

The new neuroanatomy of speech perception

Our understanding of speech recognition processes has gradually advanced over the past 50 years, from a state of almost total ignorance to one of well-informed confusion. Technical advances introduced around the middle of the last century enabled detailed description of the spectral patterns and temporal phenomena that characterize vowels and consonants, and extensive perceptual studies were undertaken to determine the relative importance of different classes of these acoustic cues. Explicit theories of how consonant and vowel percepts (phonemes) arise from such cues were developed and implemented, resulting not only in the successful artificial synthesis of naturally sounding speech, but also in the more recent development of (reasonably successful) speech-to-text transcription devices. For the most part, this progress proceeded without parallel advances in our understanding of how speech perception is actually implemented in the brain. Until recently, the conventional neuroanatomical model of speech perception had changed little from the one proposed by Wernicke in 1874, which predated the technological advances just described and included no mention of phonetic cues or phonemes (Wernicke, 1874).

This situation is now changing rapidly with the application of newer functional neuroimaging techniques, which permit relatively precise localization of brain activity associated with auditory processing. One issue of particular interest is whether or not there exist regions within the auditory cortex that are specialized for processing speech sounds. While far from universally accepted, the idea that speech sounds enjoy a special status seems plausible because of both the extraordinary acoustic complexity of the sounds and the obvious species-specific importance of speech in communication. In the past decade, scientists using functional neuroimaging have repeatedly observed a superior temporal region in both hemispheres that activates more strongly to speech than to non-speech sounds like tones and noise (Demonet *et al.*, 1992; Zatorre *et al.*, 1992; Binder *et al.*, 1996, 2000; Mummery *et al.*, 1999; Belin *et al.*, 2000). What was initially surprising was the location of this new 'speech centre': whereas the conventional neuroanatomical model of language processing emphasizes the importance of the posterior part of the superior temporal gyrus (STG), the speech-specific activation lay anterolateral to primary

auditory cortex and anterior to the mid-point of the gyrus. By contrast, the planum temporale, an area on the posterior STG that had long been considered a speech centre, did not show a preference for speech sounds (Binder *et al.*, 1996). It thus appeared as though the projections from primary to secondary auditory cortex enabling speech recognition followed an anterolateral rather than a posterior course as previously believed. In recent years, anatomical studies in monkeys provided further support for this model by showing two distinct projection systems within the auditory system, one anteriorly directed and presumably supporting recognition of complex sounds and the other posteriorly directed and presumably involved in sound localization (Romanski *et al.*, 1999).

What remains unclear from these studies is precisely why the anterolateral 'speech centre' responds preferably to speech. Is it because of the greater acoustic complexity of speech, which implies a difference in workload at the auditory level, or is it due to a special recoding of the acoustic material into abstract, linguistic codes (i.e. phonemes)? One way to decide between these alternatives is to test the brain's response to sounds that are as acoustically complex as speech but cannot be recoded as consonants or vowels, yet this enterprise has proved deceptively difficult. One such study employed reversed speech sounds, which were shown to produce activation of the 'speech centre' indistinguishable from that produced by normal speech (Binder *et al.*, 2000). Another used 'nonspeech vocalizations' such as humming, laughter, yawning, and so on, with similar results (Belin *et al.*, 2000). Thus, while these studies appear to suggest the importance of acoustic rather than phonetic factors in activation of the 'speech centre,' it is difficult to exclude the possibility of at least some recoding of the stimuli. For example, many nonspeech vocalizations can be written down ('hmmm,' 'ha ha') or named, implying a recoding process. Similarly, the non-words of reversed speech can, to some degree at least, be transcribed. What these examples illustrate is how thoroughly predisposed the speech recognition system is to phonetic analysis: given a sufficiently speech-like input, the system tends to recode the input as phonemes, making separation of these levels of processing difficult.

The study by Scott and co-workers in this issue of *Brain* provides important new evidence on this question (Scott

et al., 2000). The authors dissociated auditory processes from phonetic recoding using a technique called spectral rotation, a filtering process that preserves the acoustic complexity of the original speech but renders most phonemes virtually unrecognizable (Blessner, 1972). The results show what appears to be a further subdivision within the left temporal lobe 'speech centre.' On the lateral STG, anterolateral to primary auditory cortex, responses were as strong for spectrally rotated speech as for normal speech, suggesting processing at an auditory level. Further ventrally, in the anterior superior temporal sulcus (STS), responses were stronger for speech than for spectrally rotated speech, suggesting neural activity related to phoneme recognition. That this STS activation did not depend on acoustic complexity was further demonstrated using noise vocoding, a filtering technique that removes much of the spectral complexity from speech but preserves some phoneme recognizability (Shannon *et al.*, 1995). Unlike the spectrally rotated speech, noise-vocoded speech activated the left anterior STS comparably to normal speech.

Because Scott and co-workers used sentences as stimuli, phoneme recognition would have produced some degree of word recognition, semantic and syntactic processing. It is thus difficult to say which of these processes, or which combination of processes, is represented by the left anterior STS activation. To put these findings in a larger perspective, then, it may be useful to note that functional neuroimaging studies have also considerably advanced our understanding of lexical-semantic processing in the brain (Grabowski and Damasio, 2000). Findings from these studies indicate participation of multiple areas, including middle and inferior temporal gyri, fusiform gyrus, angular gyrus, and frontal lobe, during auditory word recognition. The superior temporal system specialized for speech sound recognition is but an early stage in a processing stream that ultimately projects to all components of this distributed system.

Jeffrey Binder

*Associate Professor of Neurology,
The Medical College of Wisconsin,
Milwaukee, Wisconsin,
USA*

References

- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature* 2000; 403: 309–12.
- Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW. Function of the left planum temporale in auditory and linguistic processing. *Brain* 1996; 119: 1239–47.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, et al. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 2000; 10: 512–28.
- Blessner B. Speech perception under conditions of spectral transformation: I. Phonetic characteristics. *J Speech Hear Res* 1972; 15: 5–41.
- Demonet J-F, Chollet F, Ramsay S, Cardebat D, Nespoulous J-L, Wise R, et al. The anatomy of phonological and semantic processing in normal subjects. *Brain* 1992; 115: 1753–68.
- Grabowski TJ, Damasio AR. Investigating language with functional neuroimaging. In: Toga AW, Mazziotta JC, editors. *Brain mapping: the systems*. San Diego (CA): Academic Press; 2000: p. 425–61.
- Mummery CJ, Ashburner J, Scott SK, Wise RJ. Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J Acoust Soc Am* 1999; 106: 449–57.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 1999; 2: 1131–6.
- Scott SK, Blank C, Rosen S, Wise RJS. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 2000; 123: 000–000.
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995; 270: 303–4.
- Wernicke C *Der Aphasische Symptomenkomplex. Eine Psychologische Studie auf Anatomischer Basis*. Breslau: M. Cohn und Weigart; 1874.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 1992; 256: 846–9.