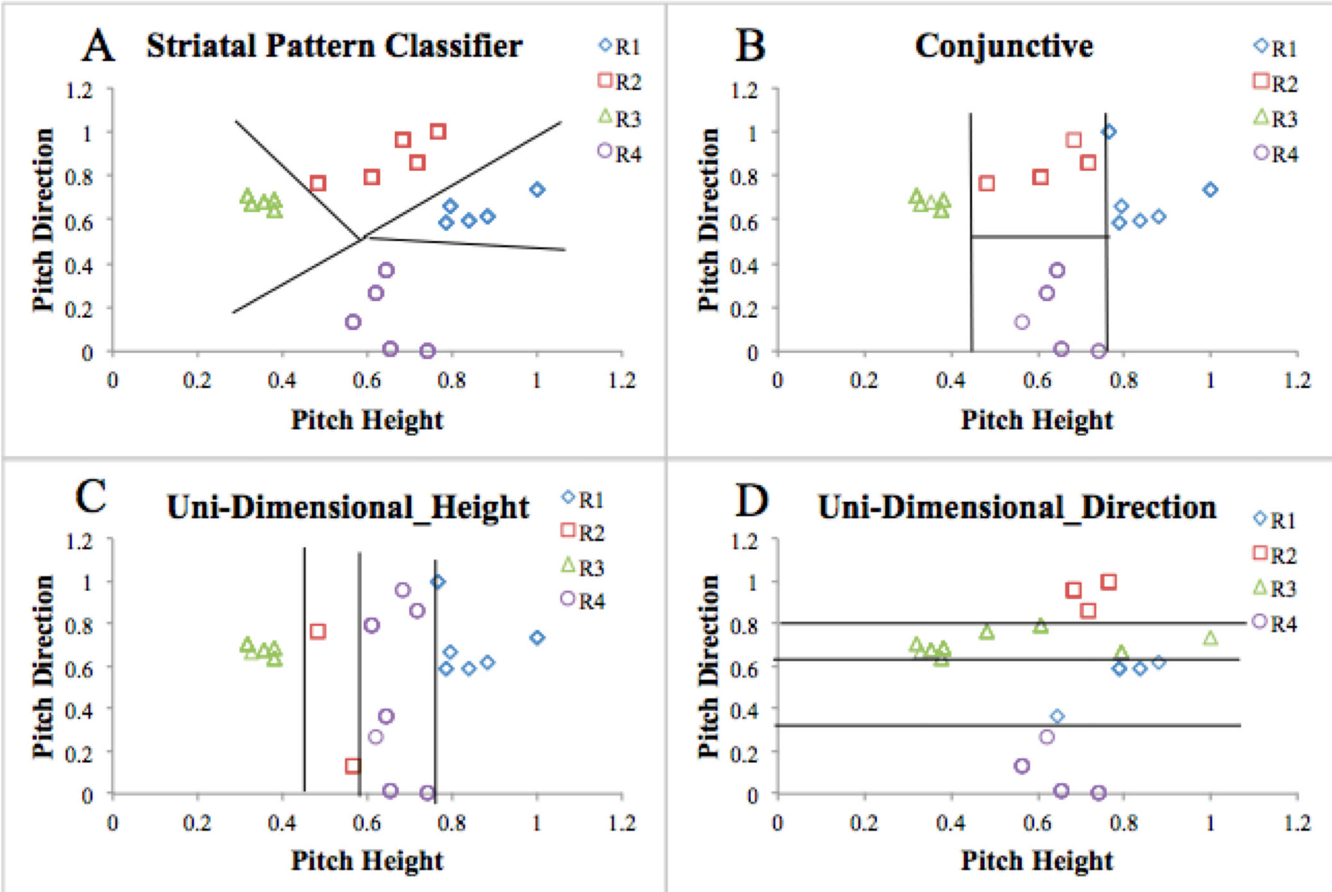

Figure A1.
Scatterplots of the responses along with the decision boundaries that separate response regions from a hypothetical participant using a version of the (A) Striatal Pattern Classifier, (B) Conjunctive rule-based, (C) Uni-Dimensional_Height, and (D) Uni-Dimensional_Direction models as applied to the female talker stimuli shown in Fig 4C.

Table 1

Z-scores Summary for each Neuropsychological Exam

A

Neuropsychological Test Experiment 1	Mean (SD)	Range
WAIS Vocabulary	0.54 (0.78)	−0.3–2.0
Digit Span	−0.02 (0.74)	−1.0–1.3
CVLT Delayed Recall (Free)	0.82 (0.93)	−0.5–2.0
CVLT Immediate Recall (Free)	0.97 (0.91)	−1.0–2.0
CVLT Delayed Recall (Cued)	0.50 (0.97)	−1.0–2.0
CVLT Immediate Recall (Cued)	0.71 (0.90)	−1.0–2.5
CVLT Recognition False Positives	−0.38 (0.72)	−1.0–1.0
CVLT Recognition True Positives	0.03 (0.84)	−1.5–1.0
FAS	−0.21 (1.06)	−2.0–1.7
Trails A	−0.36 (0.93)	−1.3–1.9
Trails B	−0.47 (0.50)	−1.1–1.0
WCST Errors	0.48 (1.14)	−1.5–2.1
WCST Perseveration	0.75 (1.11)	−1.0–3.0
Demographic Information		
Age	67.59 (5.18)	60–82
Years of Education	16.71 (3.02)	10–25

B

Neuropsychological Test Experiment 2	Mean (SD)	Range
WAIS Vocabulary	0.67 (0.88)	−1.3–1.7
Digit Span	0.39 (1.02)	−1.3–2.7
CVLT Delayed Recall (Free)	0.44 (0.93)	−1.0–2.0
CVLT Immediate Recall (Free)	0.53 (0.78)	−0.5–2.0
CVLT Delayed Recall (Cued)	0.38 (0.84)	−1.0–2.0
CVLT Immediate Recall (Cued)	0.38 (0.89)	−1.5–2.0
CVLT Recognition False Positives	−0.18 (1.12)	−1.0–3.0
CVLT Recognition True Positives	0.00 (0.77)	−1.5–1.0
FAS	−0.00 (1.07)	−2.6–1.3
Trails A	−0.32 (0.77)	−1.3–1.2
Trails B	−0.31 (0.72)	−1.2–1.2
WCST Errors	0.11 (1.19)	−2.3–1.8
WCST Perseveration	0.42 (0.69)	−0.5–2.1

B

Neuropsychological Test Experiment 2		
Demographic Information		
Age	68.41 (4.65)	60–78
Years of Education	16.76 (1.71)	12–19

C

Neuropsychological Test Experiment 3		
	Mean (SD)	Range
WAIS Vocabulary	0.72 (0.85)	−1.3–2.0
Digit Span	0.48 (0.92)	−1.3–2.7
CVLT Delayed Recall (Free)	0.43 (0.91)	−1.0–2.5
CVLT Immediate Recall (Free)	0.43 (0.81)	−1.0–2.5
CVLT Delayed Recall (Cued)	0.33 (0.76)	−1.0–2.0
CVLT Immediate Recall (Cued)	0.36 (0.85)	−1.5–2.5
CVLT Recognition False Positives	−0.14 (1.03)	−1.0–3.0
CVLT Recognition True Positives	−0.13 (1.02)	−3.0–1.0
FAS	0.02 (0.98)	−2.6–1.7
Trails A	−0.41 (0.95)	−1.8–3.1
Trails B	−0.43 (0.70)	−2.1–1.2
WCST Errors	0.26 (0.96)	−2.3–2.1
WCST Perseveration	0.46 (0.68)	−0.5–2.5
Demographic Information		
Age	68.14 (5.78)	60–83
Years of Education	16.71 (2.27)	10–22

Note: Mean z-scores for each exam with standard deviation in parenthesis and z-score range.

Table 2

Number of younger and older adults best fit by a non-separation or separation model in each block.

		Block				
		1	2	3	4	5
Younger	Non-Separation	24	11	14	8	9
	Separation	14	27	24	30	29
Older	Non-Separation	27	17	16	9	10
	Separation	8	18	19	26	25

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Correlations between category learning performance and neuropsychological measures

Exp	Digit Span	Trails A	Trails B	WCST Errors	Stroop Interference
1	**−.52***	.31	.37	−.14	.14
2	.23	−.08	−.29	.33	.12
3	**.41***	−.16	−.24	.24	.01

Note:

* = p<.05