Unlocking the role of the superior temporal gyrus for speech sound categorization

Mitchell Steinschneider

Unlocking the role of the superior temporal gyrus for speech sound categorization

Mitchell Steinschneider
Department of Neurology, Albert Einstein College of Medicine, Rose F. Kennedy Center, Bronx, New York

THE ABILITY TO ALMOST EFFORTLESSLY encode the phonemic content of running speech is a remarkable capacity of the human brain. This remarkable capacity is emphasized by the brain's maintenance of phonemic stability despite pronounced variability in the spectral and temporal characteristics of a given phoneme and a phoneme's frequent acoustic overlap with other speech sounds. For instance, vowels map out into discrete regions of acoustic space when the second formant frequency (F2) is plotted against the first formant frequency (F1). However, F2 and F1 values for a given vowel vary widely across speakers, and distributions of F1 vs. F2 values for one vowel often overlap significantly with those from others (e.g., /ae/ as in "head" and /ε/ as in "hayed") (Hillenbrand et al. 1995). In total, a whole host of sources, including a dynamic environmental background, increase variability and diminish reliable mapping of phonemes based on consistently available acoustic cues. Instead of this one-to-one assignment of specific acoustic cues with a given phoneme, the brain's task must be one of rapid categorization, placing acoustically variable speech sounds into discrete phonemic categories (see Holt and Lotto 2010 for review).

The neural network underlying phonemic categorization is slowly being clarified. Emboldened by behavioral studies demonstrating that many parallels exist between phonemic perception as seen in humans and those observed in experimental animals (Kuhl 1986; Kluender et al. 1987; Sinnott and Brown 1997), multiple investigations have examined speech processing in primary auditory cortex (A1) (Steinschneider et al. 2003; Engineer et al. 2008; Mesgarani et al. 2008). In general, A1 can best be described as performing relatively fine-grained analyses of the speech signal that facilitates, but not determines, phonemic categorization (Rauschecker and Scott 2009). These findings, obtained in experimental animals, have been supported by results obtained through direct recordings within more posterior medial portions of Heschl's gyrus, the putative location of human primary auditory cortex (Steinschneider et al. 2005; Bitterman et al. 2008; Nourski et al. 2009).

Beyond A1, the posterior lateral region of the superior temporal gyrus (PLST) represents an intermediate stage of speech processing that is envisioned to play a fundamental role in phonemic processing and categorization (Poeppel et al. 2008; Hickok 2009; Price 2010). Electrical stimulation of the posterior medial portion of Heschl's gyrus elicits very short latency responses in this area, suggesting direct connects with primary auditory cortex (Brugge et al. 2003). Potential analogs of this region in the macaque monkey include the portion of auditory cortex termed the anterolateral (AL) belt region

(Hackett 2007). Previous studies have shown that neurons in this region are both highly responsive to conspecific vocalizations and show specificity in their responses to vocalization type (Tian et al. 2001; Russ et al. 2008a). These and other physiological studies, coupled to known anatomical connections of AL, indicate that this region is an important, intermediate node in the processing stream encoding species-specific vocalizations, and might thereby serve as a useful model area for investigating phonemic encoding mechanisms (Romanski and Averbeck 2009).

Recently, Tsunada et al. (2011) provided evidence that AL is a highly relevant model area for examining the categorization process of phonemes. Two male macaque monkeys were trained to make a leftward eye saccade when two sequential speech sounds were perceived as the same, and a rightward eye saccade when the two syllables were perceived as different. The speech sounds were the syllables /bad/ and /dad/ recorded from a female speaker. These prototypic syllables were sequentially morphed such that /b/ at syllable onset systematically took on the acoustic characteristics of /d/ in 20% increments, with an additional syllable placed at the midpoint (50%) of acoustic differences. Prototypic and intermediate syllables were presented to the animals while they performed the match-to-category task described above and while the investigators recorded from single neurons in area AL.

Excellent categorization was obtained in the behavioral performance of the animals. For example, when the first of the sequentially presented syllables (reference syllable) was a /bad/ that was 40% different from the prototype, a 20% acoustic change towards the prototype /bad/ in the second of the sequentially presented syllables (test syllable) evoked a reliable behavioral response of "same" in the two monkeys, whereas the identical 20% acoustic change in the other direction (towards the /dad/ prototype) reliably elicited a behavioral response of "different" in the animals. Thus, a clear categorical-like boundary at the 50% midpoint syllable was obtained (Kuhl 1986).

Categorical-like patterns in neural firing were assessed by computing a category index (CI) on a neuron-by-neuron basis. This index represented the difference in firing rates between a "between-category difference" (BCD) and a "within-category difference" (WCD) divided by their sum. The WCD was the average of the absolute differences between test syllable firing rates of morph pairs that resided on the same side of the perceptual boundary, while the BCD represented the average of the absolute differences between test syllable firing rates of equivalent morph differences, but when the syllables resided on opposite sides of the perceptual boundary. For example, averaged CIs included the comparison of WCD firing rate differences between the 20% and 40% morphs averaged with the 60% and 80% morphs (both 20% differences but residing on the same perceptual boundary side), and the BCD firing rate

Address for reprint requests and other correspondence: M. Steinschneider, Dept. of Neurology, Albert Einstein College of Medicine, Rose F. Kennedy Center, Rm. 322, 1300 Morris Park Ave., Bronx, NY 10461 (e-mail: mitchell.steinschneider@einstein.yu.edu).

differences between the 40% and 60% morphs (20% difference now crosses the perceptual boundary). CI values greater than 0 indicate neural responses that respected the categorical boundary. Measurements were taken incrementally in 5-ms bins to assess the temporal dynamics of the responses.

Neural responses comparing the activity evoked by syllables with equal acoustic differences, but which either straddled or remained on the same side of the perceptual boundary, showed significant "categorical-like" behavior. The CI was consistently greater than 0 over the time span of the syllables. Categorical-like activity was rapid, with two peaks occurring at ~100 and 200 ms after stimulus onset. In contrast, the authors demonstrated through use of several computational methods that the activity in AL was not related to decision choice. These findings contrast with earlier works by Tsunada et al. (2011) using identical stimuli, which identified a role for an auditory-responsive region of the ventral prefrontal cortex in the behavioral decisions made by the monkeys (Russ et al. 2008b; Lee et al. 2009).

Findings in AL of the monkey complement human intracranial and functional neuroimaging studies demonstrating categorical-like speech-evoked activity in PLST. In the most directly related paper, Chang and colleagues (2010) linearly morphed the syllables /ba/, /da/, and /ga/ by parametrically modifying the starting frequency of F2 in 14 equal steps. Categorical boundaries were perceptually identified and compared with neural activity concurrently recorded from subdural grid electrodes placed over PLST. Neural responses within a latency range of 110–150 ms, as measured by the amplitude and distribution of auditory-evoked potentials, physiologically clustered in a distributed but categorical-like manner that paralleled the perceptual categorization. Thus, neural response distributions could accurately predict the perceptual category of the consonants and were not simply based on acoustic differences across the syllables. Another intracranial study of PLST found that in the 100–150 ms time frame, differences in the amplitude of high-gamma activity elicited by the voiced syllables /ba/, /da/, and /ga/ could predict differences in the responses elicited by their unvoiced counterparts /pa/, /ta/, and /ka/ (Steinschneider et al. 2011). Thus, despite differences in the voice onset time (VOT) of the consonants and their corresponding acoustic correlates, responses reflected the place of articulation of the stop consonants. In parallel, the VOT of the syllables were categorically represented in the high-gamma activity despite differences in the consonant place of articulation. By manipulating the signal-to-noise ratio of /ba/ and /da/ using varying levels of white noise, Binder and colleagues (2004) modulated the difficulty of syllable discrimination from an easy task to one where subjects performed at chance. Brain regions showing an enhanced BOLD signal specifically related to the accuracy of discrimination included PLST and the adjacent anterior lateral portion of Heschl's gyrus bilaterally. In contrast, activity related to behavioral reaction time, a measure of decision choice, was localized to the inferior frontal regions bilaterally.

While these studies strongly support the role of the lateral superior temporal gyrus in speech categorization, many questions remain. For instance, the mechanisms by which AL in monkeys and PLST in humans shape categorical-like activity is unknown. Both areas integrate convergent input from tonotopically organized core auditory cortex (Rauschecker and Tian 2004). Thus, it is reasonable to suggest that unique, distributed patterns of activity respecting the differential spectral content of the speech sounds promote categorical-like responses observed in AL and PLST (Obleser et al. 2007; Leaver and Rauschecker 2010). Integration with inputs from cells in A1 encoding a phoneme's spectral bandwidth or temporal characteristics would allow multidimensional scaling of the neural activity and further enhance phonemic categorization (Mesgarani et al. 2008; Leaver and Rauschecker 2010). The studies by Tsunada et al. (2011), Chang et al. (2010), and Steinschneider et al. (2011) did not examine if the categorical-like activity could be partially explained by differential selectivity to underlying acoustic attributes of the speech sounds, and thus this issue remains undetermined. Likewise, the roles of learning and neural plasticity in shaping categorical-like activity in AL and PLST are also open questions (Ohl and Scheich 2005; Fritz et al. 2007). The latter is especially relevant for language-learning in the young (Kuhl 2010), and it is exciting that future studies will undoubtedly build on the model system reported by Tsunada and colleagues (2011) to address these high impact concerns.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. *Nature Neurosci* 7: 295–301, 2004.

Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I. Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451: 197–201, 2008.

Brugge JF, Volkov IO, Garell PC, Reale RA, Howard MA III. Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *J Neurophysiol* 90: 3750–3763, 2003.

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT. Categorical speech representation in human superior temporal gyrus. *Nature Neurosci* 13: 1428–1433, 2010.

Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP. Cortical activity patterns predict speech discrimination ability. *Nature Neurosci* 11: 603–608, 2008.

Fritz JB, Elhilali M, Shamma SA. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J Neurophysiol* 98: 2337–2346, 2007.

Hackett TA. Organization and correspondence of the auditory cortex of humans and nonhuman primates. In: *Evolution of the Nervous System*, edited by JH Kaas. Oxford: Elsevier pp. 109–119, 2007.

Hickok G. The functional neuroanatomy of language. *Phys Life Rev* 6: 121–143, 2009.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J Acoust Soc Am* 97: 3099–3111, 1995.

Holt LL, Lotto AJ. Speech perception as categorization. *Attn Perc Psychophys* 72: 1218–1227, 2010.

Kluender KR, Diehl RL, Killeen PR. Japanese quail can learn phonetic categories. *Science* 237: 1195–1197, 1987.

Kuhl PK. Theoretical contributions of tests on animals to the special mechanisms debate in speech. *Exp Biol* 45: 233–265, 1986.

Kuhl PK. Brain mechanisms in early language acquisition. *Neuron* 67: 713–727, 2010.

Leaver AM, Rauschecker JP. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci* 30: 7604–7612, 2010.

Lee JH, Russ BE, Orr LE, Cohen YE. Prefrontal activity predicts monkeys' decisions during an auditory category task. *Front Integrative Neurosci* 3: 2009.

Mesgarani N, David SV, Fritz JB, Shamma SA. Phoneme representation and classification in primary auditory cortex. *J Acoust Soc Am* 123: 899–909, 2008.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA III, Brugge JF. Temporal envelope of time-compressed speech represented in the human auditory cortex. *J Neurosci* 29: 15564–15574, 2009.

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP. Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex* 17: 2251–2257, 2007.

Ohl FW, Scheich H. Learning-induced plasticity in animal and human auditory cortex. *Curr Opin Neurobiol* 15: 470–477, 2005.

Poeppel D, Idsardi WJ, van Wassenhove V. Speech perception at the interface of neurobiology and linguistics. *Phil Trans R Soc B* 363: 1071–1086, 2008.

Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann NY Acad Sci* 1191: 62–88, 2010.

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neurosci* 12: 718–724, 2009.

Rauschecker JP, Tian B. Processing of band-pass noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91: 2578–2589, 2004.

Romanski LM, Averbeck BB. The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu Rev Neurosci* 32: 315–346, 2009.

Russ BE, Ackelson AL, Baker AE, Cohen YE. Coding of auditory stimulus identity in the auditory non-spatial processing stream. *J Neurophysiol* 99: 87–95, 2008a.

Russ BE, Orr LE, Cohen YE. Prefrontal neurons predict choices during an auditory same-different task. *Curr Biol* 18: 1483–1488, 2008b.

Sinnott JM, Brown CH. Perception of the American English liquid /ra-la/ contrast by humans and monkeys. *J Acoust Soc Am* 102: 588–602, 1997.

Steinschneider M, Fishman YI, Arezzo JC. Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *J Acoust Soc Am* 114: 307–321, 2003.

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA III. Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cerebral Cortex* 15: 170–186, 2005.

Steinschneider M, Nourski KV, Kawasaki H, Oya H, Brugge JF, Howard MA III. Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. *Cerebral Cortex.* In press doi:10.1093/cercor/bhr014, 2011.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science* 292: 290–293, 2001.

Tsunada J, Lee JH, Cohen YE. Representation of speech categories in the primate auditory cortex. *J Neurophysiol* jn.00037.2011; published ahead of print February 23, 2011, doi:10.1152/jn.00037.2011.