The nature of non-native speech sound representations[a]

Melissa M. Baese-Berk,[1,b] Bharath Chandrasekaran,[2,c] and Casey L. Roark[2,c]

[1]*Department of Linguistics, University of Oregon, Eugene, Oregon 97403, USA*

[2]*Department of Communication Sciences and Disorders, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*

ABSTRACT:

Most current theories and models of second language speech perception are grounded in the notion that learners acquire speech sound categories in their target language. In this paper, this classic idea in speech perception is revisited, given that clear evidence for formation of such categories is lacking in previous research. To understand the debate on the nature of speech sound representations in a second language, an operational definition of "category" is presented, and the issues of categorical perception and current theories of second language learning are reviewed. Following this, behavioral and neuroimaging evidence for and against acquisition of categorical representations is described. Finally, recommendations for future work are discussed. The paper concludes with a recommendation for integration of behavioral and neuroimaging work and theory in this area.
© 2022 Acoustical Society of America. https://doi.org/10.1121/10.0015230

I. INTRODUCTION

For decades, researchers have asked how non-native speakers learn novel categories in a new language. Native listeners are thought to typically demonstrate categorical perception of speech sounds (Liberman *et al.*, 1957). That is, they demonstrate increased sensitivity for sounds across a category boundary, but reduced sensitivity for sounds within the category. However, non-native speakers typically do not demonstrate such perception of contrasts in their second language (L2), especially early in learning (e.g., Miyawaki *et al.*, 1975). Instead, they are typically unable to differentiate between sounds that are not contrastive in their native language (L1). The bulk of previous work in L2 speech sound learning has directly examined the extent to which non-native speakers are able to learn to differentiate between two sounds that do not exist in their L1. Years of work have demonstrated that, indeed, listeners are capable of learning to differentiate between speech sounds in their L2 after training in the lab, in the classroom, or from naturalistic experience (e.g., Bradlow *et al.*, 1997; Lively *et al.*, 1993; Logan *et al.*, 1991).

While significant ground has been covered in our investigations of non-native speech sound category learning in adults, a question remains: are these learners actually acquiring native-like *categories*?[1] While most models of L2 speech perception rely on the notion of emergent categorical representations, very few directly define how "categories"

are operationalized for L2 listeners and what native-like "categorical" performance would entail.

In native language contexts, speech perception is thought to be "categorical" in that adult listeners demonstrate not only increased sensitivity to sounds across category boundaries in their L1, but also reduced sensitivity within a category (Kuhl, 1991; Xu *et al.*, 2006). That is, sensitivity to equivalent acoustic steps is non-linear. For some "steps," the sensitivity is much lower than one would expect if perception were continuous (i.e., reduced sensitivity), and for others, the sensitivity is much higher than one would expect (i.e., increased sensitivity). Adult L1 users flexibly utilize their category representations to deploy position- and context-sensitive versions of these speech sounds (i.e., allophones) in production and interpret acoustically similar versions in perception (e.g., Mitterer and Reinisch, 2017). Native listeners can flexibly generalize this knowledge to novel speakers and novel lexical items (Munson, 2011). Further, according to some theories of category learning, learners extract features from their learned categories and use this information in categorizing other sounds (e.g., Xu *et al.*, 2006). That is, features of the sounds could be used to generalize to novel talkers, novel words, novel contexts, and perhaps even other sounds that share some features.

However, these factors of "categorical" perception of speech sounds are typically not systematically examined in L2 speech category learning. Instead, the focus is somewhat reductionist, a small number of tasks designed to examine discrimination and/or identification of novel speech sounds. Recent evidence from neuroimaging suggests that even when behavior on tasks typically used to investigate L2 category learning stabilizes, the neural signatures of category formation differ substantially between L2 learners and L1 language users (Feng *et al.*, 2021; Reetzke *et al.*, 2018).

Therefore, it is possible that non-native speakers are not acquiring categories, per se or at least that their behavior is not consistent with acquisition of a categorical, abstract representation. If this is the case, most models of L2 speech perception would need to be drastically revised, as "category learning" is currently the basis of the assumptions of these models. Indeed, this suggestion is in line with a growing body of work questioning the role of categories in perception during L1 learning and in L1 speech perception more broadly (Feldman *et al.*, 2021; Jusczyk, 1992; McMurray, 2022; Schatz *et al.*, 2021; Toscano *et al.*, 2010).

In this perspective article, we investigate speech sound learning in L2 and the evidence for and against native-like category representations.[2] We draw from both behavioral and neuroimaging work in the area, investigating how these two approaches could be used synergistically to better understand speech sound learning and how it may (or may not) connect to other types of learning later in life. We conclude with recommendations for future work in this area.

II. OPERATIONAL DEFINITION OF CATEGORY

The discussion of the extent to which L2 learners are acquiring *categories* begs the question "What is a *category*?" Before continuing our review of behavioral and neuroscientific evidence for and against the formation of novel sound categories in adulthood, we provide an operational definition of category and what features learned categories, from our perspective, should demonstrate. Broadly, we define a category as an abstract and generalizable representation that enables listeners to perceive highly variable acoustic input and efficiently process it through a more parsimonious representation with fewer perceptual dimensions than the raw sensory input. That is, rather than having to process every individual feature to perceive the item, we can rely instead on compact, abstract category representations (Niv, 2019; Tang *et al.*, 2019).[3] Below, we outline four facets of how behavior or neural representations may reflect truly categorical representations.

A. Categorical representations should reflect both acquired distinctiveness and acquired equivalence

If non-native speech sounds are represented categorically, learners should demonstrate both distinctiveness between categories *and* equivalence within categories. A hallmark of categorical perception is the heightened dissimilarity between categories and the reduced dissimilarity within categories (Harnad, 2003). Native listeners are both better able to detect differences between categories and less able to detect differences within categories (e.g., Liberman *et al.*, 1957). If the representations of non-native learners are truly categorical, then this bidirectional relationship should also be present.

B. Categorical representations should enable robust and flexible generalization

Categorical representations enable generalization to novel exemplars never encountered during training. Truly categorical representations should enable robust (i.e., consistent and widespread) generalization that is flexible in different contexts (i.e., variable talker or acoustic contexts) and stable over time. Native listeners can accommodate much variability to generalize to novel talkers and contexts (e.g., Maye *et al.*, 2008). If non-native speech sound representations are categorical, they should enable consistent perception across novel and/or atypical category exemplars. Generalization is crucially important for understanding the formation of categorical representations because it suggests a level of abstraction that is assumed necessary in most models of speech perception. That is, to handle the variability a listener receives, they must be able to generalize from prior experience to novel experiences that often do not have direct acoustic matches to the listener's previous input.

C. Categorical representations should be independent of the learner, learner's strategy, and learning pathway

Categorical representations should reflect more about the category and less about the perceiver. While there are some individual differences in native language speech perception and categorization, there may be fewer individual differences than in non-native language speech perception (Feng *et al.*, 2021; Schertz *et al.*, 2015) (Fig. 1). During non-native language learning, learners may rely on different cognitive processes or strategies to form representations and learn the categories (Chandrasekaran *et al.*, 2015; Yi *et al.*, 2016). It is not well understood how these different paths of non-native speech learning influence learners' psychological or neural representations of the sounds. For example, while learners who can show high-dimensional, compact representations demonstrate improved learning compared to learners who do not (Tang *et al.*, 2019), the impact of this on strategy or learning outcome in L2 speech is unclear. That is, it is possible that individual differences are not simply differences in individual performance on the same task, but rather are indicative that some learners are learning one thing while others are learning another. We propose that truly categorical representations of speech should be independent of the learner.

D. Category representations should be able to drive plasticity at multiple levels of the speech network

In a native language, listeners can leverage abstract, high-level speech sound representations across the auditory hierarchy to aid perception in challenging listening conditions (Chandrasekaran *et al.*, 2009; Krishnan *et al.*, 2005; Näätänen *et al.*, 1997). Per the reverse hierarchy theory (Ahissar and Hochstein, 2004), expertise is characterized by the ability of an individual to categorize using top-down neural mechanisms as well as the ability to reach down to
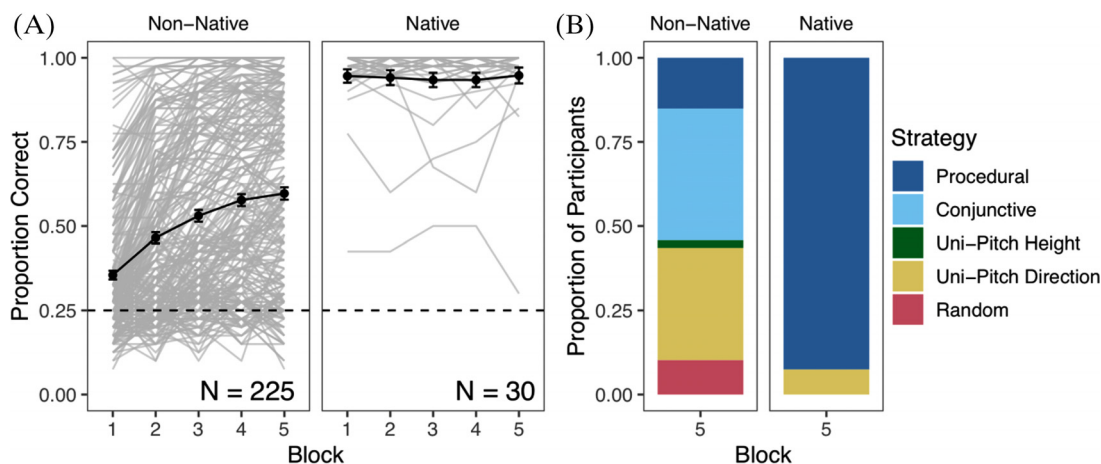
FIG. 1. (Color online) (A) Performance on Mandarin tone learning task for non-native learners (native English listeners) and native Mandarin listeners. Error bars reflect standard error of the mean. Individual subject performance is shown in gray and group mean in black. (B) Decision strategies, assessed with decision bound models (Ashby and Maddox, 1992, 1993) for non-native learners and native Mandarin listeners from the final block in (A). Most native listeners use procedural-based strategies, whereas non-native learners use a variety of strategies, including conjunctive rules and unidimensional rules based on pitch height and pitch direction of the stimulus. A subset of non-native participants randomly guess the category identity even in the final block of learning.

lower sensory levels during perceptual challenges (e.g., due to low signal-to-noise ratio). If speech representations in L2 reflect categorical representations, listeners should demonstrate the same ability to flexibly utilize multiple levels of the speech network to improve performance in challenging listening conditions.

E. Summary

Finally, it is important to note that the definition of a category and how it is reflected in psychological and neural representations is not widely agreed upon, and many other researchers have posited other facets of a category that we do not directly consider here. For example, we do not consider the nature of categorical representations (e.g., exemplar vs prototype vs boundary representations). We believe that any of these models are not necessarily in conflict with the definition presented here and could be incorporated into the definition proposed above.

In Secs. III–VII, we review evidence for categorical perception in L1 to better contextualize the hypothetical goal of L2 speech sound learning, discuss theories of categorical representations in current theories of L2 speech learning, present evidence for and against category representations in L2 speech learning, and finally present recommendations for future research in better consideration of the nature of speech sound representations.

III. CATEGORICAL PERCEPTION

Categorical perception is frequently considered a hallmark of native speech perception. An oft-used example to assess is voice onset time (VOT), the amount of time between the release of a stop consonant and the onset of periodic voicing of the vowel (Abramson and Whalen, 2017). This contrast differentiates /t/ from /d/ in English (Lisker and Abramson, 1964). In English, certain words

are distinguished only by whether they contain one of these two sounds (e.g., "tab" and "dab" are different words in English). Along this acoustic–phonetic continuum, listeners show good discrimination when contrasting sounds belong to two categories in their language but reduced discrimination when the sounds are within a single category (Liberman et al., 1957). That is, listeners demonstrate discontinuities in their perception. Rather than perceiving each acoustic step as equally distant from the previous step, some acoustic steps are perceived as more distinct than others.

Identification and discrimination of sounds by non-native listeners mirrors meaningful, behaviorally relevant distinctions in the listener's native language. For example, native English listeners are able to categorize and discriminate tokens from an /r/–/l/ continuum, a contrast that is meaningful in their native language. However, Japanese listeners demonstrate poor discrimination between those same sounds, since the distinction is not meaningful in their native language (Iverson et al., 2003; Lotto, Sato, and Diehl, 2004; Yamada and Tohkura, 1992). That is, listeners' perception relies on the category structure of their language (e.g., Best et al., 1988; Kuhl et al., 1992; Werker et al., 1981; Werker and Tees, 1984). How these categories are acquired is a matter of substantial debate. While some theoretical approaches to language acquisition have hypothesized that "category" is the unit of acquisition, other work has suggested that categories are simply emergent properties that develop as a function of exposure (Goldinger and Azuma, 2003; Hall et al., 2018; Samuel, 2020). Indeed, the notion of phonetic categories in infant learning, which has long been held as a given in the field, has been recently brought into question (Feldman et al., 2021; Schatz et al., 2021). That is, the type of learning most commonly thought of as "categorical" learning in infant perception may be better accounted for with explanations that rely on infants learning a perceptual

space that appropriately represents the speech they are exposed to, without reliance on categories, per se.

Specifically, under many circumstances, listeners demonstrate more gradient perception in their L1 than discrimination and identification results alone might suggest. For example, employing tasks that allow for more gradient responses or measurements (e.g., eye-tracking, event-related potentials, and visual analog scales, among others; Kapnoula *et al.*, 2017; McMurray *et al.*, 2002; Toscano *et al.*, 2010) results in listeners demonstrating gradient sensitivity to categories. Indeed, the observation that "categorical perception" may be relatively task dependent has been discussed for many years (e.g., Pisoni and Lazarus, 1974; Pisoni and Tash, 1974). In the provocatively titled "The end of categorical perception as we know it," Schouten and colleagues note that categorical perception as it is typically conceived may be an artefact of bias within tasks (Schouten *et al.*, 2003). However, despite these challenges, which we also address below, categorical perception is often used as a benchmark task for successful acquisition of non-native speech sounds. That is, by comparing discrimination and/or identification performance by native speakers to this same performance by learners or by comparing performance at different points in learning, investigators have inferred whether language learners are successful or unsuccessful.

With this question in mind, we turn to a related, critical question: what is the extent to which adult learners can acquire novel non-native categories, given the apparent robustness of categorical perception in one's native language? How flexible are perceptual categories in adulthood? Much previous work has addressed this issue in a variety of modalities. In many perceptual modalities, it is clear that flexibility allows for new categories to be acquired with exposure (e.g., olfaction and taste or bird watching; Royet, 2013; Tanaka and Curran, 2001). However, many perceptual categories, including novel speech sound categories, are notoriously difficult to acquire in adulthood (Flege, 1991; Fox *et al.*, 1995; Wang *et al.*, 1999), as sensitivity to these contrasts typically declines during childhood (e.g., Pegg and Werker, 1997; Sundara *et al.*, 2008; Tsao *et al.*, 2006). Even after substantial experience with a non-native language, perception and production of non-native speech sounds remain a challenge for many learners (Ingvalson *et al.*, 2011). Previous work suggests several possible explanations for this difficulty, and some studies have examined whether, and how, novel sound categories can be acquired in adulthood, focusing largely on perceptual learning.

IV. CURRENT THEORIES OF L2 SPEECH SOUND LEARNING

Various theories have provided explanations for the difficulty in non-native speech sound learning along with predictions about how such learning of sound categories ought to proceed (e.g., Best, 1995; Best and Tyler, 2007; Elvin and Escudero, 2019; Escudero, 2009; Flege, 1995; Flege and Bohn, 2021; Iverson and Kuhl, 1995; Kuhl *et al.*, 2008;

van Leussen and Escudero, 2015). Most of these theories forecast how non-native listeners will learn novel contrasts. One commonality among them is their assumption that new languages will be learned through the lens of the learner's L1. Our ability to learn new contrasts is shaped by the contrasts we already know.

While each of these theories predicts that perception of novel contrasts, and subsequent acquisition of these contrasts, is shaped by the learner's L1, the specifics of how the L1 will impact the second vary. For example, Flege's speech learning model (Flege, 1995; Flege and Bohn, 2021) predicts that sounds in the L2 that are similar to sounds in the L1 may be more difficult for a learner to acquire than other sounds in the learner's L2. The perceptual assimilation model (Best, 1995; Best and Tyler, 2007) similarly suggests that the learner's L1 will impact their perception of novel speech sounds and that this perception is likely to vary depending on how listeners assimilate novel sounds into their existing L1 categories. In contrast, the native language magnet model (Iverson and Kuhl, 1995; Kuhl *et al.*, 2008) focuses primarily on perception *within* categories. This model suggests that, while listeners are likely to use their L1 as a lens for perception of L2 sounds, this lens "pulls" perception toward L1 categories. Predictions are made for how listeners not only categorize L1 sounds but also identify the best exemplars of a target category. In contrast, the L2 linguistic perception model (Elvin and Escudero, 2019; Escudero, 2009; van Leussen and Escudero, 2015) proposes that not only must listeners shift their L1 phoneme category boundaries, but they may also need to increase or decrease the number of categories they deploy during perception of their L2. That is, in addition to influence from the L1 in terms of exact phonetic implementation of categories in their L2, listeners must also contend with the issue that the number of categories may differ across their L1 and target languages, requiring more drastic reorganization than simple shifting of boundaries between two categories.

Supporting these theoretical stances, previous research has contributed to knowledge of how novel speech sound categories are formed in both L1 and L2. For example, several researchers have demonstrated that in the laboratory, listeners are able to learn novel speech sounds after a relatively short period of exposure (e.g., Bradlow *et al.*, 1997; Bradlow *et al.*, 1999; Iverson *et al.*, 2005; Lively *et al.*, 1993; Logan *et al.*, 1991). Learning in these cases is typically defined as above-chance discrimination or identification or as a significant change from a participant's performance before training to their performance after training. Therefore, the bulk of previous work and the aforementioned theories focus primarily on the extent to which a learner is able to differentiate between two contrasting sounds in their non-native language better than chance. The models do not directly address the issue of whether learners are acquiring categorical representations or how these representations might be reflected in participant behavior across a variety of tasks. That is, our understanding of speech sound learning in L2 is largely limited to acquired dissimilarity

3028   J. Acoust. Soc. Am. **152** (5), November 2022

Baese-Berk *et al.*

across a small number of perceptual and production tasks. More importantly, however, the primary models that have been used to describe category formation have relied extensively on the notion of categories and that novel categories are the target of learning.

V. EVIDENCE FOR AND AGAINST CATEGORIES IN L2 SPEECH SOUND LEARNING

As mentioned above, previous work in this area has focused largely on a limited number of tasks. For example, in discrimination tasks, participants are asked to tell one sound (or set of sounds) apart from another. These tasks take a variety of specific forms (e.g., AX, ABX, 4IAX, etc.), but the intuition embedded in all of them is that before training, learners should be quite poor at discriminating between sounds across the (new) category boundary. However, after training, participants should demonstrate improved performance when discriminating between these sounds. Similarly, for identification tasks, participants are asked to classify or label sounds that they hear. Before training, it is expected that learners should demonstrate poor performance on these tasks but should improve after training. This improvement, or increased discrimination, from pre- to post-test for example, is often taken as evidence that participants have acquired novel categories. While this change from pre- to post-test certainly demonstrates some type of learning, it is unclear whether these tasks (often implemented separately and not in connection with any other measures of category formation) paint a complete picture of what participants are actually learning. That is, it is unclear whether these tasks can speak to the notion of whether learners are acquiring a category at all, especially given that L1 perception has demonstrated that tasks can impact how "categorical" perception is (Kapnoula *et al.*, 2017; McMurray *et al.*, 2002; Pisoni and Tash, 1974; Toscano *et al.*, 2010).

However, even under the assumption that these tasks can demonstrate *what* a learner is acquiring, the issues with demonstrations of categorical perception (or lack thereof) remain. A specific challenge to the pre-to-post-test learning improvements being evidence of the emergence of categorical representations is that non-native learners often perform at lower levels than native listeners, even with extensive training (Bradlow *et al.*, 1997, 1999; Ingvalson *et al.*, 2011; Lim and Holt, 2011; Lively *et al.*, 1993; Lotto *et al.*, 2004; McCandliss *et al.*, 2002; Yamada and Tohkura, 1990). Additionally, after initial stages of learning, while some learners demonstrate evidence of acquired distinctiveness (i.e., increased dissimilarity between categories), evidence of acquired equivalence (i.e., increased similarity within a category) is often not investigated. Indeed, while it seems as though these two changes should go hand in hand, most studies only report evidence of improved discrimination for across-category comparisons and do not report whether listeners acquire similarity over time. This is critically important because while, in general, listeners do not accurately differentiate among tokens from novel categories, there are some contrasts that listeners do more successfully

differentiate between, even without any significant training (Best *et al.*, 1988; Flege, 1995). That is, it is unclear whether the bulk of the data in this area can speak to whether learners are acquiring categorical representations, even in cases where this is claimed to be the case [see, e.g., M.M.B.-B.'s own work (Baese-Berk, 2019) for claims about the categorical nature of L2 speech sound representations].

An even clearer sign that perhaps listeners are not acquiring categories comes from generalization to novel instances not encountered during training. In general, this type of learning is quite specific to the stimuli used during training (Logan *et al.*, 1991). That is, even if learners demonstrate robust performance on the trained stimuli, they often fail to generalize this learning to novel stimuli. Even if generalization is somewhat successful, performance is often below that of the trained stimuli (Iverson *et al.*, 2005). This is true for new words containing the same contrast, trained words spoken by new talkers, related contrasts that differ in their exact phonetic realization (i.e., a trained contrast word-initially now presented word-finally), and related contrasts that share some features of the trained contrast (i.e., a voicing contrast at a new place of articulation). This lack or reduction in generalization could be taken as a suggestion that listeners are failing to develop categories under the definition that a category ought to be flexibly and robustly generalizable to new instances that a learner has not previously encountered. This is not to say that generalization is impossible. Indeed, under some circumstances, learners do generalize to novel items and novel talkers. However, to achieve this generalization, learners may require extensive amounts of variability in both talkers and trained items during their training input (e.g., Logan *et al.*, 1991; but see Brekelmans *et al.*, 2022).

However, this type of high variability training is not a successful strategy for all learners (e.g., Perrachione *et al.*, 2011). This observation brings us to an additional major challenge in the category learning literature, which is that individuals demonstrate substantial differences in terms of both quantity of learning and strategies used to learn. There are large individual differences in how well individuals learn non-native speech categories (Golestani and Zatorre, 2009; Heffner and Myers, 2021; Llanos *et al.*, 2020; McHaney *et al.*, 2021). In one prior study, a large sample of native English listeners learned to categorize Mandarin tones produced by multiple talkers (Chandrasekaran *et al.*, 2015) using trial-by-trial feedback. On average, participants showed above-chance performance [mean (M) = 60% accuracy by the final block, standard deviation (SD) = 28%], but different individual learners span the entire range of performance from at-chance levels to perfect performance [Fig. 1(A)]. The same variability is not evident in the performance of native listeners of Mandarin on the same sound-to-category mapping task [Fig. 1(A)]. What underlies different levels of success in speech category learning in non-native listeners is not well understood.

Additionally, it is possible that learners may ultimately achieve similar levels of performance but use entirely different strategies. Even in highly controlled nonspeech category

J. Acoust. Soc. Am. **152** (5), November 2022

Baese-Berk *et al.* 3029

learning tasks, learners differ widely in the strategies they use to separate the sounds into categories (Chandrasekaran *et al*., 2016; Roark and Chandrasekaran, 2021; Roark and Holt, 2019a, 2019b, 2022). During Mandarin tone learning, there was large variability in the strategies that non-native learners used even in the final block of training [Fig. 1(B)]. The same strategy differences are not apparent in native listeners' categorizations [Fig. 1(B)]. Learners differ in the acoustic cues or dimensions they use for categorization and how they weigh the cues or dimensions. Learners also can change strategies over the course of early learning, as they get more and more accurate. Even high performing learners can use different strategies relative to experts. Further, in some cases, the same listeners demonstrate different cue-weighting, and different amounts of cross-listener variability in their cue-weighting strategies, across their L1 and L2 (Schertz *et al*., 2015).

There is also evidence to suggest that experiences or identities—beyond L1 language experience—that the learner brings to a category learning problem can also influence their learning strategies and outcomes. For example, individuals with elevated depressive symptoms learned non-native Mandarin tone categories better than those without elevated depressive symptoms, using category-optimal strategies earlier and more frequently during learning (Maddox *et al*., 2014). Similarly, musicians learned better than non-musicians and used category-optimal strategies earlier and more frequently (Smayda *et al*., 2015), though it is important to note that using a single term "musicians" for a wide range of experiences also likely does not capture the substantial individual differences within this group (Tervaniemi, 2009). Individual differences in cognitive abilities like working memory have also sometimes been shown to relate to the ability to learn non-native speech categories (McHaney *et al*., 2021). However, in other studies, working memory has been shown to be unrelated to non-native speech learning (Heffner and Myers, 2021; Ingvalson *et al*., 2017; Perrachione *et al*., 2011). Given the operational definition of category presented above, this level of individual variability in performance and learning strategy is problematic for the notion of category learning in general. That is, if category learning is foundational either to L2 acquisition (i.e., is the clear target of learning) or is foundational to L1 performance (i.e., is a hallmark of native-like perception), these categories and how they are learned ought to be robust across learners. While there may be individual differences in learning speeds or strategies for novel categories, these differences should not be so robust that some learners acquire the structure while others remain at chance-level performance. That is, these individual differences must be considered to understand the nature of learned representations of L2 speech sounds—whether these representations are categorical or not.

VI. NEUROIMAGING EVIDENCE FOR AND AGAINST CATEGORY LEARNING

Recent advances in neuroimaging analytic approaches provide insights into the nature of emergent speech representations in learners, how these representations are acquired, and the extent to which these representations resemble native listeners (e.g., Bidelman *et al*., 2013; Feng *et al*., 2021; Yi *et al*., 2021; Song *et al*., 2008). An emergent perspective is that unlike native acquisition of speech categories, category learning in adulthood requires some amount of supervision (Vallabha and McClelland, 2007). The underlying hypothesis is that unsupervised learning processes are less effective in adults. Adults, therefore, rely on domain-general neural systems to assist in shaping emergent representations (Feng *et al*., 2019, Feng *et al*., 2021). In particular, the dual-learning systems (DLS) model states that initial learning is facilitated by a sound-to-rule mapping network (termed "reflective" learning) involving the executive cortico-striatal loop, the anterior cingulate cortex, and the medial temporal lobe (Yi *et al*., 2016). Such learning is characterized by generation of hypotheses on the basis of the underlying cues and validation of these hypotheses via an error monitoring process that validates rules (or promotes the generation of new rules). In contrast, later learning is facilitated by a sound-to-reward mapping network (termed "reflexive learning") involving the associative cortico-striatal loop. Such learning is characterized by stimulus-to-response mapping, facilitated by the reward value induced by motor response. Per the DLS model (Chandrasekaran *et al*., 2014; Yi and Chandrasekaran, 2016), individual differences can emerge via the balance between the two learning systems; some learners can get stuck with their exclusive reliance on the reflective learning network, when the multidimensional nature of the categories renders learning via the reflexive network more optimal (Feng *et al*., 2021).

Neuroimaging studies have yielded support for key components of the DLS model (Feng *et al*., 2021). Specifically, learners tend to activate both the reflective and reflexive network when they are processing feedback information during the sound-to-category learning task (Yi *et al*., 2016). Crucially, there appears to be a shift in the balance between these systems for successful learners. By the end of a session of training, learners who tend to activate the putamen, a part of the reflexive learning network, tend to use more multidimensional (reflexive) strategies during learning and achieve better learning performance. These results suggest a significant involvement of domain-general learning systems during L2 speech acquisition. These systems have a protracted developmental timeline and are unlikely to play a key role in native acquisition (Reetzke *et al*., 2016). Feng *et al*. (2018) showed that "representations" of tone categories can emerge in left auditory association cortex within a few hundred trials of sound-to-category training. These representations are tolerant of talker and segmental variability, suggesting some level of "abstractness." When learners encounter an error via incorrect feedback, there is greater coupling between the auditory associative cortex and the reflexive network, suggesting a role for domain-general cortico-striatal network in shaping the emergent category representations in the left temporal lobe. A question that is

relevant to the premise of this article is the nature of the emergent representation and the extent to which these representations are native-like. Using an analytic approach called representational similarity analyses (Kriegeskorte *et al.*, 2008), Feng *et al.* (2021) demonstrate that emergent representations in successful learners show similarities with those in native listeners in that they are high-dimensional and emerge in similar brain regions within the speech perceptual network. Crucially, there are striking differences between successful and less successful learners. In contrast to successful learners, less successful learners show less efficient and low-dimensional representations. Indeed, low-dimensional representations may be enough for accurate performance under some task conditions; a higher-dimensional representation may be crucial for accurate performance across a range of task conditions (Niv, 2019).

Reetzke *et al.* (2018) examined the impact of longer-term training on the sensory encoding of Mandarin tone categories in non-native listeners. In this study, learning stages were operationally defined by directly comparing learning performance to native listeners. Reetzke *et al.* binned learning stages into "novice" (after the first session), "experienced" (performance is "native-like" for three consecutive sessions), "over-trained" (ten additional sessions beyond the "experienced" stage), and "retention" (2 months after cessation of training). A categorical perception task involving a continuum from level to rising pitch trajectories was used to assess perceptual changes as a function of training across sessions. The frequency-following response (FFR), a neural measure that reflects the representational fidelity of early sensory processing, was used to assess the impact of training on sensory representation. Despite significant inter-individual differences during the early stages of learning, learners were able to categorize non-native tone categories with native-like accuracies and retained this high performance in the retention session. Native-like categorical perception emerged by the "experienced" stage of processing, indicating a shift in category identification as a function of training. In contrast, the FFRs were subject to change only by the over-trained session, indicating that low-level sensory plasticity emerges at a slower time scale (after extensive training). Crucially, plastic changes to the FFRs were retained during the retention session, suggesting that experience-dependent changes were not transient.

Taken together, neuroimaging evidence suggests that the process of category learning and the nature of emergent category representations in L2 learners show distinct differences relative to native learners. These may partially reflect fundamentally different brain dynamics in perceptual learning in adulthood. L2 category acquisition in adulthood follows a reverse hierarchy that is more supervised and top-down, whereas L1 acquisition is likely to be more unsupervised and "bottom-up" (Vallabha *et al.*, 2007). L2 acquisition may be subserved by feedback-dependent, domain-general cortico-striatal learning systems, whereas L1 acquisition may reflect unsupervised statistical-based learning processes that are minimally dependent on feedback.

Representations in native listeners are robust to talker and segmental variability, multidimensional, and efficient. While there are indications that some L2 learners may show similarities to native listeners, many other learners do not. These large-scale individual differences indicate that L2 acquisition is shaped by the individual factors (e.g., musicianship, working memory capacity) in the manner that L1 is not.

VII. FUTURE DIRECTIONS AND RECOMMENDATIONS: RETURNING TO THE OPERATIONAL DEFINITION OF CATEGORY

Given this evidence, we return to the operational definition of "category" we presented above to outline recommendations that would enable researchers to more directly answer whether learners are acquiring categories in their L2, especially early in learning.

A. Categorical representations should reflect both acquired distinctiveness and acquired equivalence

Learners should demonstrate both distinctiveness between categories *and* equivalence within categories. **Recommendations:** To probe whether learned representations are truly categorical, researchers should assess both acquired dissimilarity and equivalence. This would require additional methods in many cases where researchers commonly rely on tests of dissimilarity only (i.e., ability to group stimuli into separate groups). Tests of equivalence could include discrimination tasks, sampling along a continuum. Truly categorical representations should invoke both distinctiveness between categories and equivalence within categories.

B. Categorical representations should enable robust and flexible generalization

Truly categorical representations should enable robust (i.e., consistent and widespread) generalization that is flexible in different contexts (i.e., variable talker or acoustic contexts) and stable over time. **Recommendations:** Researchers should thoughtfully include variability during both training and generalization. This may include acoustic descriptions of the stimuli being presented to participants and a characterization of the variability being presented to participants. If researchers plan to make claims about the categorical nature of the representations being learned, they should include tests of generalization and should be clear about what these results demonstrate. Researchers should use longitudinal approaches to understand the emergence of representations from initial exposure to true expert levels.

C. Categorical representations should be independent of the learner, learner's strategy, and learning pathway

Categorical representations should reflect more about the category and less about the perceiver. **Recommendations:** One challenge to understanding the underlying causes of

J. Acoust. Soc. Am. **152** (5), November 2022

Baese-Berk *et al.*   3031

variability in strategies and performance is that they are often not reported in the literature. That is, authors tend to report group means and variability, rather than showing individual subject data. We recommend that part of the solution to understanding non-native learners' representations is acknowledging and embracing the individual variability. This approach will better enable examining the different processes or resulting representations that different learners may acquire. For instance, it is possible that different learners may acquire different structures of representations, with some forming highly abstract and categorical representations and others relying on low-level acoustic representations without underlying categorical representations.

D. Category representations should be able to drive plasticity at multiple levels of the speech network

Categorical representations should enable flexible utilization of multiple levels of the speech network to improve speech perception in noise. **Recommendations**: Neuroimaging approaches in humans rarely provide information at multiple hierarchical levels. Functional magnetic resonance imaging (fMRI), extensively used to document emergent representations of L2 categories, often disregards lower levels of the auditory hierarchy. These lower-level structures are small, physiologically noisy, and often disregarded due to a corticocentric bias in fMRI literature (Ress and Chandrasekaran, 2013). In contrast, electroencephalography (EEG) offers access to multiple hierarchical levels, but without the spatial precision of fMRI. One recommendation to overcome these methodological challenges is to conduct *multimodal* neuroimaging studies that are capable of assaying neuroplasticity across the auditory hierarchy.

E. Additional considerations

In addition to these recommendations, it is critical that researchers clearly differentiate between perceptual learning and category learning. That is, while we do not dispute the fact that leaners are changing behavior from pre- to post-test or are able to demonstrate above-chance discrimination on some trained stimuli, we caution that, given the definition of categorical representations described above and the problems laid out by L1 research in this area regarding the problematic definitions of categorical perception, this learning cannot always be described as category learning. Given this, theories and models of L2 speech sound learning should be modified such that the reliance on category structure is an emergent property during learning over a longer period of time, rather than an early-stage learning outcome that could be easily testable after only an hour or two of training.

Finally, we would like to suggest that future work can incorporate insights from a wide range of methodological approaches. By including linguistic, cognitive, computational modeling, and neuroscientific approaches, we will be better able to address questions of what learners are acquiring during training and will be able to devise better theories and models to test predictions around how learning proceeds.

[1]As we discuss later in this paper, the notion of categorical perception in L1 speech perception is also perhaps not as robust as early work in this area has implied. However, for the purposes of the current discussion, and for reconsidering the issue of learning novel speech categories in a L2, we set that discussion to the side, but we will return to it later in the paper.

[2]As a reviewer notes, in the approach of many investigators in this area, the question of "are learners acquiring categories" is conflated with the question of "are learners acquiring native-like perception." That is, by assuming that native listeners would demonstrate categorical perception on some task or tasks, researchers assume that performance mirroring is both evidence for categorical perception and evidence for native-like perception. Of course, one of these properties could exist without the other, especially given the relative "fuzziness" around categorical perception outside of basic discrimination and categorization tasks, even within a L1. Here, we attempt to problematize both notions, though we recognize in doing so we also occasionally conflate the two issues.

[3]While true category representations in native listeners may be high-dimensional but compact (i.e., not requiring attention), it is unlikely that they are as high-dimensional as the stimuli themselves. Deployment of attention-dependent, low-dimensional representations can be used to solve tasks (e.g., categorical perception tasks using synthetically generated stimuli; Niv, 2019), which may be the type of representation used by many learners in the tasks described below.

Abramson, A. S., and Whalen, D. H. (**2017**). "Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions," J. Phon. **63**, 75–86.

Ahissar, M., and Hochstein, S. (**2004**). "The reverse hierarchy theory of visual perceptual learning," Trends Cognit. Sci. **8**(10), 457–464.

Ashby, F. G., and Maddox, W. T. (**1992**). "Complex decision rules in categorization: Contrasting novice and experienced performance," J. Exp. Psychol. Hum. Percept. Perform. **18**(1), 50–71.

Ashby, F. G., and Maddox, W. T. (**1993**). "Relations between prototype, exemplar, and decision bound models of categorization," J. Math. Psychol. **37**(3), 372–400.

Baese-Berk, M. M. (**2019**). "Interactions between speech perception and production during learning of novel phonemic categories," Atten. Percept. Psychophys. **81**(4), 981–1005.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., McRoberts, G. W., and Sithole, N. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementaries," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O. S. Bohn (John Benjamins, Amsterdam), pp. 13–34.

Bidelman, G. M., Moreno, S., and Alain, C. (**2013**). "Tracing the emergence of categorical speech perception in the human auditory system," Neuroimage **79**, 201–212.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/:

3032   J. Acoust. Soc. Am. **152** (5), November 2022

Baese-Berk *et al.*

Long-term retention of learning in perception and production," Percept. Psychophys. **61**(5), 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**(4), 2299–2310.

Brekelmans, G., Lavan, N., Saito, H., Clayards, M., and Wonnacott, E. (**2022**). "Does high variability training improve the learning of non-native phoneme contrasts over low variability training? A replication," J. Mem. Lang. **126**, 104352.

Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (**2009**). "Sensory processing of linguistic pitch as reflected by the mismatch negativity," Ear Hear. **30**(5), 552–558.

Chandrasekaran, B., Yi, H. G., and Maddox, W. T. (**2014**). "Dual-learning systems during speech category learning," Psychon. Bull. Rev. **21**(2), 488–495.

Chandrasekaran, B., Yi, H.-G., Blanco, N. J., McGeary, J. E., and Maddox, W. T. (**2015**). "Enhanced procedural learning of speech sound categories in a genetic variant of FOXP2," J. Neurosci. **35**(20), 7808–7812.

Chandrasekaran, B., Yi, H.-G., Smayda, K. E., and Maddox, W. T. (**2016**). "Effect of explicit dimensional instruction on speech category learning," Atten. Percept. Psychophys. **78**(2), 566–582.

Elvin, J., and Escudero, P. (**2019**). "Cross-Linguistic influence in second language speech: Implications for learning and teaching," in *Cross-Linguistic Influence: From Empirical Evidence to Classroom Practice*, edited by M. J. Gutierrez-Mangado, M. Martínez-Adrián, and F. Gallardo-del-Puerto (Springer, Cham, Switzerland), pp. 1–20.

Escudero, P. (**2009**). "Linguistic perception of SIMILAR L2 sounds," in *Phonology in Perception*, edited by P. Boersma and S. Hamann (De Gruyter Mouton, Berlin), pp. 151–190.

Feldman, N. H., Goldwater, S., Dupoux, E., and Schatz, T. (**2021**). "Do infants really learn phonetic categories?," Open Mind **5**, 113–131.

Feng, G., Gan, Z., Wang, S., Wong, P. C., and Chandrasekaran, B. (**2018**). "Task-general and acoustic-invariant neural representation of speech categories in the human brain," Cereb. Cortex **28**(9), 3241–3254.

Feng, G., Gan, Z., Yi, H. G., Ell, S. W., Roark, C. L., Wang, S., Wong, P. C., and Chandrasekaran, B. (**2021**). "Neural dynamics underlying the acquisition of distinct auditory category structures," Neuroimage **244**, 118565.

Feng, G., Li, Y., Hsu, S.-M., Wong, P. C., Chou, T.-L., and Chandrasekaran, B. (**2021**). "Emerging native-similar neural representations underlie non-native speech category learning success," Neurobiol. Lang. **2**(2), 280–307.

Feng, G., Yi, H. G., and Chandrasekaran, B. (**2019**). "The role of the human auditory corticostriatal network in speech learning," Cereb. Cortex **29**(10), 4077–4089.

Flege, J. E. (**1991**). "Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language," J. Acoust. Soc. Am. **89**(1), 395–411.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.

Flege, J. E., and Bohn, O.-S. (**2021**). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by R. Wayland (Cambridge University, New York), pp. 3–83.

Fox, R. A., Flege, J. E., and Munro, M. J. (**1995**). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," J. Acoust. Soc. Am. **97**(4), 2540–2551.

Goldinger, S., and Azuma, T. (**2003**). "Puzzle-solving science: The quixotic quest for units in speech perception," J. Phon. **31**(3–4), 305–320.

Golestani, N., and Zatorre, R. J. (**2009**). "Individual differences in the acquisition of second language phonology," Brain Lang. **109**(2–3), 55–67.

Hall, K. C., Hume, E., Jaeger, T. F., and Wedel, A. (**2018**). "The role of predictability in shaping phonological patterns," Linguist. Vanguard **4**(s2), 20170027.

Harnad, S. (**2003**). "Categorical perception," in *Encyclopedia of Cognitive Science* (Wiley, New York).

Heffner, C. C., and Myers, E. B. (**2021**). "Individual differences in phonetic plasticity across native and nonnative contexts," J. Speech Lang. Hear. Res. **64**(10), 3720–3733.

Ingvalson, E. M., McClelland, J. L., and Holt, L. L. (**2011**). "Predicting native English-like performance by native Japanese speakers," J. Phon. **39**(4), 571–584.

Ingvalson, E. M., Nowicki, C., Zong, A., and Wong, P. (**2017**). "Non-native speech learning in older adults," Front. Psychol. **8**, 148.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278.

Iverson, P., and Kuhl, P. K. (**1995**). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," J. Acoust. Soc. Am. **97**(1), 553–562.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., and Diesch, E. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Jusczyk, P. W. (**1992**). "Developing phonological categories from the speech signal," in *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 17–64.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., and McMurray, B. (**2017**). "Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach," J. Exp. Psychol. Hum. Percept. Perform. **43**, 1594–1611.

Kriegeskorte, N., Mur, M., and Bandettini, P. A. (**2008**). "Representational similarity analysis-connecting the branches of systems neuroscience," Front. Syst. Neurosci. **2**, 4.

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (**2005**). "Encoding of pitch in the human brainstem is sensitive to language experience," Cogn. Brain Res. **25**(1), 161–168.

Kuhl, P. K. (**1991**). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," Percept. Psychophys. **50**, 93–107.

Kuhl, P. K., Conboy, B. T., Conboy, B. T., Coffey-Corina, S., Coffey-Corina, S., Padden, D., Padden, D., Rivera-Gaxiola, M., Rivera-Gaxiola, M., Nelson, T., and Nelson, T. (**2008**). "Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e)," Philos. Trans. R. Soc. B **363**(1493), 979–1000.

Kuhl, P. K., Williams, K., Lacerda, F., Stevens, K., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**(5044), 606–608.

Liberman, A. M., Harris, K., Hoffman, H., and Griffith, B. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**(5), 358–368.

Lim, S., and Holt, L. L. (**2011**). "Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization," Cogn. Sci. **35**(7), 1390–1405.

Lisker, L. I., and Abramson, A. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," J. Acoust. Soc. Am. **35**(11), 384–422.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**(3), 1242–1255.

Llanos, F., McHaney, J. R., Schuerman, W. L., Yi, H. G., Leonard, M. K., and Chandrasekaran, B. (**2020**). "Non-invasive peripheral nerve stimulation selectively enhances speech category learning in adults," npj Sci. Learn. **5**(1), 12.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**(2), 874–886.

Lotto, A. J., Sato, M., and Diehl, R. L. (**2004**). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," in *Proceedings of From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, June 11–13, Cambridge, MA, pp. C381–C386.

Maddox, W. T., Chandrasekaran, B., Smayda, K., Yi, H.-G., Koslov, S., and Beevers, C. G. (**2014**). "Elevated depressive symptoms enhance reflexive but not reflective auditory category learning," Cortex **58**, 186–198.

Maye, J., Aslin, R. N., and Tanenhaus, M. K. (**2008**). "The weckud wetch of the wast: Lexical adaptation to a novel accent," Cogn. Sci. **32**(3), 543–562.

McCandliss, B. D., Fiez, J. A., Protopapas, A., and Conway, M. (**2002**). "Success and failure in teaching the [r]-[l] contrast to Japanese Adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception," Cogn. Affect. Behav. Neurosci. **2**(2), 89–108.

McHaney, J. R., Gnanateja, G. N., Smayda, K. E., Zinszer, B. D., and Chandrasekaran, B. (**2021**). "Cortical tracking of speech in delta band relates to individual differences in speech in noise comprehension in older adults," Ear Hear. **42**(2), 343–354.

McHaney, J. R., Tessmer, R., Roark, C. L., and Chandrasekaran, B. (**2021**). "Working memory relates to individual differences in speech category learning: Insights from computational modeling and pupillometry," Brain Lang. **222**, 105010.

McMurray, B. (**2022**). "The myth of categorical perception," PsyArXiv.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (**2002**). "Gradient effects of within-category phonetic variation on lexical access," Cognition **86**, B33–B42.

Mitterer, H., and Reinisch, E. (**2017**). "Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops," Lang. Cogn. Neurosci. **32**, 1133–1147.

Miyawaki, K., Jennings, J. J., Strange, W., Libermann, A. M., Verbrugge, R., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**(5), 331–340.

Munson, C. (**2011**). "Perceptual learning in speech reveals pathways of processing," Ph.D. dissertation, University of Iowa, Iowa City, IA.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., and Luuk, A. (**1997**). "Language-specific phoneme representations revealed by electric and magnetic brain responses," Nature **385**(6615), 432–434.

Niv, Y. (**2019**). "Learning task-state representations," Nat. Neurosci. **22**(10), 1544–1553.

Pegg, J. E., and Werker, J. F. (**1997**). "Adult and infant perception of two English phones," J. Acoust. Soc. Am. **102**(6), 3742–3753.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. M. (**2011**). "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," J. Acoust. Soc. Am. **130**(1), 461–472.

Pisoni, D. B., and Lazarus, J. H. (**1974**). "Categorical and noncategorical modes of speech perception along the voicing continuum," J. Acoust. Soc. Am. **55**(2), 328–333.

Pisoni, D. B., and Tash, J. (**1974**). "Reaction times to comparisons within and across phonetic categories," Percept. Psychophys. **15**(2), 285–290.

Reetzke, R., Maddox, W. T., and Chandrasekaran, B. (**2016**). "The role of age and executive function in auditory category learning," J. Exp. Child Psychol. **142**, 48–65.

Reetzke, R., Xie, Z., Llanos, F., and Chandrasekaran, B. (**2018**). "Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood," Curr. Biol. **28**(9), 1419–1427.

Ress, D., and Chandrasekaran, B. (**2013**). "Tonotopic organization in the depth of human inferior colliculus," Front. Hum. Neurosci. **7**, 586.

Roark, C. L., and Chandrasekaran, B. (**2021**). "Variability within and across individuals during auditory category learning," PsyArXiv.

Roark, C. L., and Holt, L. L. (**2019a**). "Auditory information-integration category learning in young children and adults," J. Exp. Child Psychol. **188**, 104673.

Roark, C. L., and Holt, L. L. (**2019b**). "Perceptual dimensions influence auditory category learning," Atten. Percept. Psychophys. **81**(4), 912–926.

Roark, C. L., and Holt, L. L. (**2022**). "Long-term priors constrain category learning in the context of short-term statistical regularities," Psych. Bull. Rev. **29**, 1925–1937.

Royet, J.-P., Plailly, J., Saive, A.-L., Veyrac, A., and Delon-Martin, C. (**2013**). "The impact of expertise in olfaction," Front. Psychol. **4**, 928.

Samuel, A. G. (**2020**). "Psycholinguists should resist the allure of linguistic units as perceptual units," J. Mem. Lang. **111**, 104070.

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., and Dupoux, E. (**2021**). "Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input," Proc. Natl. Acad. Sci. U.S.A. **118**(7), e2001844118.

Schertz, J., Cho, T., Lotto, A., and Warner, N. (**2015**). "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," J. Phon. **52**, 183–204.

Schouten, B., Gerrits, E., and van Hessen, A. (**2003**). "The end of categorical perception as we know it," Speech Commun. **41**(1), 71–80.

Smayda, K. E., Chandrasekaran, B., and Maddox, W. T. (**2015**). "Enhanced cognitive and perceptual processing: A computational basis for the musician advantage in speech learning," Front. Psychol. **6**, 682.

Song, J. H., Skoe, E., Wong, P. C., and Kraus, N. (**2008**). "Plasticity in the adult human auditory brainstem following short-term linguistic training," J. Cogn. Neurosci. **20**(10), 1892–1902.

Sundara, M., Polka, L., and Molnar, M. (**2008**). "Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers," Cognition **108**(1), 232–242.

Tanaka, J. W., and Curran, T. (**2001**). "A neural basis for expert object recognition," Psychol. Sci. **12**, 43–47.

Tang, E., Mattar, M. G., Giusti, C., Lydon-Staley, D. M., Thompson-Schill, S. L., and Bassett, D. S. (**2019**). "Effective learning is accompanied by high-dimensional and efficient representations of neural activity," Nat. Neurosci. **22**(6), 1000–1009.

Tervaniemi, M. (**2009**). "Musicians—Same or different?" Ann. N.Y. Acad. Sci. **1169**(1), 151–156.

Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (**2010**). "Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech," Psychol. Sci. **21**(10), 1532–1540.

Tsao, F.-M., Liu, H.-M., and Kuhl, P. K. (**2006**). "Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants," J. Acoust. Soc. Am. **120**(4), 2285–2294.

Vallabha, G. K., and McClelland, J. L. (**2007**). "Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps," Cogn. Affect. Behav. Neurosci. **7**(1), 53–73.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., and Amano, S. (**2007**). "Unsupervised learning of vowel categories from infant-directed speech," Proc. Natl. Acad. Sci. U.S.A. **104**(33), 13273–13278.

van Leussen, J.-W., and Escudero, P. (**2015**). "Learning to perceive and recognize a second language: The L2LP model revised," Front. Psychol. **6**, 1000.

Wang, Y., Spence, M., Jongman, A., and Sereno, J. (**1999**). "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am. **106**(6), 3649–3658.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). "Developmental aspects of cross-language speech perception," Child Dev. **52**, 349–355.

Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.

Xu, Y., Gandour, J. T., and Francis, A. L. (**2006**). "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am. **120**(2), 1063–1074.

Yamada, R. A., and Tohkura, Y. (**1990**). "Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese," in *Proceedings of the First International Conference on Spoken Language Processing*, November 18–22, Kobe, Japan, pp. 757–760.

Yamada, R. A., and Tohkura, Y. (**1992**). "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," Percept. Psychophys. **52**(4), 376–392.

Yi, H. G., and Chandrasekaran, B. (**2016**). "Auditory categories with separable decision boundaries are learned faster with full feedback than with minimal feedback," J. Acoust. Soc. Am. **140**(2), 1332–1335.

Yi, H.-G., Maddox, W. T., Mumford, J. A., and Chandrasekaran, B. (**2016**). "The role of corticostriatal systems in speech category learning," Cereb. Cortex **26**(4), 1409–1420.

Yi, H., Chandrasekaran, B., Nourski, K. V., Rhone, A. E., Schuerman, W. L., Howard, M. A., Chang, E. F., and Leonard, M. K. (**2021**). "Learning nonnative speech sounds changes local encoding in the adult human cortex," Proc. Natl. Acad. Sci. U.S.A. **118**(36), 2101777118.

3034   J. Acoust. Soc. Am. **152** (5), November 2022

Baese-Berk *et al.*