



Editorial: Neural Mechanisms of Perceptual Categorization as Precursors to Speech Perception

Einat Liebenthal^{1*} and Lynne E. Bernstein²

¹ Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, ² Department of Speech and Hearing Science, George Washington University, Washington, DC, USA

Keywords: speech perception, phonemic perception, categorization, category learning, auditory processing, audiovisual processing, neural mechanism, neuroimaging

Editorial on the Research Topic

Neural Mechanisms of Perceptual Categorization as Precursors to Speech Perception

This research topic describes recent advances in understanding the brain functional organization for sensory categorization along with its implications for speech perception. Among the 14 papers, one theme is how neural representations of auditory and visual input are transformed across different scales of neural organization to enable speech perception, and another is the neural mechanisms of category learning.

In the first theme, several animal and human studies delve into the complex hierarchical organization of auditory ventral pathways for speech perception. Prior work has established an important role for the auditory ventral stream in complex sound categorization (Rauschecker and Scott, 2009; Romanski and Averbeck, 2009). In humans, a preference has convincingly been demonstrated for phonemic over non-phonemic sounds in non-primary auditory fields in the middle of the ventrolateral superior temporal cortex (mSTG/S) (Liebenthal et al., 2005; Leaver and Rauschecker, 2010). The present papers contribute novel insights about the function of dorsal areas in and near the auditory core, the functional specificity of the mSTG/S, and the role of non-auditory areas, for phonemic perception. Collectively, they suggest that multiple stages of abstraction from the original form of speech occur in low-level sensory cortices. In the mSTG/S, the neural representations are highly specific to phonemic categories.

Tsunada and Cohen's review of research in the monkey suggests that single neurons in the auditory core encode categories for simple sounds (e.g., direction of spectral changes), whereas neurons in the auditory belt encode more complex categories (including speech phonemes) based on input from the entire population of core neurons. At the cellular level, they report the intriguing finding that different classes of neurons within the auditory belt may have different sensitivity to category information: The more common pyramidal neurons encode auditory categories with less sensitivity than the less common interneurons (Tsunada et al., 2012). Astikainen et al. also show that in anesthetized rats' primary auditory cortex, neurons automatically encode structural patterns (order of syllable repetition) from a fast paced speech stream and generalize to novel patterns.

Based on intracranial high-gamma electrophysiological recordings in subjects with intractable epilepsy, Steinschneider et al. propose that within 200 ms, activity in the human primary and non-primary auditory cortices reflects non-categorical spectrotemporal sound attributes. Only later, activity in non-primary auditory areas receiving modulatory input from higher-order, lexico-semantic associative cortex represents phoneme categories.

OPEN ACCESS

Edited by:

Isabelle Peretz,
Université de Montréal, Canada

Reviewed by:

Ingrid Johnsrude,
University of Western Ontario, Canada

*Correspondence:

Einat Liebenthal
eliebenthal@partners.org

Specialty section:

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

Received: 26 October 2016

Accepted: 31 January 2017

Published: 14 February 2017

Citation:

Liebenthal E and Bernstein LE (2017)
Editorial: Neural Mechanisms of
Perceptual Categorization as
Precursors to Speech Perception.
Front. Neurosci. 11:69.
doi: 10.3389/fnins.2017.00069

Using multivariate pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data, Joanisse and Souza suggest that primary and non-primary areas in the human auditory cortex encode the direction of frequency modulations of complex non-speech sounds. Using an fMRI adaptation paradigm, Humphries et al. show that a relatively large area in the dorsolateral superior temporal cortex is sensitive to complex acoustic patterns in phonemic and non-phonemic sounds, whereas a small portion of the ventrolateral superior temporal cortex responds specifically to phonemic sounds, with relatively little overlap between the areas. In addition, an area of the medial superior temporal plane shows a preference for non-phonemic sounds. The results support a multi-stage hierarchical stream for speech perception extending from the superior temporal plane to the superior temporal sulcus.

Liebenthal et al. present a large meta-analysis of neuroimaging studies of the left superior temporal cortex, and find a strong preference for speech perception over other language functions in the mSTG/S. This area preferred linguistic over non-linguistic input and auditory over visual processing, prompting the suggestion that a high functional specificity of the left mSTS for auditory speech may be an important means by which the human brain achieves its exquisite affinity and efficiency for native speech perception.

Bernstein and Liebenthal's review of visual speech proposes a neural model of speech perception according to which visual aspects of speech are represented hierarchically in ascending visual pathways, with a functional organization similar to that of auditory pathways. Central to the model is the proposal that a visual area in the left posterior temporal cortex represents visual phoneme categories.

The second theme concerns how altered experience and training regimes affect perceptual categorization and neural processes. Current understanding of the normative organization of speech categories is based mostly on experiments with adults who have experienced normal language acquisition and who listen in their native language. Experiments that use natural or artificial factors that perturb and change the system help to further define the organization and mechanisms of categorization.

Heald et al. report on pitch categorization. They suggest that individuals vary in the extent to which they rely on an internal systematic tone organization. Absolute pitch (AP) possessors may be more analogous to speech perceivers than non-AP musical experts, and musical novices are expected to be least able to categorize tones based on internal organization. All three types of participants were influenced by the structure of the stimulus set and possessed useful prior pitch knowledge. Increased expertise was associated with greater influence of internal category structure.

Myers reviews the literature on normative category processing and suggests that second-language learning involves remapping the native language perceptual space to the perceptual space of the second language. Training studies typically use explicit category training, and Myers points to a wide network of frontal and temporal areas that is recruited as a result of such training. She suggests that learned sensitivity to categories is

first observable in the frontal lobe and with greater expertise is observable in temporal areas. This shift is consistent with the reverse hierarchy theory (Ahissar and Hochstein, 1997; Ahissar et al., 2008) and with frontal-to-temporal feedback as a mechanism that assists in warping category representations for the second language.

Callan et al. report an fMRI study comparing English and Japanese speakers listening to native and accented English /r-/l/. The accented English of Japanese natives is difficult for native English speakers and the English /r-/l/ is a difficult distinction for native Japanese speakers. In their results, temporal cortex areas are not significantly modulated by expertise. Instead, more difficult distinctions recruit the right cerebellum and left premotor cortex (PMC) in both groups. Second language listening additionally recruits the right PMC and left cerebellum.

Ley et al. discuss the value of MVPA for revealing high plasticity of sound representation in auditory temporal areas as a function of experience and learning. They suggest that sensory plasticity and attention processes interact to mediate category learning. They review findings within predictive coding models of perceptual learning and categorization that support a hierarchical architecture in which variation in sensory information confronts top-down signals that update bottom-up representations.

Scharinger et al. discuss the role of auditory attention in realistic listening conditions, when perception needs to adapt to dynamic degradation of certain stimulus cues. They use multimodal neuroimaging of oscillatory activity in the alpha band to study auditory categorization and highlight the role of posterior auditory areas and the inferior parietal cortex for optimal utilization of informative stimulus cues and inhibition of uninformative cues.

Lim et al. approach speech categorization through the perspective of cognitive neuroscience models that attempt to account for multiple learning systems and corresponding neural structures. These authors frame questions about the relationships between frontal and temporal cortices during learning within larger networks that include the basal ganglia. They discuss different types of feedback and task structure that may eventuate in different types of learning, declarative vs. procedural (Ashby et al., 1998). Category training tasks that encourage trainees to engage in *explicit* attempts to discover categorization rules or structure (declarative learning) result in limited generalization for speech categories, which are inherently multidimensional and incommensurate. Speech category learning appears to require procedural learning that involves bottom-up integration of stimulus features and dopaminergic reward signals.

Bernstein et al. behavioral study demonstrates an advantage to training with audiovisual speech in order to obtain improvements in the auditory perception of vocoded speech. Training used a paired associates task for which participants attempted to learn the associations between disyllabic non-sense words and non-sense pictures. Feedback was for association choices and not the phonemic content of the training stimuli. The audiovisual advantage

is interpreted within a multisensory extension of reverse hierarchy theory (Ahissar and Hochstein, 1997; Ahissar et al., 2008): Higher-level visual speech representations during audiovisual training may guide the top-down search for to-be-learned acoustic phonetic features. The training task may also promote procedural learning of the type described by Lim et al.

Future research should build on these insights to advance understanding of the neural basis of speech perception and learning.

REFERENCES

- Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature* 387, 401–406. doi: 10.1038/387401a0
- Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2008). Reverse hierarchies and sensory learning. *Philos. Trans. R. Soc. B* 364, 285–299. doi: 10.1098/rstb.2008.0253
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., and Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychol. Rev.* 105, 442–481. doi: 10.1037/0033-295X.105.3.442
- 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
- 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
- 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
- Romanski, L. M., and Averbeck, B. B. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu. Rev. Neurosci.* 32, 315–346. doi: 10.1146/annurev.neuro.051508.135431
- 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

AUTHOR CONTRIBUTIONS

EL and LB contributed equally to the conceptualization and editing of the research topic. EL and LB wrote the editorial together.

FUNDING

The work was supported by NIH R01 DC006287 and NIH R21 DC012634.

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.

Copyright © 2017 Liebenthal and Bernstein. 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.