

Neural Dynamics of Phonological Processing in the Dorsal Auditory Stream

Einat Liebenthal,^{1,2,3,4} Merav Sabri,¹ Scott A. Beardsley,^{2,3} Jain Mangalathu-Arumana,^{1,3} and Anjali Desai¹

¹Department of Neurology, and ²Clinical Translational Science Institute, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, ³Department of Biomedical Engineering, Marquette University, Milwaukee, Wisconsin 53201, and ⁴Department of Psychiatry, Brigham & Women's Hospital, Chestnut Hill, Massachusetts 02467

Neuroanatomical models hypothesize a role for the dorsal auditory pathway in phonological processing as a feedforward efferent system (Davis and Johnsrude, 2007; Rauschecker and Scott, 2009; Hickok et al., 2011). But the functional organization of the pathway, in terms of time course of interactions between auditory, somatosensory, and motor regions, and the hemispheric lateralization pattern is largely unknown. Here, ambiguous duplex syllables, with elements presented dichotically at varying interaural asynchronies, were used to parametrically modulate phonological processing and associated neural activity in the human dorsal auditory stream. Subjects performed syllable and chirp identification tasks, while event-related potentials and functional magnetic resonance images were concurrently collected. Joint independent component analysis was applied to fuse the neuroimaging data and study the neural dynamics of brain regions involved in phonological processing with high spatiotemporal resolution. Results revealed a highly interactive neural network associated with phonological processing, composed of functional fields in posterior temporal gyrus (pSTG), inferior parietal lobule (IPL), and ventral central sulcus (vCS) that were engaged early and almost simultaneously (at 80–100 ms), consistent with a direct influence of articulatory somatomotor areas on phonemic perception. Left hemispheric lateralization was observed 250 ms earlier in IPL and vCS than pSTG, suggesting that functional specialization of somatomotor (and not auditory) areas determined lateralization in the dorsal auditory pathway. The temporal dynamics of the dorsal auditory pathway described here offer a new understanding of its functional organization and demonstrate that temporal information is essential to resolve neural circuits underlying complex behaviors.

Introduction

Under adverse listening conditions when speech is distorted, noisy, or delivered with a foreign accent, phonemic perception is effortful and facilitated by phonological processing. Phonological processing consists of short-term maintenance of sound sequences in auditory memory during analysis of their auditory, somatosensory, and motor properties to support phonemic categorization (Wise et al., 2001; Buchsbaum et al., 2005; Hickok and Poeppel, 2007). Categorization of ambiguous syllables engages a dorsal pathway, from primary auditory cortex to posterior temporal gyrus (pSTG) and ventral parietal regions, associated with auditory short-term memory and interaction with somatosensory and motor areas (Callan et al., 2004; Golestani and Zatorre, 2004; Dehaene-Lambertz et al., 2005; Desai et al., 2008; Liebenthal et al., 2010; Kilian-Hütten et al., 2011a).

The functional organization of auditory regions in the dorsal pathway during phonological processing, particularly the time course of interactions with somatosensory and motor regions and the hemispheric lateralization pattern, is largely unknown. The notion of a simple hierarchical organization in posterior temporal cortex has been challenged by findings of phonemic neural representations not only in left pSTG (Dehaene-Lambertz et al., 2005; Chang et al., 2010; Liebenthal et al., 2010; Kilian-Hütten et al., 2011a), but also near the auditory core (Kilian-Hütten et al., 2011a), though the relative timing of phonemic activity in these areas is unknown. Interhemispheric differences in sensitivity to temporal and spectral sound properties (Zatorre and Belin, 2001; Poeppel, 2003; Boemio et al., 2005), or in resting-oscillatory properties of neurons (Giraud et al., 2007), have been suggested to predispose the left auditory cortex for processing of fast spectral transitions characteristic of phonemes. In the ventral auditory pathway, the first left-lateralized speech-specific stage of processing may be when phonemic representations in middle STG are accessed (Liebenthal et al., 2005; Obleser et al., 2007). However, it has been argued that phonological processing in pSTG is bilateral (Hickok and Poeppel, 2007).

Here, ambiguous duplex syllables, with elements presented dichotically at varying interaural asynchronies [stimulus-onset asynchronies (SOAs)], were used to parametrically modulate phonological processing and associated neural activity in the dorsal auditory pathway. Subjects performed a syllable and a chirp identification task with identical stimuli, while event-related pos-

Received April 9, 2013; revised July 29, 2013; accepted Aug. 1, 2013.

Author contributions: E.L. designed research; M.S. performed research; M.S., S.A.B., J.M.-A., and A.D. analyzed data; E.L. wrote the paper.

The authors declare no competing financial interests.

This research was supported by National Institute on Deafness and other Communication Disorders Grant R01 DC006287 (E.L.) and the Medical College of Wisconsin Clinical Translational Science Institute (E.L., S.A.B.).

Correspondence should be addressed to Einat Liebenthal, Medical College of Wisconsin, Department of Neurology, 8701 West Watertown Plank Road, Milwaukee, WI 53226. E-mail: einat@mcw.edu.

DOI:10.1523/JNEUROSCI.1511-13.2013

Copyright © 2013 the authors 0270-6474/13/3315414-11\$15.00/0

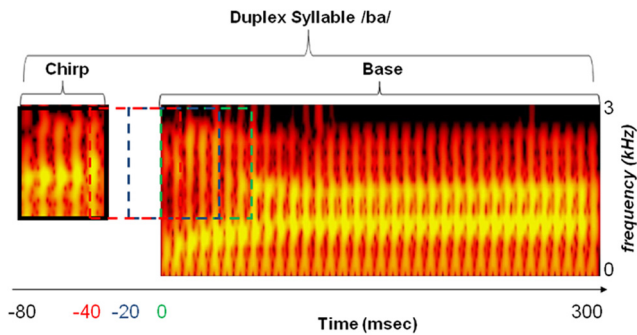


Figure 1. Schematic of duplex stimulus and onset time manipulation. Spectrograms of the chirp and base portions of the duplex syllable /ba/ are shown, with intensity of acoustic energy color coded from dark red to bright yellow. The two horizontal yellow bands represent peaks in acoustic energy with steady states centered at 0.8 kHz (first formant) and 1.4 kHz (second formant). The base, consisting of the entire syllable except the chirp portion, is 295 ms long and presented to one ear. The chirp, consisting of the initial second formant rising spectral transition, is 50 ms long and presented to the other ear with an SOA of 80–0 ms before the base. The chirp is depicted here at 80 ms SOA, with dotted traces in red, blue, and green indicating the presentation time at SOAs of 40, 20, and 0 ms, respectively. For the duplex syllable /ga/ (data not shown), the base is identical and the chirp second formant spectral transition is falling.

tentials (ERPs) and functional magnetic resonance (fMRI) images were concurrently collected. The premise for the experimental design was that duplex syllable identification would degrade with increasing SOA, whereas duplex chirp identification would not, permitting characterization of the neural network associated with phonological processing independent of general auditory analysis. Joint independent component analysis (jICA) was applied to the neuroimaging data to study the neural dynamics of brain regions involved in phonological processing with high spatiotemporal resolution (Calhoun et al., 2006; Mangalathu-Arumana et al., 2012).

Results showed that phonological processing is highly interactive, with functional fields in pSTG, inferior parietal (IPL), and ventral central sulcus (vCS) engaged early and almost simultaneously (at 80–100 ms latency), and with activity rebounding in a similar network after 300 ms. Left hemispheric lateralization was observed in IPL and vCS at 120–130 ms, but in pSTG only at 380 ms, suggesting that dorsal stream lateralization stems from a functional specialization of articulatory somatomotor (and not auditory) fields.

Materials and Methods

Subjects. Participants were 25 adults (15 females; mean age, 24 years) with no history of neurological or hearing impairments, native speakers of English, and right handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Data from five participants were excluded from ERP analysis, and data from one participant was excluded from fMRI image analysis, due to excessive artifact contamination (determined as $\geq 15\%$ of trials in a condition that was affected). In four additional participants, ERP data were not obtained due to equipment malfunction. In total, the report is based on behavioral and fMRI results from 24 participants, and ERP and jICA results from 15 participants. Informed consent was obtained in accordance with the Medical College of Wisconsin Institutional Review Board.

Stimuli. The duplex stimuli were derived from a natural utterance of /ga/, resynthesized to a two-formant syllable using Multispeech 3700 (Kay Elemetrics) as described in prior work (Liebenthal et al., 2005), and edited to 295 ms duration. The second formant (F2) spectral transition of /ba/ and /ga/ was separated from the remaining sound structure to create an isolated chirp (containing the distinctive cue for syllable identification) and a base (identical for the two syllables), each presented to one of the ears (with equal frequency) at four SOAs (0, 20, 40, and 80 ms; Fig. 1).

Sounds were delivered using a pneumatic audio system (Avotec) at ~ 70 dB, adjusted individually to accommodate differences in hearing and in positioning of the ear tips. Stimulus delivery was controlled with Presentation (Neurobehavioral Systems).

Task design and procedure. Trials, consisting of a sequence of three identical duplex stimuli presented at 1 Hz, were presented in eight runs during which simultaneous ERP/fMRI data were acquired. In half of the runs (presented in random order), participants were instructed to identify the duplex syllables as ba or ga (syllable task), and in the other half they were instructed to identify the initial chirp as falling or rising (chirp task). Participants responded by pressing one of two keys at the end of each trial. In each run, 20 trials per SOA and 20 silence baseline trials were presented in random order, for a total of 80 trials per SOA and task condition.

In a prescan session, participants practiced the syllable identification task with the original (nonduplex) syllables and the duplex syllables at an SOA of 0 ms, and the chirp identification task with the isolated chirps and duplex syllables at an SOA of 80 ms. This practice set was selected for each task because performance with it was expected to be the highest. Participants were required to achieve 90% accuracy in two consecutive practice runs to proceed to the neuroimaging session, and all were able to reach this accuracy level in two to six practice runs.

fMRI image acquisition and analysis. Images were acquired on a 3 T Excite scanner (GE Medical Systems). Functional data consisted of T2*-weighted, gradient echo, echoplanar images acquired using clustered acquisition at 7 s intervals (echo time = 20 ms, flip angle = 77° , acquisition time = 2 s). The sound sequences were positioned to start 500 ms after the end of each image acquisition to avoid perceptual masking by the acoustic noise of the scanner and to synchronize the next image acquisition with the estimated peak of the BOLD response to the sounds (Vagharchakian et al., 2012). Functional images consisted of 35 axially oriented 3.5 mm slices with a 0.5 mm interslice gap (field of view = 192 mm, 64×64 matrix), covering the whole brain. A total of 72 images were acquired per run. High-resolution anatomical images of the entire brain were obtained, using a 3-D spoiled gradient-recalled acquisition in a steady state, as a set of 130 contiguous axial slices with $0.938 \times 0.938 \times 1.0$ mm voxel dimensions.

Image analysis was conducted in AFNI (Cox, 1996). Within-subject analysis consisted of spatial coregistration of the functional images to minimize motion artifacts and registration to the anatomical images. Voxelwise multiple linear regression was applied to analyze individual time series, with reference functions representing the task (syllable, chirp), and the four SOA levels coded as a linear progression (1, 2, 3, 4). The trial reaction time (RT) was also modeled to remove activity related to behavioral performance, and six motion parameters were included as covariates of no interest. In each analysis, general linear tests were conducted between conditions.

The individual statistical maps and the anatomical scans were projected into standard stereotaxic space (Talairach and Tournoux, 1988) by linear resampling and then smoothed with a Gaussian kernel of 4 mm FWHM. Group maps were created in a random-effects analysis. The group condition maps were thresholded at a voxelwise $p < 0.01$ and were corrected for multiple comparisons by removing clusters $< 380 \mu\text{l}$, resulting in a mapwise two-tailed $p < 0.01$. The group contrast maps were thresholded at a voxelwise $p < 0.05$ and were corrected for multiple comparisons by removing clusters $< 1641 \mu\text{l}$, resulting in a mapwise two-tailed $p < 0.05$. The cluster thresholds were determined through Monte-Carlo simulations that provide the chance probability of spatially contiguous voxels exceeding the voxelwise p threshold.

ERP acquisition and analysis. Sixty-four-channel EEG activity was acquired using the Maglink system (Compumedics) in a continuous mode, with Quik-Cap electrodes positioned according to the International 10–20 System, and CPz serving as the reference. Activity was recorded at full bandwidth and digitally sampled at 500 Hz per channel. Vertical eye movements and electrocardiogram activity were monitored with bipolar recordings. Interelectrode resistance was kept below 5 k Ω .

EEG analysis was conducted in Scan 4.4 (Compumedics) and consisted of bandpass filtering at 0.1–30 Hz; ballistocardiogram artifact removal (Ellingson et al., 2004); creation of epochs of -100 to $+500$ ms

from sound onset, baseline correcting each epoch by removing the mean voltage value of the whole sweep; and rejection of epochs with voltage values exceeding $\pm 100 \mu\text{V}$. The remaining epochs were sorted and averaged according to task and SOA condition.

jICA of ERP and fMRI. In a variation of previous jICA approaches to multimodal neuroimaging data (Calhoun et al., 2006), condition-wise fMRI images and ERP epochs at 62 electrode sites were integrated in a within-subject analysis (Mangalathu-Arumana et al., 2012). This version of jICA is powerful in that it is sensitive to nonlinear (and linear) patterns of dependence on an experimental variable, and is readily amenable to incorporation of the full array of spatiotemporal information in the ERPs. Here, jICA was applied in each subject across the four SOA levels in each task. The fMRI images and ERP data in each task were restructured into a joint matrix where each row corresponds to the flattened t score functional image for one SOA level relative to baseline, concatenated with the ERP temporal sequence (-100 to $+500$ ms from stimulus onset) and flattened across electrodes for one SOA level. Principal component analysis was applied to the joint matrix to whiten the data without reducing its dimensionality, and jICA was then applied to the principal components using the Fusion ICA Toolbox (<http://mialab.mrn.org/software/fit/>). Four components were returned, each containing a flattened array of fMRI/ERP activity covarying across SOA levels. The components were expanded into their native spaces, resulting in four jICA-fMRI maps and four corresponding jICA-ERP field map time series in each task. The jICA-fMRI maps with the highest positive amplitude values in the syllable and chirp tasks were considered to represent the bulk of the activity related to phonological and general auditory processing, respectively, and were selected for further analysis. In a second step, to consider the possibility that relevant activity was also represented in other components, jICA-fMRI maps with amplitude values reaching at least 50% of those in the map selected in the first step were added. Using this two-step procedure, one joint component per task was selected in 12 subjects, and two components were selected and summed together in one of the tasks in 3 subjects.

Grouping of the individual jICA-fMRI maps was performed using a random-effects model and mapwise correction levels as used in the multiple regression fMRI analysis (computed relative to the distribution of the nonselected jICA-fMRI components). To gain greater sensitivity in the temporal cortex, the jICA-fMRI task-contrast map was also corrected by removing clusters smaller than $800 \mu\text{l}$ in the left and right superior and middle temporal gyri, resulting in two-tailed $p < 0.05$ in these areas (see Fig. 5, bottom row). Grand average (across subjects) jICA-ERP waveforms were computed for each task condition.

jICA-ERP source reconstruction. Source reconstruction of the grand average ($n = 15$) jICA-ERP waveform was performed using the weighted minimum norm estimate to solve the inverse problem (Brainstorm 3.0), and a template head model created from T1-weighted MR images of the Colin brain available in Brainstorm, using a three-shell sphere Berg approximation representing the brain, scalp, and skull. The cortical surface was parsed and represented as a high-density mesh of 15,000 vertices, with sample electric dipoles positioned at each vertex perpendicular to the cortical surface. Electrode positions were approximated based on a template electrode position file. Activity in vertices at the base of the temporal pole was masked because the assumption of uniform skull thickness is violated and a spherical head model is therefore inadequate in this region (Teale et al., 2002; Hamalainen et al., 2007).

Current source density estimates in each task (see Fig. 6A,B, top and middle rows) were expressed as time point by time point z -scores relative to the mean activity in the baseline period (-100 to -2 ms before stimulus onset). Contrast maps between tasks (see Fig. 6A,B, bottom rows) were computed by subtraction of the z -score source maps. The task maps were thresholded vertexwise at z -scores of ± 4 , and the contrast maps at z -scores of ± 3 , corresponding to $p < 10^{-5}$ and $p < 10^{-3}$, respectively, relative to the poststimulus distribution of all vertices across time. Cluster thresholding at seven contiguous vertices, spatial smoothing with a Gaussian kernel SD of five vertices, and low-pass temporal filtering at 15 Hz were applied to remove spatially and temporally spurious activity.

A region-of-interest (ROI) analysis of the jICA-ERP source maps was performed to examine the temporal course of processing in brain areas

involved in the two tasks. The ROIs were seeded in areas in which fMRI activity related to SOA was stronger in one of the tasks (see Figs. 3, 5, bottom rows), but was not correlated with RT. Using a conservative approach guided by general anatomical (arrangement of functional regions according to gyral patterns) and methodological (lower spatial resolution of ERP compared with fMRI) considerations, the ROIs were expanded along gyri so as to cover the posterior superior temporal, parietal, and precentral cortex, without overlap. Seven ROIs were created in the left hemisphere and then mirrored on the right. The regions (Fig. 7A,B, middle) consisted of the pSTG (blue), IPL (magenta), superior parietal lobule (SPL; cyan), ventral post-central gyrus (vPostCG; red), dorsal post-central gyrus (dPostCG; yellow green), ventral pre-central gyrus (vPreCG; green), and dorsal pre-central gyrus (dPreCG; yellow brown). Note that, based on these considerations, IPL included only the anterior part of the supramarginal gyrus. The pSTG included the portion of STG ventral and posterior to Heschl's gyrus. The temporal course of jICA-ERP mean source activity across vertices in each ROI was computed in each task and expressed as z -scores at each time point relative to the mean activity in the baseline period (-100 to -2 ms). Differences in time course between the two tasks and hemispheres in each ROI were considered significant at pointwise t scores corresponding to $p < 0.0002$, sustained for a period of at least 30 ms, resulting in a corrected $p < 0.005$ relative to a temporally randomized distribution of points (Guthrie and Buchwald, 1991).

Results

Ambiguous duplex syllables, consisting of a brief spectral transition containing the cue for syllable identification (the chirp) and a base containing the remainder of the syllable, were each presented to one ear at interaural stimulus onset intervals (i.e., SOA) ranging from 0 to 80 ms (Fig. 1). Subjects were asked to identify the syllables as /ba/ or /ga/ (syllable task), or the chirps as rising or falling (chirp task). In this paradigm, only syllable identification is expected to depend on SOA, thereby allowing disentanglement of the neural processes associated with phonological processing (engaged in the syllable task) from those associated with general auditory (nonlinguistic, engaged in the chirp task) analysis of the speech sounds.

Behavioral performance

The effects of task (syllable, chirp) and SOA (0, 20, 40, 80 ms) on the behavioral accuracy and RT of duplex stimulus identification were examined in an ANOVA and are summarized in Figure 2. Overall, the results confirm that the performance accuracy and RT were dependent on SOA only in the syllable task.

Mean accuracy across SOAs was significantly lower in the syllable (78%) than in the chirp (92%) task ($F_{(1,23)} = 34.66$, $p < 0.00001$). Across tasks, accuracy was lower at an SOA of 80 ms than at SOAs of 0, 20, and 40 ms ($F_{(3,69)} = 9.10$, $p < 0.00005$). There was also an interaction between task and SOA, with accuracy in the syllable task lower at SOAs of 80 ms (66%, SE = 3%) and 40 ms (78%, SE = 3%) than at SOAs of 20 ms (86%, SE = 3%) and 0 ms (83%, SE = 2%); at an SOA of 80 ms than at an SOA of 40 ms; and no accuracy change as a function of SOA in the chirp task ($F_{(3,69)} = 15.37$, $p < 0.00001$).

For RT, there was a trend for an overall longer RT in the syllable (1025 ms) than in the chirp task (965 ms; $p = 0.07$). Across tasks, the RT was longer at an SOA of 80 ms than at SOAs of 40, 20, and 0 ms ($F_{(3,69)} = 6.01$, $p < 0.001$). There was also a significant interaction between task and SOA, with RT in the syllable task significantly longer at SOAs of 80 ms (1073 ms, SE = 40 ms) and 40 ms (1034 ms, SE = 37 ms) than at SOAs of 20 ms (1001 ms, SE = 35 ms) and 0 ms (992 ms, SE = 34 ms); at an SOA of 80 ms than at an SOA of 40 ms; and no RT change as a function of SOA in the chirp task ($F_{(3,69)} = 3.74$, $p < 0.01$).

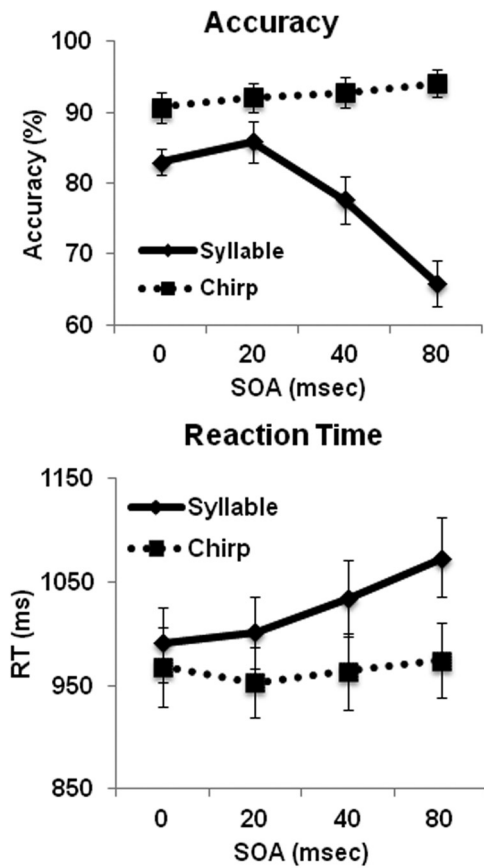


Figure 2. Behavioral accuracy and reaction time. Mean accuracy and RT at each SOA in the syllable (full trace) and chirp (dotted trace) identification tasks. Error bars represent the SEM across subjects.

the longer RT in the syllable relative to chirp task was largely due to an increase with SOA in the duration of the initial portion of the syllable (consisting of the chirp and corresponding portion of the base) containing the relevant information for syllable identification (Fig. 1), and not to an inherent longer neural processing time for phonological relative to nonlinguistic auditory processing.

Functional magnetic resonance imaging

The effects of task and SOA on the fMRI activity were investigated in a voxelwise multiple regression coding the trial SOA as a linear progression, and including regressors representing trial RT to model activity related to behavioral performance.

In the syllable task, activity in bilateral pSTG, IPL, and vCS increased proportionally with SOA (Fig. 3, top row). In the chirp task, only activity negatively related with SOA was found, in left angular gyrus (AG), right inferior frontal gyrus (IFG), and the postcentral gyrus, and in bilateral supplementary motor area (SMA; Fig. 3, middle row). The areas of negative activation in the chirp SOA correlation map largely coincided with areas of positive activation seen in the correlation map with RT, suggesting that the chirp SOA correlation map primarily reflected a decrease in executive functions with SOA in this task. Activity systematically related to behavioral performance may have been less effectively modeled (and removed) by the RT regressor in the chirp condition because of the limited variability in mean performance measures in that condition (Fig. 2). A direct comparison of the SOA effect in the two tasks revealed a stronger linear relationship with SOA in the syllable task in left pSTG, IPL, and vCS (Fig. 3, bottom row). Other foci of activation seen in this contrast were either in areas related with RT and negatively activated in the chirp condition (right IFG, bilateral SMA, and anterior cingulate) or reflected stronger activity during the baseline period (left anterior temporal pole; see Fig. 5). Importantly, the activity positively and linearly related to SOA in the syllable task, in left pSTG, IPL, and vCS was in areas in which activity was not associated with behavioral performance. The size, mean, and peak amplitude, and the peak location of activation clusters in each contrast in Figure 3 are given in Table 1.

Other foci of activation seen in this contrast were either in areas related with RT and negatively activated in the chirp condition (right IFG, bilateral SMA, and anterior cingulate) or reflected stronger activity during the baseline period (left anterior temporal pole; see Fig. 5). Importantly, the activity positively and linearly related to SOA in the syllable task, in left pSTG, IPL, and vCS was in areas in which activity was not associated with behavioral performance. The size, mean, and peak amplitude, and the peak location of activation clusters in each contrast in Figure 3 are given in Table 1.

Event-related potentials

The ERP waveforms in the two tasks, shown averaged across SOA conditions in Figure 4, were characterized by a frontocentral negativity peaking at ~160 ms followed by a frontocentral positivity peaking at ~250 ms, consistent with the spatiotemporal characteristics of N1 and P2 responses evoked by syllables (Näätänen and Picton, 1987; Martin et al., 1997). In the syllable task, this sequence of ERPs was followed by a prolonged frontal negativity peaking at ~350 ms, coinciding with the range of the N320 and N350 components previously associated with phonological processing (Bentin et al., 1999).

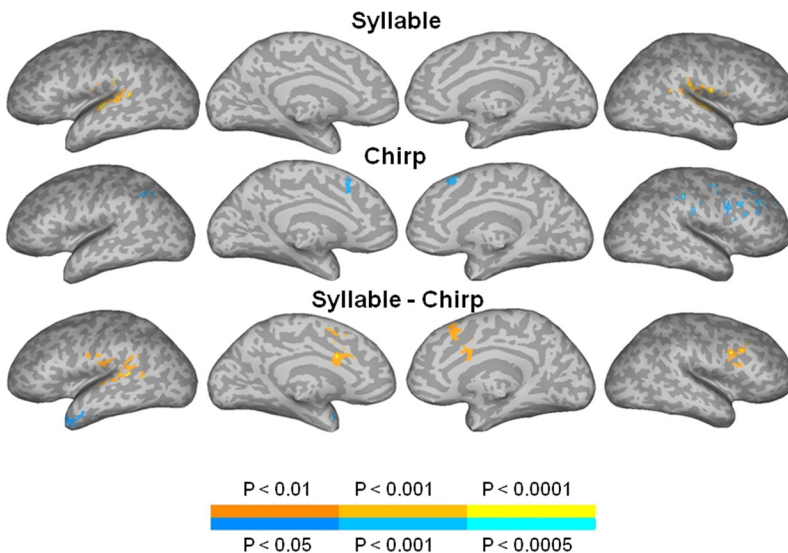


Figure 3. fMRI linear SOA effects. Multiple regression analysis showing group fMRI activity linearly related to SOA in the syllable (top row) and chirp (middle row) tasks, and in the contrast between the tasks (bottom row). The color scale indicates the voxelwise probability for false positives, for positive (orange–yellow) and negative (blue–cyan) activations, in each task (legend above color scale) and in the contrast between the tasks (legend below color scale). The maps were corrected for multiple comparisons, as indicated in Materials and Methods.

Importantly, the mean RT in the syllable task was related to SOA, such that the increase in RT at each SOA (relative to SOA 0 ms) was generally consistent with the SOA (81 ms at SOA 80 ms, 43 ms at SOA 40 ms, and 10 ms at SOA 20 ms). This suggests that

Joint independent component analysis of fMRI and ERP

Integration of the fMRI and ERP results was conducted using a within-subject variant of jICA (Mangalathu-Arumana et al.,

Table 1. Size, mean and peak amplitude, and approximate peak location in standard stereotaxic space and Brodmann area, of activation clusters in multiple regression maps showing group fMRI activity linearly related to SOA in the syllable and chirp tasks versus rest, and in the contrast between the tasks

Cluster size (mm ³)	Mean amplitude (z-scores)	Peak amplitude (z-scores)	Peak x-coordinate (mm)	Peak y-coordinate (mm)	Peak z-coordinate (mm)	Structure	BA
Syllable task versus rest							
3256	3.0876	4.7889	56	−10	6	R STG	22
3121	3.0659	5.4576	−63	−15	8	L STG	22
Chirp task versus rest							
946	−2.9914	−4.4785	34	31	30	R IFG	44/45
700	−3.0859	−4.6969	52	−24	32	R PostCG	2
641	−2.9057	−3.7492	8	18	47	R SMA	6
413	−3.0559	−4.5292	−36	−51	38	L AG	39
Syllable task versus chirp task							
4478	2.4697	4.4798	−67	−30	19	L pSTG	42/22
2097	2.2952	3.9870	−4	11	28	L aC	24
1810	2.4784	3.9056	43	10	31	R IFG	44/45
2615	−2.4060	−3.9634	−43	19	−23	L aMTG	21

Peaks listed were at least 30 mm apart. Clusters with negative amplitude values are listed in separate tables and reflect activation stronger in the second relative to first condition in the contrast. For some of the larger clusters, multiple peaks are listed (Talairach and Tournoux, 1988). R, Right; L, left; BA, Brodmann area; a, anterior; C, cingulate; MTG, middle temporal gyrus.

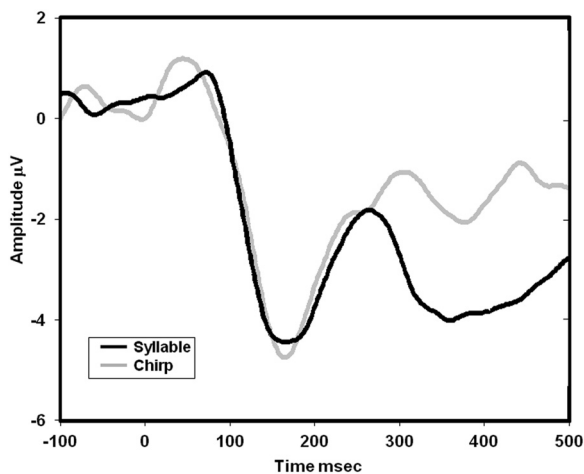


Figure 4. Grand average ERP waveforms. The grand average ERP waveforms in the syllable (black trace) and chirp (gray trace) tasks, displayed here averaged across SOAs and across frontocentral electrodes.

2012), to examine the temporal course of processing in brain regions showing linear or nonlinear variation with SOA. In each subject and task, the mean fMRI and ERP signals covarying as a function of SOA were associated with a joint component consisting of an fMRI spatial map (jICA-fMRI) and an ERP topographical map time series (jICA-ERP). Group jICA-fMRI maps in each task were computed using a random-effects model (Fig. 5), and the neural generators of the grand average jICA-ERP in each task were estimated using a minimum norm solution (Fig. 6). The temporal course of neural activity, in seven brain regions in each hemisphere that were more strongly activated in one of the tasks, was assessed in an ROI analysis of the jICA-ERP sources (Fig. 7).

jICA-fMRI

The syllable and chirp tasks (compared with baseline) induced overall similar patterns of activation, with greater signal (Fig. 5, two top rows, orange-yellow colors) in bilateral STG, IPL, IFG, SMA, and thalamus, and in left SPL, and pre- and post-central gyri. Greater signal in the baseline (blue-cyan colors) was observed in bilateral anterior temporal pole, AG, parieto-occipital cortex, middle frontal gyrus, anterior cingulate, and precuneus. A direct comparison between the two tasks showed greater activation in the chirp task (Fig. 5, third row, blue-cyan colors) in IPL and SPL bilaterally. Stronger activity for the syllable over the

chirp task was observed in the left pSTG, albeit only at a more lenient threshold using a small-volume temporal lobe mask correction (Fig. 5, bottom row). Importantly, there was not significant activity in the right temporal cortex at this lenient correction level. Other small foci of activity in the task contrast, in left AG, bilateral precuneus, medial superior frontal gyrus, and anterior temporal pole, were due to greater activation in the baseline relative to task conditions (observed as negative activation in Fig. 5, two top rows), consistent with task-induced deactivations possibly reflecting suspension of spontaneous semantic processing during rest (McKiernan et al., 2006). The small foci in the task contrast in bilateral SMA were in areas where activity was related to RT. The size, mean, and peak amplitude, and the peak location of activation clusters in each contrast are given in Table 2.

Together, the linear regression (Fig. 3) and jICA (Fig. 5) fMRI maps suggest that neural activity in the syllable task in left pSTG, IPL, and vCS varied linearly with SOA and independently of performance RT. This activity was better described in the linear regression analysis (which was based on trial-by-trial variations with SOA and included regressors for trial RT) than in the jICA (which was based on average variations with SOA across trials). In contrast, neural activity in the chirp task in bilateral IPL and SPL varied nonlinearly with SOA and was therefore better described in the jICA-fMRI map (which was sensitive to nonlinear variations).

jICA-ERP

ERP activity in both tasks was observed during two main time periods, an early period ~80–230 ms and a late period ~300 ms and onward. Neural source reconstructions of the grand average jICA-ERP waveforms in the early and late time periods (Fig. 6A,B, respectively) are shown for the syllable (Fig. 6A,B, top row) and chirp (Fig. 6A,B, middle row) tasks, and for the task difference (Fig. 6A,B, bottom row). The main activity stronger in the syllable task was seen in bilateral pSTG, and left ventral parietal and posterior frontal areas (IPL, vPostCG, and vPreCG). Activity stronger in the chirp task was seen primarily in bilateral STG and right parietal areas (SPL, IPL, and dPostCG). Other activation foci in the task contrast (in bilateral SMA, left anterior STG, left parieto-occipital cortex, and right IFG) were in areas in which activity was found to be stronger in the baseline condition (Fig. 5, jICA-fMRI maps) or was related to RT.

The temporal course of ERP activity in each ROI, represented as mean z-scores across all vertices in the ROI in the period after stimulus presentation (0–500 ms) relative to the baseline (−100

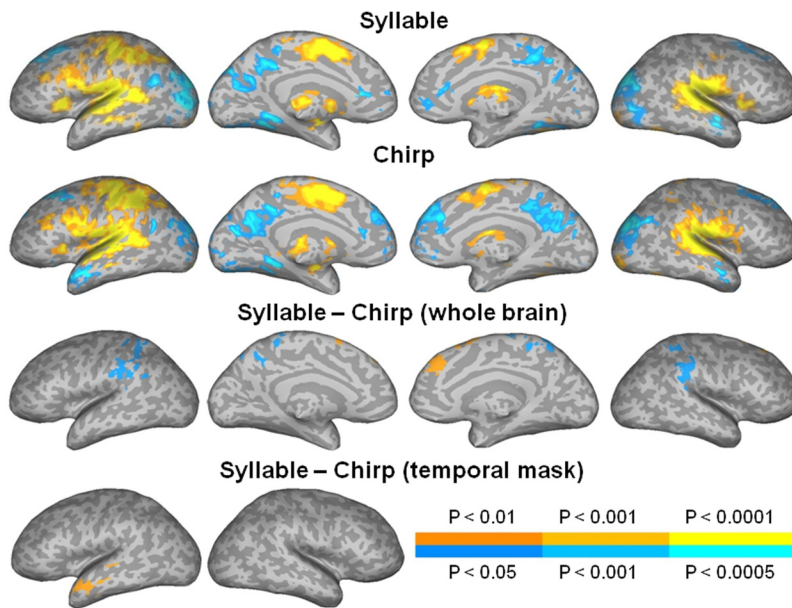


Figure 5. jICA-fMRI maps. jICA showing group fMRI activity varying with SOA in the syllable (top row) and chirp (second row) tasks, and in the contrast between the tasks (third and bottom rows). The color scale indicates the voxelwise probability for false positives, for positive (orange-yellow) and negative (blue-cyan) activations, in each task (legend above color scale) and in the contrast between the tasks (legend below color scale). The maps were corrected for multiple comparisons by removing small clusters in the whole brain (three top rows) or in the temporal cortex (bottom row), as indicated in Materials and Methods.

to -2 ms), is shown for the syllable (full trace) and chirp (dotted trace) tasks in the left and right hemispheres (Fig. 7*A, B*, respectively). The pattern of lateralization in each ROI and task is shown in Figure 7*C*. Overall, neural activity was stronger in the left hemisphere and centered in ventral parietal areas in the syllable task, and it was stronger in the right hemisphere and centered in dorsal parietal areas in the chirp task. In the syllable task, the earliest activity was observed in bilateral pSTG starting at 80 ms and peaking at ~ 100 ms, and was closely followed and dominated by activity in left IPL starting at 96 ms and peaking at ~ 140 ms, and somewhat weaker activity occurred in left vPostCG and vPreCG with a similar temporal profile to left IPL. The activity was significantly left lateralized in the ventral parietal regions (IPL, vPostCG), vPreCG, and SPL in the early time window, and in pSTG, vPostCG, and vPreCG in the late time window. In the chirp task, the earliest and strongest activity was observed in right SPL, starting at 90 ms and peaking at ~ 160 ms, with activity in bilateral STG exceeding the significance threshold only at ~ 120 ms. Activity in the chirp task was right lateralized in all ROIs except STG and dPreCG in both the early and late time windows.

Discussion

The neural dynamics of phonological processing were examined independently of those of auditory processing, by parametrically modulating the perception of ambiguous speech stimuli. The behavioral and neuroimaging results, showing a significant interaction between the effects of task type and interaural SOA on both behavioral performance measures and functional brain maps, confirm that different neural processing of identical duplex stimuli was elicited in each task. The dependence of behavioral and neural measures on interaural SOA specifically in the syllable task, suggests that dichotic fusion of the chirp and base portions of the syllable was required for syllable identification, consistent with prior reports (Repp et al., 1983; Bentin and Mann, 1990). Phonemic perception in the duplex syllable task

emerged from dichotic fusion of temporally misaligned spectral elements of the syllable. Under these listening conditions with distorted speech input, a dorsal auditory stream associated with phonological processing (Wise et al., 2001; Buchsbaum et al., 2005; Hickok and Poeppel, 2007) was activated, with relatively stronger activity in the syllable task in a perisylvian network including left pSTG, IPL, and vCS.

Activity in left pSTG, IPL, and vCS increased linearly with interaural SOA specifically in the syllable task and independently of fluctuations in behavioral performance (Fig. 3), consistent with a role for these areas in phonological processing. jICA of the EEG and fMRI data indicated that activity correlated with SOA in the syllable task arose in bilateral pSTG at 80–90 ms after stimulus onset, was quickly followed and dominated by strong activity in left IPL, vPostCG, and vPreCG at 95–230 ms; and rebounded in bilateral STG, IPL, left vPostCG, and vPreCG after 300 ms (Figs. 6, 7). This pattern of neural dynamics is incompatible with a simple hierarchical organization in an afferent pathway, whereby neural activity flows from primary to higher fields that process increasingly complex auditory patterns, and on to cognitive and executive areas. Rather, the finding that the left IPL, an area associated with short-term phonological storage (Paulesu et al., 1993; Buchsbaum and D'Esposito, 2009), possibly subvocalization (Koelsch et al., 2009; Price, 2010) and somatosensory articulatory feedback (Tourville et al., 2008), and the left vCS, associated with orofacial somatomotor control (Corfield et al., 1999; Fesl et al., 2003), are active early during the period of perceptual analysis of the sounds in superior temporal cortex points to interactive processing with efferent feedback from somatomotor to auditory cortex. The period 80–230 ms after sound onset coincides with the N1 and P2 electrophysiological responses, reflecting neural processes related to auditory analysis and object perception in superior temporal cortex, including analysis of speech spectrotemporal features relevant to phonemic perception (Liégeois-Chauvel et al., 1999; Eggermont and Ponton, 2002; Ahveninen et al., 2006; Chang et al., 2010; Liebenthal et al., 2010; Steinschneider et al., 2011; Tsunada et al., 2011). The strong and early activity in left ventral parietal and central sulcus regions observed here provides firm evidence that the somatomotor cortex associated with orofacial movement control plays a direct role in phonemic perception, at least when speech is ambiguous. Later activity observed after 300 ms in the duplex syllable task, in the same left temporoparietal–posterior frontal perisylvian network, could reflect the activation and maintenance of categorical neural representations of the syllables for task-related response selection. Indeed, this time range coincides with the N2 and P3 ERP components, shown to accurately reflect speech categorization in discrimination tasks (Maiste et al., 1995; Martin et al., 1997; Toscano et al., 2010), and with the negative N320 and N350 responses to pronounceable letter strings, associated with the activation of phonological representations from print (Bentin et al., 1999).

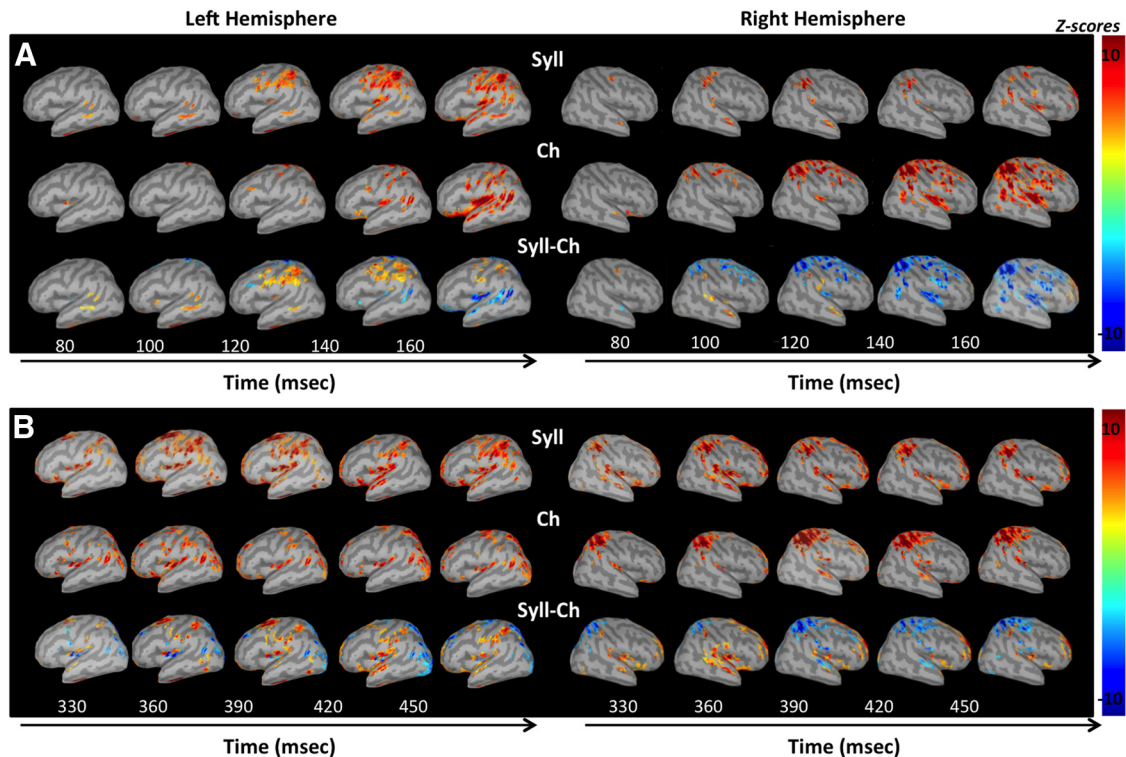


Figure 6. jICA-ERP source maps. **A, B**, Source reconstruction maps of the grand average jICA-ERP components in the early (**A**) and late (**B**) time windows, in the syllable (Syll; top rows) and chirp (Ch; middle row) tasks, and in the contrast between the tasks (Syll-Ch, bottom row). Maps in each task (Syll, Ch) are expressed as z-scores computed from current source density values at each time point relative to the mean values in the baseline period (-100 to -2 ms), and as the difference between z-scores in the task contrast (Syll-Ch). Maps in each task were thresholded at z-scores of ± 4 , and the contrast maps are thresholded at z-scores of ± 3 , as detailed in Materials and Methods.

Despite the mounting evidence for involvement of the dorsal auditory stream in perceptual processes, the neural mechanisms underlying auditory and somatomotor interactions are far from resolved. Recent neuroanatomical models (Rauschecker and Scott, 2009; Hickok et al., 2011) postulate a primary role for the dorsal auditory stream in speech perception as a feedforward efferent system, from frontal premotor to posterior temporal auditory areas via inferior parietal cortex, whereby predictive motor signals modulate sensory processing, at least when the speech input is degraded. Motor influences on phonemic categorization performance have been described in the ventral precentral gyrus at the level of the premotor cortex (Wilson et al., 2004; Meister et al., 2007; Chang et al., 2011; Osnes et al., 2011) and primary motor cortex (Mottonen et al., 2009). In ventral parietal cortex, neurons representing the somatosensory (tactile and proprioceptive) articulatory properties of speech sounds have been hypothesized to exert a modulatory influence on phonemic perception (Guenther, 2006; Tourville et al., 2008). Feedback from IPL to left posterior auditory areas has been suggested to play an important role in perceptual learning of ambiguous phonemic categories (Kilian-Hütten et al., 2011b). Indeed, the left IPL is activated during overt categorization tasks of phonemic and also of trained nonphonemic sounds, with the level of activity positively related with categorization ability (Caplan and Waters, 1995; Celsis et al., 1999; Jacquemot et al., 2003; Dehaene-Lambertz et al., 2005; Raizada and Poldrack, 2007; Desai et al., 2008). Feedback to the pSTG from areas representing the somatomotor articulatory properties of speech sounds may act as a top-down selection mechanism to tune the auditory areas to the set of possible phonemic inputs (Hickok et al., 2011) and to facilitate activation of phonemic representations in pSTG. Neurocomputa-

tionally, the role of bottom-up and top-down interactions in perceptual processing can be understood in terms of predictive coding, whereby forward connections conveying prediction errors and reciprocal backward connections mediating predictions are balanced to minimize prediction errors and optimize the probabilistic representation of sensory input (Friston, 2010). Such an inference scheme can be used to model the neural computations underlying perceptual categorization, corresponding to the mapping of noisy and dynamic sensory input to a fixed point in perceptual space (Friston and Kiebel, 2009).

The present findings, revealing the temporal dynamics of the temporoparietal–posterior frontal perisylvian network during a phonemic categorization task, permit the updating of our current understanding of auditory dorsal pathway function in phonemic perception in several important ways. The early timing of the IPL and vCS activity, during the phase of auditory perceptual analysis and a mere 15–20 ms after the pSTG, alleviates any concerns regarding potential confounding effects of subarticulatory or behavioral decision-making processes (Callan et al., 2004; Hickok et al., 2011) and confirms a genuine role for these regions in phonemic perception. The location of the BOLD activity in the ventral tip of the post- and pre-central gyri is consistent with that of primary sensorimotor cortex associated with orofacial motor control (Corfield et al., 1999; Fesl et al., 2003), although involvement of premotor fields cannot be ruled out. The strong activity in inferior parietal cortex further suggests that somatosensory feedback contributed significantly to phonemic perception in the syllable task. On the other hand, the ventrolateral prefrontal cortex was not differentially activated during phonological processing despite the use of an overt categorization task, consistent with

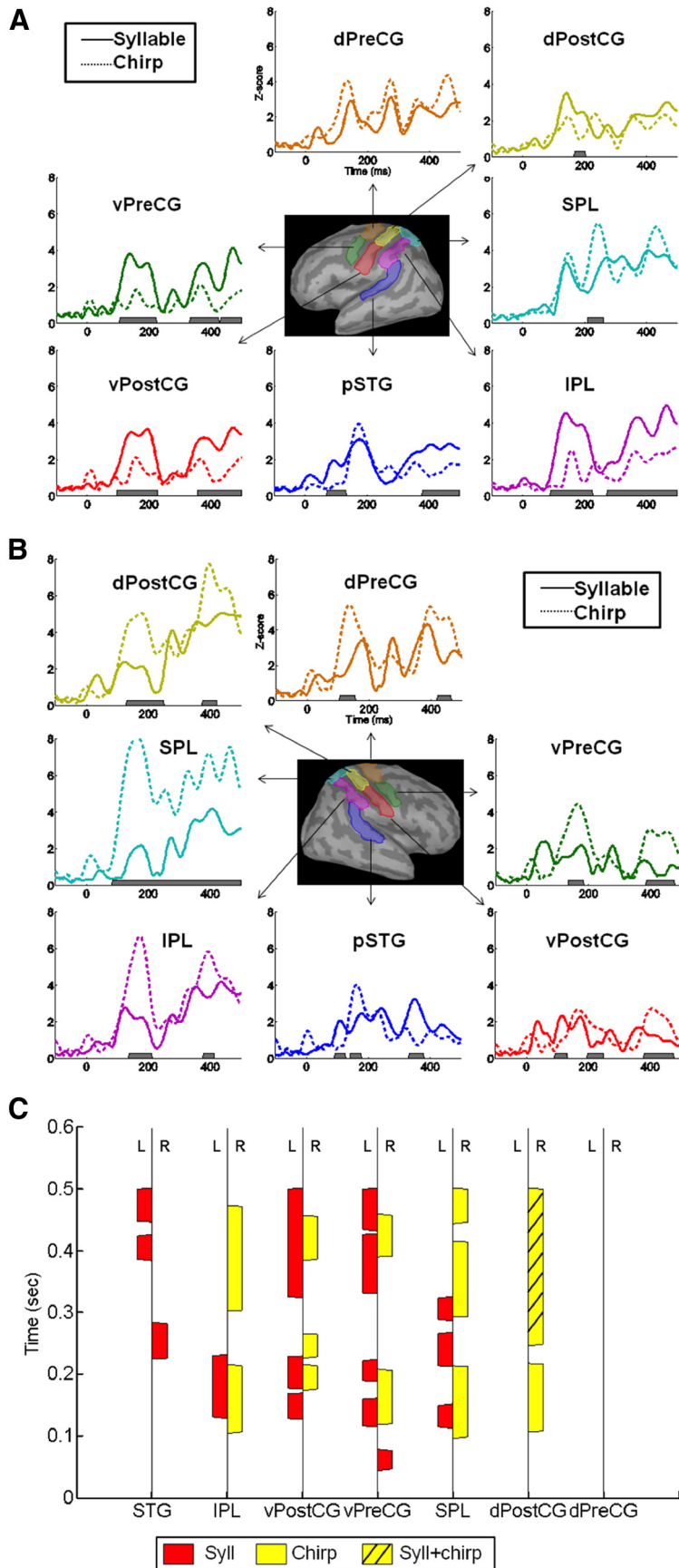


Figure 7. Region-of-interest analysis of the jICA-ERP source maps. **A, B**, Temporal course of the grand average jICA-ERP source activity in seven temporal and parietal ROIs in the syllable (full trace) and chirp (dotted trace) tasks, in the left (**A**) and right (**B**) hemispheres. The source activity is expressed as z-scores computed relative to the mean activity in the baseline period (−100 to

a domain-nonspecific role for this region. Together, the results suggest the existence of a direct feedback loop from ventral parietal and ventral central sulcus regions to posterior temporal cortex, representing an influence of somatosensory and articulatory representations of speech sounds on phonemic perception. In the monkey (Seltzer and Pandya, 1978; Petrides and Pandya, 1984, 2009) and more recently in the human (Frey et al., 2008; Makris et al., 2009), the anterior part of the IPL was demonstrated to have strong reciprocal connections with ventral premotor (and ventrolateral prefrontal) areas controlling orofacial musculature via the superior longitudinal fasciculus, and with the posterior superior temporal cortex via the middle longitudinal fasciculus. These anatomical connections could form the basis for the functional phonological loop activated here. Feedback from discrete somatomotor representations of speech may serve to narrow the range of possible sound inputs, and to activate categorical phoneme representations in pSTG. The activation and maintenance of categorical phonemic representations may correspond to the second phase of neural activity observed in this study. According to this view, neural representations of both graded and categorical properties of sounds are present within the same general posterior temporal region, but are activated at different time phases, consistent with the existence of feedforward and feedback processes in this region.

An important aspect of the findings is that left lateralization in the syllable task was observed in IPL and vCS in the early time period, but in the pSTG only at ~380 ms latency. This pattern implies that left lateralization in the phonological dorsal pathway is due initially to stronger activation of left IPL and vCS, and is imposed only later on the pSTG through feedback interactions. The results are generally in-

←
 −2 ms). The z-score values exceeding 1.65 were considered significant deviations from the baseline. Periods of significant task differences in each ROI are indicated by shaded rectangles. The ROIs (**A** and **B**, middle) correspond to regions activated more strongly in one of the tasks, and include the pSTG (blue), IPL (magenta), SPL (cyan), vPostCG (red), dPostCG (yellow green), vPreCG (green), and dPreCG (brown). **C**, Lateralization pattern along time (*y*-axis) in each ROI (*x*-axis), represented as periods of significantly stronger activity in the left (L; left of vertical line) or right (R; right of vertical line) hemisphere, is shown for the syllable (red) and chirp (yellow) tasks, and for periods of overlapping lateralization between tasks (striped yellow).

Table 2. Activation clusters in multiple regression maps showing group jICA fMRI activity varying with SOA in the syllable and chirp tasks versus rest, and in the contrast between the tasks, using a whole-brain and temporal lobe correction

Cluster size (mm ³)	Mean amplitude (z-scores)	Peak amplitude (z-scores)	Peak x-coordinate (mm)	Peak y-coordinate (mm)	Peak z-coordinate (mm)	Structure	BA
Syllable task versus rest							
67,140	3.7938	7.6776	−59	−23	14	L STG	41/42
		6.299	−39	−20	48	L pCG	3
		5.125	−32	21	14	L insula	13
28,906	4.3227	7.4931	36	−30	16	R STG	41/42
24,454	3.1576	4.9798	37	−57	−24	R cerebellum	
		4.337	−53	−51	−19	L ITG	37
13,627	3.6721	5.8282	−5	−5	52	L SMA	6
1486	3.0261	3.9805	57	−11	46	R pCG	3
813	3.2498	3.9794	−28	−89	−12	L IOG	17/18
493	2.7659	3.4431	29	−94	−6	R IOG	17/18
455	2.914	3.5029	−1	−36	54	L PL	5
36,991	−3.2547	−5.371	−44	−77	11	L MOG	19
		−4.842	−24	−42	−8	L PH	36/37
		−4.301	−15	−97	10	R MOG	17/18
6950	−3.1506	−4.2970	3	−47	39	R pC	31
3873	−3.1827	−4.3460	−23	19	42	L MFG	8
3305	−3.0464	−4.3736	16	−79	−11	R IOG	17/18
		−4.710	33	−84	18	R MOG	19
		−4.070	26	−43	−9	R PH	36/37
2798	−2.9790	−4.1373	−2	30	12	L aC	24
1175	−3.3262	−4.6389	50	−6	−13	R MTG	21
385	−2.8291	−3.4027	26	11	50	R SFG	6
Chirp task versus Rest							
109,730	3.7135	7.4443	−59	−24	13	L STG	41/42
		6.547	−6	−4	50	L SMA	6
		6.266	−49	−35	44	L SPL	2
		4.563	11	−37	63	R PL	5
37,748	3.8121	7.4224	61	−16	5	R STG	22
		4.981	33	−29	12	R insula	13
		4.391	19	−2	24	R caudate	
9618	3.3176	5.0629	9	−51	−16	R cerebellum	
3422	2.9151	3.883	−49	−48	−32	L cerebellum	
1468	3.1435	4.005	21	−89	−15	R IOG	17/18
970	3.3666	5.881	−26	−13	−6	L hippocampus	
400	3.0663	3.8728	−26	−91	−12	L IOG	17/18
38,427	−3.1726	−5.0905	33	−79	21	R MOG	19
		−4.8003	18	36	34	R SFG	9
18,821	−3.0881	−3.800	−2	56	19	L SFG	9/10
		−3.8834	−26	−77	−6	L IOG	17/18
		−4.427	−5	−65	15	L pC	31
4109	−3.0150	−4.259	−7	−45	39	L pC	31
3486	−3.1294	−4.3608	−60	0	−15	L aMTG	21
2990	−3.133	−4.5389	−21	27	43	L SFG	8
1410	−3.2806	−4.685	−23	−40	−7	L PH	36/37
965	−3.2411	−4.3769	53	−2	−19	R aMTG	21
651	−2.8234	−3.3943	17	−80	−10	R IOG	17/18
624	−2.8712	−3.4028	−23	−72	−38	L cerebellum	
561	−2.8232	−3.4613	23	8	−30	R PH	38
424	−3.1688	−4.0731	−52	−58	22	L AG	39
Syllable task versus chirp task (whole-brain analysis)							
3695	2.3653	4.256	13	31	52	R SFG	8
4989	−2.2825	−3.7420	−4	−38	60	L PL	5
4768	−2.3816	−3.4364	54	−33	41	R IPL	40
3247	−2.3895	−3.1479	−51	−43	31	L IPL	40
Syllable task versus chirp task (temporal lobe mask)							
2722	0.8027	3.4377	−50	3	−12	L aMTG	21/38
801	2.2772	2.786	−59	−23	2	L pSTG	21/22

R, Right; L, left; a, anterior; p, posterior; v, ventral; d, dorsal; SFG, superior frontal gyrus; MFG, middle, frontal gyrus; MTG, middle temporal gyrus; ITG, inferior temporal gyrus; C, cingulate; MOG, middle occipital gyrus; IOG, inferior occipital gyrus; PL, paracentral lobule; PH, parahippocampus; BA, Brodmann area.

compatible with theories emphasizing hemispheric differences in auditory processing (Boemio et al., 2005; Giraud et al., 2007), at least as the basis for left hemispheric dominance during phonological processing. Instead, functional specialization of somato-

sensory and motor areas may determine lateralization in the dorsal auditory stream.

In summary, the neural dynamics of phonological processing in the dorsal auditory pathway described here are consistent with

reciprocal activity in pSTG, IPL, and vCS, and with left lateralization originating in IPL and vCS.

Notes

Supplemental material for this article is available at <http://www.jneurosci.org>. This material has not been peer reviewed.

References

- Ahveninen J, Jääskeläinen IP, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin FH, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006) Task-modulated “what” and “where” pathways in human auditory cortex. *Proc Natl Acad Sci U S A* 103:14608–14613. [CrossRef Medline](#)
- Bentin S, Mann V (1990) Masking and stimulus intensity effects on duplex perception: a confirmation of the dissociation between speech and non-speech modes. *J Acoust Soc Am* 88:64–74. [CrossRef Medline](#)
- Bentin S, Mouchetant-Rostaing Y, Giard MH, Echallier JF, Pernier J (1999) ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J Cogn Neurosci* 11:235–260. [CrossRef Medline](#)
- Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395. [CrossRef Medline](#)
- Buchsbaum BR, D’Esposito M (2009) Repetition suppression and reactivation in auditory-verbal short-term recognition memory. *Cereb Cortex* 19:1474–1485. [CrossRef Medline](#)
- Buchsbaum BR, Olsen RK, Koch P, Berman KF (2005) Human dorsal and ventral auditory streams subsolve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48:687–697. [CrossRef Medline](#)
- Calhoun VD, Adali T, Pearlson GD, Kiehl KA (2006) Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data. *Neuroimage* 30:544–553. [CrossRef Medline](#)
- Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orsensory internal models. *Neuroimage* 22:1182–1194. [CrossRef Medline](#)
- Caplan D, Waters GS (1995) On the nature of the phonological output planning processes involved in verbal rehearsal: evidence from aphasia. *Brain Lang* 48:191–220. [CrossRef Medline](#)
- Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, Chollet F (1999) Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9:135–144. [CrossRef Medline](#)
- Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010) Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13:1428–1432. [CrossRef Medline](#)
- Chang EF, Edwards E, Nagarajan SS, Fogelson N, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2011) Cortical spatio-temporal dynamics underlying phonological target detection in humans. *J Cogn Neurosci* 23:1437–1446. [CrossRef Medline](#)
- Corfield DR, Murphy K, Josephs O, Fink GR, Frackowiak RS, Guz A, Adams L, Turner R (1999) Cortical and subcortical control of tongue movement in humans: a functional neuroimaging study using fMRI. *J Appl Physiol* 86:1468–1477. [Medline](#)
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173. [CrossRef Medline](#)
- Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear Res* 229:132–147. [CrossRef Medline](#)
- Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S (2005) Neural correlates of switching from auditory to speech perception. *Neuroimage* 24:21–33. [CrossRef Medline](#)
- Desai R, Liebenthal E, Waldron E, Binder JR (2008) Left posterior temporal regions are sensitive to auditory categorization. *J Cogn Neurosci* 20:1174–1188. [CrossRef Medline](#)
- Eggermont JJ, Ponton CW (2002) The neurophysiology of auditory perception: from single units to evoked potentials. *Audiol Neurootol* 7:71–99. [CrossRef Medline](#)
- Ellingson ML, Liebenthal E, Spanaki MV, Prieto TE, Binder JR, Ropella KM (2004) Ballistocardiogram artifact reduction in the simultaneous acquisition of auditory ERPs and fMRI. *Neuroimage* 22:1534–1542. [CrossRef Medline](#)
- Fesl G, Morigg B, Schmid UD, Naidich TP, Herholz K, Yousry TA (2003) Inferior central sulcus: variations of anatomy and function on the example of the motor tongue area. *Neuroimage* 20:601–610. [CrossRef Medline](#)
- Frey S, Campbell JS, Pike GB, Petrides M (2008) Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J Neurosci* 28:11435–11444. [CrossRef Medline](#)
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138. [CrossRef Medline](#)
- Friston K, Kiebel S (2009) Cortical circuits for perceptual inference. *Neural Netw* 22:1093–1104. [CrossRef Medline](#)
- Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RS, Laufs H (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56:1127–1134. [CrossRef Medline](#)
- Golestani N, Zatorre RJ (2004) Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* 21:494–506. [CrossRef Medline](#)
- Guenther FH (2006) Cortical interactions underlying the production of speech sounds. *J Commun Disord* 39:350–365. [CrossRef Medline](#)
- Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. *Psychophysiology* 28:240–244. [CrossRef Medline](#)
- Hämäläinen JA, Leppänen PH, Guttorm TK, Lyytinen H (2007) N1 and P2 components of auditory event-related potentials in children with and without reading disabilities. *Clin Neurophysiol* 118:2263–2275. [CrossRef Medline](#)
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402. [CrossRef Medline](#)
- Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69:407–422. [CrossRef Medline](#)
- Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003) Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J Neurosci* 23:9541–9546. [Medline](#)
- Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011a) Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci* 31:1715–1720. [CrossRef Medline](#)
- Kilian-Hütten N, Vroomen J, Formisano E (2011b) Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *Neuroimage* 57:1601–1607. [CrossRef Medline](#)
- Koelsch S, Schulze K, Sammler D, Fritz T, Müller K, Gruber O (2009) Functional architecture of verbal and tonal working memory: an fMRI study. *Hum Brain Mapp* 30:859–873. [CrossRef Medline](#)
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. *Cereb Cortex* 15:1621–1631. [CrossRef Medline](#)
- Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR (2010) Specialization along the left superior temporal sulcus for auditory categorization. *Cereb Cortex* 20:2958–2970. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Maiste AC, Wiens AS, Hunt MJ, Scherg M, Picton TW (1995) Event-related potentials and the categorical perception of speech sounds. *Ear Hear* 16:68–90. [Medline](#)
- Makris N, Papadimitriou GM, Kaiser JR, Sorg S, Kennedy DN, Pandya DN (2009) Delineation of the middle longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. *Cereb Cortex* 19:777–785. [CrossRef Medline](#)
- Mangalathu-Arumana J, Beardsley SA, Liebenthal E (2012) Within-subject joint independent component analysis of simultaneous fMRI/ERP in an auditory oddball paradigm. *Neuroimage* 60:2247–2257. [CrossRef Medline](#)
- Martin BA, Sigal A, Kurtzberg D, Stapells DR (1997) The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds/ba/and/da. *J Acoust Soc Am* 101:1585–1599. [CrossRef Medline](#)
- McKiernan KA, D’Angelo BR, Kaufman JN, Binder JR (2006) Interrupting the “stream of consciousness”: an fMRI investigation. *Neuroimage* 29:1185–1191. [CrossRef Medline](#)
- Meister IG, Wilson SM, Deblick C, Wu AD, Iacoboni M (2007) The essen-

- tial role of premotor cortex in speech perception. *Curr Biol* 17:1692–1696. [CrossRef Medline](#)
- Mottonen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. *J Neurosci* 29:9819–9825. [CrossRef Medline](#)
- Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sounds: a review and an analysis of the component structure. *Psychophysiology* 24:375–425. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113. [CrossRef Medline](#)
- Osnes B, Hugdahl K, Specht K (2011) Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage* 54:2437–2445. [CrossRef Medline](#)
- Paulesu E, Frith CD, Frackowiak RS (1993) The neural correlates of the verbal component of working memory. *Nature* 362:342–345. [CrossRef Medline](#)
- Petrides M, Pandya DN (1984) Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J Comp Neurol* 228:105–116. [CrossRef Medline](#)
- Petrides M, Pandya DN (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol* 7:e1000170. [CrossRef Medline](#)
- Poeppl D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling time.” *Speech Commun* 41:245–255. [CrossRef](#)
- Price CJ (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci* 1191:62–88. [CrossRef Medline](#)
- Raizada RD, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726–740. [CrossRef Medline](#)
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724. [CrossRef Medline](#)
- Repp BH, Milburn C, Ashkenas J (1983) Duplex perception: confirmation of fusion. *Percept Psychophys* 33:333–337. [CrossRef Medline](#)
- Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149:1–24. [CrossRef Medline](#)
- Steinschneider M, Nourski KV, Kawasaki H, Oya H, Brugge JF, Howard MA 3rd (2011) Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. *Cereb Cortex* 21:2332–2347. [CrossRef Medline](#)
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical.
- Teale P, Benkers T, Rojas D, Reite M (2002) Determination of the sphere origin for MEG source modelling in temporal regions. *Phys Med Biol* 47:1161–1166. [CrossRef Medline](#)
- Toscano JC, McMurray B, Dennhardt J, Luck SJ (2010) Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychol Sci* 21:1532–1540. [CrossRef Medline](#)
- Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39:1429–1443. [CrossRef Medline](#)
- Tsunada J, Lee JH, Cohen YE (2011) Representation of speech categories in the primate auditory cortex. *J Neurophysiol* 105:2634–2646. [CrossRef Medline](#)
- Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S (2012) A temporal bottleneck in the language comprehension network. *J Neurosci* 32:9089–9102. [CrossRef Medline](#)
- Wilson SM, Iacoboni M (2006) Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* 33:316–325. [CrossRef Medline](#)
- Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural subsystems within “Wernicke's area.” *Brain* 124: 83–95. [CrossRef](#)
- Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953. [CrossRef Medline](#)