Tonal Aspects of Languages 2016
24-27 May 2016, Buffalo, New York

Stability and plasticity in the neural representation of linguistic pitch patterns

Rachel Tessmer[1], *Bharath Chandrasekaran*[1, 2]

[1] Department of Communication Sciences and Disorders, The University of Texas at Austin, Texas, USA

[2] Department of Psychology, The University of Texas at Austin, Texas, USA

bchandra@utexas.edu

Abstract

In this paper we discuss evidence that early, long-term experience with a tonal language leaves a stable and lasting influence on the neural encoding of linguistic pitch patterns. We will also demonstrate that plastic changes in the neural representation of linguistic pitch are possible during adulthood, after just a few hundred trials of sound-to-category mapping training that establish behavioral relevance of the incoming stimuli. Notably, such category training creates large individual differences in learning success, the reasons for which are unclear. Within the context of a novel theoretical approach, the dual-learning systems (DLS) model, we will discuss the mechanisms underlying tone learning in adulthood, sources of individual differences, and the potential to design optimized training approaches that reduce individual differences.

Index Terms: tone, individual differences, speech perception, fMRI, EEG, plasticity

1. Introduction

The concept of an 'auditory critical period' is extremely popular in language and speech learning literature [1]. The critical period view posits that language learning competency dramatically declines with age [2, 3]. Consequently, adults who try to learn a second language may have disproportionate difficulty achieving native-like fluency and proficiency. Related to speech specifically, studies show that adults, relative to infants, are less sensitive to the critical acoustic dimensions that disambiguate non-native speech sounds [4, 5].

Several studies from our lab have examined at adult speech learning, well beyond the 'critical period'. We have found that, while individual variability is large, adults can learn to categorize foreign speech sounds within a few blocks of auditory training [6, 7, 8, 9]. Some adults learn with consummate ease, as if they are not limited by a critical period. Others struggle and do not accurately categorize novel speech sounds even after substantial training. The long-term goal of this line of research is to identify the neural mechanisms underlying individual differences in learning success and leverage this knowledge in designing optimal training programs. The key premise is that adults are wired differently than children and infants, and therefore use different strategies to learn. Current training approaches are not tailored to the adult brain and learning strategies. In this article we discuss behavioral and neuroimaging studies that investigate the neural representation of speech signals and assess the impact of long-term and short-term training on neural representation. We then discuss the source of interindividual differences in learning, and discuss how we can optimize speech and language training programs. The common thread is our utilization of linguistic tones, specifically Mandarin lexical tone categories, as a speech category learning problem for native speakers of American English.

2. Neural representation of tones

Pitch is a primary cue in disambiguating tone categories. How pitch is extracted, encoded, and mapped to behaviorally-relevant categories has been studied in both animal and human models. A non-invasive method of investigating the representation of pitch patterns is through scalp-recorded electrophysiology. The frequency-following response (FFR) is a scalp-recorded electrophysiological response that reflects phase-locked activity from neural ensembles in the auditory system [10, 11]. The FFR faithfully represents fundamental frequency and harmonic components [12]. Filtering the FFR signal to isolate the contribution from the auditory brainstem can help form a neural reconstruction of the auditory signal ('a neurophonic') [10, 13]. The FFR also can represent time-varying pitch patterns, and therefore has been extensively utilized in studying lexical tone representation [14, 15, 16, 17, 18]. Krishnan and colleagues first examined FFRs to index neural pitch tracking in Mandarin Chinese and English speakers [16]. They found that Mandarin Chinese speakers showed stronger pitch representation, better pitch tracking, and stronger second harmonic representation than English speakers. Their findings demonstrate long-term plasticity in the fundamental representation of pitch. Song and colleagues measured FFRs to pitch patterns before and following sound-to-meaning auditory training and found enhanced neural pitch tracking after training [17].

Recent studies conducted in my lab use FFR to examine the impact of training on the representation of linguistic pitch patterns. In one experiment, five days of passive exposure did not change the neural representation of Mandarin tones in native English speakers. In a second experiment, training native English speakers with no prior Mandarin exposure with just 480 trials where they received feedback while categorizing tones enhanced their neural representation of these tones. In a third experiment, native English speakers underwent long-term training. With training, their feature representation became more native-like. Even following 2 months post-training, behavioral performance remained at a native-like accuracy; however, the neural representation reverted to language-specific norms, suggesting a complex relationship between behavior and neural plasticity. Dramatic neural changes accompany behavioral training, but once categories are established in the brain, native-like behavior is subserved by minimal changes to neural function.

 http://dx.doi.org/10.21437/TAL.2016-1

3. Training to categorize lexical tones

Adult English speakers have difficulty perceiving tonal contrasts [19, 20]. The computational challenge arises from having to map continuous, multidimensional signals to meaningful, discrete categories [21, 22, 23, 24, 25, 26]. Previous research posits that interference from native language speech categories and perceptual warping due to prior language experience may contribute to difficulty learning non-native speech categories [27, 28, 29, 30, 31, 32]. However, despite these challenges, native English speakers with no tonal language experience can acquire pitch categories with extensive training [19, 7, 8, 9, 33].

Auditory training approaches need to overcome the inherent perceptual warping induced by the first language. For example, multidimensional scaling studies have revealed that at least two dimensions (pitch height and pitch direction) are critical for distinguishing tones across languages. While the pitch height dimension is weighted similarly by native and non-native speakers, pitch direction is weighted more by tonal language speakers, likely due to the greater relevance of this acoustic dimension in disambiguating tone categories across multiple speakers [34, 31, 35, 36]. Chandrasekaran, Sampath and Wong showed that attention to pitch direction increases in native English speakers with sound-to-meaning auditory training. FFRs to dynamic, and not static, portions of tones are more robust following training [37]. A recent study showed that explicitly instructing learners on relevant dimensions can enhance speech category learning [38]. Instructing participants to focus on pitch direction, not pitch height, resulted in more accurate tone categorization. Computational modeling results from the study suggest that pitch direction instruction led to faster and more frequent use of task-optimal multidimensional strategies (using both height and directional cues), as well as greater perceptual selectivity for pitch direction.

We have conducted several training studies that significantly improved (at the group level) native English speakers' tone classification ability. In this category learning task, participants are presented with natural speech productions by native Mandarin Chinese speakers and must learn to categorize these tones (map tones to button-presses) using trial-by-trial feedback. This task leverages prior work that has shown that generalizable learning requires 1) natural stimuli produced by multiple talkers to assist in 'talker normalization' and 2) trial-by-trial feedback to monitor errors [39, 40, 41, 42, 43, 8, 9, 44, 6]. Feedback has been shown to lead to greater learning ([45, 46]. Bradlow argues that the use of multiple talkers during training helps learners attend to talker-invariant acoustic cues that distinguish categories [47]. Multi-talker training has also been shown to lead to better generalization, suggesting variability in training may be beneficial to some participants [47, 41, 48]. Figure 1 shows the result of 5 blocks (consisting of 40 stimuli per block (2 talkers X 4 tone) of tone training using this approach [9].

4. Individual differences

A common thread seen across tone training studies is the immense individual variability (discernable in Figure 1). Some learners can categorize tones quickly and retain high accuracies across training, while others show slower progress, and still others seem to progress barely past chance performance. This large individual variability in learning success brings into question how it is that some individuals can learn these categories so optimally beyond the critical period, while others struggle (as the critical period hypothesis would predict). What strategies are successful and less successful learners using? By identifying strategy differences, could we modify training to reduce the challenges that result in individual variability? To explore sources of individual differences and develop training paradigms that reduce individual differences, we have adopted approaches from visual category learning literature. Our premise is that learning tone categories can be construed broadly as a categorization problem, one that involves mapping continuous, highly-variable sounds to discrete categories. Behavioral and neuroscience studies suggest the existence of multiple learning systems for categories [49, 50, 7]. In particular, the dual-learning systems (DLS) model, developed by our lab, posits two competing systems where learning occurs [7]. The *reflective* system is procedural, involves executive attention and working memory, and is involved in rule-based learning [33]. The *reflexive* system is procedural, does not involve working memory, and learns by associating perceptions with actions that are rewarded via feedback. The dominance of a system over the other comes from a shift in balance, as the reflective system is better suited for learning categories that are easily described by rules, while the reflexive system is better suited for learning categories that involve the integration of difficult to describe dimensions [7, 8, 44]. Explicit instructions provided early in training have shown enhanced speech category learning, suggesting reflective learning initially [7]. However, implicit training procedures that support reflexive learning have been shown to lead to more successful speech category learning [8]. The DLS model provides neurobiological sources for these learning systems. The reflective system is thought to rely upon an executive corticostriatal loop involving dorsolateral prefrontal cortex, the head of the caudate nucleus, the hippocampus, and the anterior cingulate cortex [8]. Meanwhile, the reflexive system is thought to rely upon cortical-striatal synapses in the ventral striatum, the caudate body, putamen, and primary motor cortex.

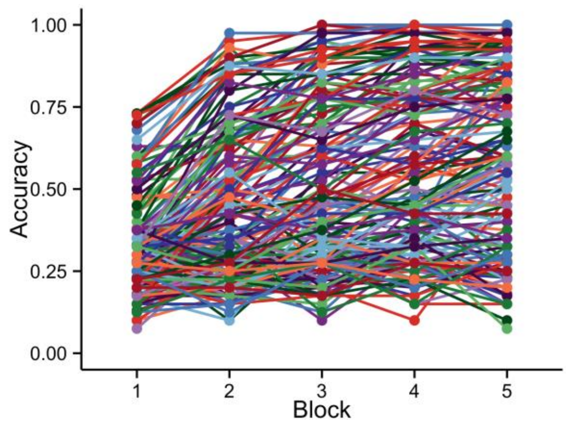

Figure 1: *Each line represents the learning curve for tone categorization accuracy across 5 training blocks, demonstrating the extent of individual variability in tone learning [9].*

The balance of control between the two learning systems may be an important reason for individual differences in successful learning of tone categories [51, 8]. Individual category learning success has been found to be associated with putamen and motor cortex activation [8]. These areas are thought to be key parts of the corticostriatal motor loop that is important for the reflexive learning system [52]. Maddox and Chandrasekaran posit that individual differences come from learners adopting and persisting with sub-optimal learning strategies involving unidimensional rules [44]. Across several studies, our lab has shown that learners who show the least gains in learning utilize simple unidimensional reflective rules (for example, only focusing on the pitch height dimension), while successful learners switch to multidimensional, reflexive strategies. Since reflexive strategies are not working-memory related, this may open up cognitive resources to attend to other key components of the task.

Research in our lab has looked at various aspects that could account for individual differences, including musical training, aging, and clinical factors. Musicians have shown an advantage in learning to categorize Mandarin tones [53, 54, 55, 56, 57]. One study investigated how aging affects tone category learning. An age-related deficit in performance was found, where older adults used multi-dimensional strategies significantly less than younger adults and were found to perseverate on unidimensional rules [51]. Younger adults were more likely than older adults to shift to more optimal, multi-dimensional strategies and to have higher accuracy. Another study found that individuals with elevated depressive symptoms showed an advantage in reflexive-optimal category learning as well as an advantage in learning non-native speech categories [58].

Candidate gene studies have also shed light on potential sources for these individual differences. *FOXP2* mutation has been linked to neurodevelopmental deficits in speech and language, thus prompting the study of this gene in spoken language in humans [59, 60]. When *FOXP2* was introduced to mice, striatal medium spiny neurons showed increased dendritic length and synaptic plasticity [61]. Additionally, different vocalizations were measured and less exploratory behavior was found. Another study found that mice with *FOXP2* introduced into their system learned stimulus-response associations faster [62]. These findings pose the question of how *FOXP2* contributes to human speech, how corticostriatal circuits have adapted, and how changes to corticostriatal circuitry may lead to more rapid switching to reflexive strategies. Individual differences in learning success for tone categorization were found to be strongly associated with variation in the *FOXP2* gene [9]. The variant associated with enhanced speech category learning was also associated with earlier and more frequent use of reflexive strategies.

5. Designing optimal training programs and future directions

Taking into consideration tone training studies and individual differences in tone categorization success, another line of research involve developing individualized, neurobiologically-informed training approaches for speech and language learning. Prior studies comparing specific manipulations to training, such as the amount of information provided in feedback, whether the feedback is immediate or delayed, and whether talkers are random or grouped, have shown that manipulations supporting reflexive learning lead to greater categorization success [8].

This leads to the question of the extent to which providing training that can enhance reflexive processing eliminates individual differences in tone learning success. We are finding that the answer is much more complex. An emerging viewpoint is that the two learning systems are initially, interactive. Paul and Ashby posit a bootstrap interaction theory, wherein successful early reflective learning can bootstrap and enhance reflexive-optimal task performance [63]. A one-way interaction between systems may occur wherein the explicit system bootstraps early learning and later uses the procedural system to refine it. Based on the bootstrap interaction theory and DLS, optimizing training for initial reflective and later transitioning to reflexive learning may leverage interactions and lead to better speech and language learning.

A recent study in our lab examined the impact of tone category training manipulations on novel word learning. Participants received training manipulations to target either the reflective or reflexive system, or both systems, with the prediction that initial reflective and later reflexive-optimal environments would result in better tone categorization. Training manipulations consisted of the amount of feedback information and whether talker presentation was randomized or grouped. Prior work has shown different types of feedback can enhance speech category learning [45, 8]. Feedback can be informationally rich (where explicit rules are provided) or minimally informative (such as a beep to signify an error). Informationally rich feedback enhances reflective learning by allowing greater opportunities to test hypotheses. Meanwhile, minimal feedback enhances reflexive learning, as it reduces available resources for reflective processing. Prior studies have argued that reducing talker variability can promote faster hypothesis testing and validation, requiring less working memory and supporting reflective learning [8]. In contrast, randomized talker presentation prevents learners from predicting the next talker, disrupting rule testing and supporting reflexive learning where they may associate talker-invariant acoustic cues with implicit rewards. We thus employed informationally rich feedback and blocked talkers to enhance reflective learning and minimally informative feedback and randomized talkers to enhance reflexive learning. Following tone categorization training with these manipulations, participants completed a sound-to-meaning training program to learn lexical items distinguished by tone categories. We found that a reflective-to-reflexive transition in tone categorization training resulted in better novel word learning. This finding provide initial support for the bootstrap interaction theory and the importance of the two learning systems within the DLS model for speech and language learning.

6. Discussion

In this paper, we discussed mechanisms underlying tone learning in adulthood, training and sources of individual differences in tone learning, and designing enhanced training approaches to reduce these differences. The mechanisms underlying tone learning is adults are still not fully known. We have learned about the neural representation of tones through noninvasive imaging studies using fMRI and auditory electrophysiology. Studies using the FFR as a metric demonstrate the important finding that, although training can have long-lasting behavioral effects, evidenced by native-like task performance, neural representations of these tones differ based on long-term language experience. Establishing behavioral relevance for tones and receiving feedback on performance aids learning. Although training has a significant impact, large individual differences in learning success persist. In attempting to understand the reasons for these differences, we have examined adult learning strategies within the framework of the DLS model. While there is support for reflexive learning of tones, a more complicated interaction, wherein initial reflective learning bootstraps later reflexive learning, has been found to benefit word learning with lexical tones. Further research is needed to develop optimal speech and language training programs that can ameliorate individual differences in tone learning.

7. Acknowledgements

This research was supported by NIDCD grant R01DC013315 awarded to Bharath Chandrasekaran. The authors thank Alex Dimakis, Patrick C. M. Wong, and Todd Maddox for collaboration and significant contributions to the work. The authors also thank the SoundBrain Lab research assistants for their role in participant scheduling and data collection.

8. References

[1] J.F. Werker and T. K. Hensch, "Critical periods in speech perception: new directions." *Psychology*, vol. 66, no. 1, pp. 173-196, Jan. 2015.

[2] W. Penfield and L. Roberts, *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press, 1959.

[3] E. Lenneberg, *Biological Foundations of Language*. New York, NY: Wiley, 1967.

[4] P. K. Kuhl, "Perception of auditory equivalence classes for speech in early infancy," *Infant Behavior and Development,* vol. 6, no. 203, pp. 263-285, 1983.

[5] P. Iverson *et al.*, "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition*, vol. 87, no. 1, pp. B47-B57, Feb. 2003.

[6] H. Yi *et al.,* "The role of corticostriatal systems in speech category learning," *Cerebral Cortex*, vol. 26, no. 4, pp. 1409-1420, Apr. 2016.

[7] B. Chandrasekaran, S. R. Koslov, and W. T. Maddox, "Toward a dual-learning systems model of speech category learning," *Frontiers in Psychology*, vol. 5, pp. 825, Jul. 2014.

[8] B. Chandrasekaran, H. Yi, and W. T. Maddox, "Dual-learning systems during speech category learning," *Psychonomic Bulletin & Review,* vol. 21, no. 2, pp. 488-495, Apr. 2014.

[9] B. Chandrasekaran *et al.*, "Enhanced procedural learning of speech sound categories in a genetic variant of *FOXP2*," *Journal of Neuroscience,* vol. 35, no. 20, pp. 7808-7812, May 2015.

[10] J. T. Marsh, W. S. Brown and J. C. Smith, "Far-field recorded frequency-following responses: correlates of low pitch auditory perception in humans," *Electroencephalography and Clinical Neurophysiology,* vol. 38, no. 2, pp. 113-119, Feb. 1975.

[11] J. C. Smith, J. T. Marsh and W. S. Brown, "Far-field recorded frequency-following responses: evidence for the locus of brainstem sources," *Electroencephalography and Clinical Neurophysiology,* vol. 39, no. 5, pp. 465-472, Nov. 1975.

[12] N. Kraus and T. Nicol, "Brainstem origins for cortical 'what' and 'where' pathways in the auditory system," *Trends in Neurosciences,* vol. 28, no. 4, pp. 176-181, Apr. 2005.

[13] E. Skoe and N. Kraus, "Auditory brainstem response to complex sounds: a tutorial," *Ear and Hearing*, vol. 31, no. 3, pp. 302-324, Jun. 2010

[14] J. Swaminathan, J. Krishnan and J. T. Gandour, "Pitch encoding in speech and nonspeech contexts in the human auditory brainstem", *Neuroreport,* vol. 19, no. 11, pp. 1163-1167, Jul. 2008.

[15] A. Krishnan *et al.*, "Human frequency-following response: representation of pitch contours in Chinese tones," *Hearing Research,* vol. 189, no. 1, pp. 1-12, Mar. 2004.

[16] A. Krishnan *et al.,* "Encoding of pitch in the human brainstem is sensitive to language experience," *Cognitive Brain Research,* vol. 25, no. 1, pp. 161-168, Sep. 2005.

[17] J. H. Song *et al.,* "Plasticity in the adult human auditory brainstem following short-term linguistic training," *Journal of Cognitive Neuroscience,* vol., 20, no. 10, pp. 1892-1902, Oct. 2008.

[18] P. C. M. Wong *et al.*, "Musical experience shapes human brainstem encoding of linguistic pitch patterns," *Nature Neuroscience,* vol. 10, no. 4, pp. 420-422, Mar. 2007.

[19] Y. Wang *et al.*, "Training American listeners to perceive Mandarin tones," *The Journal of the Acoustical Society of America,* vol. 106, no. 6, pp. 3649-3658, Dec. 1999.

[20] Y. Wang, A. Jongman, and J. A. Sereno, "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training," *The Journal of the Acoustical Society of America,* vol. 113, no. 2, pp. 1033-1043, Feb. 2003.

[21] L. Lisker, "'Voicing' in English: a catalogue of acoustic features signaling signaling /b/ versus /p/ in trochees," *Language and Speech,* vol. 29, no. 1, pp. 3-11, Jan. 1986.

[22] J. Hillenbrand *et al.*, "Acoustic characteristics of American English vowels," *The Journal of the Acoustical Society of America,* vol. 97, no. 5, pp. 3099-3111, May 1995.

[23] A. Jongman and C. B. Moore, "The role of language experience in speaker and rate normalization processes," in *INTERSPEECH,* 2000, pp. 62-65.

[24] G. K. Vallabha *et al*., "Unsupervised learning of vowel categories from infant-directed speech," in *Proceedings of the National Academy of Sciences*, 2007, pp. 13273-13278.

[25] L. L. Holt and A. J. Lotto, "Speech perception within an auditory cognitive science framework," *Current Directions in Psychological Science*, vol. 17, no. 1, pp. 42-46, Feb. 2008.

[26] L. L. Holt and A. J. Lotto. "Speech perception as categorization," *Attention, Perception, & Psychophysics,* vol. 72, no. 5, pp. 1218-1227, Jul. 2010.

[27] C. T. Best, "Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development," in *Speech and Face Processing in the First Year of Life*. Netherlands: Springer, 1993, pp. 289-304.

[28] C. T. Best, B. Morrongiello and R. Robson, "Perceptual equivalence of acoustic cues in speech and nonspeech perception," *Perception & Psychophysics*, vol. 29, no. 3, pp. 191-211, May 1981.

[29] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege,* Netherlands: John Benjamins Publishing, 2007, ch. 2, pp. 13-34.

[30] J. E. Flege, "Age of learning and second language speech," in *Second language acquisition and the critical period hypothesis*, Mahwah, NJ: Lawrence Erlbaum Associates, 1999, pp. 101-131.

[31] A. L. Francis *et al., "*Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," *Journal of Phonetics,* vol. 36, no. 2, pp. 268-294, Apr. 2008.

[32] A. L. Francis and H. C. Nusbaum, "Selective attention and the acquisition of new phonetic categories," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 28, no. 2, pp. 349-366, Apr. 2002.

[33] W. T. Maddox *et al*., "Dual systems of speech category learning across the lifespan," *Psychology and Aging,* vol. 28, no. 4, pp. 1042-1056, Dec 2013.

[34] B. Chandrasekaran, J. T. Gandour, and A. Krishnan, "Neuroplasticity in the processing of pitch dimensions: a multidimensional scaling analysis of the mismatch negativity," *Restorative Neurology and Neuroscience,* vol. 25, no. 3-4, pp. 195-210, Jun. 2007.

[35] D. W. Massaro, M. M. Cohen and C. Y. Tseng, "The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese," *Journal of Chinese Linguistics*, vol. 13, no. 2, pp. 267-289, Jun. 1985.

[36] J. T. Gandour and R. A. Harshman, "Crosslanguage differences in tone perception: a multidimensional scaling investigation," *Language and Speech,* vol. 21, no. 1, pp. 1-33, Jan. 1978.

[37] B. Chandrasekaran, P. D. Sampath and P. C. M. Wong, "Individual variability in cue-weighting and lexical tone learning," *The Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. 456-465, Jul. 2010.

[38] B. Chandrasekaran *et al.,* "Effect of explicit dimensional instruction on speech category learning," *Attention, Perception, & Psychophysics*, vol. 78, no. 2, pp. 566-582, Feb. 2016.

[39] A. R. Bradlow *et al., "*Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production," *Perception & Psychophysics*, vol. 61, no. 6, pp. 977-985, Jan. 1999.

[40] S. J. Lim and L. L. Holt, "Learning foreign sounds in an alien world: videogame training improves non-native speech categorization," *Cognitive Science*, vol. 35, no. 7, pp.1390-1405, Sep. 2011.

[41] S. E. Lively *et al.*, “Training Japanese listeners to identify English /r/ and /l/ III. Long-term retention of new phonetic categories,” *The Journal of the Acoustical Society of America*, vol. 96, no. 4, pp. 2076-2087, Oct. 1994.

[42] E. Tricomi *et al.,* “Performance feedback drives caudate activation in a phonological learning task,” *Journal of Cognitive Neuroscience,* vol. 18, no. 6, pp. 1029-1043, Jun. 2006.

[43] Y. Zhang *et al.,* “Neural signatures of phonetic learning in adulthood: a magnetoencephalography study,” *Neuroimage,* vol. 46, no. 1, pp. 226-240, May 2009.

[44] W. T. Maddox and B. Chandrasekaran, “Tests of a dual-system model of speech category learning,” *Bilingualism: Language and Cognition*, vol. 17, no. 4, pp. 709-728, Oct. 2014.

[45] J. L. McClelland, J. A. Fiez, and B. D. McCandliss, “Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects, ” *Physiology & Behavior,* vol. 77, no. 4-5, pp. 657-662, Dec. 2002.

[46] Vallabha, G. K., & McClelland, J. L., “Success and failure of new speech category learning in adulthood: consequences of learned Hebbian attractors in topographic maps, ” *Cognitive, Affective, & Behavioral Neuroscience,* vol. 7, no. 1, pp. 53-73, Mar. 2007.

[47] A. R. Bradlow, “Training non-native language sound patterns: lessons from training Japanese adults on the English,” in *Phonology and Second Language Acquisition*, 2008, ch.10, pp. 287-308.

[48] S. E. Lively, J. S. Logan and D. B. Pisoni, “Training Japanese listeners to identify English/r/and/l/. II: the role of phonetic environment and talker variability in learning new perceptual categories,” *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1242-1255, Sep 1993.

[49] E. M. Nomura and P. J. Reber, “A review of medial temporal lobe and caudate contributions to visual category learning,” *Neuroscience & Biobehavioral Reviews*, vol. 32, no. 2, pp. 279-291, Dec. 2008.

[50] Waldron, E. M., & Ashby, F. G., “The effects of concurrent task interference on category learning: evidence for multiple category learning systems,” *Psychonomic Bulletin & Review,* vol. 8, no. 1, pp. 168-176, Mar. 2001.

[51] S. J. Lim, J. A. Fiez, and L. L. Holt, “How may the basal ganglia contribute to auditory categorization and speech perception?,” *Frontiers in Neuroscience,* vol. 8, no. 230, Aug. 2014.

[52] C. A. Seger, “How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback,” *Neuroscience & Biobehavioral Reviews*, vol. 32, no. 2, pp. 265-278, Dec. 2008.

[53] J. A. Alexander, P. C. M., Wong, and A. R. Bradlow, “Lexical tone perception in musicians and non-musicians,” in *INTERSPEECH, 2005,* pp. 397-400.

[54] T. L. Gottfried and D. Riester, “Relation of pitch glide perception and Mandarin tone identification,” *The Journal of the Acoustical Society of America,* vol. 108, no. 5, pp. 2604, Nov. 2000.

[55] C. Y. Lee and T. H. Hung, “Identification of Mandarin tones by English-speaking musicians and nonmusicians,” *The Journal of the Acoustical Society of America,* vol. 124, no. 5, pp. 3235-3248, Nov. 2008.

[56] P.C. M. Wong and T. K. Perrachione, “Learning pitch patterns in lexical identification by native English-speaking adults,” *Applied Psycholinguistics,* vol. 28, no. 4, pp. 565-585, Oct. 2007.

[57] K. E. Smayda, B. Chandrasekaran and W. T. Maddox, “Enhanced cognitive and perceptual processing: a computational basis for the musician advantage in speech learning,” *Frontiers in Psychology,* vol. 6, no. 682, May 2015.

[58] W. T. Maddox *et al*., “Elevated depressive symptoms enhance reflexive but not reflective auditory category learning,” *Cortex,* vol. 58, pp. 186-198, Sep. 2014.

[59] C. S. Lai *et al*., “A forkhead-domain gene is mutated in a severe speech and language disorder,” *Nature,* vol. 413, no. 6855, pp. 519-523, Oct. 2001.

[60] S. E. Fisher and C. Scharff, “FOXP2 as a molecular window into speech and language,” *Trends in Genetics,* vol. 25, no. 4, pp. 166-177, Apr. 2009.

[61] W. Enard *et al*., “A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice,” *Cell,* vol. 137, no. 5, pp. 961-971, May 2009.

[62] C. Schreiweis *et al.*, “Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance” in *Proceedings of the National Academy of Sciences*, 2014, pp. 14253-14258.

[63] E. J. Paul and F. G. Ashby, “A neurocomputational theory of how explicit learning bootstraps early procedural learning,” *Frontiers in Computational Neuroscience,* vol. 7, no. 117, pp. 251-266, Dec. 2013.