The functional neuroanatomy of speech perception: Spatial and temporal considerations

David Poeppel
Cognitive Neuroscience of Language Laboratory
Department of Linguistics, Department of Biology, University of Maryland College Park

The cortical basis of speech perception - and phonological processing more generally – remains poorly understood and an area of controversy (Poeppel 1996). Part of the difficulty stems from the fact that the neural systems supporting speech perception vary as a function of task. Specifically, the set of cognitive and neural systems involved in performing traditional laboratory speech perception tasks, such as syllable discrimination or identification, are not necessarily the same as those involved in speech perception as it occurs during natural language comprehension. Based on a review of data from a range of methodological approaches (neuropsychology, neuroimaging, electrophysiology), as well as new experimental evidence (fMRI, PET), we propose that auditory cortical fields in the posterior half of the superior temporal gyrus, *bilaterally*, constitute the primary substrate for constructing sound-based representations of speech, and that these sound-based representations interface with different supramodal systems in a task-dependent manner (Hickok & Poeppel 2000). Tasks that require access to the mental lexicon (i.e., accessing meaning-based representations) rely on a ventral pathway in which auditory-speech representations are mapped onto meaning; tasks that require explicit access to speech segments rely on a dorsal pathway which interfaces auditory- and articulatory-based representations of speech (cf. Zatorre et al. 1992; Burton et al. 2000). Figure 1 illustrates the proposed functional anatomy of the speech processing system. What unifies the proposed model is the notion of *feature*. Featural representations are an effective way to store lexical information and articulatory commands are best understood in terms of features as well; importantly, there is evidence (Phillips et al. 2000) that auditory cortex can access features).

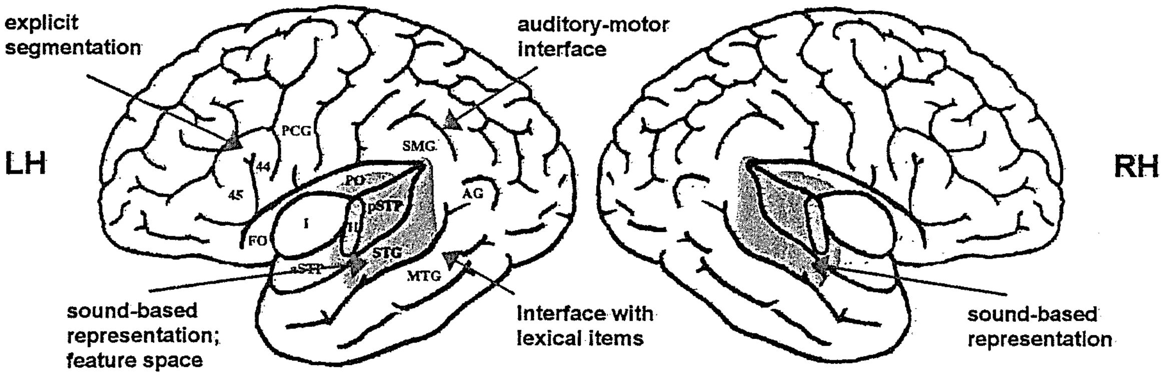


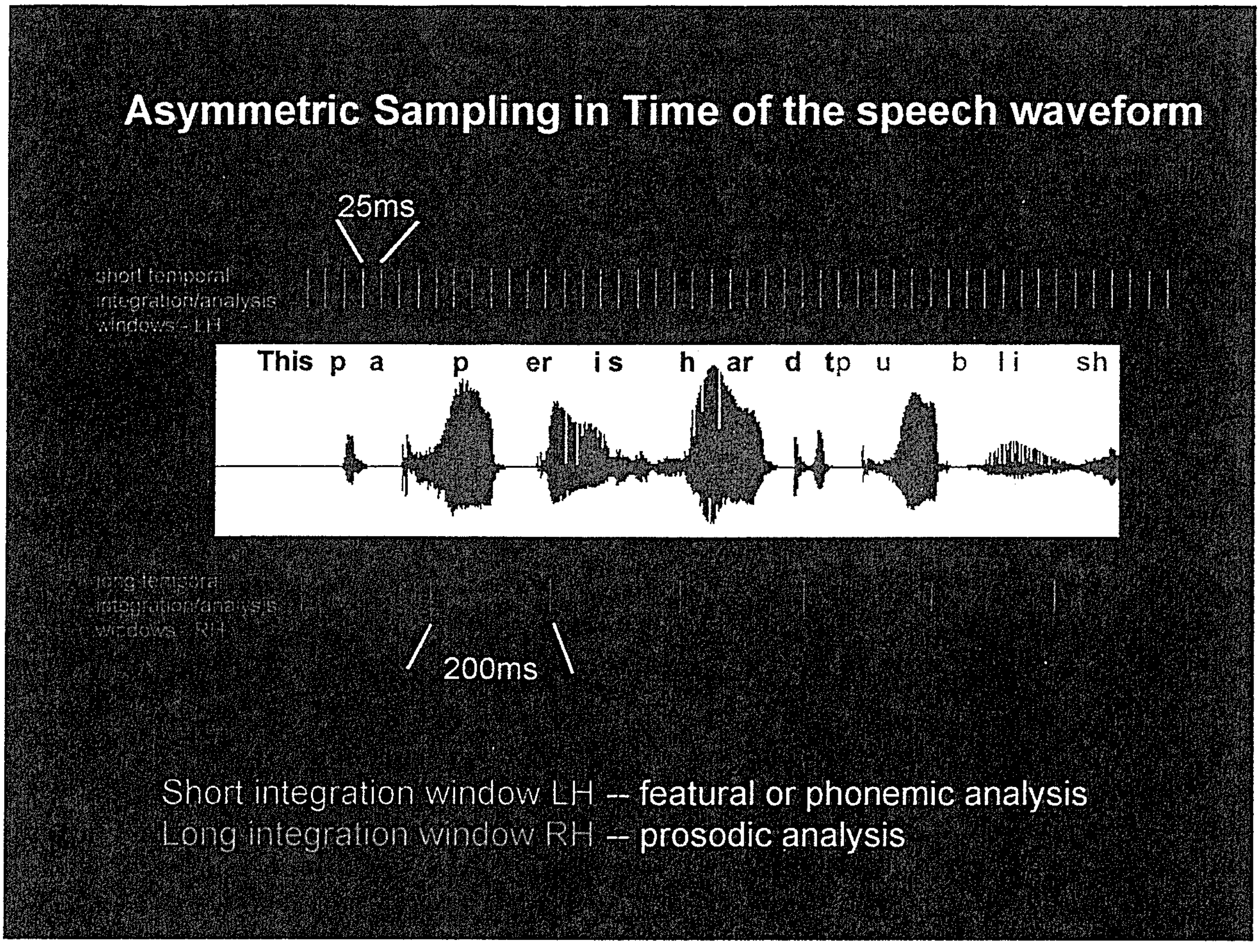
Figure 1

The model accounts well for activation data as well as the clinical profiles from fluent aphasics. To what extent, however, are right-hemisphere mechanisms also implicated, as suggested by Figure 1? An increasing body of evidence suggests that cortical fields in the right temporal lobe (in addition to primary auditory projection areas) play a critical role in the analysis of the speech signal (Belin et al. 2000; Binder et al. 2000; Boatman et al. 1998; Buchman et al. 1986; Burton et al. 2000; Hickok & Poeppel 2000; Norris & Wise 2000). An integrated model of the anatomy and physiology of speech perception will therefore need to account for the involvement of both left and right temporal areas. It is proposed here that the crucial hemispheric difference derives from the manner in which information relevant to speech is processed and represented in the time domain.

The model that is proposed, Asymmetric Sampling in Time (AST), builds on three observations. First, it is known from psychophysics and physiology that information that unfolds over time is "chunked". In particular, *temporal integration windows* provide the logistical framework to organize temporally developing information. The temporal window for which there exists the most psychophysical evidence (across methods and sensory systems) has a duration on the order of 25-40ms. Second, it is been shown in neuropsychological and imaging data that left hemisphere mechanisms are suited for the execution of tasks that require high temporal resolution, e.g. tasks in which rapidly changing (30-40ms) acoustic transients must be identified (Belin et al. 1998; Johnsrude et al. 1997; Nicholls 1996). The final observation is that acoustic and articulatory phenomena occur on several different time scales. For example, acoustic-phonetic components such as formant transitions typically occur over short time scales, say 25-50ms. On the other hand there are (perceptually robust) phenomena on the temporal order of syllables (150-250ms), such as the analysis of tonality and prosody.

The Asymmetric Sampling in Time (AST) hypothesis holds that left temporal cortical areas preferentially extract information over *25ms temporal integration windows*. Right hemisphere areas preferentially integrate over *long, 150-250ms integration windows*. By assumption, the input signal (in this case auditory) has a neural representation that is bilaterally symmetric (e.g. at the level of A1); beyond the initial representation, however, the signal gets elaborated asymmetrically in the time domain. Another way to describe the AST proposal is to say that the *sampling rate* of non-primary auditory areas is different, with LH sampling at high frequencies (~40Hz) and RH sampling at low frequencies (4-10Hz).

AST is an example of functional segregation, a standard concept in cortical processing. Specifically, it is an example of multi-resolution analysis, a signal processing strategy common in other domains, such as vision (cf. V4 versus V5/MT properties). Moreover, the concept connects in natural ways to the notion of global versus local processing in vision. On this view, there is a global, lower time-resolution analysis at, e.g., the syllabic level and a local, high temporal resolution analysis at the sub-syllabic level.


Asymmetric Sampling in Time of the speech waveform

Short integration window LH -- featural or phonemic analysis
Long integration window RH -- prosodic analysis

Figure 2

Figure 2 illustrates the intuition that motivates the AST hypothesis. Empirical predictions for AST include: (1) always bilateral activation in natural speech tasks; (2) linguistic and affective prosody (at the level of intonation contour) should both be associated with right hemisphere mechanisms; (3) music perception should lateralize to the right for most musical attributes (including pitch); (4) rapid FM sweeps lateralize to LH, slow (eg 300ms) sweeps should drive RH; (5) in electrophysiologic recordings, 40Hz activity should have more power in LH recordings.

New data from hemodynamic (PET) and electromagnetic (MEG) recordings will be shown in support of some of the above predictions. For example, the connection between temporal integration windows and oscillatory activity is explored in a whole-head MEG study (Poeppel et al. 2000). If activity in specific frequency bands is associated with the functional activation of each hemisphere (e.g. phonetic segmentation in the left versus prosodic analysis in the right), then the relevant frequency bands might be differentially salient in the two hemispheres during different functional states. We tested this basic hypothesis by performing MEG recordings during the presentation of auditory stimuli of varying spectral complexity, ripples (dynamic broadband stimuli). Ripples were constructed to approximate spectral and temporal aspects of speech. Whole-head MEG (Magnes 2500, 4D-Neuroimaging) and high-density EEG (ESI-128, Neuroscan Inc.) were used to acquire auditory evoked responses. Spectral analysis techniques provided estimates of the relevant frequency responses. High-frequency responses were robustly different for left and right regions, with gamma activity (25-60Hz) being more pronounced in left temporal cortex. The data are consistent with the view that sensory input is recorded in specific frequency bands and on different time-scales by the two hemispheres, depending on the nature of the sensory input. This physiologic interpretation of the AST model is illustrated in Figure 3.

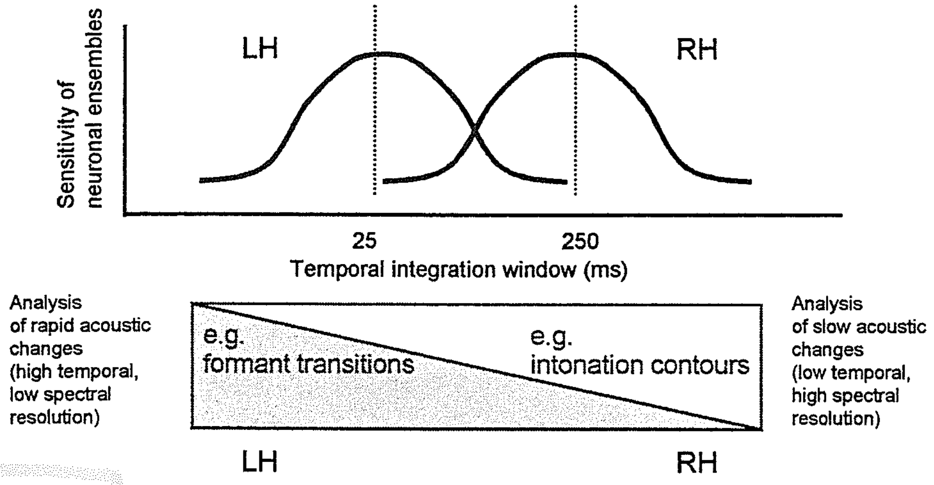


Figure 3

REFERENCES

Belin, P. et al. (1998) Lateralization of speech and auditory temporal processing. J Cogn Neurosci 10: 536-40.

Belin, P. et al. (2000) Voice-selective areas in human auditory cortex. Nature 403: 309-12.

Binder, J. et al. (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10: 512-28.

Boatman, D. et al. (1998) Right hemisphere speech perception revealed by amobarbital injection and electrical interference. Neurology 51: 458-64.

Buchman et al. (1986) Word deafness: one hundred years later. J Neurol Neurosurg Psychiatry : 489-99.

Burton, M. et al. (2000) The role of segmentation in phonological processing: an fMRI investigation. J Cogn Neurosci 12: 679-90.

Hickok, G. & Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. Trends Cogn Sci 4: 131-138.

Johnsrude, I. et al. (1997) Left-hemisphere specialization for the processing of acoustic transients. Neuroreport 8: 61-5.

Nicholls, ME. (1996) Temporal processing asymmetries between the cerebral hemispheres: evidence and implications. Laterality 1: 97-137.

Norris, D. & Wise, R. (2000). The study of prelexical and lexical processes in comprehension: psycholinguistics and functional neuroimaging, The new cognitive neurosciences, . Cambridge: MIT Press.

Phillips, C. et al. (in press) Auditory cortex accesses phonological categories: an MEG mismatch study. Journal of Cognitive Neuroscience.

Poeppel, D. (1996) A critical review of PET studies of phonological processing. Brain & Language 55: 317-351.

Poeppel, D. et al. (2000,). High-frequency response asymmetry to auditory stimuli of varying spectral complexity. Paper presented at the Society for Neuroscience, New Orleans.

Zatorre, R. et al. (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256: 846-849.

ACKNOWLEDGMENTS

This work is in collaboration with Allen Braun (NIH), Greg Hickok (UC Irvine), Alec Marantz (MIT), Colin Phillips (UMCP), Tim Roberts (UCSF), and Jonathan Simon (UMCP). DP is supported by the McDonnell-Pew Program in Cognitive Neuroscience, the National Science Foundation LIS initiative, NIH DC 0463801, and the U. Maryland Center of Comparative Neuroscience.