

Tilburg University

## How learning to abstract shapes neural sound representations

Ley, A.; Vroomen, J.; Formisano, E.

*Published in:*  
Frontiers in Neuroscience

*DOI:*  
[10.3389/fnins.2014.00132](https://doi.org/10.3389/fnins.2014.00132)

*Publication date:*  
2014

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
Ley, A., Vroomen, J., & Formisano, E. (2014). How learning to abstract shapes neural sound representations. *Frontiers in Neuroscience*, 8, [132]. <https://doi.org/10.3389/fnins.2014.00132>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# How learning to abstract shapes neural sound representations

Anke Ley<sup>1,2</sup>, Jean Vroomen<sup>1</sup> and Elia Formisano<sup>2\*</sup>

<sup>1</sup> Department of Medical Psychology and Neuropsychology, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

<sup>2</sup> Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands

## Edited by:

Einat Liebenthal, Medical College of Wisconsin, USA

## Reviewed by:

Rajeev D. S. Raizada, Cornell University, USA

Andre Brechmann, Leibniz Institute for Neurobiologie, Germany

## \*Correspondence:

Elia Formisano, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, Netherlands  
e-mail: e.formisano@maastrichtuniversity.nl

The transformation of acoustic signals into abstract perceptual representations is the essence of the efficient and goal-directed neural processing of sounds in complex natural environments. While the human and animal auditory system is perfectly equipped to process the spectrotemporal sound features, adequate sound identification and categorization require neural sound representations that are invariant to irrelevant stimulus parameters. Crucially, what is relevant and irrelevant is not necessarily intrinsic to the physical stimulus structure but needs to be learned over time, often through integration of information from other senses. This review discusses the main principles underlying categorical sound perception with a special focus on the role of learning and neural plasticity. We examine the role of different neural structures along the auditory processing pathway in the formation of abstract sound representations with respect to hierarchical as well as dynamic and distributed processing models. Whereas most fMRI studies on categorical sound processing employed speech sounds, the emphasis of the current review lies on the contribution of empirical studies using natural or artificial sounds that enable separating acoustic and perceptual processing levels and avoid interference with existing category representations. Finally, we discuss the opportunities of modern analyses techniques such as multivariate pattern analysis (MVPA) in studying categorical sound representations. With their increased sensitivity to distributed activation changes—even in absence of changes in overall signal level—these analyses techniques provide a promising tool to reveal the neural underpinnings of perceptually invariant sound representations.

**Keywords:** auditory perception, perceptual categorization, learning, plasticity, MVPA

## SOUND PERCEPTION—MORE THAN TIME-FREQUENCY ANALYSIS

Despite major advances in the past years to unravel the functional organization principles of the auditory system, the neural processes underlying sound perception are still far from being understood. Complementary research in animals and humans has revealed the properties of responses of neurons and neuronal populations along the auditory pathway from the cochlear nucleus to the cortex. Current knowledge on the neural representation of the spectrotemporal features of the incoming sound is such that the sound spectrogram can be accurately reconstructed from neuronal population responses (Pasley et al., 2012). Yet, the precise neural representation of the acoustic sound features alone cannot explain sound perception fully. In fact, how a sound is perceived may be invariant to changes of its acoustic properties. Unless the context in which a sound is repeated is absolutely identical to the first encounter—which is rather unlikely under natural circumstances—recognizing a sound is not trivial, given that the acoustic properties of the two repetitions may not entirely match. Obviously, this poses an extreme challenge to the auditory system. To maintain processing efficiency, acoustically different sounds must be mapped onto the same perceptual representation. Thus, an essential part of sound processing is the reduction

or perceptual categorization of the vast diversity of spectrotemporal events into meaningful (i.e., behaviorally relevant) units. However, despite the ease with which humans generally accomplish this task, the detection of relevant and invariant information in the complexity of the sensory input is not straightforward. This is also reflected in the performance of artificial voice and speech recognition systems for human-computer interaction, that is far below that of humans, which is mainly due to the difficulty of dealing with the naturally occurring variability in speech signals (Benzeguiba et al., 2007). In humans, the need for perceptual abstraction in everyday functioning manifests itself in pathological conditions such as the autism spectrum disorder (ASD). Next to their susceptibility to more general cognitive deficits in abstract reasoning and concept formation (Minschew et al., 2002), individuals with ASD tend to show enhanced processing of detailed acoustic information while processing of more complex and socially relevant sounds such as speech may be diminished (reviewed in Ouimet et al., 2012).

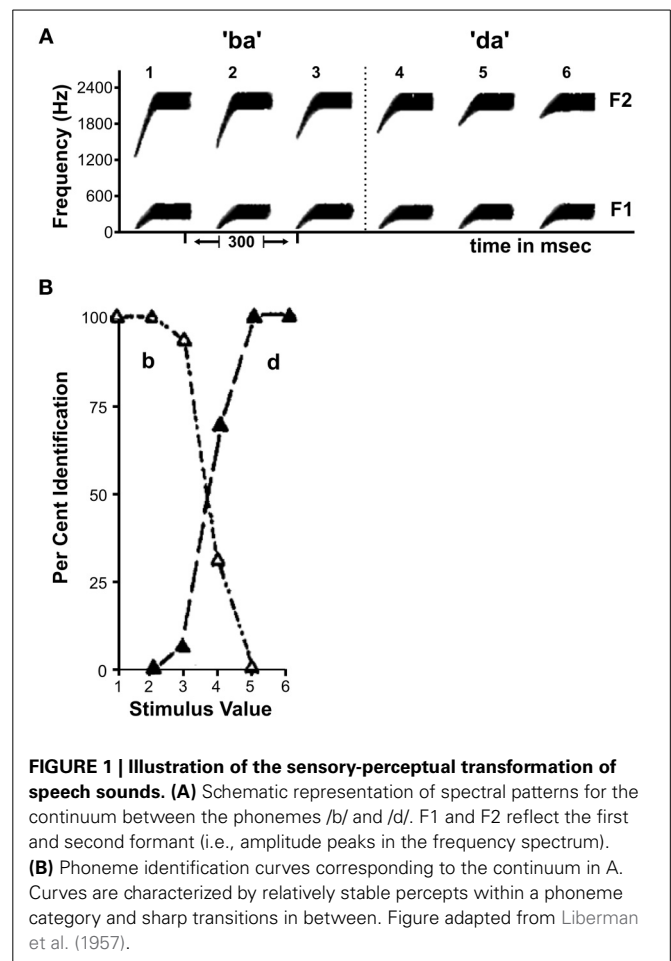
Speech sounds have been widely investigated in the context of sensory-perceptual transformation as they represent a prominent example of perceptual sound categories that comprise a large number of acoustically different sounds. Interestingly, there is not a clear boundary between two phoneme categories such as

/b/ and /d/: the underlying acoustic features vary smoothly from one category to the next (Figure 1A). Remarkably though, if people are asked to identify individual sounds randomly taken from this spectrotemporal continuum as either /b/ or /d/ their percept does not vary gradually as suggested by the sensory input. Instead, the sounds from the first portion of the continuum are robustly identified as /b/, while the sounds from the second part are perceived as /d/ with an abrupt perceptual switch in between (Figure 1B). Performance on discrimination tests further suggests that people are fairly insensitive to the underlying variation of the stimuli within one phoneme category, mapping various physically different stimuli onto the same perceptual object (Liberman et al., 1957). At the category boundary, however, the same extent of physical difference is perceived as a change in stimulus identity. This difference in perceptual discrimination also affects speech production, which strongly relies on online monitoring of auditory feedback. Typically, a self-produced error in the articulation of a speech sound is instantaneously corrected for if, e.g., the output vowel differs from the intended vowel category. An acoustic deviation of the same magnitude and direction may however be tolerated if the produced sound and the intended sound fall within the same perceptual category (Niziolek and Guenther, 2013). This suggests that the within-category differences in the physical domain are perceptually compressed to create a robust representation of the phoneme category while between-category differences are perceptually enhanced to rapidly detect the relevant change of phoneme identity. This phenomenon is termed “Categorical Perception” (CP, Harnad, 1987) and has been demonstrated for stimuli from various natural domains apart from speech, such as music (Burns and Ward, 1978), color (Bornstein et al., 1976; Franklin and Davies, 2004) and facial expressions of emotion (Etcoff and Magee, 1992), not only for humans but also for monkeys (Freedman et al., 2001, 2003), chinchillas (Kuhl and Miller, 1975), songbirds (Prather et al., 2009), and even crickets (Wytenbach et al., 1996). Thus, the formation of discrete perceptual categories from a continuous physical signal seems to be a universal reduction mechanism to deal with the complexity of natural environments.

Several recent reviews have discussed the neural representation of sound categories in auditory cortex (AC) and the role of learning-induced plasticity (e.g., Nourski and Brugge, 2011; Spierer et al., 2011). The emphasis of the current review lies on recent empirical studies using natural or artificial sounds and experimental paradigms that enable separating acoustic and perceptual processing levels and avoid interference with existing category representations (such as for speech). Additionally, we discuss the opportunities of modern analyses techniques such as multivariate pattern analysis (MVPA) in studying categorical sound representations.

### THE ROLE OF EXPERIENCE IN THE FORMATION OF PERCEPTUAL CATEGORIES

While CP has been demonstrated many times for a large variety of stimuli, the mechanisms underlying this phenomenon remain debated. Even for speech, which has most widely been investigated, the relative contribution of innate processes and learning in the formation of phoneme categories is not completely



resolved. Despite the striking consistency of perceptual phoneme boundaries across different listeners, behavioral evidence suggests that those boundaries are malleable depending on the context in which the sounds are perceived (Benders et al., 2010). Additionally, cross-cultural studies have shown that language learning influences the discriminability of speech sounds, such that phonemes in one particular language are only perceived categorically by speakers of that language and continuously otherwise (Kuhl et al., 1992). Similarly, lifelong (e.g., musical training) as well as short-term experience both affect behavioral processing—and neural encoding (see below)—of relevant speech cues, such as pitch, timber and timing (Kraus et al., 2009). In support of the claim that speech CP can be acquired through training stand experimental learning studies that successfully induced discontinuous perception of a non-native phoneme continuum through elaborate category training (Myers and Swan, 2012). Nevertheless, even after extensive training, non-native phoneme contrasts tend to remain less robust than speech categories in the native language. Apart from the age of acquisition, the complexity of the learning environment and in particular the offered stimulus variability during category learning seems to affect the ability to discriminate novel phonetic contrasts (Logan et al., 1991). A prevalent theory for the formation of speech categories in particular is the motor theory of speech perception (Liberman and

Mattingly, 1985). This theory claims that speech sounds are categorized based on the distinct motor commands for the vocal tract used for pronunciation. Further fueled by the discovery of mirror neurons, the theory still has its proponents (for review see Galantucci et al., 2006), however, today, it is disputed in its strict form in which speech processing is considered special, as the recruitment of the motor system for sound identification has been demonstrated for various forms of non-speech action-related sounds (Kohler et al., 2002). Furthermore, accumulating evidence indicates that CP can be induced by learning for a variety of non-speech stimulus material (e.g., simple noise sounds, Guenther et al., 1999 and inharmonic tone complexes, Goudbeek et al., 2009). The use of artificially constructed categories for studying CP has the advantage that the physical distance between neighboring stimuli can be controlled such that the similarity ratings of within- or between-category stimuli can be attributed to true perceptual effects, rather than the metrics of the stimulus dimensions. Nevertheless, one should bear in mind that the long-term exposure to statistical regularities of the acoustics of natural sounds might exert a lasting influence on the formation of new sound categories. In support of this claim, Scharinger et al. (2013b) revealed a strong preference for negatively correlated spectral dimensions typical for speech and other natural categories when participants learned to categorize novel auditory stimuli. In line with this behavioral documentation in humans, a recent study in rodent pups demonstrated the proneness of auditory receptive fields to the systematics of the acoustic environment shaping the tuning curves of cortical neurons. Most importantly, these neuronal changes were shown to parallel an increase in perceptual discrimination of the employed sounds, which points to a link between (early) neuronal plasticity and perceptual discrimination ability (Köver et al., 2013). In sum, these experiments demonstrated that the perceptual abilities could be modified by learning and experience, while the role of pre-existing (i.e., innate) neural structures and their early adaptation in critical phases of maturation might play a vital role.

## NEURAL REPRESENTATIONS OF PERCEPTUAL SOUND CATEGORIES

Behavioral studies have been complemented with research on the neural implementation of perceptual sound categories. Forming new sound categories or assigning a new stimulus to an existing category requires the integration of bottom-up stimulus driven information with knowledge from prior experience and memory as well as linking this information to the appropriate response in case of an active categorization task. Different research lines have highlighted the contribution of neural structures along the auditory pathway and in the cortex to this complex and dynamic process.

Functional neuroimaging studies employing natural sound categories such as voices, speech, and music have located object-specific processing units in higher level auditory areas in the superior temporal lobe (Belin et al., 2000; Leaver and Rauschecker, 2010). Particularly, native phoneme categories were shown to recruit the left superior temporal sulcus (STS) (Liebenthal et al., 2005) and the activation level of this region seems to correlate with the degree of categorical processing (Desai et al., 2008).

While categorical processes in the STS were documented by further studies, the generalization to other sound categories beyond speech remains controversial, given that the employed stimuli were either speech sounds or artificial sounds with speech-like characteristics (Leech et al., 2009; Liebenthal et al., 2010). Even if speech sounds are natural examples of the discrepancy between sensory and perceptual space, the results derived from these studies may not generalize to other categories, as humans are processing experts for speech (similar to faces) even prior to linguistic experience (Eimas et al., 1987). In addition, regions in the temporal lobe were shown to retain the sensitivity to acoustic variability within sound categories, while highly abstract phoneme representations (i.e., invariant to changes within one phonetic category) appear to depend on decision-related processes in the frontal lobe (Myers et al., 2009). These results are highly compatible with those from cell recordings in rhesus monkey (Tsunada et al., 2011). Based on the analysis of single-cell responses to human speech categories, the authors suggest that “a hierarchical relationship exists between the superior temporal gyrus (STG) and the ventral PFC whereby STG provides the ‘sensory evidence’ to form the decision and ventral PFC activity encodes the output of the decision process.” Analog to the two-stage hierarchical processing model in the visual domain (Freedman et al., 2003; Jiang et al., 2007; Li et al., 2009), the set of findings reviewed above suggests that processing areas in the temporal lobe only constitute a preparatory stage for categorization. Specifically, the model proposes that the tuning of neuronal populations in lower-level sensory areas is sharpened according to the category-relevant stimulus features, forming a task-independent reduction of the sensory input (but see below for a different view on the role of early auditory areas). In case of an active categorization task, this information is projected to higher-order cortical areas in the frontal lobe. The predominant recruitment of the prefrontal cortex (PFC) during early phases of category learning (Little and Thulborn, 2005) and in the context of an active categorization task (Boettiger and D’Esposito, 2005; Husain et al., 2006; Li et al., 2009) support the concept that it plays a major role in rule learning and attention-related processes modulating lower-level sound processing rather than being the site of categorical sound representations *per se*.

Categorical processing does however not exclusively proceed along the auditory “what” stream. To study the neural basis of CP, Raizada and Poldrack (2007) measured fMRI while subjects listened to pairs of stimuli taken from a phonetic /ba/-/da/ continuum. Responses in the supramarginal gyrus were significantly larger for pairs that included stimuli belonging to different phonetic categories (i.e., crossing the category boundary) than for pairs with stimuli from a single category. The authors interpreted these results as evidence for “neural amplification” of relevant stimulus difference and thus for categorical processing in the supramarginal gyrus. Similar analyses showed comparatively little amplification of changes that crossed category boundaries in low-level auditory cortical areas (Raizada and Poldrack, 2007). Novel findings revived the motor theory of categorical processing: Chevillet et al. (2013) provide evidence that the role of the premotor cortex (PMC) is not limited to motor-related processes during active categorization, but that the phoneme-category tuning of



premotor regions may essentially facilitate also more automatic speech processes via dorsal projections originating from pSTS. While this automatic motor route is probably limited to processing of speech and other action-related sound categories, the diversity of the categorical processing networks documented in the above cited studies demonstrates that there is not a single answer to where and how sound categories are represented. The role that early auditory cortical fields play in the perceptual abstraction from the acoustic input remains a relevant topic of current research. A recent study from Nelken's group indicated that neurons in the cat primary auditory area convey more information about abstract auditory entities than about the spectro-temporal sound structure (Chechik and Nelken, 2012). These results are in line with the proposal that neuronal populations in primary AC encode perceptual abstractions of sounds (or *auditory objects*, Griffiths and Warren, 2004) rather than their physical make up (Nelken, 2004). Furthermore, research from Scheich's group has suggested that sound representations in primary AC are largely context- and task- dependent and reflect memory-related and semantic aspects of actively listening to sounds (Scheich et al., 2007). This suggestion is also supported by the observation of semantic/categorical effects within early (~70 ms) post-stimulus time windows in human auditory evoked potentials (Murray et al., 2006).

Finding empirical evidence for abstract categorical representations in low-level auditory cortex in humans, however, remains challenging as it requires experimental paradigms and analysis methods that allow disentangling the perceptual processes from the strong dependence of these auditory neurons on the physical sound attributes. Here, carefully controlled stimulation paradigms in combination with fMRI pattern decoding (see below) could shed light on the matter. For example, Staeren et al. (2009) were able to dissociate perceptual from stimulus-driven processes by controlling the physical overlap of stimuli within and between natural sound categories. They revealed categorical sound representations in spatially distributed and even overlapping activation patterns in early areas of human AC. Similarly, studies employing fMRI-decoding to investigate the auditory cortical processing of speech/voice categories have put forward a "constructive" role of early auditory cortical networks in the formation of perceptual sound representations (Formisano et al., 2008; Kilian-Hütten et al., 2011a; Bonte et al., 2014).

Crucially, studying context-dependence and plasticity of sound representations in early auditory areas may help unraveling their nature. For example, Dehaene-Lambertz et al. (2005) demonstrated that even early low-level sound processing is susceptible to top-down directed cognitive influences. In a combination of fMRI and electrophysiological measures, they showed that identical acoustic stimuli were processed in a different fashion, depending on the "perceptual mode" (i.e., whether participants perceived the sounds as speech or artificial whistles).

This literature review illustrates that in order to understand the neural mechanisms underlying the formation of perceptual categories, it is necessary to (1) carefully separate perceptual from acoustical sound representations, (2) distinguish between lower-level perceptual representations and higher-order or feedback-guided decision- and task-related processes and also (3) avoid

interference with existing processing networks for familiar and overlearned sound categories.

## LEARNING AND PLASTICITY

Most knowledge about categorical processing in the brain is derived from experiments employing speech or other natural (e.g., music) sound categories. While providing important insights about the neural representations of familiar sound categories, these studies lack the potential to investigate the mechanisms underlying the transformation from acoustic to more abstract perceptual representations. Sound processing must however remain highly plastic beyond sensitive periods early in ontogenesis to allow efficient processing adapted to the changing requirements of the acoustic environment.

Studying these rapid experience-related neural reorganizations requires controlled learning paradigms of new sound categories. With novel, artificial sounds, the acoustic properties can be controlled, such that physical and perceptual representations can be decoupled and interference with existing representations of familiar sound categories can be avoided (but see Scharinger et al., 2013b). A comparison of pre- and post-learning neural responses provides information about the amenability of sound representations along different levels of the auditory processing hierarchy to learning-induced plasticity. Extensive research by Fritz and colleagues has provided convincing evidence for learning-induced plasticity of cortical receptive fields. In ferrets that were trained on a target (tone) detection task, a large proportion of cells in primary AC showed significant changes in spectro-temporal receptive field (STRF) shape during the detection task, as compared with the passive pre-behavioral STRF. Relevant to the focus of this review, in two-thirds of these cells the changes persisted in the post-behavior passive state (Fritz et al., 2003, see also Shamma and Fritz, 2014). Additionally, recent results from animal models and human studies have revealed evidence for similar cellular and behavioral mechanisms for learning and memory in the auditory brainstem (e.g., Tzounopoulos and Kraus, 2009).

Learning studies further provide the opportunity to look into the interaction of lower-level sensory and higher-level association cortex during task- and decision-related processes (De Souza et al., 2013). In contrast to juvenile plasticity, which is mainly driven by bottom-up input, adult learning is supposedly largely dependent on top-down control (Kral, 2013). Thus, categorical processing after short-term plasticity induced by temporary changes of environmental demands might differ from the processes formed by early-onset and long-term adaptation to speech stimuli. Even though there is evidence that with increasing proficiency in category discrimination, neural processing of newly learned speech sounds starts to parallel that of native speech (Golestani and Zatorre, 2004), a discrepancy between ventral and dorsal processing networks for highly familiar native sound categories and non-native or artificial sound categories respectively has been suggested by recent work (Callan et al., 2004; Liebenthal et al., 2010, 2013). This difference potentially limits the generalization to native speech of findings derived from studies employing artificial sound categories.

Several studies have examined the changes in the neural sound representations underlying the perceptual transformations

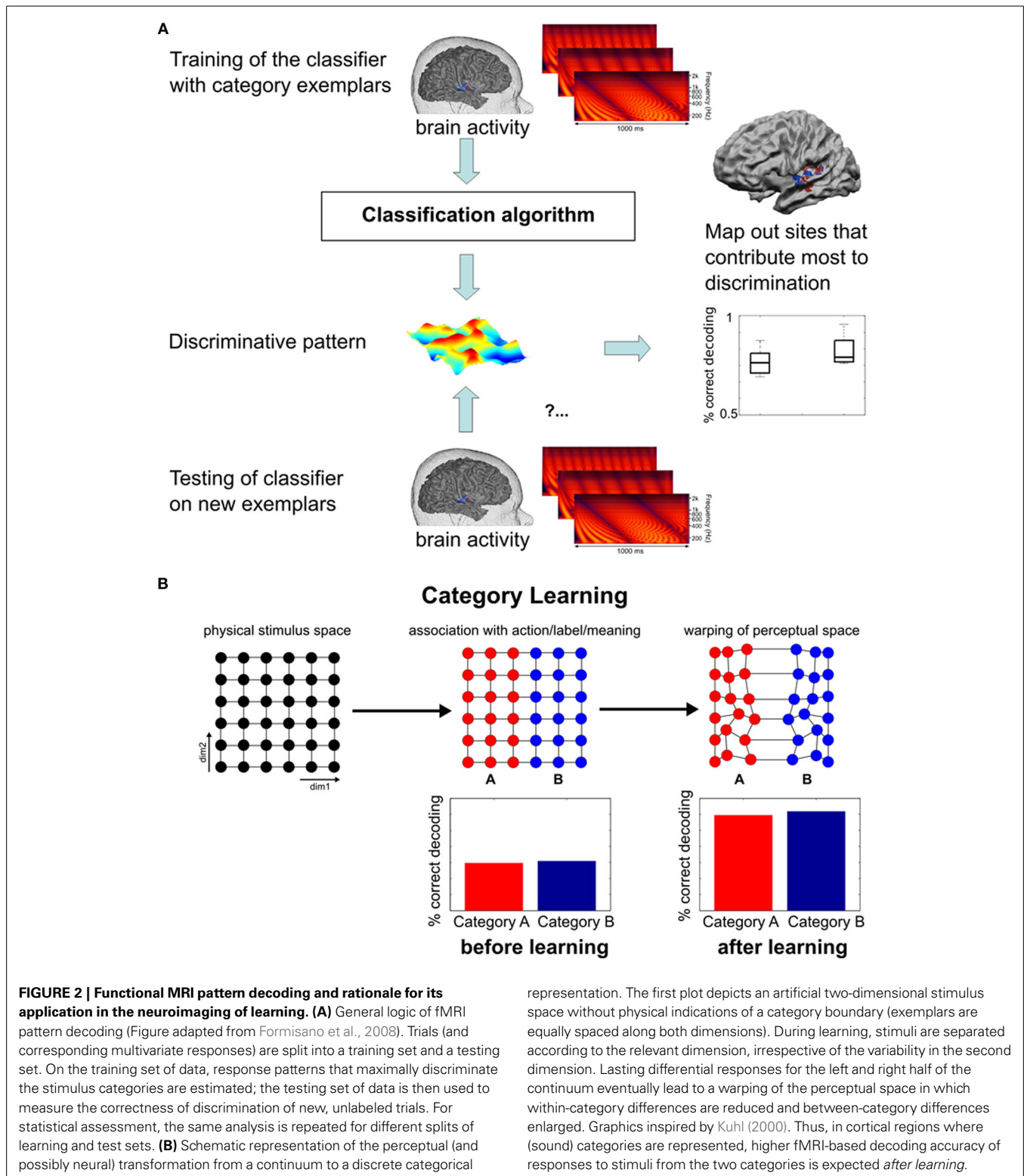
induced by category learning. A seminal study with gerbils demonstrated that learning to categorize artificial sounds in the form of frequency sweeps resulted in a transition from a physical (i.e., onset frequency) to a categorical (i.e., up vs. down) sound representation already in the primary AC (Ohl et al., 2001). In contrast to the traditional understanding of primary AC as a feature detector, this finding implicates that sound representations at the first cortical analysis stage are more abstract and prone to plastic reorganization imposed by changes in environmental demands. In fact, sound stimuli have passed through several levels of basic feature analyses before they ascend to the superior temporal cortex (Nelken, 2004). Thus, as discussed above, sound representations in primary AC are unlikely to be faithful copies of the physical characteristics. Even though the involvement of AC in categorization of artificial sounds has also been demonstrated in humans (Guenther et al., 2004), conventional subtraction paradigms typically employed in fMRI studies lack sufficient sensitivity to demarcate distinct categorical representations. Due to the large physical variability within categories and the similarity of sounds straddling the category boundary, between-category contrasts often do not reveal significant results (Klein and Zatorre, 2011). Furthermore, the effects of category learning on sound processing as demonstrated in animals were based on changes in the spatiotemporal activation pattern without apparent changes in response strength (Ohl et al., 2001; Engineer et al., 2014). Using *in vivo* two-photon calcium imaging in mice, Bathellier et al. (2012) have convincingly shown that categorical sound representations—which can be selected for behavioral or perceptual decisions—may emerge as a consequence of non-linear dynamics in local networks in the auditory cortex (Bathellier et al., 2012, see also Tsunada et al., 2012 and a recent review by Mizrahi et al., 2014).

In human neuroimaging, these neuronal effects that do not manifest as changes in overall response levels may remain inscrutable to univariate contrast analyses. Also, fMRI designs based on adaptation, or more generally, on measuring responses to stimulus pairs/sequences (e.g., as in Raizada and Poldrack, 2007) do not allow excluding generic effects related to the processing of sound sequences or potential hemodynamic confounds, as the reflection of neuronal adaptation/suppression effects in the fMRI signals is complex (Boynton and Finney, 2003; Verhoef et al., 2008).

Modern analyses techniques with increased sensitivity to spatially distributed activation changes in absence of changes in overall signal level provide a promising tool to decode perceptually invariant sound representations in humans (Formisano et al., 2008; Kilian-Hütten et al., 2011a) and detect the neural effects of learning (Figure 2). Multivariate pattern analysis (MVPA) employs established classification techniques from machine learning to discriminate between different cognitive states that are represented in the combined activity of multiple locally distributed voxels, even when their average activity does not differ between conditions (see Haynes and Rees, 2006; Norman et al., 2006; Haxby, 2012 for tutorial reviews). Recently, Ley et al. (2012) demonstrated the potential of this method to trace rapid transformations of neural sound representations, which are entirely based on changes in the way the sounds are

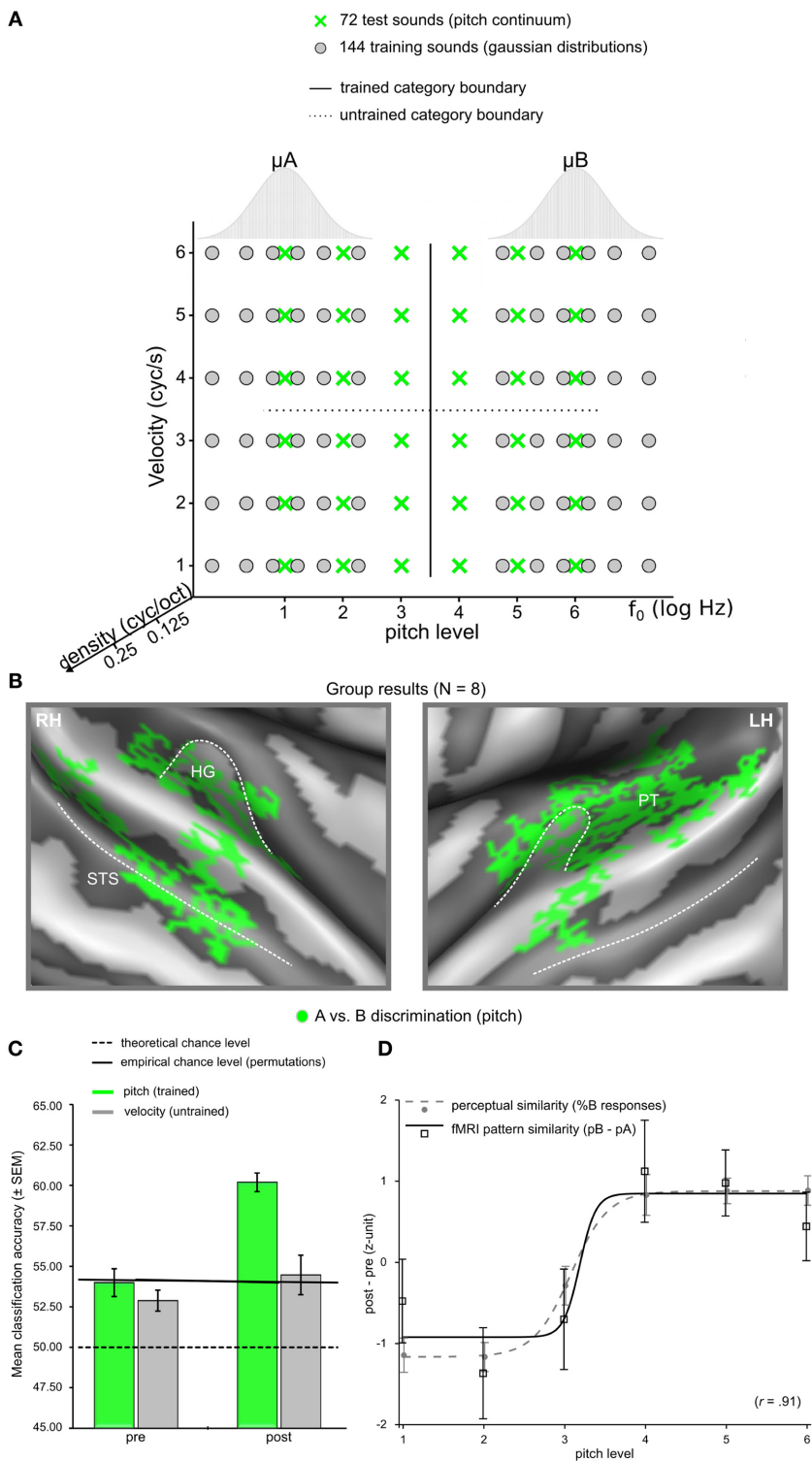
perceived induced by a few days of category learning (Figure 3). In their study, participants were trained to categorize complex artificial ripple sounds, differing along several acoustic dimensions into two distinct groups. BOLD activity was measured before and after training during passive exposure to an acoustic continuum spanned between the trained categories. This design ensured that the acoustic stimulus dimensions were uninformative of the trained sound categorization such that any change in the activation pattern could be attributed to a warping of the perceptual space rather than physical distance. After successful learning, locally distributed response patterns in Heschl's gyrus (HG) and its adjacency became selective for the trained category discrimination (pitch) while the same sounds elicited indistinguishable responses before. In line with recent findings in rat primary AC (Engineer et al., 2013), the similarity of the cortical activation patterns reflected the sigmoid categorical structure and correlated with perceptual rather than physical sound similarity. Thus, complementary research in animals and humans indicate that perceptual sound categories are represented in the activation patterns of distributed neuronal populations in early auditory regions, further supporting the role of the early AC in abstract and experience-driven sound processing rather than acoustic feature mapping (Nelken, 2004). It is noteworthy that these abstract categorical representations were detectable despite passive listening conditions. This is an important detail, as it demonstrates that categorical representations are (at least partially) independent of higher-order decision or motor-related processes. Furthermore, it suggests that some preparatory (i.e., multipurpose) abstraction of the physical input happens at the level of the early auditory cortex.

The mechanisms of neuroplasticity underlying category learning and the origin of the categorical organization of sound representations in the auditory cortex are still quite poorly understood and deserve further investigation. Hypotheses are primarily derived from perceptual learning studies in animals. These studies show that extensive discrimination training may elicit reorganization of the auditory cortical maps, selectively increasing the representation of the behaviorally relevant sound features (Recanzone et al., 1993; Polley et al., 2006). This suggests that environmental and behavioral demands lead to changes of the auditory tuning properties of neurons such that more neurons are tuned to the relevant features to achieve higher sensitivity in the relevant dimension. This reorganization is mediated by synaptic plasticity, i.e., the strengthening of neuronal connections following rules of Hebbian learning (Hebb, 1949; for recent review, see Caporale and Dan, 2008). Passive learning studies suggest that attention is not necessary for sensory plasticity to occur (Watanabe et al., 2001; Seitz and Watanabe, 2003). However, in contrast to the mostly unequivocal sound structure used for perceptual learning experiments, learning to categorize a large number of sounds differing along multiple dimensions requires either sound distributions indicative of the category structure (Goudbeek et al., 2009) or a task including response feedback in order to extract the relevant and category discriminative sound feature. This selective enhancement of features requires some top-down gating mechanism. Attention can act as such a filter, increasing feature saliency (Lakatos et al., 2013) by selectively modulating the tuning properties of neurons in the auditory cortex, eventually



leading to a competitive advantage of behaviorally relevant information (Bonte et al., 2009, 2014; Ahveninen et al., 2011). As a consequence, more neural resources would be allocated to the behaviorally relevant information at the expense of information

that is irrelevant for the decision. The adaptive allocation of neural resources to diagnostic information after category learning is supported by evidence from monkey electrophysiology (Sigala and Logothetis, 2002; De Baene et al., 2008) and human imaging,



**FIGURE 3 | Representation of the study by Ley et al. (2012). (A)** Multidimensional stimulus space spanning the two categories A and B. **(B)** Group discrimination maps based on the post-learning fMRI data for the trained stimulus division (i.e., “low pitch” vs. “high pitch”), displayed on an average reconstructed cortical surface after cortex-based realignment. **(C)** Average classification accuracies based on fMRI data prior to category training and after successful category learning for the two types of stimulus

space divisions (trained vs. untrained) and the respective trial labeling. **(D)** Changes in pattern similarity and behavioral identification curves. After category learning, neural response patterns for sounds with higher pitch (pitch levels 4, 5, 6) correlated with the prototypical response pattern for class B more strongly than class A, independent of other acoustic features. The profile of these correlations on the pitch continuum closely reflected the sigmoid shape of the behavioral category identification function.



showing decreased activation for prototypical exemplars of a category relative to exemplars near the category boundary (Guenther et al., 2004). This idea of categorical sound representations being sparse or parsimonious is also compatible with fMRI observations by Brechmann and Scheich (2005), showing an inverse correlation of auditory cortex activation and performance in an auditory categorization task. The recent discovery of a positive correlation between gray matter probability in parietal cortex and the optimal utilization of acoustic features in a categorization task (Scharinger et al., 2013a) provides further evidence for the crucial role of attentional processes in feature selection necessary for category learning. Reducing the representation of a large number of sounds to few relevant features presents an enormous processing advantage. It facilitates the read-out of the categorical pattern due to the pruned data structure and limits the neural resources by avoiding redundancies in the representation according to the concept of sparse coding (Olshausen and Field, 2004).

To date, there are several models for describing the neural circuitry between sensory and higher-order attentional processes mediating learning-induced plasticity. Predictive coding models propose that the dynamic interaction between bottom-up sensory information and top-down modulation by prior experience shapes the perceptual sound representation (Friston, 2005). This implies that categorical perception would arise from the continuous updating of the internal representation during learning to incorporate all variability present within a category, with the objective of reducing the prediction error (i.e., the difference between sensory input and internal representation). Consequently, lasting interaction between forward driven processing and backward modulation could induce synaptic plasticity and result in an internal representation that correctly matches the categorical structure and therefore optimally guides correct behavior also beyond the scope of the training period. The implementation of these Bayesian processing models rests on fairly hierarchical structures consisting of forward, backward and lateral connections entering different cortical layers (Felleman and Van Essen, 1991; Hackett, 2011). According to the Reverse Hierarchy Theory (Ahissar and Hochstein, 2004), category learning would be initiated by high-level processes involved in rule-learning, controlling via top-down modulation selective plasticity at lower-level sensory areas sharpening the responses according to the learning rule (Sussman et al., 2002; Myers and Swan, 2012). In accordance with this view, attentional modulation involving a fronto-parietal network of brain areas appears most prominent during early phases of learning, progressively decreasing with expertise (Little and Thulborn, 2005; De Souza et al., 2013). Despite recent evidence for early sensory-perceptual abstraction mechanisms in human auditory cortex (Murray et al., 2006; Bidelman et al., 2013), it is crucial to note that the reciprocal information exchange between higher-level and lower-level cortical fields happens very fast (Kral, 2013) and even within the auditory cortex, processing is characterized by complex forward, lateral and backward microcircuits (Atencio and Schreiner, 2010; Schreiner and Polley, 2014). Therefore, the origin of the categorical responses in AC is difficult to determine unless the response latencies and laminar structure are carefully investigated.

## CROSSMODAL PLASTICITY—CONSIDERATIONS FOR FUTURE STUDIES

Considering that sound perception strongly relies on the integration of information represented across multiple cortical areas, simultaneous input from the other sensory modalities presents itself as a major source of influence on learning-induced plasticity of sound representations. In fact, there is compelling behavioral evidence that the human perceptual system integrates specific, event-relevant information across auditory and visual (McGurk and MacDonald, 1976) or auditory and tactile (Gick and Derrick, 2009) modalities and that mechanisms of multisensory integration can be shaped through experience (Wallace and Stein, 2007). Together, these two facts predict that visual or tactile contexts during learning have a major impact on perceptual reorganization of sound representations.

Promising insights are provided by behavioral studies showing that multimodal training designs are generally superior to unimodal training designs (Shams and Seitz, 2008). The beneficial effect of multisensory exposure during training may last beyond the training period itself reflected in increased performance after removal of the stimulus from one modality (for review, see Shams et al., 2011). This effect has been demonstrated even for brief training periods and arbitrary stimulus pairs (Ernst, 2007), promoting the view that short-term multisensory learning can lead to lasting reorganization of the processing networks (Kilian-Hütten et al., 2011a,b). Given the considerable evidence for response modulation of auditory neurons by simultaneous non-acoustic events and even crossmodal activation of the auditory cortex in absence of sound stimuli (Calvert et al., 1997; Foxe et al., 2002; Fu et al., 2003; Brosch et al., 2005; Kayser et al., 2005; Pekkola et al., 2005; Schürmann et al., 2006; Nordmark et al., 2012), it is likely that sound representations at the level of AC are also prone to influences from the visual or tactile modality. Animal electrophysiology has suggested different laminar profiles for tactile and visual pathways in the auditory cortex indicative for forward and backward directed input respectively (Schroeder and Foxe, 2002). Crucially, the quasi-laminar resolution achievable with state-of-art ultra-high field fMRI (Polimeni et al., 2010) provides new possibility to systematically investigate—in humans—the detailed neurophysiological basis underlying the influence of non-auditory input on sound perception and on learning induced plasticity in sound representations in the auditory cortex.

## CONCLUSION

In recent years, the phenomenon of perceptual categorization has stimulated a tremendous amount of research on the neural representation of perceptual sound categories in animals and humans. Despite this large data pool, no clear answer could yet be found on where abstract sound categories are represented in the brain. Whereas animal research provides increasing evidence for complex processing abilities of early auditory areas, results from human studies tend to promote more hierarchical processing models in which categorical perception relies on higher order temporal and frontal regions. In this review, we discussed this apparent discrepancy and illustrated the potential pitfalls attached to research on categorical sound processing. Separating perceptual and acoustical processes possibly represents

the biggest challenge. In this respect, it is crucial to note that many “perceptual” effects, demonstrated in animal studies, did not manifest as changes in overall signal level. Recent research has shown that while these effects may remain inscrutable to univariate contrast analyses typically employed in human neuroimaging, modern analysis techniques—such as fMRI-decoding—is capable of unraveling perceptual processes in locally distributed activation patterns. It is also becoming increasingly evident that in order to grasp the full capacity of auditory processing in low-level auditory areas, it is necessary to consider its susceptibility to context and task, flexibly adapting its processing resources according to the environmental demands. In order to bring the advances from animal and human research closer together, future approaches on categorical sound representations in humans are likely to require an integrative combination of controlled stimulation designs, sensitive measurement techniques (e.g., high field fMRI) and advanced analysis techniques.

## ACKNOWLEDGMENTS

This work was supported by Maastricht University, Tilburg University and the Netherlands Organization for Scientific Research (NWO; VICI grant 453-12-002 to Elia Formisano).

## 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
- Ahveninen, J., Hämäläinen, M., Jääskeläinen, I. P., Ahlfors, S. P., Huang, S., Lin, F.-H., et al. (2011). Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4182–4187. doi: 10.1073/pnas.1016134108
- Atencio, C. A., and Schreiner, C. E. (2010). Laminar diversity of dynamic sound processing in cat primary auditory cortex. *J. Neurophysiol.* 192–205. doi: 10.1152/jn.00624.2009
- Bathellier, B., Ushakova, L., and Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* 76, 435–449. doi: 10.1016/j.neuron.2012.07.008
- 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
- Benders, T., Escudero, P., and Sjerps, M. (2010). The interrelation between acoustic context effects and available response categories in speech sound categorization. *J. Acoust. Soc. Am.* 131, 3079–3087. doi: 10.1121/1.3688512
- Benzeguiba, M., De Mori, R., Derou, O., Dupont, S., Erbes, T., Jouvett, D., et al. (2007). Automatic speech recognition and speech variability: a review. *Speech Commun.* 49, 10–11. doi: 10.1016/j.specom.2007.02.006
- 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. doi: 10.1016/j.neuroimage.2013.04.093
- Boettiger, C. A., and D’Esposito, M. (2005). Frontal networks for learning and executing arbitrary stimulus-response associations. *J. Neurosci.* 25, 2723–2732. doi: 10.1523/JNEUROSCI.3697-04.2005
- Bonte, M., Hausfeld, L., Scharke, W., Valente, G., and Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J. Neurosci.* 34, 4548–4557. doi: 10.1523/JNEUROSCI.4339-13.2014
- Bonte, M., Valente, G., and Formisano, E. (2009). Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *J. Neurosci.* 29, 1699–1706. doi: 10.1523/JNEUROSCI.3694-08.2009
- Bornstein, M. H., Kessen, W., and Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 115–129. doi: 10.1037/0096-1523.2.1.115
- Boynton, G. M., and Finney, E. M. (2003). Orientation-specific adaptation in human visual cortex. *J. Neurosci.* 23, 8781–8787.
- Brechmann, A., and Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cereb. Cortex* 15, 578–587. doi: 10.1093/cercor/bhh159
- Brosch, M., Selezneva, E., and Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J. Neurosci.* 25, 6797–6806. doi: 10.1523/JNEUROSCI.1571-05.2005
- Burns, E. M., and Ward, W. D. (1978). Categorical perception-phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.* 63, 456–468. doi: 10.1121/1.381737
- Callan, D. E., Jones, J. A., Callan, A. M., and Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22, 1182–1194. doi: 10.1016/j.neuroimage.2004.03.006
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science* 276, 593–596. doi: 10.1126/science.276.5312.593
- Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annu. Rev. Neurosci.* 31, 25–46. doi: 10.1146/annurev.neuro.31.060407.125639
- Chechik, G., and Nelken, I. (2012). Auditory abstraction from spectro-temporal features to coding auditory entities. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18968–18973. doi: 10.1073/pnas.1111242109
- 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
- De Baene, W., Ons, B., Wagemans, J., and Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learn. Mem.* 15, 717–727. doi: 10.1101/lm.1040508
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33. doi: 10.1016/j.neuroimage.2004.09.039
- 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
- De Souza, A. C. S., Yehia, H. C., Sato, M., and Callan, D. (2013). Brain activity underlying auditory perceptual learning during short period training: simultaneous fMRI and EEG recording. *BMC Neurosci.* 14:8. doi: 10.1186/1471-2202-14-8
- Eimas, P. D., Miller, J. L., and Jusczyk, P. W. (1987). “On infant speech perception and the acquisition of language,” in *Categorical Perception. The Groundwork of Cognition*, ed S. Harnad (Cambridge, MA: Cambridge University Press), 161–195.
- Engineer, C. T., Perez, C. A., Carraway, R. S., Chang, K. Q., Roland, J. L., and Kilgard, M. P. (2014). Speech training alters tone frequency tuning in rat primary auditory cortex. *Behav. Brain Res.* 258, 166–178. doi: 10.1016/j.bbr.2013.10.021
- Engineer, C. T., Perez, C. A., Carraway, R. S., Chang, K. Q., Roland, J. L., Sloan, A. M., et al. (2013). Similarity of cortical activity patterns predicts generalization behavior. *PLoS ONE* 8:e78607. doi: 10.1371/journal.pone.0078607
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J. Vis.* 7, 1–14. doi: 10.1167/7.5.7
- Etcoff, N. L., and Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition* 44, 227–240. doi: 10.1016/0010-0277(92)90002-Y
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (2008). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* 322, 970–973. doi: 10.1126/science.1164318
- Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J. Neurophysiol.* 88, 540–543. doi: 10.1151/jn.00694.2001
- Franklin, A., and Davies, I. R. L. (2004). New evidence for infant colour categories. *Br. J. Dev. Psychol.* 22, 349–377. doi: 10.1348/0261510041552738
- 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

- Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246.
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141
- Fu, K.-M. G., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., et al. (2003). Auditory cortical neurons respond to somatosensory stimulation. *J. Neurosci.* 23, 7510–7515.
- Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377. doi: 10.3758/BF03193857
- Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502–504. doi: 10.1038/nature08572
- 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
- Goudbeek, M., Swingle, D., and Smits, R. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1913–1933. doi: 10.1037/a0015781
- Griffiths, T. D., and Warren, J. D. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892. doi: 10.1038/nrn1538
- 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
- Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hear. Res.* 271, 133–146. doi: 10.1016/j.heares.2010.01.011
- Harnad, S. (eds.). (1987). *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.
- Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage* 62, 852–855. doi: 10.1016/j.neuroimage.2012.03.016
- Haynes, J.-D., and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534. doi: 10.1038/nrn1931
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: Wiley.
- Husain, F. T., Fromm, S. J., Pursley, R. H., Hosey, L., Braun, A., and Horwitz, B. (2006). Neural bases of categorization of simple speech and nonspeech sounds. *Hum. Brain Mapp.* 27, 636–651. doi: 10.1002/hbm.20207
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., and Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron* 53, 891–903. doi: 10.1016/j.neuron.2007.02.015
- Kayser, C., Petkov, C. I., Augath, M., and Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron* 48, 373–384. doi: 10.1016/j.neuron.2005.09.018
- Kilian-Hütten, N., Valente, G., Vroomen, J., and Formisano, E. (2011a). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J. Neurosci.* 31, 1715–1720. doi: 10.1523/JNEUROSCI.4572-10.2011
- Kilian-Hütten, N., Vroomen, J., and Formisano, E. (2011b). Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *Neuroimage* 57, 1601–1607. doi: 10.1016/j.neuroimage.2011.05.043
- Klein, M. E., and Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia* 49, 878–887. doi: 10.1016/j.neuropsychologia.2011.01.008
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848. doi: 10.1126/science.1070311
- Köver, H., Gill, K., Tseng, Y.-T. L., and Bao, S. (2013). Perceptual and neuronal boundary learned from higher-order stimulus probabilities. *J. Neurosci.* 33, 3699–3705. doi: 10.1523/JNEUROSCI.3166-12.2013
- Kral, A. (2013). Auditory critical periods: a review from system's perspective. *Neuroscience* 247, 117–133. doi: 10.1016/j.neuroscience.2013.05.021
- Kraus, N., Skoe, E., Parbery-Clark, A., and Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann. N.Y. Acad. Sci.* 1169, 543–557. doi: 10.1111/j.1749-6632.2009.04549.x
- Kuhl, P. K. (2000). A new view of language acquisition. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11850–11857. doi: 10.1073/pnas.97.22.11850
- Kuhl, P. K., and Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190, 69–72. doi: 10.1126/science.1166301
- 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
- Lakatos, P., Musacchia, G., O'Connell, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761. doi: 10.1016/j.neuron.2012.11.034
- Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612. doi: 10.1523/JNEUROSCI.0296-10.2010
- 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
- Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., and Formisano, E. (2012). Learning of new sound categories shapes neural response patterns in human auditory cortex. *J. Neurosci.* 32, 13273–13280. doi: 10.1523/JNEUROSCI.0584-12.2012
- Li, S., Mayhew, S. D., and Kourtzi, Z. (2009). Learning shapes the representation of behavioral choice in the human brain. *Neuron* 62, 441–452. doi: 10.1016/j.neuron.2009.03.016
- Lieberman, 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
- Lieberman, 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
- Liebenthal, 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
- Liebenthal, 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
- Liebenthal, 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
- Little, D. M., and Thulborn, K. R. (2005). Correlations of cortical activation and behavior during the application of newly learned categories. *Brain Res. Cogn. Brain Res.* 25, 33–47. doi: 10.1016/j.cogbrainres.2005.04.015
- 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, 874–886. doi: 10.1121/1.1894649
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Minshew, N. J., Meyer, J., and Goldstein, G. (2002). Abstract reasoning in autism: a disassociation between concept formation and concept identification. *Neuropsychology* 16, 327–334. doi: 10.1037/0894-4105.16.3.327
- Mizrahi, A., Shalev, A., and Nelken, I. (2014). Single neuron and population coding of natural sounds in auditory cortex. *Curr. Opin. Neurobiol.* 24, 103–110. doi: 10.1016/j.conb.2013.09.007
- Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., and Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *J. Neurosci.* 26, 1293–1302. doi: 10.1523/JNEUROSCI.4511-05.2006
- 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
- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* 14, 474–480. doi: 10.1016/j.conb.2004.06.005
- 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

- Nordmark, P. F., Pruszyński, J. A., and Johansson, R. S. (2012). BOLD responses to tactile stimuli in visual and auditory cortex depend on the frequency content of stimulation. *J. Cogn. Neurosci.* 24, 2120–2134. doi: 10.1162/jocn\_a\_00261
- Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005
- Nourski, K. V., and Brugge, J. F. (2011). Representation of temporal sound features in the human auditory cortex. *Rev. Neurosci.* 22, 187–203. doi: 10.1515/rns.2011.016
- Ohl, F. W., Scheich, H., and Freeman, W. J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. *Nature* 412, 733–736. doi: 10.1038/35089076
- Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007
- Ouimet, T., Foster, N. E. V., Tryfon, A., and Hyde, K. L. (2012). Auditory-musical processing in autism spectrum disorders: a review of behavioral and brain imaging studies. *Ann. N.Y. Acad. Sci.* 1252, 325–331. doi: 10.1111/j.1749-6632.2012.06453.x
- Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol.* 10:e1001251. doi: 10.1371/journal.pbio.1001251
- Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., et al. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3T. *Neuroreport* 16, 125–128. doi: 10.1097/00001756-200502080-00010
- Polimeni, J. R., Fischl, B., Greve, D. N., and Wald, L. L. (2010). Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human VI. *Neuroimage* 52, 1334–1346. doi: 10.1016/j.neuroimage.2010.05.005
- Polley, D. B., Steinberg, E. E., and Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J. Neurosci.* 26, 4970–4982. doi: 10.1523/JNEUROSCI.3771-05.2006
- Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., and Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* 12, 221–228. doi: 10.1038/nn.2246
- 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
- Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13, 87–103.
- Scharinger, M., Henry, M. J., Erb, J., Meyer, L., and Obleser, J. (2013a). Thalamic and parietal brain morphology predicts auditory category learning. *Neuropsychologia* 53C, 75–83. doi: 10.1016/j.neuropsychologia.2013.09.012
- Scharinger, M., Henry, M. J., and Obleser, J. (2013b). Prior experience with negative spectral correlations promotes information integration during auditory category learning. *Mem. Cogn.* 41, 752–768. doi: 10.3758/s13421-013-0294-9
- Scheich, H., Brechmann, A., Brosch, M., Budinger, E., and Ohl, F. W. (2007). The cognitive auditory cortex: task-specificity of stimulus representations. *Hear. Res.* 229, 213–224. doi: 10.1016/j.heares.2007.01.025
- Schreiner, C. E., and Polley, D. B. (2014). Auditory map plasticity: diversity in causes and consequences. *Curr. Opin. Neurobiol.* 24, 143–156. doi: 10.1016/j.conb.2013.11.009
- Schroeder, C. E., and Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res. Cogn. Brain Res.* 14, 187–198. doi: 10.1016/S0926-6410(02)00073-3
- Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V., and Hari, R. (2006). Touch activates human auditory cortex. *Neuroimage* 30, 1325–1331. doi: 10.1016/j.neuroimage.2005.11.020
- Seitz, A. R., and Watanabe, T. (2003). Is subliminal learning really passive? *Nature* 422, 2003. doi: 10.1038/422036a
- Shamma, S., and Fritz, J. (2014). Adaptive auditory computations. *Curr. Opin. Neurobiol.* 25C, 164–168. doi: 10.1016/j.conb.2014.01.011
- Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006
- Shams, L., Wozny, D. R., Kim, R., and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Front. Psychol.* 2:264. doi: 10.3389/fpsyg.2011.00264
- Sigala, N., and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320. doi: 10.1038/415318a
- Spieler, L., De Lucia, M., Bernasconi, F., Grivel, J., Bourquin, N. M., Clarke, S., et al. (2011). Learning-induced plasticity in human audition: objects, time, and space. *Hear. Res.* 271, 88–102. doi: 10.1016/j.heares.2010.03.086
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., and Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Curr. Biol.* 19, 498–502. doi: 10.1016/j.cub.2009.01.066
- Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., and Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Brain Res. Cogn. Brain Res.* 13, 393–405. doi: 10.1016/S0926-6410(01)00131-8
- Tsunada, J., Lee, J. H., and Cohen, Y. E. (2011). Representation of speech categories in the primate auditory cortex. *J. Neurophysiol.* 105, 2634–2646. doi: 10.1152/jn.00037.2011
- Tsunada, J., Lee, J. H., and Cohen, Y. E. (2012). Differential representation of auditory categories between cell classes in primate auditory cortex. *J. Physiol.* 590, 3129–3139. doi: 10.1113/jphysiol.2012.232892
- Tzounopoulos, T., and Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62, 463–469. doi: 10.1016/j.neuron.2009.05.002
- Verhoef, B. E., Kayaert, G., Franko, E., Vangeneugden, J., and Vogels, R. (2008). Stimulus similarity-contingent neural adaptation can be time and cortical area dependent. *J. Neurosci.* 28, 10631–10640. doi: 10.1523/JNEUROSCI.3333-08.2008
- Wallace, M. T., and Stein, B. E. (2007). Early experience determines how the senses will interact. *J. Neurophysiol.* 97, 921–926. doi: 10.1152/jn.00497.2006
- Watanabe, T., Náñez, J. E., and Sasaki, Y. (2001). Perceptual learning without perception. *Nature* 413, 844–848. doi: 10.1038/35101601
- Wytenbach, R. A., May, M. L., and Hoy, R. R. (1996). Categorical perception of sound frequency by crickets. *Science* 273, 1542–1544. doi: 10.1126/science.273.5281.1542

**Conflict of Interest Statement:** The authors declare 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: 02 March 2014; accepted: 14 May 2014; published online: 03 June 2014.  
 Citation: Ley A, Vroomen J and Formisano E (2014) How learning to abstract shapes neural sound representations. *Front. Neurosci.* 8:132. doi: 10.3389/fnins.2014.00132  
 This article was submitted to Auditory Cognitive Neuroscience, a section of the journal *Frontiers in Neuroscience*.  
 Copyright © 2014 Ley, Vroomen and Formisano. 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.