Behavioral/Systems/Cognitive

# Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech

**Daniel A. Abrams,[1] Trent Nicol,[1] Steven Zecker,[1] and Nina Kraus[1,2]**

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences, and [2]Departments of Neurobiology and Physiology, and Otolaryngology, Northwestern University, Evanston, Illinois 60208

Cortical analysis of speech has long been considered the domain of left-hemisphere auditory areas. A recent hypothesis poses that cortical processing of acoustic signals, including speech, is mediated bilaterally based on the component rates inherent to the speech signal. In support of this hypothesis, previous studies have shown that slow temporal features (3–5 Hz) in nonspeech acoustic signals lateralize to right-hemisphere auditory areas, whereas rapid temporal features (20–50 Hz) lateralize to the left hemisphere. These results were obtained using nonspeech stimuli, and it is not known whether right-hemisphere auditory cortex is dominant for coding the slow temporal features in speech known as the speech envelope. Here we show strong right-hemisphere dominance for coding the speech envelope, which represents syllable patterns and is critical for normal speech perception. Right-hemisphere auditory cortex was 100% more accurate in following contours of the speech envelope and had a 33% larger response magnitude while following the envelope compared with the left hemisphere. Asymmetries were evident regardless of the ear of stimulation despite dominance of contralateral connections in ascending auditory pathways. Results provide evidence that the right hemisphere plays a specific and important role in speech processing and support the hypothesis that acoustic processing of speech involves the decomposition of the signal into constituent temporal features by rate-specialized neurons in right- and left-hemisphere auditory cortex.

*Key words:* auditory cortex; cerebral asymmetry; speech; speech envelope; syllable; children

## Introduction

Speech processing, defined as the neural operations responsible for transforming acoustic speech input into linguistic representations, is a well established aspect of human cortical function. Classically, speech processing has been thought to be mediated primarily by left-hemisphere auditory areas of the cerebral cortex (Wernicke, 1874). This view continues to receive wide acceptance based on results from studies investigating the functional neuroanatomy of speech perception. Acoustical processing of speech involves cortical analysis of the physical features of the speech signal, and normal speech perception relies on resolving acoustic events occurring on the order of tens of milliseconds (Phillips and Farmer, 1990; Tallal et al., 1993). Because temporal processing of these rapid acoustic features has been shown to be the domain of left-hemisphere auditory cortex (Belin et al., 1998; Liégeois-Chauvel et al., 1999; Zatorre and Belin, 2001; Zaehle et al., 2004; Meyer et al., 2005), acoustic processing of speech is thought to be predominantly mediated by left-hemisphere auditory structures (Zatorre et al., 2002). Phonological processing of speech, which involves mapping speech sound input to stored phonological

representations, has been shown to involve a network in the superior temporal sulcus (STS) lateralized to the left hemisphere (Scott et al., 2000; Liebenthal et al., 2005; Obleser et al., 2007). Semantic processing of speech, which involves retrieving the appropriate meanings of words, is thought to occur in a network localized to left inferior temporal (Rodd et al., 2005) and frontal (Wagner et al., 2001) gyri.

A recent hypothesis, called the "asymmetric sampling in time" (AST) hypothesis, has challenged the classical model by proposing that acoustical processing of speech occurs bilaterally in auditory cortex based on the component rates inherent to the speech signal (Poeppel, 2003). Acoustic-rate asymmetry is thought to precede language-based asymmetries (i.e., phonological and semantic asymmetries) and is supported by results that show that slow, nonspeech acoustic stimuli (3–5 Hz) are lateralized to right-hemisphere auditory areas (Boemio et al., 2005), whereas rapid acoustic stimuli (20–50 Hz) are lateralized to left-hemisphere auditory areas (Zatorre and Belin, 2001; Zaehle et al., 2004; Schonwiesner et al., 2005).

It is not known to what extent this putative mechanism applies to the slow temporal features in speech, known as the speech envelope (Rosen, 1992). The speech envelope provides syllable pattern information and is considered both sufficient (Shannon et al., 1995) and essential (Drullman et al., 1994) for normal speech perception. A prediction of the AST hypothesis is that slow acoustic features in speech are processed in right-hemisphere auditory areas regardless of left-dominant asymmetries for language processing. To examine this question, we mea-

sured cortical-evoked potentials in 12 normally developing children in response to speech sentence stimuli and compared activation patterns measured over left and right temporal cortices.

## Materials and Methods

The research protocol was approved by the Institutional Review Board of Northwestern University. Parental consent and the child's assent were obtained for all evaluation procedures, and children were paid for their participation in the study.

*Participants.* Participants consisted of 12 children between 9 and 13 years old who reported no history of neurological or otological disease and were of normal intelligence [scores >85 on the Brief Cognitive Scale (Woodcock and Johnson, 1977)]. The reason for having children serve as subjects is that we are ultimately interested in describing auditory deficits in children with a variety of clinical disorders (Koch et al., 1999). A necessary step in describing abnormal auditory function is first describing these processes in normal children, as we have done here. Children were recruited from a database compiled in an ongoing project entitled "Listening, Learning, and the Brain." Children who had previously participated in this project and had indicated interest in participating in additional studies were contacted via telephone. All subjects were tested in one session.

*Stimuli.* Stimuli consisted of the sentence stimulus "The young boy left home," produced in three modes of speech: conversational, clear, and compressed (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). These three modes of speech have different speech envelope cues and were used as a means to elicit a variety of cortical activation patterns. Conversational speech is defined as speech produced in a natural and informal manner. Clear speech is a well described mode of speech resulting from greater diction (Uchanski, 2005). Clear speech is naturally produced by speakers in noisy listening environments and enables greater speech intelligibility relative to conversational speech. There are many acoustic features that are thought to contribute to enhanced perception of clear speech relative to conversational speech, including greater intensity of speech, slower speaking rates, and more pauses. Most importantly, with respect to the current work, an established feature of clear speech is greater temporal envelope modulations at low frequencies of the speech envelope, corresponding to the syllable rate of speech (1–4 Hz) (Krause and Braida, 2004). With respect to the particular stimuli used in the current study, greater amplitude envelope modulations are evident in the clear speech relative to the conversational stimuli. For example, there is no amplitude cue between "The" and "young" (0–450 ms) (supplemental Fig. 1, available at www.jneurosci.org as supplemental material) evident in the broadband conversational stimulus envelope; however, an amplitude cue is present in the broadband clear stimulus envelope. This phenomenon also occurs between the segments "boy" and "left" (450–900 ms) (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). Compressed speech approximates rapidly produced speech and is characterized by a higher-frequency speech envelope. Compressed speech is more difficult to perceive compared with conversational speech (Beasley et al., 1980) and has been used in a previous study investigating cortical phase-locking to the speech envelope (Ahissar et al., 2001).

Conversational and clear sentences were recorded in a soundproof booth by an adult male speaker at a sampling rate of 16 kHz. Conversational and clear speech sentences were equated for overall duration to control for slower speaking rates in clear speech (Uchanski, 2005). This was achieved by compressing the clear sentence by 23% and expanding the conversational sentence by 23%. To generate the compressed sentence stimulus, we doubled the rate of the conversational sample using a signal-processing algorithm in Adobe Audition (Adobe Systems, San Jose, CA). This algorithm does not alter the pitch of the signal. The duration of the clear and conversational speech sentences was 1500 ms, and the duration of the compressed sentence was 750 ms.

*Recording and data processing procedures.* A personal computer-based stimulus delivery system (Neuroscan GenTask) was used to output the

sentence stimuli through a 16-bit converter at a sampling rate of 16 kHz. Speech stimuli were presented unilaterally to the right ear through insert earphones (Etymotic Research ER-2) at 80 dB sound pressure level (SPL). Stimulus presentation was pseudorandomly interleaved. To test ear-of-stimulation effects, three subjects were tested in a subsequent session using unilateral left-ear stimulation. The polarity of each stimulus was reversed for half of the stimulus presentations to avoid stimulus artifacts in the cortical responses. Polarity reversal does not affect perception of speech samples (Sakaguchi et al., 2000). An interval of 1 s separated the presentation of sentence stimuli. Subjects were tested in a sound-treated booth and were instructed to ignore the sentences. To promote subject stillness during long recording sessions as well as diminish attention to the auditory stimuli, subjects watched a videotape movie of his or her choice and listened to the soundtrack to the movie in the nontest ear with the sound level set <40 dB SPL. This paradigm for measuring cortical-evoked potentials has been used in previous studies investigating cortical asymmetry for speech sounds (Bellis et al., 2000; Abrams et al., 2006) as well as other forms of cortical speech processing (Kraus et al., 1996; Banai et al., 2005; Wible et al., 2005). Although it is acknowledged that cortical activity in response to a single stimulus presentation includes contributions from both the experimental speech stimulus and the movie soundtrack, auditory information in the movie soundtrack is highly variable throughout the recording session. Therefore, the averaging of auditory responses across 1000 stimulus presentations, which serves as an essential method for reducing the impact of noise on the desired evoked response, is thought to remove contributions from the movie soundtrack. Cortical responses to speech stimuli were recorded with 31 tin electrodes affixed to an Electrocap International (Eaton, OH) brand cap (impedance, <5 kΩ). Additional electrodes were placed on the earlobes and superior and outer canthus of the left eye. These act as the reference and eye-blink monitor, respectively. Responses were collected at a sampling rate of 500 Hz for a total of 1000 repetitions each for clear, conversational, and compressed sentences.

Processing of the cortical responses consisted of the following steps. First, excessively noisy segments of the continuous file (typically associated with subject movement) were manually rejected. The continuous file was high-pass filtered at 1 Hz, and removal of eye-blink artifacts was accomplished using the spatial filtering algorithm provided by Neuroscan (Compumedics USA, Charlotte, NC). The continuous file was then low-pass filtered at 40 Hz to isolate cortical contributions, and the auditory-evoked potentials were then downsampled to a sampling rate of 200 Hz. All filtering was accomplished using zero phase-shift filters, and downsampling was accompanied by IIR low-pass filtering to correct for aliasing (Compumedics). The goal of this filtering scheme was to match the frequency range of the speech envelope (Rosen, 1992). Responses were artifact rejected at a ±75 μV criterion. Responses were then subjected to noise reduction developed by our laboratory that has been used in improving the signal-to-noise ratio of brainstem and cortical-evoked potentials. The theoretical basis for the noise reduction is that auditory-evoked potentials are essentially invariant across individual stimulus repetitions, whereas the background noise is subject to variance across stimulus repetitions. Thus, the mean evoked response is significantly diminished by the fraction of repetitions that least resembles it. If these noisy responses are removed, the signal-to-noise ratio of the cortical response improves considerably with virtually no change to morphology of the average waveform. The algorithm calculated the average response from all 1000 sweeps for each stimulus condition at each electrode then performed Pearson's correlations between each of the 1000 individual stimulus repetitions and the average response. The 30% of repetitions with the lowest Pearson's correlations from each stimulus condition were removed from subsequent analyses, and the remaining repetitions were averaged and rereferenced to a common reference computed across all electrodes. Therefore, following the noise reduction protocol, cortical responses from each subject represent the average of ~700 repetitions of each stimulus. Data processing resulted in an averaged response for 31 electrode sites and three stimulus conditions measured in all 12 subjects.

*Data analysis.* All data analyses were performed using software written in Matlab (The Mathworks, Natick, MA). Broadband amplitude envelopes were determined by performing a Hilbert transform on the broad-

band stimulus waveforms (Drullman et al., 1994). The unfiltered amplitude envelope was low-pass filtered at 40 Hz to isolate the speech envelope (Rosen, 1992) and match the frequency characteristics of the cortical responses; the envelopes were then resampled to 200 Hz. Data are presented for three temporal electrode pairs [(1) T3–T4; (2) T5–T6; and (3) Tp7–Tp8] according to the modified International 10-20 recording system (Jasper, 1958). The modification is the addition of the Tp7–Tp8 electrode pair in which Tp7 is located midway between T3 and T5 and Tp8 is located midway between T4 and T6.

Two types of analyses were performed on the data: cross-correlation and root mean square (RMS) analysis. First, cross-correlations between the broadband speech envelope and cortical responses at each temporal electrode for the "envelope-following period" (250–1500 ms for conversational and clear stimuli, 250–750 ms for the compressed stimulus) were performed using the "xcov" function in Matlab. The peak in the cross-correlation function was found at each electrode between 50 and 150 ms lags, and the *r* value and lag at each peak were recorded. The *r* values were Fisher transformed before statistical analysis. RMS amplitudes at each electrode were calculated for two different time ranges: the "onset" period was defined by the time range 0–250 ms for all stimuli; the envelope-following period was defined as 250–1500 ms for conversational and clear stimuli and 250–750 ms for the compressed stimulus.

*Statistical analysis.* The statistical design used a series of three completely "within-subjects," repeated-measures ANOVAs (RMANOVAs) to assess hemispheric effects for cross-correlation and RMS measures. A primary goal of this work was to describe patterns of cortical asymmetry across speech conditions, and because $2 \times 3 \times 3$ (hemisphere × electrode pair × stimulus condition) RMANOVAs indicated no interactions involving stimulus condition, the subsequent analysis collapsed across stimulus condition and was performed as $2 \times 3$ (hemisphere × electrode pair) RMANOVAs. This enabled a matched statistical comparison of each electrode pair (i.e., T3 vs T4, T5 vs T6, Tp7 vs Tp8) for each subject across stimulus conditions. A $2 \times 3 \times 2$ (hemisphere × electrode pair × stimulation ear) RMANOVA was used to assess whether asymmetry effects seen in the cross-correlation and RMS analyses affected stimulation ear. Paired, Bonferroni-corrected *t* tests (two-tailed) comparing matched electrode pairs (i.e., T3 vs T4, T5 vs T6, Tp7 vs Tp8) were used for all *post hoc* analyses. RMANOVA *p* values <0.05 and paired *t* test *p* values <0.01 were considered statistically significant.

## Results

Inspection of raw cortical responses measured at the six temporal lobe electrodes to the speech sentence stimuli revealed two discrete components in all temporal lobe electrodes: (1) a large negative onset peak and (2) a series of positive peaks that appeared to closely follow the temporal envelope of the stimulus. We called the former component the onset and the latter component the envelope-following portion of the response (see Fig. 1 for clear speech stimulus; see supplemental Figs. 2, 3, available at www.j-neurosci.org as supplemental material, for conversational and compressed conditions, respectively). Both speech onset (Warrier et al., 2004) and envelope-following (Ahissar et al., 2001)
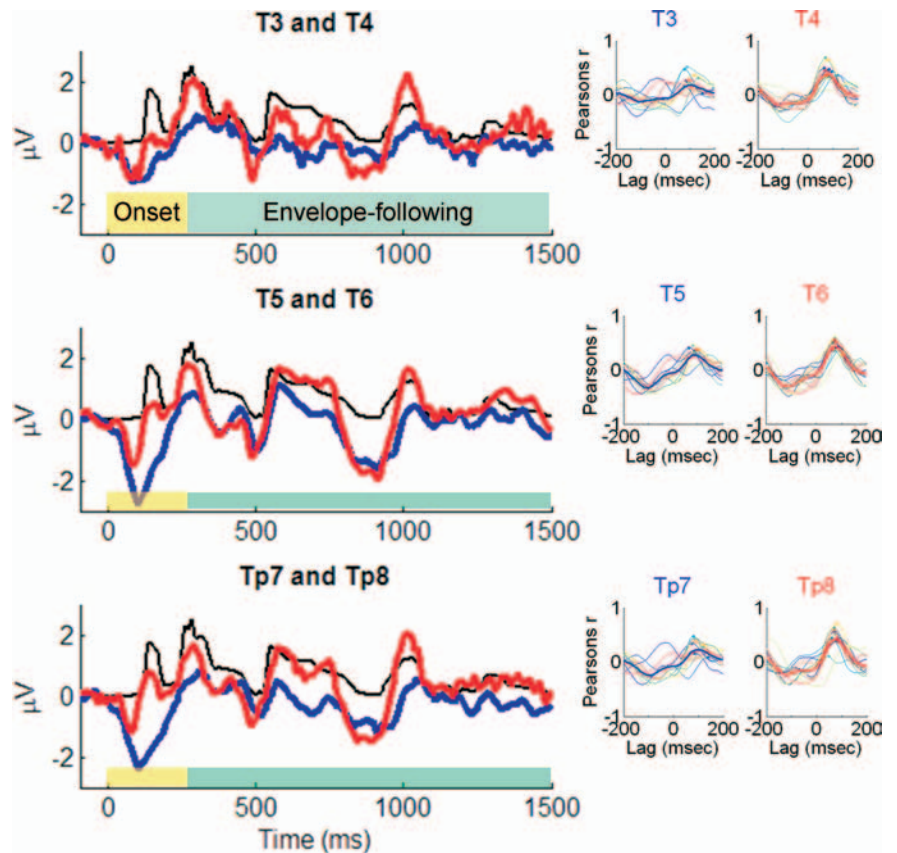


**Figure 1.** Left column, Grand average cortical responses from three matched electrode pairs and broadband speech envelope for "clear" stimulus condition. The black lines represent the broadband speech envelope for the clear speech condition, the red lines represent cortical activity measured at right-hemisphere electrodes, and the blue lines represent activity from left-hemisphere electrodes. Ninety-five milliseconds of the prestimulus period is plotted. The speech envelope was shifted forward in time 85 ms to enable comparison to cortical responses; this time shift is for display purposes only. Right column, Cross-correlograms between clear speech envelope and individual subjects' cortical responses for each electrode pair. A small dot appears at the point chosen for subsequent stimulus-to-response correlation analyses.

components have been demonstrated in previous studies of human auditory cortex; this latter study called this phenomenon speech envelope "phase-locking," and the same nomenclature will be used here.

Grand average cortical responses from three matched electrode pairs (Fig. 1, left column) and individual subject cross-correlograms (Fig. 1, right column) indicated a number of relevant features. First, a moderate linear relationship was indicated between the broadband temporal envelope of the stimulus and raw cortical responses for all temporal lobe electrodes measured across all subjects (mean peak correlation, 0.37; SD, 0.09). Second, this peak correlation occurred in the latency range of well established, obligatory cortical potentials measured from children of this age range (Tonnquist-Uhlen et al., 2003) (mean lag, 89.1 ms; SD, 7.42 ms). Cortical potentials in this time range, measured from temporal lobe electrodes, are associated with activity originating in secondary auditory cortex (Scherg and Von Cramon, 1986; Ponton et al., 2002). Third, and most importantly, there appeared to be qualitative differences between cortical responses from right-hemisphere electrodes compared with matched electrodes of the left hemisphere. Specifically, right-hemisphere cortical responses appeared to conform to the contours of the stimulus envelope in greater detail than left-hemisphere responses. This was further evidenced in the correlograms, which had more consistent and sharper peaks, as
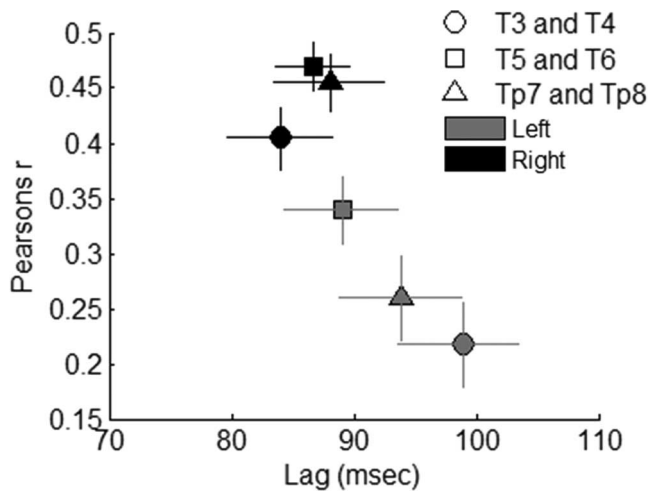
**Figure 2.** Average cross-correlogram peaks. Values represent the average peak lag and *r* value, collapsed across stimulus conditions, for each stimulus envelope–cortical response correlation at the three electrode pairs. Right-hemisphere electrodes are black, and left-hemisphere electrodes are gray. Error bars represent 1 SEM.



**Figure 3.** Average RMS amplitudes for envelope-following and onset (inset) periods. The onset period was defined as 0–250 ms of the cortical response, and the envelope-following period was defined as 250–1500 (clear and conversational conditions) or 250–750 ms (compressed condition). Right-hemisphere electrodes are black, and left-hemisphere electrodes are gray. Error bars represent 1 SEM.

well as larger overall correlations, in right-hemisphere electrodes. These particular characteristics would suggest better right-hemisphere phase-locking to the speech envelope.

**Speech envelope phase-locking analysis**
To quantify temporal envelope phase-locking, we identified the maximum in correlograms (Fig. 1, right column) for lags between 50 and 150 ms for all stimulus conditions. This time range was selected because previous studies have shown that cortical synchronization to the temporal structure of brief speech sounds occurs in this range (Sharma and Dorman, 2000), and most correlograms in the current data set indicated a positive peak in this time range. An initial $2 \times 3 \times 3$ RMANOVA (hemisphere $\times$ electrode pair $\times$ stimulus condition) indicated differences in phase-locking across stimulus conditions (main effect of stimulus condition: $F_{(2,22)} = 19.327$; $p < 0.0001$), which was expected given significant acoustical differences between the stimuli (see Materials and Methods); however, the pattern of asymmetry for cortical phase-locking was similar for the three stimulus conditions (hemisphere $\times$ stimulus condition interaction: $F_{(2,22)} < 1$; $p > 0.7$). Based on this result, and our interest in describing patterns of cortical asymmetry across speech conditions, we collapsed all additional statistical analyses on correlation *r* values across the three stimulus conditions. A $2 \times 3$ RMANOVA (hemisphere $\times$ electrode pair) statistical analysis on peak correlation values revealed a significant main effect of hemisphere ($F_{(1,35)} = 21.125$; $p < 0.0001$). All three of these electrode pairs showed this hemispheric effect (left vs right electrode; paired *t* tests: $t_{(35)} > 3.70$; $p \leq 0.001$ for all three pairs) (Fig. 2), and there was no statistical difference in the degree of asymmetry between electrode pairs (RMANOVA hemisphere $\times$ electrode interaction: $F_{(2,22)} = 1.206$; $p > 0.3$). To ensure that these results were not biased by our definition of the time frame of the envelope-following component of the response, we performed identical analyses on the entire response, including the onset component, and the results were the same [0–1500 ms for conversational and clear stimuli, 0–750 ms for compressed stimulus; $2 \times 3$ RMANOVA (hemisphere $\times$ electrode pair); main effect of hemisphere: $F_{(1,35)} = 10.658$, $p = 0.002$]. These data indicate that all three temporal electrode pairs showed a significant and similar
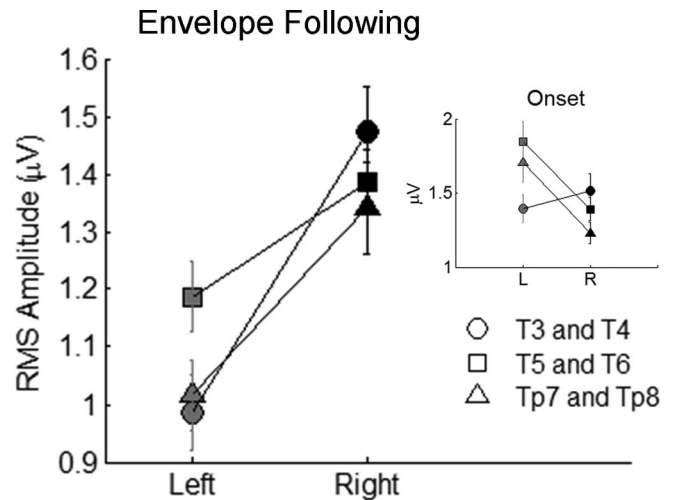
pattern of right-hemisphere asymmetry for speech envelope phase-locking.

**Response magnitude analysis: onset and envelope-following period**
In addition to asymmetry for phase-locking, inspection of the raw cortical data also revealed an interesting pattern of response amplitudes in the onset and envelope-following response components. At stimulus onset, response amplitudes appear to be consistently greater in left-hemisphere electrodes, particularly in T5–T6 and Tp7–Tp8 electrode pairs. Given that subjects received stimulation in their right ear, this finding was anticipated based on the relative strength of contralateral connections in the ascending auditory system (Kaas and Hackett, 2000). Surprisingly, during the envelope-following period of the response, right-hemisphere responses appeared to be larger than the left for all electrode pairs.

We quantified this phenomenon by calculating RMS amplitude over the onset and envelope-following periods for all stimulus conditions (Fig. 3). First, we performed a $2 \times 3 \times 3$ RMANOVA (hemisphere $\times$ electrode pair $\times$ stimulus condition) on onset RMS values, which revealed that stimulus condition did not affect asymmetry for RMS onset (hemisphere $\times$ stimulus condition interaction: $F_{(2,22)} = 1.398$; $p > 0.25$); this result enabled us to collapse all additional statistical analyses on onset RMS across the three stimulus conditions. Results from the $2 \times 3$ RMANOVA (hemisphere $\times$ electrode pair) indicated that left-hemisphere responses were significantly larger than the right over the onset period (main effect of hemisphere: $F_{(1,35)} = 4.686$; $p = 0.037$), and there were differences in this pattern of onset asymmetry across the three electrode pairs (hemisphere $\times$ electrode pair interaction: $F_{(2,70)} = 14.805$; $p < 0.001$). *Post hoc t* tests indicated that the main effect of hemisphere for onset RMS was driven by the posterior electrode pairs, whereas the anterior pair, T3–T4, did not contribute to this effect (paired *t* tests: T3 and T4, $t_{(35)} = 0.924$, $p > 0.35$; T5 and T6, $t_{(35)} = 2.892$, $p = 0.007$; Tp7 and Tp8, $t_{(35)} = 3.348$, $p = 0.002$).

For the envelope-following period, a $2 \times 3 \times 3$ RMANOVA (hemisphere $\times$ electrode pair $\times$ stimulus condition) was per-

formed on envelope-following RMS values. Results again revealed that stimulus condition did not affect asymmetry (hemisphere × stimulus condition interaction: $F_{(2,22)} = 2.244$; $p > 0.10$), enabling us to collapse all additional statistical analyses on envelope-following RMS across the three stimulus conditions. Results from the 2 × 3 RMANOVA (hemisphere × electrode pair) for the envelope-following RMS indicated that right-hemisphere responses were significantly larger than the left at all three electrode pairs [2 × 3 RMANOVA (hemisphere × electrode pair); main effect of hemisphere: $F_{(1,35)} = 32.768$, $p < 0.00001$; paired $t$ tests: T3 and T4, $t_{(35)} = 5.565$, $p < 0.00001$; T5 and T6, $t_{(35)} = 3.385$, $p = 0.002$; Tp7 and Tp8, $t_{(35)} = 4.767$, $p < 0.0001$]. These data indicate that the right hemisphere has significantly larger response amplitudes during the envelope-following period despite being ipsilateral to the side of acoustic stimulation.

### Individual subject analysis

To quantify phase-locking and RMS amplitude asymmetries within individual subjects, we entered $r$ values from the cross-correlation analysis and RMS amplitudes from the envelope-following period, respectively, into the asymmetry index $(R - L)/(R + L)$ using matched electrode pairs (T3 and T4, T5 and T6, and Tp7 and Tp8). Using this index, values approaching 1 indicate a strong rightward asymmetry, values approaching $-1$ indicate a strong leftward asymmetry, and a value of 0 indicates symmetry. Results from this analysis indicate that greater right-hemisphere phase-locking, defined as asymmetry values $>0$, occurred in 78% of the samples (binomial test: $z = 5.96$, $p < 0.0001$) and right-hemisphere $r$ values were more than twice as great as those seen for the left hemisphere (mean asymmetry index, 0.35). For RMS amplitude, 82% of the samples indicated greater envelope-following amplitude in the right hemisphere (binomial test: $z = 6.74$, $p < 0.0001$), and right-hemisphere amplitudes were ~33% greater than those seen in the left hemisphere (mean asymmetry index, 0.14) during the envelope-following period.

### Ear-of-stimulation analysis

To ensure that the right-hemisphere asymmetries for envelope phase-locking and RMS amplitude were not driven by the use of right-ear stimulation, we measured cortical responses to the speech sentences in three of the subjects using left-ear stimulation, which again enabled a completely within-subjects statistical analysis. Results indicate that when subjects were stimulated in their left ear, envelope phase-locking was again greater in the right hemisphere [2 × 3 RMANOVA (hemisphere × electrode pair); main effect of hemisphere: $F_{(1,8)} = 15.532$, $p = 0.004$]. Moreover, when compared directly to responses elicited by right-ear stimulation, envelope phase-locking asymmetries were statistically similar regardless of the ear of stimulation (Fig. 4) [2 × 3 × 2 RMANOVA (hemisphere × electrode pair × stimulation ear); interaction (hemisphere × stimulation ear): $F_{(1,8)} = 0.417$, $p > 0.5$]. For the RMS analysis, left-ear stimulation resulted in larger onset responses in the right hemisphere, again consistent with contralateral dominance for onsets (Fig. 5, inset) [2 × 3 RMANOVA (hemisphere × electrode pair); main effect of hemisphere: $F_{(1,8)} = 6.40$, $p = 0.035$]. In addition, the asymmetry pattern for onset RMS with left-ear stimulation was statistically different from the pattern seen for right-ear stimulation [2 × 3 × 2 RMANOVA (hemisphere × electrode pair × stimulation ear); interaction of (hemisphere × stimulation ear): $F_{(1,8)} = 24.390$, $p = 0.001$]. Importantly, the RMS of the envelope-following period remained greater in the right hemisphere with left-ear stim-
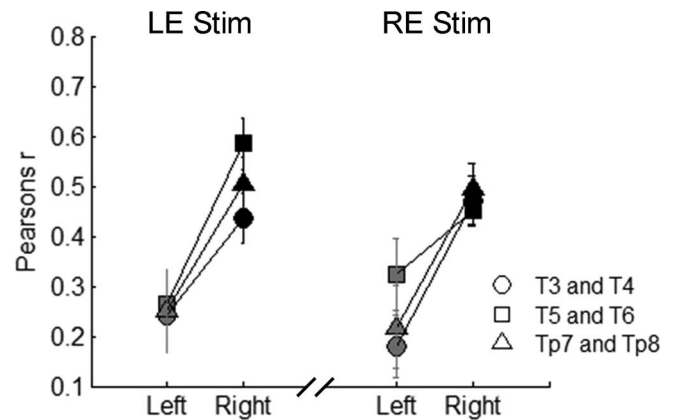
**Figure 4.** Left-ear versus right-ear stimulation comparison: speech envelope phase-locking. Right-hemisphere electrodes are black, and left-hemisphere electrodes are gray. Error bars represent 1 SEM. LE Stim, Left-ear stimulation; RE Stim, right-ear stimulation.
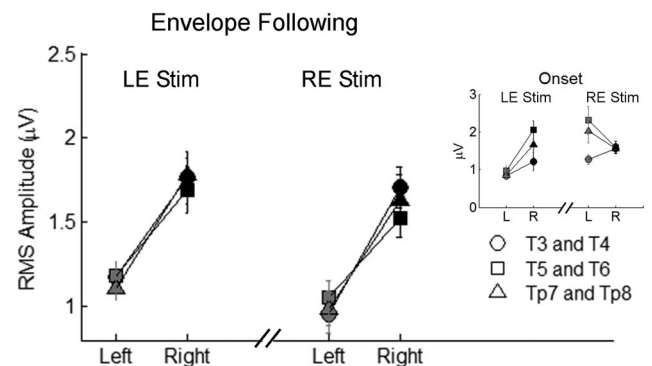
**Figure 5.** Left-ear versus right-ear stimulation comparison: RMS amplitude of the envelope-following period. Right-hemisphere electrodes are black, and left-hemisphere electrodes are gray. Error bars represent 1 SEM. Inset, RMS comparison of the onset period. LE Stim, Left-ear stimulation; RE Stim, right-ear stimulation.

ulation (Fig. 5) [2 × 3 RMANOVA (hemisphere × electrode pair); main effect of hemisphere: $F_{(1,8)} = 36.028$, $p < 0.001$] and was statistically similar to the pattern of asymmetry resulting from right-ear stimulation [2 × 3 × 2 RMANOVA (hemisphere × electrode pair × stimulation ear); interaction (hemisphere × stimulation ear): $F_{(1,8)} = 0.047$, $p > 0.8$]. Together, these data indicate that changing the ear of stimulation from right to left does not affect right-hemisphere asymmetry for envelope phase-locking or envelope RMS amplitude. In contrast, onset RMS amplitudes are always larger in the hemisphere contralateral to the ear of stimulation.

### Discussion

Biologically significant acoustic signals contain information on a number of different time scales. The current study investigates a proposed mechanism for how the human auditory system concurrently resolves these disparate temporal components. Results indicate right-hemisphere dominance for coding the slow temporal information in speech known as the speech envelope. This form of asymmetry is thought to reflect acoustic processing of the speech signal and was evident despite well known leftward asymmetries for processing linguistic elements of speech. Furthermore, rightward asymmetry for the speech envelope was unaffected by the ear of stimulation despite the dominance of contralateral connections in ascending auditory pathways.

## Models of speech perception and the AST hypothesis

The neurobiological foundation of language has been a subject of great interest for well over a century (Wernicke, 1874). Recent studies using functional imaging techniques have enabled a detailed description of the functional neuroanatomy of spoken language. The accumulated results have yielded hierarchical models of speech perception consisting of a number of discrete processing stages, including acoustic, phonological, and semantic processing of speech (Hickok and Poeppel, 2007; Obleser et al., 2007).

It is generally accepted that each of these processing stages is dominated by left-hemisphere auditory and language areas. The acoustic basis of speech perception is typically investigated by measuring cortical activity in response to speech-like acoustic stimuli that have no linguistic value but contain acoustic features that are necessary for normal speech discrimination. Acoustic features lateralized to left-hemisphere auditory areas include rapid frequency transitions (Belin et al., 1998; Joanisse and Gati, 2003; Meyer et al., 2005) and voice-onset time (Liégeois-Chauvel et al., 1999; Zaehle et al., 2004), both of which are necessary for discriminating many phonetic categories. The cortical basis for phonological processing of speech has been investigated by measuring neural activation in response to speech phoneme (Obleser et al., 2007), syllable (Liebenthal et al., 2005; Abrams et al., 2006), word (Binder et al., 2000), and sentence (Scott et al., 2000; Narain et al., 2003) stimuli while carefully controlling for the spectrotemporal acoustic characteristics of the speech signal. Results from these studies have consistently demonstrated that a region of the left-hemisphere STS underlies phonological processing of speech. Studies of cortical processing of semantic aspects of speech have measured brain activation while the subject performed a task in which semantic retrieval demands were varied. Results from these studies have shown that activation of inferior temporal (Rodd et al., 2005) and frontal (Wagner et al., 2001) gyri, again biased to the left hemisphere, underlie semantic processing. It should be noted that right-hemisphere areas are also activated in studies of acoustical, phonological, and semantic speech processing; however, left-hemisphere cortical structures have typically shown dominant activation patterns across studies.

Results from the current study are among the first to show that the right hemisphere of cerebral cortex is dominant during speech processing. These data contradict the conventional thinking that language processing consists of neural operations primarily confined to the left hemisphere of the cerebral cortex. Moreover, results from the current study show right-dominant asymmetry for the speech envelope despite these other well established forms of leftward asymmetry.

Results add to the literature describing hierarchical models of speech processing by providing important details about the initial stage of cortical speech processing: prelinguistic, acoustic processing of speech input. Results support the notion that the anatomical basis of speech perception is initially governed by the component rates present in the speech signal. This statement raises a number of interesting question regarding hierarchical models of speech perception. What is the next stage of processing for syllable pattern information in right-hemisphere auditory areas? Does slow temporal information in speech follow a parallel processing route relative to phonological processing? It is hoped that these questions will receive additional consideration and investigation.

Right-hemisphere dominance for slow temporal features in speech supports the AST hypothesis, which states that slow temporal features in acoustic signals lateralize to right-hemisphere auditory areas, whereas rapid temporal features lateralize to the left (Poeppel, 2003). Results extend the AST hypothesis by providing a new layer of detail regarding the nature of this asymmetric processing. Beyond showing asymmetry for the magnitude of neural activation (RMS amplitude results) (Fig. 3), which might have been predicted from previous studies, our results show that right-hemisphere auditory neurons follow the contours of the speech envelope with greater precision compared with the left hemisphere (Fig. 2). This is an important consideration, because this characteristic of right-hemisphere neurons had not been proposed in previous work and could represent an important cortical mechanism for speech envelope coding.

An influential hypothesis that predates AST states that there is a relative trade-off in auditory cortex for representing spectral and temporal information in complex acoustic signals such as speech and music (Zatorre et al., 2002). It is proposed that temporal resolution is superior in left-hemisphere auditory cortex at the expense of fine-grained spectral processing, whereas the superior spectral resolution of the right hemisphere is accompanied by reduced temporal resolution. The current results suggest that there is, in fact, excellent temporal resolution in the right hemisphere, but it is limited to a narrow range of low frequencies. However, it is not known to what extent the asymmetries demonstrated here might reflect the preference of the right hemisphere for spectral processing.

## Previous studies investigating envelope representations

Previous studies of the human auditory system have described cortical encoding of slow temporal information in speech. In one study, it was shown that cortical phase-locking and frequency matching to the speech envelope predicted speech comprehension using a set of compressed sentence stimuli (Ahissar et al., 2001). There are a few important differences between the current work and that of Ahissar et al. (2001). First, hemispheric specialization was not reported in their work. Second, the analyses (i.e., phase-locking, frequency matching) were conducted on the average of multiple speech sentences with similar envelope patterns, which was necessary given the parameters of the simultaneous speech comprehension task. In contrast, cortical responses in the current study represent activity measured to isolated sentence stimuli and enable a more detailed view of cortical following to individual sentences (Fig. 1 and supplemental Figs. 2, 3, available at www.jneurosci.org as supplemental material).

The current results also show similarities to findings from a recent study that investigated rate processing in human auditory cortex in response to speech (Luo and Poeppel, 2007). In this study, it was shown that different speech sentence stimuli elicited cortical activity with different phase patterns in the theta band (4–8 Hz) and theta-band dissimilarity was lateralized to the right hemisphere. A limitation of this work is that cortical responses were not compared with the stimulus; the analysis only compared cortical responses elicited by the various speech stimuli. Therefore, it was not transparent that the theta-band activity was driven by phase-locking to the speech envelope. Although many of the conclusions are the same as those described here, to our knowledge, our experiment is the first to explicitly show right-hemisphere dominance for phase-locking to the speech envelope.

Single-unit studies of auditory cortex in animal models suggest potential mechanisms underlying right-hemisphere dominance for coding the speech envelope. Across a variety of animal models, a sizable population of auditory cortical neurons is synchronized to the temporal envelope of species-specific calls (Wang et al., 1995; Nagarajan et al., 2002; Gourevitch and Egger-

mont, 2007), which show many structural similarities to human speech; one such study called these neurons "envelope peak-tracking units" (Gehr et al., 2000). One possible explanation for right-dominant asymmetry for envelope phase-locking is that a disproportionate number of envelope peak-tracking units exist in the right-hemisphere auditory cortex of humans. Additional studies with near-field recordings in humans may be able to address this question.

A potential limitation of this work is that children served as subjects, and it is not known whether right-hemisphere speech envelope effects also occur in adults. Whereas the current data cannot discount this possibility, we believe this is unlikely based on the fact that adults show cortical phase-locking to the speech envelope (Ahissar et al., 2001) and have previously demonstrated a right-hemisphere preference for slow, nonspeech acoustic stimuli (Boemio et al., 2005). An interesting possibility is that children have pronounced syllable-level processing relative to adults, reflecting a stage in language acquisition. Future studies may be able be better delineate the generality of this hemispheric asymmetry as well as possible interactions with language development in normal and clinical populations.

Across languages, the syllable is considered a fundamental unit of spoken language (Gleason, 1961), although there is debate as to its phonetic definition (Ladefoged, 2001). The speech envelope provides essential acoustic information regarding syllable patterns in speech (Rosen, 1992), and psychophysical studies of the speech envelope have demonstrated that it is an essential acoustic feature for speech intelligibility (Drullman et al., 1994). Results described here provide evidence that a cortical mechanism for processing syllable patterns in ongoing speech is the routing of speech envelope cues to right-hemisphere auditory cortex. Given the universality of the syllable as an essential linguistic unit and the biological significance of the speech signal, it is plausible that discrete neural mechanisms, such as those described here, may have evolved to code this temporal feature in the human central auditory system.

## References

Abrams DA, Nicol T, Zecker SG, Kraus N (2006) Auditory brainstem timing predicts cerebral asymmetry for speech. J Neurosci 26:11131–11137.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci USA 98:13367–13372.

Banai K, Nicol T, Zecker SG, Kraus N (2005) Brainstem timing: implications for cortical processing and literacy. J Neurosci 25:9850–9857.

Beasley DS, Bratt GW, Rintelmann WF (1980) Intelligibility of time-compressed sentential stimuli. J Speech Hear Res 23:722–731.

Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y (1998) Lateralization of speech and auditory temporal processing. J Cogn Neurosci 10:536–540.

Bellis TJ, Nicol T, Kraus N (2000) Aging affects hemispheric asymmetry in the neural representation of speech sounds. J Neurosci 20:791–797.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528.

Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395.

Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053–1064.

Gehr DD, Komiya H, Eggermont JJ (2000) Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. Hear Res 150:27–42.

Gleason HA (1961) An introduction to descriptive linguistics, revised edition. New York: Holt, Rinehart and Winston.

Gourevitch B, Eggermont JJ (2007) Spatial representation of neural responses to natural and altered conspecific vocalizations in cat auditory cortex. J Neurophysiol 97:144–158.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Jasper HH (1958) The ten-twenty electrode system of the international federation. Electroencephalogr Clin Neurophysiol 10:371–375.

Joanisse MF, Gati JS (2003) Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. NeuroImage 19:64–79.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci USA 97:11793–11799.

Koch DB, McGee TJ, Bradlow AR, Kraus N (1999) Acoustic-phonetic approach toward understanding neural processes and speech perception. J Am Acad Audiol 10:304–318.

Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB (1996) Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science 273:971–973.

Krause JC, Braida LD (2004) Acoustic properties of naturally produced clear speech at normal speaking rates. J Acoust Soc Am 115:362–378.

Ladefoged P (2001) A course in phonetics, Ed 4. Fort Worth, TX: Harcourt College Publishers.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631.

Liégeois-Chauvel C, de Graaf JB, Laguitton V, Chauvel P (1999) Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb Cortex 9:484–496.

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010.

Meyer M, Zaehle T, Gountouna VE, Barron A, Jancke L, Turk A (2005) Spectro-temporal processing during speech perception involves left posterior auditory cortex. NeuroReport 16:1985–1989.

Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. J Neurophysiol 87:1723–1737.

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362–1368.

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple stages of auditory speech perception reflected in event-related FMRI. Cereb Cortex 17:2251–2257.

Phillips DP, Farmer ME (1990) Acquired word deafness, and the temporal grain of sound representation in the primary auditory cortex. Behav Brain Res 40:85–94.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time." Speech Commun 41:245–255.

Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M (2002) Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. Clin Neurophysiol 113:407–420.

Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex 15:1261–1269.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373.

Sakaguchi S, Arai T, Murahara Y (2000) The effect of polarity inversion of speech on human perception and data hiding as an application. Int Conf Acoust Speech Signal Process 2:917–920.

Scherg M, Von Cramon D (1986) Evoked dipole source potentials of the human auditory cortex. Electroencephalogr Clin Neurophysiol 65:344–360.

Schonwiesner M, Rubsamen R, von Cramon DY (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur J Neurosci 22:1521–1528.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Sharma A, Dorman M (2000) Neurophysiologic correlates of cross-language phonetic perception. J Acoust Soc Am 107:2697–2703.

Tallal P, Miller S, Fitch RH (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann NY Acad Sci 682:27–47.

Tonnquist-Uhlen I, Ponton CW, Eggermont JJ, Kwong B, Don M (2003)

Maturation of human central auditory system activity: the T-complex. Clin Neurophysiol 114:685–701.

Uchanski RM (2005) Clear speech. In: Handbook of speech perception (Pisoni DB, Remez RE, eds), pp 207–235. Malden, MA: Blackwell Publishers.

Wagner AD, Pare-Blagoev EJ, Clark J, Poldrack RA (2001) Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron 31:329–338.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685–2706.

Warrier CM, Johnson KL, Hayes EA, Nicol T, Kraus N (2004) Learning impaired children exhibit timing deficits and training-related improvements in auditory cortical responses to speech in noise. Exp Brain Res 157:431–441.

Wernicke C (1874) Der Aphasische Symptomencomplex. Breslau, Poland: Cohn and Weigert.

Wible B, Nicol T, Kraus N (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 128:417–423.

Woodcock R, Johnson M (1977) Woodcock-Johnson psychoeducational battery: tests of cognitive ability. Allen, TX: DLM Teaching Resources.

Zaehle T, Wustenberg T, Meyer M, Jancke L (2004) Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur J Neurosci 20:2447–2456.

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946–953.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46.