



Emergence of category-level sensitivities in non-native speech sound learning

Emily B. Myers^{1,2,3*}

¹ Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT, USA

² Department of Psychology, University of Connecticut, Storrs, CT, USA

³ Haskins Laboratories, New Haven, CT, USA

Edited by:

Einat Liebenthal, Medical College of Wisconsin, USA

Reviewed by:

Matthew H. Davis, MRC Cognition and Brain Sciences Unit, UK

Caroline A. Niziolek, University of California, San Francisco, USA

*Correspondence:

Emily B. Myers, University of Connecticut, 850 Bolton Rd., Unit 1085, Storrs, CT 06269, USA
e-mail: emily.myers@uconn.edu

Over the course of development, speech sounds that are contrastive in one's native language tend to become perceived categorically: that is, listeners are unaware of variation within phonetic categories while showing excellent sensitivity to speech sounds that span linguistically meaningful phonetic category boundaries. The end stage of this developmental process is that the perceptual systems that handle acoustic-phonetic information show special tuning to native language contrasts, and as such, category-level information appears to be present at even fairly low levels of the neural processing stream. Research on adults acquiring non-native speech categories offers an avenue for investigating the interplay of category-level information and perceptual sensitivities to these sounds as speech categories emerge. In particular, one can observe the neural changes that unfold as listeners learn not only to perceive acoustic distinctions that mark non-native speech sound contrasts, but also to map these distinctions onto category-level representations. An emergent literature on the neural basis of novel and non-native speech sound learning offers new insight into this question. In this review, I will examine this literature in order to answer two key questions. First, where in the neural pathway does sensitivity to category-level phonetic information first emerge over the trajectory of speech sound learning? Second, how do frontal and temporal brain areas work in concert over the course of non-native speech sound learning? Finally, in the context of this literature I will describe a model of speech sound learning in which rapidly-adapting access to categorical information in the frontal lobes modulates the sensitivity of stable, slowly-adapting responses in the temporal lobes.

Keywords: speech perception, phonetic category, second language acquisition, inferior frontal gyrus, superior temporal gyrus

INTRODUCTION

Phonetic categories, the basic perceptual units of language, are defined over distributions in acoustic space. For any phonetic category (e.g., /d/) there will be a range of acoustic tokens that will all be computed as acceptable members of a given phonetic category. To take a classic example, voiced and voiceless stops (e.g., /d/ vs. /t/) are primarily distinguished in initial position by the acoustic/articulatory parameter known as voice onset time, or VOT. For a native English speaker, VOTs less than about 30 ms are heard as /d/ sounds and those greater than 30 ms are perceived as /t/ sounds. The process of learning phonetic categories requires that the listener learn the boundaries of this acoustic space in order to understand how any given acoustic token maps to the phonology of his/her native language. To take the example given above, the English-learning child will learn that the voicing boundary falls at about 30 ms VOT in her language, but the Spanish-learning child will learn a boundary at about 0 ms VOT (Lisker and Abramson, 1964). This learning process is complicated by the fact that phonetic categories are typically defined by multiple acoustic parameters (e.g., VOT, vowel length, closure

duration, burst amplitude). In this sense, we may think of the process of learning phonetic category boundaries as one of defining a hyperplane through multi-dimensional acoustic space.

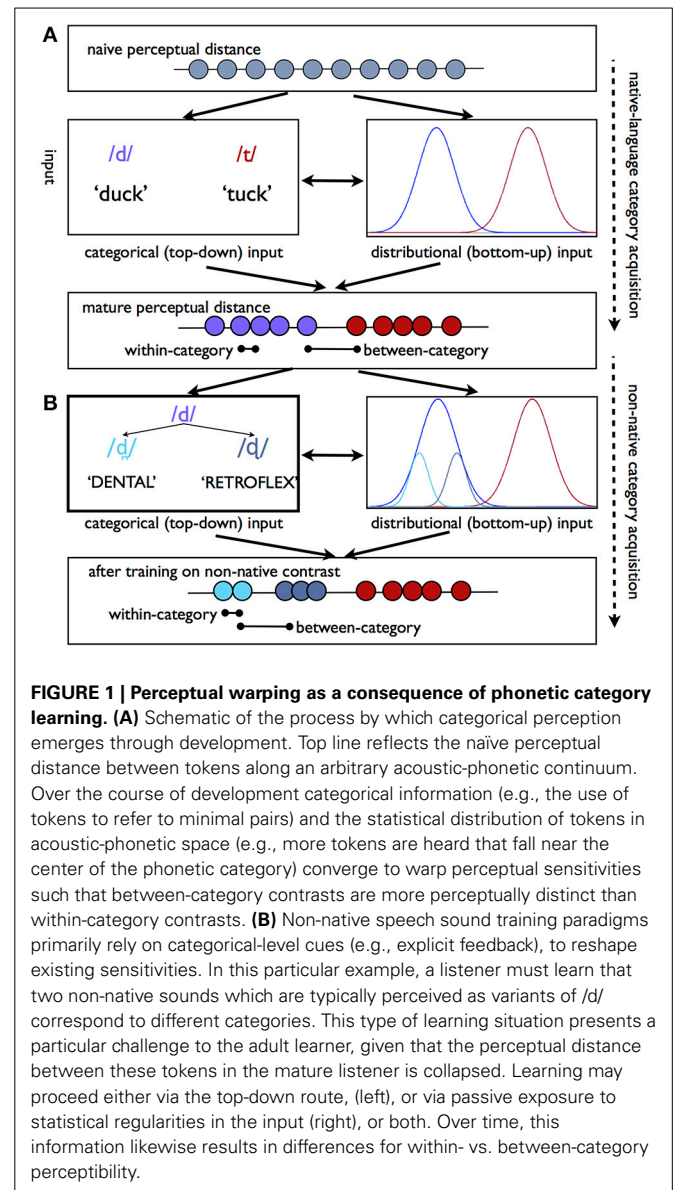
In theory, all that is necessary for successful phonetic processing is the discovery of the location of phonetic boundaries in acoustic space. However, human speech perception is more complex than this. Over the course of development, acoustic differences that are contrastive in the child's native language become perceived as more distinctive, while those that are non-contrastive (i.e., they fall within the same phonetic category) become perceived as less distinctive (Eimas et al., 1971; Werker and Tees, 1999; Polka et al., 2001; Best and McRoberts, 2003; Kuhl et al., 2008). This perceptual pattern, namely excellent discrimination of items that fall between categories in the face of poor discrimination of items within phonetic categories, is referred to as categorical perception (Liberman et al., 1957). Through early childhood, this trajectory continues, with native-language contrasts becoming perceived more categorically and non-native contrasts becoming less categorical between ages 2 and 6 (e.g., Burnham et al., 1991). By the time listeners reach adulthood,

many phonetic categories are perceived categorically, and as such the mature phonetic processing system is not only sensitive to the boundaries of phonetic space, but exhibits perceptual warping such that certain portions of that space are easier to discriminate than others.

It is a matter of significant debate as to how categorical perception emerges. One proposal is that the statistical distribution of phonetic tokens in acoustic-phonetic space may provide sufficient information to reshape perceptual sensitivities even before functional phonetic categories have developed in the learner (Kuhl et al., 1992; Guenther and Gjaja, 1996; Maye et al., 2002, 2008). This view stems from the observation that the speech tokens that listeners are exposed to are not evenly distributed in acoustic space. For instance, the listener will hear many more examples of /t/ with a VOT near 60 ms than with a VOT of 120 ms, although both are considered to be members of /t/ category (Figure 1A). Some evidence suggests that infant and adult listeners alike may be able to take advantage of distributional/statistical information in order to amplify acoustic distinctions that fall between different distributions and minimize those within the distribution (Maye et al., 2002, 2008; Hayes-Harb, 2007; Emberson et al., 2013). Crucially, this perceptual reshaping can happen even when listeners know nothing about the functional use of phonetic categories—that is, when listeners are only passively exposed to the input, and never hear speech sounds used referentially.

Nonetheless, young and old learners alike are exposed to additional sources of information regarding the sounds that are contrastive in their language. The use of phonetic categories to refer to different visual objects has been shown to result in better discrimination of those sounds (Yeung and Werker, 2009), and the appearance of different sounds in different lexical contexts may have a similar effect (Feldman et al., 2013). Ultimately, it is clear that the language learner must eventually learn the phonology of his or her own language. This sort of top-down information may continue to reshape perceptual sensitivities to these same sounds as the language user matures (Figure 1A). Given that the warping of perceptual space seen in adults may have arisen both as a consequence of passive, bottom-up data derived from the statistical distribution of tokens in the input, as well as the acquisition of functional, category-level information about the phonology of one's language, it is challenging to attribute behavioral and neural patterns we observe in adult phonemic perception to either bottom-up sensitivities to the acoustic input or top-down knowledge of phonetic category status.

This obstacle is particularly evident when discussing the neural systems that are responsive to phonetic category identity. For instance, if it is the case that statistical/distributional information in the signal is sufficient to guide the emergence of phonetic category identity, neural structures that are responsive to phonetic category structure may be those that have formed as a function of these bottom-up properties of the signal rather than as a response to the functional, linguistic use of phonetic categories. To the extent to which we believe passive mechanisms may also be sufficient to reshape sensitivities to complex acoustic information in auditory and auditory association cortex (Pallier et al., 1997; Zhou and Merzenich, 2007), mature, native language



neural sensitivity may in large part reflect these bottom-up mechanisms.

In order to develop plausible hypotheses about the nature of phonetic category formation in adult, non-native acquisition, it is first important to discuss current evidence regarding the neural processing of native-language phonetic category structure.

NATIVE LANGUAGE PHONETIC CATEGORY STRUCTURE IN THE BRAIN

It has been well established that the bilateral superior temporal lobes are preferentially responsive to intelligible speech sounds compared to identifiable non-speech sounds (e.g., Belin et al., 2000, 2002), and compared to acoustically-matched sounds which are unintelligible as speech. (Okada et al., 2010; Evans et al., 2013). Recent evidence from direct cortical recording has revealed populations of neurons that code for dimensions of the phonetic

inventory, including place of articulation and manner of articulation, showing that the human temporal lobes are well-equipped to distinguish between the sounds of speech (Chang et al., 2010; Mesgarani et al., 2014). What is less clear is the extent to which these systems are specifically tuned to native-language contrasts or whether they show a more general sensitivity to, or preference for, many classes of speech sounds (for a more complete review, see Turkeltaub and Branch Coslett, 2010).

In order to answer this question, the review below is restricted to evidence in which the neural response reflects specific sensitivity to the internal structure of native-language speech categories. In particular, studies which show different responses to variability within and between categories can be said to show this kind of sensitivity.

MID-TO-POSTERIOR SUPERIOR TEMPORAL GYRUS TUNING TO NATIVE-LANGUAGE CATEGORY STRUCTURE

As a seat of complex acoustic processing, the bilateral temporal lobes play a primary role in processing the auditory details of the speech signal. Evidence suggests that there is a gradient of sensitivity along the temporal lobe from finer-grained acoustic processing near Heschl's gyrus (HG) to increasing specificity in tuning to one's native language as the processing stream flows in both the anterior and posterior directions along the STG/STS. In particular, middle portions lateral to HG have been shown to respond to native speech sounds compared to well-controlled non-speech sounds (Liebenthal et al., 2005; see Turkeltaub and Branch Coslett, 2010; DeWitt and Rauschecker, 2012 for meta-analyses). In contrast, regions including middle-STG territory lateral to HG and extending posterior along the STG/STS have been more tightly linked to phonological processing, and in particular have been shown to be sensitive to phonetic category structure. For instance, the bilateral superior temporal gyrus and superior temporal sulcus (STG and STS) are sensitive to how typical a speech sound is a member of its phonetic category (Guenther et al., 2004; Myers, 2007). This gradient response reflects the non-uniform structure of phonetic categories, suggesting that the temporal lobes are tuned to the internal perceptual structure of native-language categories, and are not merely sensitive to all speech sound dimensions.

The sensitivity of left posterior temporal areas in the perception of contrasts between- and within-category is supported by a series of studies using repetition suppression or habituation designs. While these studies differ in their details, all share a design in which a repeated presentation of a phonetic stimulus is followed by either an identical stimulus or a change in stimulus. Neural sensitivity to changes between and within the category are assessed by comparing activation for "change" trials to "repeat" trials. More categorical responses, as reflected by selective sensitivity to between-category compared to either repeated or within-category contrasts, were found in the left supramarginal gyrus, and in left posterior superior temporal sulcus (Joanisse et al., 2007; Myers et al., 2009).

Evidence that the temporal lobes respond to native-language contrasts also comes from the mismatch negativity paradigm. Larger MMN responses are seen to deviant tokens which cross a phonetic category boundary than those that change within

the category (Phillips, 2001). Of interest, the MMN source is thought to arise from bilateral temporal cortex, shows greater left-lateralization for native language contrasts (see Naatanen et al., 2007 for review; Zevin et al., 2010), and MMN responses over the left temporal lobe are larger to phonetic than non-phonetic contrasts when employing direct cortical recording (Molholm et al., 2014), particularly in or near the STS. This MMN response is not restricted to temporal lobes however; the MMN response is thought to have a secondary source in left prefrontal cortex (Paavilainen et al., 2003, see further discussion of frontal contributions in section "Left inferior frontal involvement in categorical responses to native-language contrasts").

Discussion above has been limited to studies which specifically show differences in responsiveness to within vs. between-category contrasts. Nonetheless, converging evidence from other types of designs suggests that posterior portions of the left STG/STS are responsive to the category identity of native-language speech sounds (e.g., Desai et al., 2008; Chang et al., 2010; Liebenthal et al., 2010; Mesgarani et al., 2014). Of interest, speech category sensitivity in temporal regions is not limited to purely perceptual paradigms but it is also evident in auditory feedback for speech motor control. In particular, when speakers receive perturbations to auditory feedback that fall near the phonetic category boundary, greater compensation is seen in the speech production response, with concomitant greater activation for near-boundary compared to far-boundary shifts in the bilateral posterior STG (Niziolek and Guenther, 2013). Taken together, these results suggest that the posterior superior temporal lobes, particularly on the left, show fine-grained tuning to the acoustic properties of one's native language, with greater (or perhaps selective) neural sensitivity to acoustic distinctions that result in a change in phonetic category. It is of note that responses in the posterior STG/STS are not driven solely by bottom-up characteristics of the acoustic signal, but are also modulated by shifts in phonetic category boundary, and by changes in the perceptual status of the stimulus (e.g., non-speech to speech) (e.g., Desai et al., 2008; Gow et al., 2008; Myers and Blumstein, 2008).

LEFT INFERIOR FRONTAL INVOLVEMENT IN CATEGORICAL RESPONSES TO NATIVE-LANGUAGE CONTRASTS

While the temporal lobes no doubt shoulder much of the burden in processing the sounds of speech, evidence suggests that left prefrontal cortex also plays a role in the computation of phonetic identity. In two passive repetition suppression studies, responses to category-level information (e.g., greater responses to between-category than within-category shifts, yet no difference between within-category and repeated trials) were seen in premotor areas (Chevillet et al., 2013), and in an "invariant" response in the precentral gyrus and pars opercularis (Myers et al., 2009). Pre-motor areas which had been identified as sensitive to between-category changes showed significant task-related functional connectivity during passive listening to sites in the posterior temporal lobes (Chevillet et al., 2013), which led to the interpretation that phonetic category computations rely on forward projections between the temporal and frontal lobes along the dorsal route (Hickok and Poeppel, 2004). A recent analysis by Lee et al. (2012) examined category-level sensitivity of several brain regions using new

data in which participants passively listened to syllables along a ba—da continuum as well as using existing data from a repetition suppression paradigm (Raizada and Poldrack, 2007). In this study, the authors employed a moving searchlight technique with whole-brain multi-voxel pattern analysis (MVPA, Kriegeskorte et al., 2006) to search for clusters of voxels in which the patterns of activation could discriminate between two different phoneme categories (da vs. ba). Sensitivity to category-level information was seen in the left pars opercularis and pre-supplementary motor region as well as in the left superior temporal lobe. Converging evidence from studies in which cortical processing is disrupted using TMS also points to a role for frontal structures in computing category membership: stimulation of motor cortex sites slightly alters categorical perception in phoneme categorization and discrimination tasks (Mottonen and Watkins, 2009; D'Ausilio et al., 2012).

What is less clear is the precise role or roles of these frontal structures, which may indeed constitute functionally distinct sub-regions within the frontal lobes. The implication of premotor areas has led to the hypothesis that articulatory codes for speech may be activated to either guide perceptual hypotheses generated in the temporal lobes, or, more radically, to act as the contents of the abstract speech sound category (Lieberman and Mattingly, 1985). At the same time, the influence of frontal areas may not be limited to access to articulatory information, nor, indeed, is category-sensitive activation limited to premotor cortex. Anterior to premotor cortex, regions in Broca's homolog have been found to be sensitive to category-level information in a domain-general sense, and evidence from single-cell recordings in non-human primates suggests that invariant responses to category membership may arise in frontal areas (e.g., Freedman et al., 2001). As such, the involvement of frontal areas may not reflect motor-related activity, but may reflect access to a more abstract category representation. In general, these results suggest that a complex of information arising from prefrontal regions generally may guide perception (Davis and Johnsrude, 2007; Liebenthal et al., 2013).

At the same time, the role of frontal structures in speech intelligibility “in the wild” has been questioned (Hickok and Poeppel, 2007; Hickok et al., 2011). It has been observed that lesions to left inferior frontal areas need not impair explicit decisions of phonetic category identity, and rarely create errors in phonemic perception (Basso et al., 1977; Rogalsky et al., 2011), and that while stimulation of premotor sites may impair categorization decisions, there is no evidence of deficits in comprehension as a result of such stimulation (Krieger-Redwood et al., 2013). Engagement of frontal structures for speech perception has been especially observed in the presence of ambiguity or noise in the signal (Binder et al., 2004; D'Ausilio et al., 2012), and as such frontal areas are argued to be peripheral to processing the sounds of speech. Some (D'Ausilio et al., 2012) while agreeing that frontal involvement for perception seems especially important in the context of noise in the signal, point out that noisy signals and imperfect productions are actually the norm rather than the exception in the typical language environment, and that we should resist the temptation to view frontal influences in speech perception as epiphenomenal. As such the types of activation patterns observed in studies of categorical perception can be

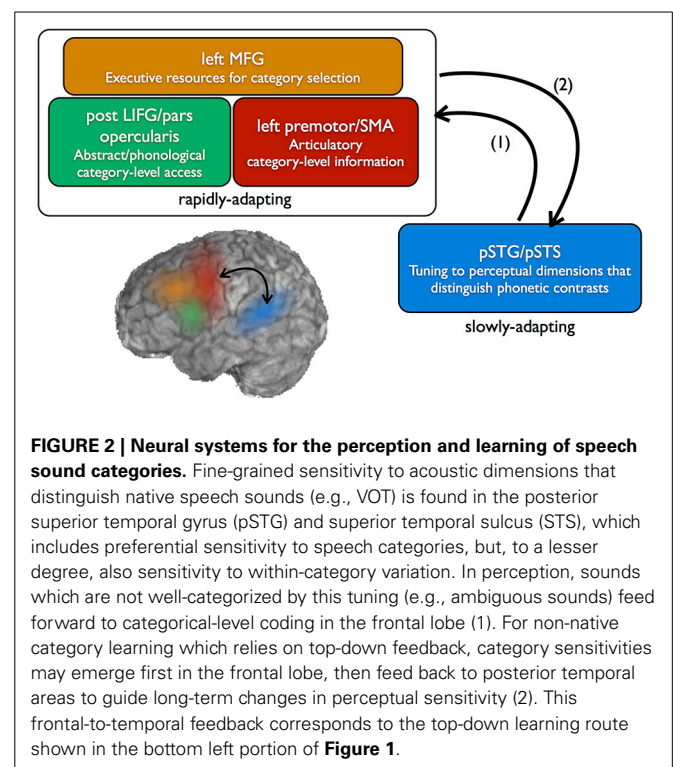
accommodated by assuming that frontal structures are consulted in less optimal listening conditions.

Whether the codes accessed in the inferior frontal lobes are articulatory or abstract in nature, evidence suggests that coding in the left prefrontal areas is more categorical than that represented in the temporal lobe. This suggests an architecture whereby fine-grained acoustic-phonetic details of the speech stream are processed in the left STG/STS, and this information is then projected forward to prefrontal regions to consult with categorical-level codes in a complex of frontal areas (Figure 2).

NON-NATIVE PHONETIC CATEGORY ACQUISITION: A CASE OF FUNCTIONAL PLASTICITY

As discussed above, the mature language learner comes to the second-language learning process with a set of pre-established perceptual sensitivities which may either facilitate or hinder the acquisition of a new category. In particular, to learn a new phonetic contrast which falls within the acoustic territory occupied by native language sounds, the listener must learn to either (a) shift an existing phonetic category boundary, as in the case of an English speaker learning the a VOT boundary which corresponds to the Spanish d/t contrast, or (b) divide an existing phonetic category into two, as in the case of the English listener learning to perceive the Hindi dental vs. retroflex stop contrast (Figure 1B). This latter scenario seems particularly challenging, as an entire native-language architecture has developed which prevents the listener from perceiving distinctions within the phonetic category.

By the time adulthood is reached, one's sensitivities to native-language phonetic categories have reached a stability point. In fact, non-native categories, particularly those that fall within



an existing native-language category, are notoriously difficult to acquire in adulthood (Best et al., 1988). The fact that even motivated adults can struggle to distinguish certain non-native contrasts has led to conclusion that there is a critical period for phonetic category learning. This critical period may result from losses in neuroplasticity which prevent the adult listener from altering perceptual sensitivities in order to accommodate these into native language category structure (Pallier et al., 1997).

Nonetheless, with sufficient training, many individuals are able to learn to perceive non-native contrasts (Logan et al., 1991; Lively et al., 1993; Bradlow et al., 1997; Golestani and Zatorre, 2009), with some learners achieving native-like proficiency. Individuals who learn to speak a second language “in real life” (as opposed to in laboratory training conditions) have multiple sources of information which can guide the formation of new phonetic categories. Much as infants may be able to use information regarding the statistical distribution of phonetic tokens in acoustic space to reshape sensitivities, adults who are exposed to a non-native language will likely hear the same kinds of distributional information, whether they are able to take advantage of it or not (Figure 1B). Crucially for the adult learner, top-down information about phonetic category identity, either in the form of referential information (e.g., using two sounds to refer to two different words) or even through explicit classroom instruction, is often very salient in the environment. Unfortunately, almost all studies regarding the emergence of non-native phonetic sensitivity in the brain have used training paradigms where top-down information about category identity is provided to participants (see left side of Figure 1B). As such, we can draw limited conclusions regarding the emergence of neural sensitivities to non-native contrasts via more passive, bottom-up mechanisms in which listeners capitalize on distributional properties of the input.

PERCEPTUAL WARPING FROM NON-NATIVE CATEGORY TRAINING

Before considering the neural structures that are sensitive to phonetic category training, it is first important to assess whether top-down (e.g., categorization) training results in a perceptual pattern that resembles native language perception. As discussed above, acquisition of a native-language contrast appears to involve not only learning the boundary between categories, but also results in changes in perception of acoustic contrasts within and between these categories. Given that the types of training paradigms used in many studies bear a scant relationship to the authentic language acquisition environment, it would not be surprising to find that participants might successfully be able to complete a categorization task using non-native stimuli (that is, learn the location of the category boundary) while showing no difference in the relative perceptibility of between and within-category contrasts. Fortunately, converging evidence suggests that training participants on category-level information results in changes in discriminability of tokens across the trained continuum. Studies investigating training on the /l/ vs. /r/ contrast in native-Japanese listeners (McCandliss et al., 2002), and the Hindi dental vs. retroflex stop contrast (/ɖ/ vs. /ɖʱ/) in English listeners (Golestani and Zatorre, 2009) show that training on categorization tasks transfers to discrimination tasks, and specificity of the discrimination peak appears to be closely linked to both the location

of the learned category boundary for each participant as well as to the relative success of each listener in acquiring the new contrast (Guenther et al., 1999; Wade and Holt, 2005; Golestani and Zatorre, 2009; Swan and Myers, 2013).

FUNCTIONAL BRAIN CHANGES RESULTING FROM NON-NATIVE CATEGORY TRAINING

When adults learn a non-native contrast, either via explicit category training, or from more naturalistic experience, brain structures which show specific sensitivity to native language contrasts must somehow reshape responses in order to accommodate a new categorical division of acoustic space. In general, we may ask whether the same neural resources are recruited for non-native speech sound perception following training as are implicated for native-language perception. Non-native phonetic training often takes the form of categorization training on either syllables or minimal pairs with explicit feedback to participants, often using a perceptual fading design, in which participants initially categorize maximally distinct tokens, then proceed to finer distinctions in a stepwise fashion (Golestani and Zatorre, 2004; Liebenthal et al., 2010; Myers and Swan, 2012). In this situation the availability of category-level information can be said to be at its maximum, as participants receive feedback regarding the accuracy of the categorical decision. When examining task-related activation before and after training, a wide network of regions are recruited, including bilateral temporal and left inferior frontal structures (Callan et al., 2003; Golestani and Zatorre, 2004) which show greater task-related activation to non-native sounds after compared to before training. Concordant evidence using a similar training paradigm yielded greater activation for non-native categorization post-training in a series of frontal regions and left inferior parietal regions (Ventura-Campos et al., 2013). Given the explicit nature of the categorization task, these studies are vulnerable to the criticism that the activation in inferior frontal regions is related to the metalinguistic task, rather than to the perception of phonetic category differences *per se* (see Section, “Left inferior frontal involvement in categorical responses to native-language contrasts,” above).

Nonetheless, a study from our lab supported the involvement of a separate set of frontal structures, namely the left and right middle frontal gyri in categorical perception of learned speech sounds (Myers and Swan, 2012). In this study, participants were trained to categorize a three-way phonetic continuum (voiced stops ranging from dental to retroflex to velar place of articulation: /d/ vs. /dʱ/ vs. /g/) according to two different boundary locations, with one group trained to place the category boundary between the dental and retroflex tokens, and a separate group trained to place the category boundary between the retroflex and velar tokens. Participants were trained over two sessions, and neural sensitivity post-training was assessed using an short-interval habituation design which did not require participants to categorize speech sounds (see Joanisse et al., 2007; Myers et al., 2009). Despite the fact that the task required no judgments of phonetic category identity during scanning, activity in the bilateral middle frontal gyri reflected differential sensitivity to between vs. within-category contrasts according to the training of the participants. Of interest, no difference in activation for between

vs. within-category contrasts was seen in the temporal lobes, suggesting that differential responsiveness to learned category structure need not rely on retuning of sensitivities in the temporal lobe.

Support for the involvement of inferior frontal regions for non-native category learning can be seen in other passive paradigms. An analysis of resting-state functional data before and after intensive (one day) and distributed (six sessions) of non-native category training suggested that a decrease in degree of functional connectivity between two regions of interest in the left frontal operculum and left superior parietal lobule was significantly correlated with participant accuracy (Ventura-Campos et al., 2013). To unpack this result further, this suggests that individuals who were more successful in learning the non-native contrast showed a decrease in the degree of coherence between frontal and parietal structures, perhaps reflecting a decreased reliance on the frontal-to-parietal connection over the course of learning.

Nonetheless, training-related activity is not exclusive to these frontal regions. A series of training studies have shown significant involvement of temporal structures in sensitivity to trained speech and complex non-speech sounds. Lieberthal et al. (2010) trained participants over four sessions to identify non-speech sounds which resembled speech sounds in their spectral and temporal properties. Activation in the left posterior STS increased for trained non-speech sounds following training, with additional small clusters in left inferior frontal areas. Similarly, Leech et al. (2009) used an implicit training method which paired complex non-speech sounds with unique characters in a video game. After several sessions playing the game, the degree of increased activation within a speech-selective ROI in the left STS posterior to HG correlated with the degree of training success. Notably, this pattern did not emerge in a whole-brain analysis, and it may be the case that the creation of the speech-selective ROI may have eliminated the consideration of regions that would not respond to the speech vs. environmental sound contrast. Left posterior STS/STG activation has also been shown to correlate with training success in pitch pattern learning (Wong et al., 2007).

It is possible that the asymmetry between studies which have shown involvement of temporal regions in novel contrast sensitivity and those which have not may be attributed to the duration and/or intensity of training. Our study (Myers and Swan, 2012) employed only two 45-min sessions of training, whereas other studies have employed multiple intense training sessions. One proposal is that sensitivity to category-level information emerges early in the frontal lobe and only later is evident in temporal structures. This pattern would be consistent with a variety of proposals outside the language literature which suggest a shift from executive or category-level processing to sensory-based processing as expertise is gained (Ahissar and Hochstein, 2004; Nahum et al., 2008).

In order to address this question, we performed a replication of Myers and Swan (2012) in which we extended the training to ten 45-min sessions over 2 weeks (Myers et al., under review). Participants in this study were now trained to just distinguish dental and retroflex voiced stop consonants. Pre-

and post-training scans were performed using the short-interval habituation design (Myers and Swan, 2012), and during scanning participants were asked to perform a pitch detection task in which they responded to high-pitched syllables on infrequent catch trials. Rather than search for areas which show global changes in activation as a function of training, we targeted regions which showed a differential sensitivity to between-category compared to within-category contrasts. Similar to other studies investigating categorical perception, the logic was that regions which showed sensitivity to the learned category structure following training could not be said to be influenced merely by changes in attention, motivation, or familiarity with the stimuli. At pre-test, only the left middle frontal gyrus showed differences in activation for between- compared to within-category stimuli. After training, activation differences were seen in a bilateral network including the left precentral gyrus, right and left STG, left IPL, and right insula. Importantly, both left and right posterior STG were shown to be correlated at post-test with participants' behavioral accuracy at post-test, suggesting that temporal activation resulting from 10 days of training was not only sensitive to the "categorical" nature of the stimuli (between vs. within) but also was predictive of learning.

INDIVIDUAL VARIABILITY IN SPEECH SOUND LEARNING

Many of the above-mentioned studies have searched for the neural correlates of variability in the perception of non-native contrasts. Variability in non-native perception is evident not only in training studies, but also in the varying degrees of proficiency that second-language learners attain (e.g., Bradlow et al., 1997; Flege et al., 1999). Studies which have examined the neural correlates of these differences among learners have come to differing conclusions regarding the source of this variability. Diaz et al. (2008) report that poorer perceivers of non-native contrasts showed an attenuated MMN response compared to better perceivers. The source of the MMN was inferred from the latency and distribution across electrodes, and was hypothesized by the authors to be the frontal component. The authors interpreted this response as reflecting engagement of an attentional network in better perceivers, whereas the lack of difference in the temporal component reflected similar fidelity in acoustic-phonetic processing across better and poorer perceivers. By contrast, a study by Raizada showed that the patterns of activation within the right Heschl's gyri of Japanese L2 learners were predictive of that population's ability to discriminate /l/ vs. /r/ contrasts (Raizada et al., 2010). In the end, it is likely that functional variation at multiple points in the phonetic processing stream contribute to differences in learning success, with some learners excelling because of superior acoustic processing, and others achieving success due to the appropriate deployment of auditory attention, for instance.

Individual differences in brain structure are also predictive of phonetic learning success. Work by Golestani et al. (2002) and Golestani and Pallier (2007) showed that better learners showed differences in brain morphology in the left HG and a greater leftwards asymmetry in parietal cortex which was evident in WM volume. This asymmetry may reflect more efficient or precise coding of acoustic information which is especially relevant in speech sound learning (although see Burgaleta et al., 2014 for

a null finding relating brain morphology to speech perception abilities in a bilingual population). An advantage for processing the fine-grained aspects of sound might have surprising professional consequences as well. A unique study (Golestani et al., 2011) found that individuals who were employed as phoneticians showed differences in the morphology of left Heschl's gyrus compared to a control group. Of interest, there was also a correlation between the surface area and structure of the left pars opercularis and years of experience working as a phonetician, providing a hint that frontal differences in morphology may have arisen through experience-induced plasticity rather than from innate differences in brain structure.

This finding raises the question of whether experience learning a non-native phonetic contrast might actually induce structural changes in the brain. This type of plasticity is not unprecedented. Changes in brain morphology have been found following training on a variety of tasks (see Zatorre et al., 2012 for a review) and relevant for the current discussion, following a semester of intensive second-language learning (Stein et al., 2012). In our study of intensive non-native speech sound training (Myers et al., under review), changes in gray matter volume were seen in a region deep to the left supramarginal gyrus comparing pre-training scans to post-training scans. This same region is among the set of regions in which individual variation is associated with successful phonetic category learning (Golestani et al., 2007), and with individual differences in non-native sound production (Golestani and Pallier, 2007). Moreover, in our study, the coherence of white matter pathways (as measured by DTI) near the arcuate fasciculus in this same vicinity was seen to correlate with learning success, suggesting that the strength of frontal-to-posterior connections along the dorsal route contributes to non-native category learning. Taken together, these results suggest that even relatively short-term training can serve to strengthen connections that are necessary for non-native speech sound learning.

A FRONTAL TO TEMPORAL ROUTE FOR PHONETIC CATEGORY LEARNING

The extant literature on non-native speech sound learning suggests that the long-term consequence of speech category training is the retuning of posterior temporal regions such that they show increased sensitivity to the dimensions of the learned speech sounds. Of note, this same region also shows sensitivity to phonetic category structure in native speech perception which is presumably acquired slowly over the course of development. Broadly speaking, this is consistent with most models of the neural bases of speech perception (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). However, data suggests that short-term adjustments to learned phonetic category structure may be seen first in the frontal lobe (Myers and Swan, 2012), and only after sustained or more intensive training do these same sensitivities appear in the posterior temporal lobe (e.g., Leech et al., 2009; Myers et al., under review). Moreover, individual training success correlates with the coherence of white matter pathways at pre-training (Myers et al., under review), in an area that is consistent with the dorsal stream route connecting posterior temporoparietal regions to frontal structures (Hickok and Poeppel, 2007). Of note, this frontal-to-temporoparietal route is not the only

connection which has been shown to correlate with non-native training success. Resting-state functional connectivity before and after training reflects a decreased reliance on frontal-to-superior parietal connections after training (Ventura-Campos et al., 2013) which has been attributed to a decreased reliance on a "salience" network. Of note, Ventura-Campos and colleagues also show strong resting-state connectivity between the frontal operculum and the SMG, but this connectivity did not show any significant correlation with training success. The authors speculate that this lack of correlation may in part reflect the lower individual variability shown in the frontal-to-SMG connectivity findings.

This pattern of results leads us to propose that early learning of non-native speech categories in the context of explicit top-down information involves first feed-forward connections from posterior temporal cortex to ventrolateral prefrontal cortex (Garell et al., 2013), where acoustic representations access category-level (articulatory, phonological, or abstract) information (Figure 2). Categorical sensitivity to non-native speech sounds emerges first in the inferior frontal lobe as participants learn the boundaries through acoustic space which define functional categories. This allows for rapid learning of category boundaries without fundamentally reshaping neural sensitivity to low-level details of the signal. Over time, frontal-to-temporal feedback connections may serve as an error signal on auditory sensitivities to these speech sounds, reshaping the sensitivity of auditory association cortex. The view that frontal-to-temporal feedback signals may play a role in rapid auditory plasticity finds support from animal models (Winkowski et al., 2013), and human data suggests that stimulation of frontal sites may facilitate auditory perceptual learning (Sehm et al., 2013). We suggest that the process of retuning sensitivities in the temporal lobes unfolds more slowly, over the course of minimally several days of training or experience.

Notably, our findings suggest that learners can achieve at least moderate success in training without any detectable change in the responsiveness of the temporal lobes (Myers and Swan, 2012). One open question is whether training which only recruits frontal lobe is retained over time. It may be the case that temporal lobe encoding is actually necessary for long-term learning of the speech contrast (Myers et al., under review). It is also unknown whether short-term learning in the frontal lobes reflects a different perceptual status of the stimulus as compared to when this sensitivity emerges in the temporal lobes. For instance, it is possible that frontal encoding relies more heavily on domain-general systems for perceptual categorization whereas temporal encoding reflects a more genuine status of the stimuli as phonetic categories.

A system which allowed for rapid, on-the-fly adaptation to new phonetic category structure might present several advantages not only for learning new speech contrasts, but also for processing details of native language speech. As listeners, we are exposed to speech variants that differ significantly from our native language phonetic categories, for instance, in the case of foreign accents, yet we are also able to quickly adapt to non-standard speech sounds (Bradlow and Bent, 2008; Kraljic et al., 2008). A neural system which likewise showed rapid, contextually-sensitive flexibility to shift phonetic category boundaries would facilitate this kind of adaptation. At the same time, unconstrained

flexibility in processing non-standard speech sounds could be disadvantageous—for instance, one’s phonetic category boundaries should not be continuously perturbed by every exposure to a new talker or accent. As such, a separate neural system which shows more stable, slowly-adapting responses would be also advantageous.

Several testable predictions fall out of this type of model. First, if frontal-to-temporal feedback is necessary for non-native phonetic category learning, patients with frontal lobe pathology (e.g., individuals with Broca’s aphasia) would have significant deficits in the acquisition and retention of new category information, while retaining sensitivities to native language phonetic category information learned pre-insult. Second, under the assumption that frontal systems are only engaged when category-level information is required for acquisition, it should be the case that incidental learning of phonetic categories, whether via sensitivity to statistical properties of the input (Hayes-Harb, 2007), or through other implicit methods (e.g., Lim and Holt, 2011; Vlahou et al., 2012) should be spared in this same population. Finally, if this frontal-to-temporal pathway is directed along the arcuate fasciculus, the coherence of this pathway should predict better speech sound learning at an individual level (see Myers et al., under review), and category training should be difficult for patients whose lesions implicate this pathway. Finally, as shown by Ventura-Campos et al. (2013), functional connectivity between frontal and posterior sites should inversely correlate with learning success as listeners transfer category-level learning to reshape perceptual sensitivities in the posterior temporal lobe.

CONCLUSION

The model described here is motivated largely through training studies which have used explicit, metalinguistic tasks in order to induce phonetic category sensitivities. There is still much to learn regarding phonetic category acquisition. First, little is known regarding the mechanisms which support encoding of statistical/distributional information which may reshape sensitivities “for free” as listeners are passively exposed to a new language. In the visual and auditory (non-speech) modalities, evidence suggests that medial temporal lobe and subcortical structures, in particular the caudate, may play a crucial role in encoding statistical regularities in the input (e.g., Turk-Browne et al., 2009; Durrant et al., 2013). Yet it is unknown whether the same structures mediate statistical learning for non-native speech sounds. At least one study (Golestani and Zatorre, 2004) showed engagement of the caudate for non-native speech sounds after training, although this result was attributed by these authors to the role of the caudate in motor speech control rather than in statistical learning.

Relatedly, the process of learning a non-native contrast involves encoding speech sounds in memory, but also protecting these newly-learned sounds from interference from existing similar speech sounds in one’s native language. Recent work from our lab (Earle and Myers, under review) suggests that consolidation during sleep plays a significant role in this process. Participants who learned a non-native speech contrast in the evening showed improvements in discrimination of this contrast after an overnight interval and 24 h after learning, whereas

participants who learned the same contrast in the morning did not show retention of the contrast after sleep. A follow-up suggested that the morning group’s failure to retain the contrast was due to interference from exposure to similar native-language speech sounds over the course of the day. Taken together, this evidence suggests that (a) sleep plays a stabilizing role in the perceptual learning of speech sounds and (b) interference before sleep can serve to disrupt perceptual learning. This finding joins a literature on perceptual learning of synthetic speech sounds (Fenn et al., 2003, 2013) and on lexical learning which point to a crucial role for sleep in either abstracting away from the episodic details of the input, or to protection of learning from decay. While the neural bases of sleep-related consolidation for speech sounds have yet to be investigated, following a complementary systems memory framework (McClelland et al., 1995; O’Reilly and Rudy, 2001), one might predict that immediate encoding of novel speech sounds would implicate the hippocampus, while the overnight interval would serve to transfer this learning to cortical systems (e.g., Davis et al., 2009). This hippocampal-to-cortical transfer is thought to support abstraction from the episodic details of the signal to a more abstract representation of the input.

Perhaps most importantly it has yet to be determined whether second-language learning in immersion or in the classroom induces the same types of neural responses observed here. To fully understand the boundaries of plasticity in adult phonetic category learning, future research will need to be directed at these topics.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD grants R03 DC009495 and R01 DC013064 to Emily B. Myers, and NIH NICHD grant P01 HD001994 (Rueckl, PI). Thanks to Sayako Earle and Alexis Johns for helpful comments on an earlier version of this manuscript. The content is the responsibility of the author and does not necessarily represent official views of the NIH, NIDCD, or NICHD.

REFERENCES

- Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464. doi: 10.1016/j.tics.2004.08.011
- Basso, A., Casati, G., and Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex* 13, 85–95. doi: 10.1016/S0010-9452(77)80057-9
- Belin, P., Zatorre, R. J., and Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res. Cogn. Brain Res.* 13, 17–26. doi: 10.1016/S0926-6410(01)00084-2
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078
- Best, C. C., and McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Lang. Speech* 46(pt 2–3), 183–216. doi: 10.1177/00238309030460020701
- Best, C. T., McRoberts, G. W., and Sithole, N. M. (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. doi: 10.1037/0096-1523.14.3.345
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., and Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301. doi: 10.1038/nn1198
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. doi: 10.1016/j.cognition.2007.04.005

- 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, 2299–2310. doi: 10.1121/1.418276
- Burgaleta, M., Baus, C., Díaz, B., and Sebastián-Gallés, N. (2014). Brain structure is related to speech perception abilities in bilinguals. *Brain Struct. Funct.* 219, 1405–1416. doi: 10.1007/s00429-013-0576-9
- Burnham, D. K., Earnshaw, L. J., and Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *J. Child Lang.* 18, 231–260. doi: 10.1017/S0305000900011041
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., and Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage* 19, 113–124. doi: 10.1016/S1053-8119(03)00020-X
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432. doi: 10.1038/nn.2641
- Chevillet, M. A., Jiang, X., Rauschecker, J. P., and Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *J. Neurosci.* 33, 5208–5215. doi: 10.1523/JNEUROSCI.1870-12.2013
- D'Ausilio, A., Bufalari, I., Salmas, P., and Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex* 48, 882–887. doi: 10.1016/j.cortex.2011.05.017
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., and Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *J. Cogn. Neurosci.* 21, 803–820. doi: 10.1162/jocn.2009.21059
- Davis, M. H., and Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147. doi: 10.1016/j.heares.2007.01.014
- Desai, R., Liebenthal, E., Waldron, E., and Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *J. Cogn. Neurosci.* 20, 1174–1188. doi: 10.1162/jocn.2008.20081
- DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 109, E505–E514. doi: 10.1073/pnas.1113427109
- Díaz, B., Baus, C., Escera, C., Costa, A., and Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16083–16088. doi: 10.1073/pnas.0805022105
- Durrant, S. J., Cairney, S. A., and Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb. Cortex* 23, 2467–2478. doi: 10.1093/cercor/bhs244
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306. doi: 10.1126/science.171.3968.303
- Emberson, L. L., Liu, R., and Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition* 128, 82–102. doi: 10.1016/j.cognition.2012.12.006
- Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., et al. (2013). The pathways for intelligible speech: multivariate and univariate perspectives. *Cereb. Cortex*. doi: 10.1093/cercor/bht083. [Epub ahead of print].
- Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., and Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition* 127, 427–438. doi: 10.1016/j.cognition.2013.02.007
- Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2013). Sleep restores loss of generalized but not rote learning of synthetic speech. *Cognition* 128, 280–286. doi: 10.1016/j.cognition.2013.04.007
- Fenn, K. M., Nusbaum, H. C., and Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature* 425, 614–616. doi: 10.1038/nature01951
- Flège, J. E., MacKay, I. R., and Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *J. Acoust. Soc. Am.* 106, 2973. doi: 10.1121/1.428116
- Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316. doi: 10.1126/science.291.5502.312
- Garell, P. C., Bakken, H., Greenlee, J. D. W., Volkov, I., Reale, R. A., Oya, H., et al. (2013). Functional connection between posterior superior temporal gyrus and ventrolateral prefrontal cortex in human. *Cereb. Cortex* 23, 2309–2321. doi: 10.1093/cercor/bhs220
- Golestani, N., Molko, N., Dehaene, S., LeBihan, D., and Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cereb. Cortex* 17, 575–582. doi: 10.1093/cercor/bhk001
- Golestani, N., and Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cereb. Cortex* 17, 929. doi: 10.1093/cercor/bhl003
- Golestani, N., Paus, T., and Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron* 35, 997–1010. doi: 10.1016/S0896-6273(02)00862-0
- Golestani, N., Price, C. J., and Scott, S. K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *J. Neurosci.* 31, 4213–4220. doi: 10.1523/JNEUROSCI.3891-10.2011
- Golestani, N., and Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* 21, 494–506. doi: 10.1016/j.neuroimage.2003.09.071
- Golestani, N., and Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain Lang.* 109, 55–67. doi: 10.1016/j.bandl.2008.01.005
- Gow, D. W., Segawa, J. A., Ahlfors, S. P., and Lin, F.-H. (2008). Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. *Neuroimage* 43, 614–623. doi: 10.1016/j.neuroimage.2008.07.027
- Guenther, F. H., and Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *J. Acoust. Soc. Am.* 100(2 pt 1), 1111–1121. doi: 10.1121/1.416296
- Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *J. Acoust. Soc. Am.* 106, 2900–2912. doi: 10.1121/1.428112
- Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., and Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hear. Res.* 47, 46–57. doi: 10.1044/1092-4388(2004)005
- Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Lang. Res.* 23, 65–94. doi: 10.1177/0267658307071601
- Hickok, G., Costanzo, M., Capasso, R., and Miceli, G. (2011). The role of Broca's area in speech perception: evidence from aphasia revisited. *Brain Lang.* 119, 214–220. doi: 10.1016/j.bandl.2011.08.001
- Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Joanisse, M. F., Zevin, J. D., and McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cereb. Cortex* 17, 2084–2093. doi: 10.1093/cercor/bhl124
- Kraljic, T., Brennan, S. E., and Samuel, A. G. (2008). Accommodating variation: dialects, idiolects, and speech processing. *Cognition* 107, 54–81. doi: 10.1016/j.cognition.2007.07.013
- Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., and Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *J. Cogn. Neurosci.* 25, 2179–2188. doi: 10.1162/jocn_a_00463
- Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3863–3868. doi: 10.1073/pnas.0600244103
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., 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. Lond. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364
- Lee, Y. S., Turkeltaub, P., Granger, R., and Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J. Neurosci.* 32, 3942–3948. doi: 10.1523/JNEUROSCI.3814-11.2012
- Leech, R., Holt, L. L., Devlin, J. T., and Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *J. Neurosci.* 29, 5234–5239. doi: 10.1523/JNEUROSCI.5758-08.2009

- Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417
- Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6
- Liebethal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., and Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621–1631. doi: 10.1093/cercor/bhi040
- Liebethal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., and Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cereb. Cortex* 20, 2958–2970. doi: 10.1093/cercor/bhq045
- Liebethal, E., Sabri, M., Beardsley, S. A., Mangalathu-Arumana, J., and Desai, A. (2013). Neural dynamics of phonological processing in the dorsal auditory stream. *J. Neurosci.* 33, 15414–15424. doi: 10.1523/JNEUROSCI.1511-13.2013
- Lim, S., and Holt, L. L. (2011). Learning foreign sounds in an alien world: videogame training improves non-native speech categorization. *Cogn. Sci.* 35, 1390–1405. doi: 10.1111/j.1551-6709.2011.01192.x
- Lisker, L., and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–422.
- Lively, S., Logan, J., and Pisoni, D. (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, 1242–1255. doi: 10.1121/1.408177
- Logan, J., Lively, S., and Pisoni, D. (1991). Training Japanese listeners to identify English/r/ and /l/: a first report. *J. Acoust. Soc. Am.* 89, 874–886. doi: 10.1121/1.1894649
- Maye, J., Weiss, D. J., and Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Dev. Sci.* 11, 122–134. doi: 10.1111/j.1467-7687.2007.00653.x
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3
- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (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, 89–108. doi: 10.3758/CABN.2.2.89
- McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457.
- Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994
- Molholm, S., Mercier, M. R., Liebethal, E., Schwartz, T. H., Ritter, W., Foxe, J. J., et al. (2014). Mapping phonemic processing zones along human perisylvian cortex: an electro-corticographic investigation. *Brain Struct. Funct.* 219, 1369–1383. doi: 10.1007/s00429-013-0574-y
- Mottron, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819. doi: 10.1523/JNEUROSCI.6018-08.2009
- Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: an fMRI investigation. *Neuropsychologia* 45, 1463–1473. doi: 10.1016/j.neuropsychologia.2006.11.005
- Myers, E. B., and Blumstein, S. E. (2008). The neural bases of the lexical effect: an fMRI investigation. *Cereb. Cortex* 18, 278. doi: 10.1093/cercor/bhm053
- Myers, E. B., Blumstein, S. E., Walsh, E., and Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol. Sci.* 20, 895–903. doi: 10.1111/j.1467-9280.2009.02380.x
- Myers, E. B., and Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *J. Cogn. Neurosci.* 24, 1695–1708. doi: 10.1162/jocn_a_00243
- Naatnen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026
- Nahum, M., Nelken, I., and Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS Biol.* 6:e126. doi: 10.1371/journal.pbio.0060126
- Niziolek, C. A., and Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *J. Neurosci.* 33, 12090–12098. doi: 10.1523/JNEUROSCI.1008-13.2013
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex* 20, 2486–2495. doi: 10.1093/cercor/bhp318
- O'Reilly, R. C., and Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol. Rev.* 108, 311–345. doi: 10.1037/0033-295X.108.2.311
- Paavilainen, P., Mikkonen, M., Kilpeläinen, M., Lehtinen, R., Saarela, M., and Tapola, L. (2003). Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: a human auditory event-related potential study. *Neurosci. Lett.* 349, 79–82. doi: 10.1016/S0304-3940(03)00787-0
- Pallier, C., Bosch, L., and Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition* 64, B9–B17. doi: 10.1016/S0010-0277(97)00030-9
- Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cogn. Sci.* 25, 711–731. doi: 10.1207/s15516709cog2505_5
- Polka, L., Colantonio, C., and Sundara, M. (2001). A cross-language comparison of /d l-/ perception: evidence for a new developmental pattern. *J. Acoust. Soc. Am.* 109, 2190–2201. doi: 10.1121/1.1362689
- Raizada, R. D., and Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56, 726–740. doi: 10.1016/j.neuron.2007.11.001
- Raizada, R. D. S., Tsao, F. M., Liu, H. M., and Kuhl, P. K. (2010). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: prediction of individual differences. *Cereb. Cortex* 20, 1–12. doi: 10.1093/cercor/bhp076
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- Rogalsky, C., Love, T., Driscoll, D., Anderson, S. W., and Hickok, G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase* 17, 178–187. doi: 10.1080/13554794.2010.509318
- Sehm, B., Schnitzler, T., Obleser, J., Groba, A., Ragert, P., Villringer, A., et al. (2013). Facilitation of inferior frontal cortex by transcranial direct current stimulation induces perceptual learning of severely degraded speech. *J. Neurosci.* 33, 15868–15878. doi: 10.1523/JNEUROSCI.5466-12.2013
- Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., et al. (2012). Structural plasticity in the language system related to increased second language proficiency. *Cortex* 48, 458–465. doi: 10.1016/j.cortex.2010.10.007
- Swan, K. S., and Myers, E. B. (2013). Category labels induce boundary-dependent perceptual warping in learned speech categories. *Second Lang. Res.* 29, 391–411. doi: 10.1177/0267658313491763
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., and Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* 21, 1934–1945. doi: 10.1162/jocn.2009.21131
- Turkeltaub, P. E., and Branch Coslett, H. (2010). Localization of sub-lexical speech perception components. *Brain Lang.* 114, 1–15. doi: 10.1016/j.bandl.2010.03.008
- Ventura-Campos, N., Sanjuán, A., González, J., Palomar-García, M.-Á., Rodríguez-Pujadas, A., Sebastián-Gallés, N., et al. (2013). Spontaneous brain activity predicts learning ability of foreign sounds. *J. Neurosci.* 33, 9295–9305. doi: 10.1523/JNEUROSCI.4655-12.2013
- Vlahou, E. L., Protopapas, A., and Seitz, A. R. (2012). Implicit training of nonnative speech stimuli. *J. Exp. Psychol. Gen.* 141, 363–381. doi: 10.1037/a0025014
- Wade, T., and Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *J. Acoust. Soc. Am.* 118, 2618–2633. doi: 10.1121/1.2011156
- Werker, J. F., and Tees, R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annu. Rev. Psychol.* 50, 509–535. doi: 10.1146/annurev.psych.50.1.509
- Winkowski, D. E., Bandyopadhyay, S., Shamma, S. A., and Kanold, P. O. (2013). Frontal cortex activation causes rapid plasticity of auditory cortical processing. *J. Neurosci.* 33, 18134–18148. doi: 10.1523/JNEUROSCI.0180-13.2013

- Wong, P. C. M., Perrachione, T. K., and Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Hum. Brain Mapp.* 28, 995–1006. doi: 10.1002/hbm.20330
- Yeung, H. H., and Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition* 113, 234–243. doi: 10.1016/j.cognition.2009.08.010
- Zatorre, R. J., Fields, R. D., and Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat. Neurosci.* 15, 528–536. doi: 10.1038/nn.3045
- Zevin, J. D., Datta, H., Maurer, U., Rosania, K. A., and McCandliss, B. D. (2010). Native language experience influences the topography of the mismatch negativity to speech. *Front. Hum. Neurosci.* 4:212. doi: 10.3389/fnhum.2010.00212
- Zhou, X., and Merzenich, M. M. (2007). Intensive training in adults refines A1 representations degraded in an early postnatal critical period. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15935–15940. doi: 10.1073/pnas.0707348104

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 March 2014; accepted: 20 July 2014; published online: 08 August 2014.

Citation: Myers EB (2014) Emergence of category-level sensitivities in non-native speech sound learning. *Front. Neurosci.* 8:238. doi: 10.3389/fnins.2014.00238

This article was submitted to *Auditory Cognitive Neuroscience*, a section of the journal *Frontiers in Neuroscience*.

Copyright © 2014 Myers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.