

# Brain Imaging Correlates of Temporal Quantization in Spoken Language

David Poeppel

Department of Linguistics and Department of Biology  
University of Maryland College Park  
dpoeppel@deans.umd.edu

## Abstract

Psychophysical research has established that temporal-integration windows of several different sizes are critical for the analysis of any acoustic speech signal. Recent work from our laboratory has examined speech processing in the human auditory cortex using both hemodynamic (fMRI, PET) and electromagnetic (MEG, EEG) recording techniques. These studies provide evidence for at least two distinct temporal scales relevant to the integration and processing of speech at the cortical level – a relatively short window of 25-50 ms and a longer window of 150-300 ms. In addition to support for processing on these time scales, there is also evidence for hemispheric asymmetry in temporal quantization. Left auditory cortex shows enhanced sensitivity to rapid temporal changes (possibly associated with segmental and subsegmental perceptual analysis), while right auditory cortex is more sensitive to slower changes (possibly associated with syllabic rate processing and dynamics of pitch).

## 1. Introduction

The functional organization of human auditory cortex and its role in the representation and processing of speech have recently been investigated with the currently available noninvasive imaging techniques. These methods include the hemodynamically based imaging methods positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) as well as the electromagnetically based neurophysiological methods electroencephalography (EEG) and magneto-encephalography (MEG). Such noninvasive techniques provide complementary views to the data obtained in more clinical contexts (e.g. deficit-lesion correlation in neuropsychology, intracranial recordings performed for surgical evaluations, etc.). One compelling advantage conferred by the ability to record non-invasively in vivo is that one can acquire psychophysical data during recording, permitting relatively constrained interpretations of the observed activations with respect to the behavioral tasks. The relative advantages and disadvantages of these techniques, both from a signal recording and an experimental-design perspective, have been reviewed at length [e.g. 1, 2]. In the context of the exploration of human auditory processing, the relevant generalizations are (i) that the spatial resolution of the hemodynamic techniques is good (on the order of 0.5-5mm) whereas their temporal resolution is limited (by the hemodynamic response) to approximately 1s (fMRI) or lower (PET); and (ii) that the spatial resolution of the electromagnetic techniques is limited to approximately 5-10mm (by the nature of the inverse problem as it applies to source localization) whereas their

temporal resolution is commensurate with functionally relevant neuronal activity (1 ms or better).

Given these attributes of the recording techniques, research on temporal properties – whether properties of the speech signal or properties of the recorded response – must be sensitive to the specific limitations and advantages afforded by a given approach. Building on a model of the cortical functional anatomy of speech sound processing (Section 2), I summarize psychophysical (Section 3), electro-physiological (Section 4.1), and hemodynamic (Section 4.2) experiments that are consistent with the following hypothesis on the cortical mediation of speech perception (Section 5). First, speech sound processing is mediated bilaterally in the superior temporal gyrus; second, there exist privileged temporal processing regimes that form the basis of processing on the syllabic (~150-300 ms) and (sub)segmental (~20-50 ms) scales; and third, there is a hemispheric asymmetry associated with these time scales such that processing based on the shorter temporal integration window is left-hemisphere biased and processing based on the longer integration window right-hemisphere biased (asymmetric sampling in time, AST). This perspective (multiresolution processing in different temporal integration windows, instantiated at the neuronal ensemble circuit level) accounts well for observations deriving from both neuropsychology and imaging that point to *anatomic symmetry* but *functional asymmetry*: the analysis of suprasegmental prosodic phenomena and dynamic pitch occurs over longer time scales and is right lateralized, the analysis of short-time-scale information such as rapid formant transitions and is left lateralized.

## 2. Functional anatomy

PET and fMRI studies provide a range of relevant insights into the functional anatomy of the speech perception system, and there is emerging consensus regarding a large-scale functional anatomy of speech perception and its interface with speech production [3, 4]. In particular, (i) the analysis of acoustic-phonetic information is mediated by the superior temporal gyrus (STG), (ii) a ‘ventral’ pathway originating in dorsal STG interfaces acoustically derived input with lexical representations in the middle and inferior temporal gyri (MTG and ITG), (iii) a ‘dorsal’ route pathway forms the basis for an auditory-motor (acoustic-to-articulatory) interface area (Sylvian parieto-temporal), and (iv) the ‘dorsal’ route incorporates left inferior frontal areas in which articulatory representations appear to be mediated.

One feature of the functional anatomy about which there exists considerable agreement, and to which imaging studies have contributed significantly [5, 6], is that the contribution of the early stages is strongly bilateral (STG in particular). The striking ‘dominance’ of the left

hemisphere associated with language processing – an attribute often extended to speech processing - appears beyond the analysis of the speech signal, which occurs bilaterally in core, belt, and parabelt human areas. The bilateral nature of the response is, in fact, consistent with data from neuropsychology, for example work on the syndrome ‘pure word deafness’ [7, 8]. The deficit-lesion data from such cases show that word deafness (i.e. relatively preserved ability to process non-speech signals compared to speech) occurs only in cases in which the integrity of both left and right STG are compromised.

### 3. Temporal integration

Psychophysical and physiological experiments suggest that information unfolding over time is quantized or ‘chunked.’ In particular, temporal integration windows provide the logistical framework to organize temporally developing information. The temporal windows for which there exists a body of psychophysical evidence (across methods and sensory systems) have durations on the order of 25-40ms and 150-250ms. Two brief examples highlight the relevance of the slow, syllabic-scale integration [9, 10] and show the interaction between the information carried on the different time scales.

#### 3.1. Audio-visual integration

Participants were tested in two audio-visual (AV) experiments that focused on temporal coincidence in AV speech. Recordings of /pa/ and /ba/ were dubbed onto video recordings of /ka/ or /ga/, respectively to produce the illusory fusion percepts /ta/, or /da/ [11]. First, an identification task using AV pairs with asynchronies ranging from -467 ms (auditory lead) to +467 ms was conducted. Fusion responses occurred over temporal asynchronies from -30ms to +170ms audio lag. Second, simultaneity judgments for incongruent and congruent audiovisual tokens were collected. McGurk pairs were more readily judged as asynchronous than congruent pairs. However, characteristics of the temporal window over which simultaneity and fusion responses were maximal were quite similar, suggesting the existence of a 200ms duration asymmetric bimodal temporal integration window [12].

#### 3.2. Multiple scales and their interaction

We created, based on natural sentences, test items in which the slow or rapid modulations were selectively extracted and played back in order to examine to what extent intelligibility is modulated by information on different time scales. The original wide band speech signal was split into 14 frequency bands with an FIR filter bank spanning the range 0-6000Hz spaced in 1/3 octave steps along the cochlear frequency map. The amplitude envelope from each band was computed by means of a Hilbert transform and then either low- (0-3Hz) or high- (22- Hz) passed before being reconstituted again with the original carrier signal. The result for each original signal (S) is an S\_low and an S\_high version, containing only low (below 5Hz) or high (above 20Hz) modulation frequencies. Each of these signals, when presented separately in intelligibility tasks, shows limited intelligibility (S\_low 40%, S\_high 17%). However, the dichotic presentation of S\_low with S\_high results in high (65%) intelligibility. In fact, the performance is significantly better than expected on an

additive model and hints at a real interaction between the information carried at these time scales. The study demonstrates that intelligibility crucially depends on both slowly modulating and rapidly modulating components of speech and hints at a binding process, in which a conjunction of these creates an emergent representation that is critical for successful speech processing [13].

## 4. Imaging experiments

Building on the psychophysical work that suggests the relevance of two time scales corresponding roughly to a 20-50ms time constant (electrophysiological gamma band range) and a 150-250ms constant (electrophysiological theta range), we turn to non-invasive methods. We briefly examine studies pointing to the primacy of rapid integration as well as hemispheric asymmetry.

#### 4.1. Electrophysiology

##### *Magnetoencephalography (MEG) studies*

If coherent activity in specific physiological frequency bands is associated with the functional characteristics of each hemisphere, then the relevant frequency bands might be differentially salient in the two hemispheres. We performed MEG recordings during presentation of auditory stimuli of varying spectral complexity, including (continuous) speech and ripples (auditory analogue of visual gratings). The power ratios between spectral frequency bands were computed, revealing pronounced asymmetries. In particular, the gamma/theta ratio is different for left and right hemispheres, with gamma activity being more pronounced in the left temporal areas. The effect is robust in that it is observed in MEG recordings for simple and complex stimuli. The characteristics of the spectral power ratios differ among conditions for the complex auditory stimuli and differ between single-trial and averaged data, suggesting the importance of analyzing both time-locked and non-time-locked activity. These data are consistent with the view that information is analyzed in specific frequency bands and different time-scales in the hemispheres and supports the left/short-right/long integration hypothesis.

#### 4.2. Hemodynamic (fMRI) imaging

When we turn to PET or fMRI studies, the underlying measurement approach and inherently limited temporal resolution prevent one from studying the cortical electrophysiological responses that are directly elicited by the stimulus parameter of interest. Rather, one is looking at a rather distal response, mediated by the hemodynamic lag. Insofar as one wants to take advantage of the superior spatial resolution afforded by, for example, fMRI – and insofar as one remains interested in the timing dimension – one must build the relevant temporal distinctions into the stimulus and then test whether the manipulations in the acoustic signal are reflected in a systematic way in the fMRI-recorded signal.

Here, a novel acoustic stimulus with parametrically varying temporal structure, reminiscent of Huffman sequences, is introduced to study time-dependent auditory cortical processing. Stimuli were nine seconds in duration, and were generated from segments consisting of a sum of sinusoids with randomized amplitude, phase, and frequency components drawn from a Gaussian distribution.

Six conditions were created by choosing Gaussian distributed segment duration means of 12, 25, 45, 85, 160, and 300 ms. Manipulating the starting ( $f_1$ ) and ending ( $f_2$ ) frequency of each segment and the segment 'offset' relative to its neighbor within a half-octave range (1000-1500 Hz), produced three qualitatively different types of stimuli: (i) 'Constant;'  $f_1 = f_2$  with no offset between segments, (ii) 'Tonal;'  $f_1 = f_2$  with Gaussian distributed offsets, and (iii) 'FM;'  $f_1$  swept linearly (and randomly) upwards/downwards to  $f_2$  over the entire half-octave range. All three stimulus types possess the same RMS power, and power spectral density (over the nine-second stimulus duration), permitting analysis of explicitly temporal processing. A single-trial sparse acquisition functional magnetic resonance imaging (fMRI) design was employed.

The results reveal at least two phenomena germane to the present considerations. (1) In both left and right superior temporal areas, there is a striking sensitivity to the temporal structure of the sound, with increasing segment duration associated with increased local activation. This result was observed throughout auditory cortex, including putative human primary auditory cortex in the medial portion of Heschl's gyrus (transverse temporal gyrus; core), a substantive extent of the superior temporal gyrus (belt), and superior temporal sulcus. From the perspective of single-unit physiology, of course, such a result is not at all surprising – i.e. neurons throughout the auditory pathway are highly sensitive to temporal properties of the stimulating signal, say by phase-locking to the envelope of the signal. (2) Sitting on top of this main effect of bilateral sensitivity to timing structure, we observed a remarkably clear interaction between segment duration (or segment SOA) and lateralization in a non-primary area. Specifically, for our stimuli comprised of long duration segments, there was a strong rightward bias in the STS; in contrast, stimuli made from short-duration segments (12ms, 25ms, 45ms) elicited stronger left temporal responses. Put differently, when the modulation spectrum peak is relatively low, right STS is significantly more activated; when the modulation spectrum peak is higher, the response is more bilateral with a tendency to be left lateralized. In summary, there was no hemispheric effect of time-structure in core/primary auditory cortex, and only a mild interaction between hemisphere and timing/segment duration in STG. However, in STS, there was a strong interaction between timing/segment duration such that slower modulations were right lateralized whereas more rapid modulations were bilateral or slightly left lateralized [14].

## 5. The asymmetric sampling in time view

Cumulatively, the psychophysical, patient, electrophysiological, and hemodynamic data we mentioned point to three features: (i) the relevance of temporal information – both in the speech signal and in the neuronal signature – for the successful analysis and representation of speech sounds in cortex, (ii) the relevance of temporal integration constants of two sizes, and (iii) the anatomically bilateral nature of speech sound processing. Here I briefly outline a timing-based hypothesis that captures a range of these effects, the asymmetric sampling in time (AST) hypothesis [8, 10].

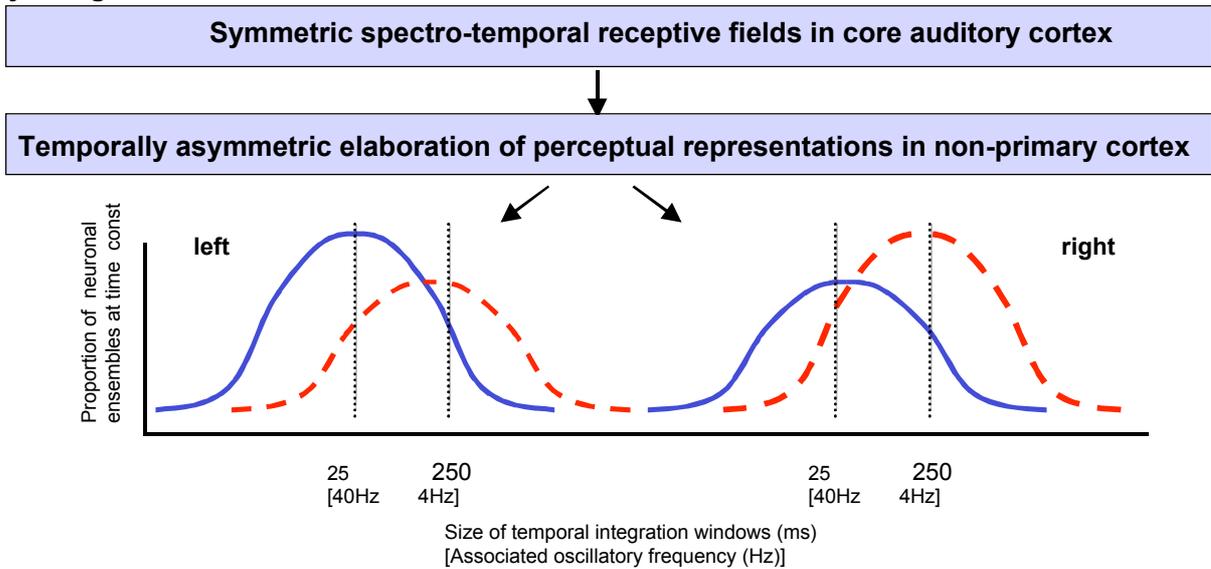
The AST hypothesis (illustrated in Figure 1 below) assumes that the input speech signal has a spectro-

temporal representation that is bilaterally symmetric at the primary/core cortical level. Beyond the initial representation, however, the signal is analysed asymmetrically in the time domain: left non-primary auditory areas preferentially extract information from shorter 20-50ms temporal integration windows; right hemisphere homologues preferentially extract information from longer 150-250ms integration windows. Whereas both hemispheres have available the neuronal machinery to perform analyses on different time scales, the main difference between left and right non-primary auditory areas is in terms of the time constants used to process the input, with right areas having a propensity for longer time constants than left areas.

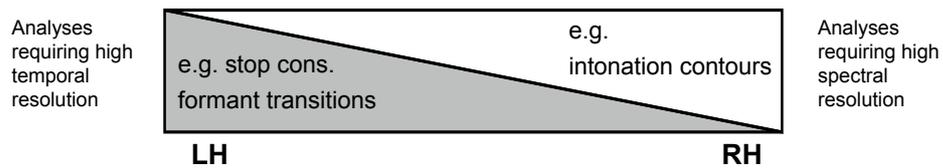
AST exemplifies functional segregation and multi-resolution analysis, processing strategies common in other domains. Figure 1 outlines some of the properties. The same input signal will be subjected to two types of analysis that yield complementary information. Insofar as rapidly changing information is relevant for a given perceptual task, left cortical regions provide the more appropriate substrate; more gradually changing information or information that requires fine-grained spectral distinctions will be predominantly analyzed by right cortical mechanisms. This proposal is very close in spirit to Zatorre et al.'s [15] view that there is a tradeoff between spectral and temporal sensitivity that is reflected in cortical lateralization. The AST proposal differs from their hypothesis in that I assume that both left and right temporal cortices can execute either computation (because neuronal populations with the corresponding time constants exist in both hemispheres). However, I stipulate a small asymmetry in the proportion of neuronal ensembles that exhibit a given time constant – and it is the slight anatomic asymmetry (or rather, the slight asymmetry in proportion of ensembles with certain circuit property) that forms the basis for the timing based processing asymmetries.

A range of predictions can be examined, including: (1) always bilateral activation in natural speech tasks; (2) linguistic and affective prosody at the phrasal level should both be associated with right hemisphere mechanisms; (3) the analysis of stop consonant formant transitions should be left lateralized; (4) suprasegmental phenomena that occur at the level of syllables should be more driven by right hemisphere mechanisms; (5) music perception should lateralize to the right for most musical attributes, including pitch; (6) if temporal integration windows are physiologically reflected as oscillatory brain activity, then the shorter time windows associated with the left hemisphere should yield oscillations in the gamma band, which should have more power in the left hemisphere.

### a. Physiological lateralization



### b. Functional lateralization



## 6. Acknowledgements

Supported by NIH DC 05660 and NIH DC 0463801 to DP. Experiments by Anthony Boemio, Maria Chait, Huan Luo, and Virginie van Wassenhove. Allen Braun, Ken Grant, Steve Greenberg, Greg Hickok, and Jonathan Simon have contributed to many aspects of this work.

## 7. References

- [1] Jezzard, P., Matthews, P., and Smith S. (eds.). *Functional MRI*, Oxford University Press, Oxford, 2001.
- [2] Roberts, T.P.L., Poeppel, D., Rowley, "Magnetoencephalography and magnetic source imaging", *Neuropsych, Neuropsychol, Behav Neurol*, 11: 49-64, 2000.
- [3] Hickok, G. and Poeppel, D., "Towards a new functional anatomy of speech perception", *Trends Cog Sci* 4: 131-139, 2000.
- [4] Scott, S. and Johnsrude, I., "The neuroanatomical and functional organization of speech perception", *Trends Neurosci* 26:100-107, 2003.
- [5] Mummery, C.J., Ashburner, Scott, S., and Wise, "Functional neuroimaging of speech perception in six normal and two aphasic subjects", *J Acoust Soc Am* 106:449-457, 1999.
- [6] Poeppel, D. Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., Braun, A., "Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex", *Neuropsychologia*, in press.
- [7] Buchman A., Garron D., Trost-Cardamone J., Wichter M., & Schwartz M., "Word deafness: one hundred years later", *J Neurol Neurosurg Psychiatry*, 49:489-499, 1986.
- [8] Poeppel, D., "Pure word deafness and the bilateral nature of the speech code", *Cognitive Science* 25:679-693, 2001.
- [9] Grant, K.W., and Greenberg, S., "Speech intelligibility derived from asynchronous processing of auditory-visual information," *Proc. AVSP 2001 International Conference on Auditory-Visual Speech Processing*, Scheelsminde, Denmark, 132-137, 2001.
- [10] Poeppel, D., "The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'", *Speech Communication*, in press.
- [11] McGurk H., McDonald J., "Hearing lips and seeing voices", *Nature* 264: 746-747, 1976.
- [12] van Wassenhove, V., Grant, K., Poeppel, D., Cognitive Neuroscience Society Annual Meeting, San Francisco, 2002.
- [13] Chait, M. Greenberg, S., Arai, T., Poeppel, D., Cognitive Neuroscience Society Annual Meeting, New York, 2003.
- [14] Boemio, A., Fromm, S., Braun, A., Poeppel, D. Cognitive Neuroscience Society Annual Meeting, New York, 2003.
- [15] Zatorre R.J., Belin, P., Penhune, V.B., "Structure and function of auditory cortex: music and speech", *Trends Cogn Sci.* 6:37-46, 2002.