

## Localization of sublexical speech perception components

Peter E Turkeltaub\*, H. Branch Coslett

Department of Neurology, University of Pennsylvania, 3400 Spruce Street, 3 West Gates Building, Philadelphia, PA 19104, USA

### ARTICLE INFO

#### Article history:

Accepted 28 March 2010

Available online 21 April 2010

#### Keywords:

Auditory  
Categorical perception  
Phoneme  
Phonetic  
Phonology  
Speech perception  
fMRI  
Meta-analysis  
Neuroimaging  
Language

### ABSTRACT

Models of speech perception are in general agreement with respect to the major cortical regions involved, but lack precision with regard to localization and lateralization of processing units. To refine these models we conducted two Activation Likelihood Estimation (ALE) meta-analyses of the neuroimaging literature on sublexical speech perception. Based on foci reported in 23 fMRI experiments, we identified significant activation likelihoods in left and right superior temporal cortex and the left posterior middle frontal gyrus. Sub-analyses examining phonetic and phonological processes revealed only left mid-posterior superior temporal sulcus activation likelihood. A lateralization analysis demonstrated temporal lobe left lateralization in terms of magnitude, extent, and consistency of activity. Experiments requiring explicit attention to phonology drove this lateralization. An ALE analysis of eight fMRI studies on categorical phoneme perception revealed significant activation likelihood in the left supramarginal gyrus and angular gyrus. These results are consistent with a speech processing network in which the bilateral superior temporal cortices perform acoustic analysis of speech and non-speech auditory stimuli, the left mid-posterior superior temporal sulcus performs phonetic and phonological analysis, and the left inferior parietal lobule is involved in detection of differences between phoneme categories. These results modify current speech perception models in three ways: (1) specifying the most likely locations of dorsal stream processing units, (2) clarifying that phonetic and phonological superior temporal sulcus processing is left lateralized and localized to the mid-posterior portion, and (3) suggesting that both the supramarginal gyrus and angular gyrus may be involved in phoneme discrimination.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

Daily, we rely on our ability to perceive and interpret a near continuous stream of aural speech. Most of us perform this computationally burdensome task effortlessly, regardless of background noise or spectrotemporal degradation of the speech signal. Converging evidence from multiple disciplines using a variety of techniques has provided insights into the neural basis of this remarkable perceptual ability (Hickok & Poeppel, 2007; Obleser & Eisner, 2009; Price, 2000; Scott & Johnsrude, 2003; Zatorre, Belin, & Penhune, 2002). Sublexical speech perception (i.e. phonemes and syllables) is of particular interest because altered sublexical perception is associated with developmental language disorders, including specific language impairment and dyslexia (Bogliotti, Serniclaes, Messaoud-Galusi, & Sprenger-Charolles, 2008; Serniclaes, Sprenger-Charolles, Carre, & Demonet, 2001; Szenkovits & Ramus, 2005). Despite extensive research into the neural basis of sublexical speech perception, basic questions remain regarding the brain structures involved.

Before discussing the neuroanatomical correlates of speech perception, it is worth defining some terms we will use throughout this paper with regard to subprocesses of speech perception. These

definitions are based on those of Colin Phillips (Phillips, 2001), but have been operationalized to apply to the types of imaging studies considered here. We use the term *acoustic processing* to refer to the spectrotemporal analysis of speech or non-speech auditory signals that is independent of language experience. We use the term *phonetic processing* to refer to auditory processing that is shaped by language experience and is specific to speech. We use the term *phonological processing* to refer to the use of abstract symbolic mental representations of speech sounds specific to one's language experience. Note that both acoustic and phonetic processes operate on a continuous, "analog" auditory signal, whereas phonological processes employ discrete abstract representations that are used for lexical access or explicit judgments on speech sounds. In many studies, stimuli and tasks that recruit phonological processing do not control for phonetic demands, and these two processes are sometimes difficult to disambiguate. Finally, we use the term *categorical phoneme perception* to refer to the assignment of phoneme labels to speech sounds and the discrimination between them. For our purposes this specifically refers to imaging experiments in which discrimination of speech sounds across phoneme boundaries is compared to discrimination within phoneme boundaries.

The superior temporal sulcus (STS), which is commonly associated with phonetic and phonological aspects of speech perception, is often divided into anterior, middle, and posterior portions. We

\* Corresponding author. Fax: +1 215 349 5579.

E-mail address: [Peter.turkeltaub@uphs.upenn.edu](mailto:Peter.turkeltaub@uphs.upenn.edu) (P.E Turkeltaub).

use the term anterior STS to refer to that portion extending anterior to Heschl's gyrus (HG), corresponding roughly to a Talairach *Y* greater than  $-15$  (Leonard, Puranik, Kuldau, & Lombardino, 1998). We use the term middle STS to refer to the portion ventral to HG, roughly *Y* of  $-15$  to  $-30$ . Posterior STS then refers to the portion posterior to HG, roughly *Y* less than  $-30$ .

Most modern models agree that auditory speech is processed via parallel streams specialized for analyzing different aspects of the speech signal. Belin and Zatorre proposed that a ventral stream functions in auditory object recognition based on invariant auditory features, and a dorsal stream analyzes spectral motion, perceiving time-varying aspects of signals emitted by auditory objects (Belin & Zatorre, 2000). In this model, phonetic processing, which requires evaluation of rapidly changing spectral content, is performed by the dorsal stream, running caudally from primary auditory cortex. The precise gyral anatomy is not specified. Hickok and Poeppel proposed a ventral stream which functions in lexico-semantic access and comprehension (Hickok & Poeppel, 2007). In this model, the dorsal stream functions in sensory-motor integration for articulatory preparation. Neuroanatomically, this model is fairly specific: initial acoustic analysis occurs in the bilateral dorsal superior temporal cortices (including the superior temporal gyri (STG), Heschl's gyri, and planum temporale), which send output to the bilateral mid-posterior superior temporal sulci (STS) for phonetic and phonological processing. The STS then project to the dorsal and ventral processing streams. The ventral stream proceeds anteriorly, along the bilateral middle temporal gyri. The dorsal stream links the left posterior planum temporale (termed the Sylvian-parietal-temporal) with a left frontal articulatory network. All projections in this model are reciprocal, allowing for significant top-down influence on lower level processes. These models differ in the exact functions performed by dorsal and ventral streams, but generally predict that sublexical phonetic and phonological processes are primarily instantiated posterior to primary auditory cortex, either within the dorsal stream, or prior to the separation of major pathways.

Scott and Johnsrude proposed multiple anterior and posterior processing streams, with anterior streams important for mapping phonetic cues to lexical representations, and posterior streams processing articulatory-gestural representations (Scott & Johnsrude, 2003). In this model, the auditory association cortex is structured hierarchically, extending outward from primary cortex. A region of lateral middle STG and middle STS processes complex acoustic signals including, but not specific to speech. Phonetic components of intelligible speech are processed in anterior STS (Scott, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006). In contrast, the posterior STS is associated with auditory-visual integration, input buffering of speech, and interaction with motor systems for speech in this model. Rauschecker and Scott (2009) recently presented a new adaptation of this model in which the dorsal stream links auditory percepts with articulatory acts using "internal models" via the inferior parietal lobule. This model maintains the hierarchical structure of auditory association cortex from the Scott and Johnsrude (2003) model, and suggests that regions of middle and anterior STS process intelligible speech, including syllables, with phonetic maps possibly implemented anteriorly.

Binder et al. (2000) also described a hierarchical arrangement of auditory association cortex extending outward from primary cortex. Analysis of complex feature combinations that define phoneme identity (corresponding to our definitions of phonetic and phonological processing) are processed in a long extent of the STS, spanning anterior, middle, and posterior portions. Areas extending ventrally, anteriorly, and posteriorly in the middle temporal gyrus, temporal pole, and angular gyrus (AG) provide access to lexical information. Both the Scott and Johnsrude (2003) and Binder et al. (2000) models are based largely on studies using word

and sentence level stimuli, but imply that the findings also apply to sublexical speech with regard to localization of phonetic processes. Posing a similar hierarchical arrangement, a recent review specifically addressing sublexical speech processing proposed a serial processing stream in temporoparietal cortex, which probabilistically abstracts information from speech at multiple levels (Obleser & Eisner, 2009). The dorsal STG, including Heschl's gyrus, performs initial acoustic analysis, and an STS region spanning anterior, middle, and posterior portions performs "pre-lexical abstraction," a probabilistic version of phonetic and phonological processing. The supramarginal gyrus (SMG) is implicated in phonological short-term memory, which is needed for detection of categorical differences between speech sounds.

Thus, these models predict slightly different localization of phonetic and phonological processing in sublexical speech perception. The first two (Belin & Zatorre, 2000; Hickok & Poeppel, 2007) predict that phonetic and phonological processing occur posteriorly, specifically in the mid-posterior STS in the Hickok and Poeppel (2007) model. The second two models (Rauschecker & Scott, 2009; Scott & Johnsrude, 2003) predict that middle to anterior STS is primarily responsible for phonetic processing. Finally, the last two models (Binder et al., 2000; Obleser & Eisner, 2009) predict a long span of STS including portions of anterior, middle, and posterior STS performs these processes. These models generally describe left lateralization of phonetic and phonological processes, although Binder (2000) and Hickok and Poeppel (2007) note involvement of the right hemisphere as well.

These subtly different predictions of localization may emerge in part as we attempt to merge differently conceived cognitive computations required for speech perception into our constructs of phonetic and phonological processing. Some differences may also result as authors of these models synthesize research using a wide variety of auditory stimuli in both humans and animals. Here, we aim to provide a precise specification of the human anatomic bases of some sublexical speech perception processes based on prior neuroimaging research.

Most neuroimaging studies of sublexical speech processing compare auditory perception of isolated phonemes or syllables to non-speech auditory control stimuli such as tones, music, or manipulated speech analogs. These studies activate the superior temporal cortex reliably, but the precise localization within this large region is variable. Some report anterior temporal activity (Obleser et al., 2006), while others report posterior activity (Desai, Liebenthal, Waldron, & Binder, 2008). Some of this variability undoubtedly results from intersubject variability, methodological differences in image acquisition and analysis, or other sources of localization uncertainty in neuroimaging research. Some of this variability may also reflect recruitment of different computational subprocesses within the speech perception network related to differences in the stimuli or experimental tasks. For example, activity in an fMRI study comparing perception of speech with non-speech control stimuli that are poorly matched on spectrotemporal variables (e.g. tones) could represent phonetic processing related to the speech content or acoustic processing due to the acoustic differences between the task and control stimuli. Activity in comparisons between speech and closely matched non-speech stimuli are much more likely to represent phonetic processing than acoustic processing. In these studies, phonological processing is engaged in relation to the use of abstract phonological representations, such that explicit tasks requiring judgments on speech sounds (e.g. same-different decisions) should recruit phonological systems to a greater degree than passive listening paradigms. Qualitatively, studies using explicit decision tasks and closely matched control stimuli result in more frequent activation of the left posterior STS than

other studies of sublexical speech perception (Desai et al., 2008), confirming models predicting posterior localization of phonological processes in the STS. Given the degree of variance in the data from imaging studies addressing this issue, a quantitative, statistical confirmation of this trend is desirable.

The degree of lateralization of temporal lobe activity is still a matter of debate as well. The initial spectrotemporal acoustic analysis of speech and other complex auditory signals is clearly performed bilaterally, with each hemisphere handling different aspects of the signal (Boemio, Fromm, Braun, & Poeppel, 2005; Jamison, Watkins, Bishop, & Matthews, 2006; Obleser, Eisner, & Kotz, 2008; Zatorre et al., 2002). The lateralization of later processing stages, including phonetic and phonological processes, is more controversial, but is generally thought to be left lateralized, at least to some degree (Rauschecker & Scott, 2009). Some analyses find bilateral activity, with greater left lateralization for words than pseudowords (Binder et al., 2000); others find left lateralization for sublexical speech units (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005). Presuming a left lateralization for sublexical perception, some studies only report results from the left temporal lobe (Mottonen et al., 2006). Hickok and Poeppel (2007), making a case for the bilateral basis of sublexical phonological processing, presented the results of several prior neuroimaging studies comparing perception of sublexical speech units to non-speech auditory stimuli. These studies activated the STG and STS bilaterally. Still, the activity was more extensive in the left compared to the right in some studies, and others failed to activate the right at all.

These basic questions of localization are not purely academic. The power of neuroimaging studies is limited by the necessity to correct for multiple comparisons in voxelwise statistical tests. Limiting analyses to regions consistently active during sublexical speech perception could greatly improve the sensitivity of fine-grained studies into specific subprocesses of perception. This will become increasingly important as speech researchers extend their methods to include powerful analysis techniques now used almost exclusively in vision research (Drucker, Kerr, & Aguirre, 2009). Clinically, localization of consistent findings in large groups of subjects may be important for future treatments of language disorders using brain stimulation techniques such as transcranial magnetic stimulation (Naeser et al., 2005).

The overarching goal of this analysis is to quantitatively specify the localization of sublexical speech perception components based on prior neuroimaging studies. We have four specific aims: (1) to specify the locations of consistent activity when perception of sublexical speech is compared to non-speech auditory stimuli, a paradigm that recruits acoustic, phonetic, and phonological processes; (2) to examine the localization of phonetic and phonological processes in these studies by examining specific subsets of speech versus non-speech imaging studies; (3) to assess the degree of lateralization of sublexical speech perception in the temporal lobes, and (4) to specify areas involved in categorical phoneme perception.

We present two meta-analyses using the Activation Likelihood Estimation (ALE) method, an objective, quantitative, voxelwise technique that has been validated, and is widely used in the neuroimaging community (Laird et al., 2005, 2005; Turkeltaub, Eden, Jones, & Zeffiro, 2002). Meta-analysis 1 addresses Aims 1–3 above using ALE and basic statistical tests on stereotactic coordinates collected from neuroimaging studies comparing perception of sublexical speech sounds to other non-speech auditory stimuli. Meta-analysis 2 addresses Aim 4 using ALE to analyze neuroimaging studies examining categorical phoneme perception using sublexical speech sounds.

## 2. Materials and methods

### 2.1. Selection of literature studies for meta-analysis 1

#### 2.1.1. Literature searches

We identified several studies comparing auditory perception of sublexical speech with non-speech stimuli from two recent review articles on speech processing (Hickok & Poeppel, 2007; Obleser & Eisner, 2009). We then searched the Pubmed and PsychINFO online databases for additional studies using the keywords “sublexical,” “pre-lexical,” “phonetic,” “phoneme,” “syllable,” “speech,” “perception,” “fMRI,” “PET,” “neuroimaging,” and “non-speech” in appropriate combinations. References of relevant publications were also reviewed for additional studies missed during the database searches. Directly searching for prominent researchers in the speech perception field yielded additional studies.

#### 2.1.2. Inclusion criteria

We selected studies for inclusion based on the following criteria: (1) imaging modality was fMRI or PET; (2) subjects were normal, healthy, right-handed adults; (3) experiment intended to identify areas more active during perception of auditory sublexical speech stimuli as compared to auditory non-speech stimuli; (4) task stimuli were natural or synthetic stimuli perceived as speech consisting of vowels (V), consonants (C), or non-word syllables; (5) control stimuli were auditory stimuli not perceived as speech; and (6) results were reported in stereotactic 3-dimensional coordinate system.

#### 2.1.3. Exclusion criteria

Studies were excluded from the meta-analysis using the following criteria: (1) case reports of single subjects; (2) control condition was rest or silence; and (3) analysis used pre-specified anatomically limited ROIs (ROI analyses based on functional activity obtained within the study were not excluded).

#### 2.1.4. Selection of experiments from included studies

Some studies meeting the above selection criteria reported multiple experiments using a single group of subjects comparing different sublexical speech conditions with different non-speech conditions (e.g. CV versus tones, and CV versus sine wave analogs). In these cases, only one experiment was selected from each group of subjects to avoid over-representation of these subjects in the meta-analysis. We selected these experiments using a simple heuristic. When multiple experiments were reported using different speech stimuli, we selected the most complex speech stimulus (e.g. CV rather than V). When experiments were reported using combinations of speech stimuli, we selected the experiment with the most speech content (e.g. CV + V + C rather than C only). When multiple experiments were reported using different control stimuli, we selected the control stimulus that most closely matched the speech stimulus acoustically (e.g. spectrally rotated speech analogs rather than pure tones). If these rules did not clearly favor a single experiment, we selected the experiment that most directly addressed the question “how does brain activity during perception of sublexical speech differ from perception of auditory non-speech sounds?”

### 2.2. Selection of literature studies for meta-analysis 2

#### 2.2.1. Literature search

We identified several studies on categorical phoneme perception during searches performed for meta-analysis 1. We identified additional studies using additional keyword search terms, including “categorical perception,” “categorization,” and “discrimina-

tion” in appropriate combinations. We then reviewed references of relevant publications for additional studies.

### 2.2.2. Inclusion criteria

We selected experiments for meta-analysis 2 based on the following inclusion criteria: (1) imaging modality was fMRI or PET; (2) subjects were normal, healthy, right-handed adults; (3) experiments intended to identify brain areas involved in categorical phoneme perception (i.e. areas that respond more to stimuli in different phoneme categories than stimuli in the same category); (4) task stimuli were natural or synthetic stimuli perceived as speech consisting of vowels (V), consonants (C), or non-word syllables (CV, VC, CVC, or CVVC); and (5) results were reported in stereotactic 3-dimensional coordinate system.

### 2.2.3. Exclusion criteria

Studies were excluded from the meta-analysis using the following criteria: (1) case reports of single subjects; (2) analysis used pre-specified anatomically limited ROIs (ROI analyses based on functional activity obtained within the study were not excluded); and (3) experiments that compared perception of different speech categories in which a phoneme discrimination experiment would be agnostic as to the assignment of which condition is task and which is control (e.g. voiced > unvoiced consonants).

### 2.2.4. Selection of experiments from included studies

As for meta-analysis 1, we included only one experiment from each subject group selected for the meta-analysis. Because the experimental design of these studies was quite varied, we selected the single experiment from each study that best addressed the question “which areas of the brain respond more to stimuli in different phoneme categories than stimuli in the same category?”

## 2.3. ALE methods for meta-analyses 1 and 2

The ALE technique is a widely used, validated, automated, quantitative method for voxelwise meta-analysis of neuroimaging foci reported in the literature. It is based on the idea that, although neuroimaging activity is reported as discrete foci in X Y Z coordinate space, there is some uncertainty as to the exact location of activity. The ALE technique models this localization uncertainty as a 3-dimensional Gaussian probability density distribution, and performs a logical OR function at each voxel in the brain. This voxelwise calculation essentially asks “what is the likelihood that at least one of the foci in the literature should have been reported at this voxel?” The result of this calculation is called the ALE value. Statistical significance is assessed against Monte Carlo style permutations using sets of foci distributed randomly within the brain volume. See Laird, McMillan et al. (2005), Laird et al. (2005), Turkeltaub et al. (2002) for a detailed descriptions of the ALE technique, and validation against fMRI and standard label-based meta-analysis.

ALE analysis was implemented using GingerALE 1.1 ([www.brainmap.org](http://www.brainmap.org)) and in-house C++ programs. We converted all MNI coordinates into Talairach space using the *icbm2tal* transform (Lancaster et al., 2007). A Gaussian full-width-at-half-maximum of 12 mm was used for all analyses. Twenty thousand permutations were performed for significance thresholding of all analyses. We determined significance using a conservative three-stage critical thresholding approach intended to limit type 1 error. First, a false discovery rate  $q$  threshold of .01 was applied. Next, to ensure that significant results truly represented coherence across multiple experiments rather than clustering of foci from a few studies, we eliminated voxels in which significant ALE values were driven by only one or two studies. To this end, we used in-house C++ programs to calculate the single-experiment contributions to

the ALE values at each voxel. Then we tabulated the number of experiments contributing at least  $1/N$  of the total ALE value at each voxel, where  $N$  is the number of experiments included in the analysis. A threshold of  $1/N$  was selected because under conditions of ideal coherence in the literature, each experiment would contribute  $1/N$  of the total ALE value at a given voxel (Turkeltaub et al., 2002). We then masked out voxels at which only one or two studies contributed to the total ALE value. Finally, we applied a cluster extent threshold of  $100 \text{ mm}^3$ . A promising alternative method for ensuring concordance across multiple studies in ALE analysis using a random effects statistics was recently published (GingerALE 2.0) (Eickhoff et al., 2009). We elected to use standard ALE analysis (GingerALE 1.1) here for several reasons: (1) We have not found the new method to be as stringent in requiring concordance across multiple studies as the thresholding method described here, (2) The new method does not allow for subtraction ALE analysis yet, (3) The new method calculates the width of Gaussian distributions, and may underestimate uncertainty in localization, and (4) The new method uses an analytic method for calculating statistical significance that has not yet been published.

Subtraction ALE analysis was used for *Phonetic/Phonological Analysis 2* and the *Laterality of Temporal Lobe Activity Analysis* (Laird et al., 2005). This technique allows direct statistical comparison between two sets of foci by subtracting the two ALE maps on a voxelwise basis, and then running permutations of that subtraction with sets of random foci to assess significance. FDR  $q < .01$  and cluster extent  $>100 \text{ mm}^3$  thresholds were used for these analyses. Single-experiment contributions cannot be calculated for a subtraction map, so we ensured concordance across three or more experiments in *Phonetic/Phonological Analysis 2* based on single experiment ALE maps of only the positive group in the subtraction (e.g. experiments 11–23 in *Phonetic/Phonological Analysis 2a*).

Other statistical tests on coordinates were implemented in SPSS 16. If Levene’s test for equality of variance was significant (at an alpha of .05),  $T$ -tests assumed unequal variance. Otherwise equal variances were assumed. An alpha of .05 was used for all tests. All  $P$ -values are 2-tailed. Visualizations were implemented with MRICron, using the Colin brain template in Talairach space. Renderings are maximum intensity projections with a search depth of 16 mm. Gyral anatomical labels were assigned based on the LONI Probabilistic Brain Atlas (LPBA40) maximum probability atlas (Shattuck et al., 2008). Activation likelihoods located deep in gray matter at boundaries between gyrus labels on the LPBA40 atlas were also labeled with the appropriate sulcus. For example, if the maximum probability label at a given peak was STG according to the LPBA40, and the activity was located deep in the gray matter at the boundary between the STG and MTG label fields, we used the label STG/STS in Table 2 and in the text. Brodmann Areas were assigned based on the Talairach Client using a  $\pm 2 \text{ mm}$  search cube or nearest gray matter if no Brodmann label was assigned within the cube range (Lancaster et al., 2000).

## 3. Results

### 3.1. Meta-analysis 1

#### 3.1.1. Localization of sublexical speech perception compared to non-speech (Aim 1)

We identified 23 experiments that met our inclusion and exclusion criteria for the speech versus non-speech analysis (Table 1). These experiments were published in 19 different papers, and included a total of 300 subjects. All experiments used either block design or event related fMRI. A variety of sublexical speech stimuli were used, including natural and synthetic vowels, consonants, and syllables. Likewise, a variety of control stimuli were used,

**Table 1**  
Meta-analysis 1 experiments.

ID#	Reference	N	Language	Speech stimulus	Control stimulus	Task type	Foci	Source of foci in paper
1	Benson et al. (2001)	12	American English	Natural and synthetic V, CV, CVC	Music	Monitoring	13	Table 3a increases
2	Benson, Richardson, Whalen, and Lai (2006)	12	American English	Synthetic Sine Wave CVC	Analog, music	Passive	2	Table 2A
3	Heinrich et al. (2008)	19	Canadian English	Synthetic V	Noise	Monitoring	2	Table 5 speech ROI
4	Jancke et al. (2002) Exp 1	9	German	Natural CV	Tones	Monitoring	5	Table 1 Exp 1
5	Jancke et al. (2002) Exp 2	6	German	Natural CV	Noise, Tones	Passive	5	Table 1 Exp 2 CV > (T + N)
6	Jancke et al. (2002) Exp 3	6	German	Natural CV	Noise	Passive	8	Table 1 Exp 3 CV > N
7	Obleser et al. (2006)	13	German	Natural V	Noise	Monitoring	4	Table 2 all vowels > BPN
8	Uppenkamp, Johnsrude, Norris, Marslen-Wilson, and Patterson (2006)	10	British English	Natural and Synthetic V	Analog	Monitoring	23	Table 2
9	Vouloumanos et al. (2001)	15	Canadian English	Natural CVC	Analog	Monitoring	3	Table 1 (a-ii)
10	Obleser, Zimmermann, Van Meter, and Rauschecker (2007)	13	American English	Natural C	Analog	Passive	3	Table 1 intelligible consonants > spectrally rotated analogues
11	Dehaene-Lambertz et al. (2005)	19	French	Synthetic Sine Wave CV	Analog <sup>a</sup>	Phonetic Decision	2	Table 1 1
12	Liebenthal et al. (2005)	25	American English	Synthetic CV	Analog	Phonetic Decision	7	Table 1
13	Desai et al. (2008)	28	American English	Synthetic Sine Wave CV	Analog <sup>a</sup>	Phonetic Decision	3	Appendix (b) post P > pre P
14	Joanisse and Gati (2003)	7	Canadian English	Synthetic CV	Tones	Phonetic Decision	9	Appendix speech > non-speech
15	Golestani and Zatorre (2004)	10	Canadian English	Synthetic CV	Noise	Phonetic Decision	5	Table 1 pretraining
16	Blumstein, Myers, and Rissman (2005)	13	American English	Synthetic CV	Tones	Phonetic Decision	1	Table 2 endpoint > tones
17	Hutchison et al. (2008)	14	American English	Synthetic CV	Tones	Phonetic Decision	5	Table 1 syllables > tones
18	Zaehle, Geiser, Alter, Jancke, and Meyer (2008)	16	Swiss German	Synthetic CV	Noise	Phonetic Decision	6	Table 1 speech > non-speech
19	Gandour et al. (2002a) Group 1	10	Thai	Natural Thai CVC, CWC	Hum	Phonetic Decision	5	Table 2 Thai VL versus D: speech versus non-speech
20	Gandour et al. (2002a) Group 2	10	Chinese	Natural Thai CVC, CWC	Hum	Phonetic Decision	2	Table 2 Chinese VL versus D: speech versus non-speech
21	Gandour et al. (2002b) Group 1	8	Thai	Natural Thai CVC, CWC	Hum	Phonetic Decision	4	Table 3 Thai vowel duration versus duration-hums
22	Gandour et al. (2002b) Group 2	8	English	Natural Thai CVC, CWC	Hum	Phonetic Decision	2	Table 3 English vowel duration versus duration-hums
23	Rimol et al. (2005)	17	Norwegian	Natural CV	Noise	Phonetic Decision	3	Table 2 CV > noise(CV)

"Analog" control stimuli are spectrally or temporally manipulated versions of speech stimuli perceived as non-speech.

<sup>a</sup> These experiments used the same sine wave stimuli for task and control conditions. They are initially perceived as non-speech, but subjects perceive them as speech after brief training. The comparison is stimuli perceived as speech versus stimuli perceived as non-speech.

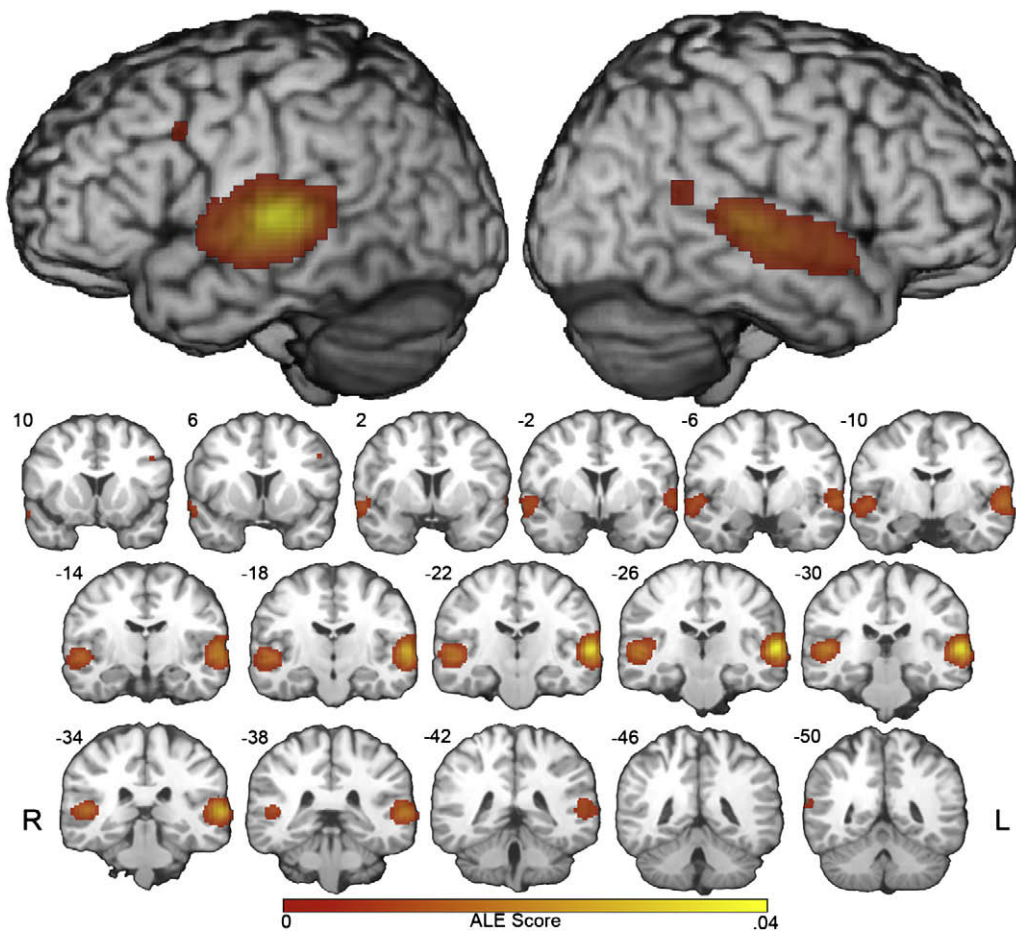
including pure tones, band-passed noise, music, hums, and manipulated speech analogs. Three general task types were used: (1) explicit decision tasks, including same-different judgments and phonological judgments (e.g. phoneme identification); (2) monitoring tasks to maintain subject attention; and (3) passive listening. These experiments reported a total of 122 foci of activity. We sorted the experiments based on task type (explicit decision versus passive or monitoring), control type (speech-analog versus lower level), and speech type (synthetic versus natural). All three experiments that used explicit decision tasks and speech-analog controls also used synthetic speech, confounding these variables somewhat.

ALE analysis of these 23 experiments resulted in four clusters of significant ALE values, representing non-random clustering of activation foci reported in the literature: large clusters in the bilateral STG extending into the STS, a small cluster in posterior right middle temporal gyrus and STS, and a small cluster in the middle frontal gyrus (Fig. 1, Table 2). Of these significant areas, the left STG achieved the highest activation likelihood. The anterior-posterior extent of this cluster ranged from Y of 2 to -46, including parts of the STG, STS, Heschl's gyrus, and planum temporale. The peak ALE value

was 0.039, which lay in the STG proper. This high ALE score denotes a striking degree of concordance in the 122 foci included in the analysis. For comparison, the maximum ALE value achieved in the original ALE analysis on single word reading was only 0.015, despite the inclusion of 172 foci (Turkeltaub et al., 2002). This left STG ALE peak derived from true coherence among experiments included in the meta-analysis. Ten out of 23 experiments met our threshold for a significant contribution to the left STG peak ALE value and seven other experiments contributed to a lesser degree (Table 2). The experimental design of the experiments did not impact their contributions greatly, indicating that this area was activated regardless of task type, stimulus type, or control type (Table 2, "Mean ALE Value Contribution Per Experiment").

The peak voxel in the right STG/STS was nearly homotopic to the left STG peak (52 -28 2 versus -58 -26 4), but the ALE value was only half that of the left (0.019 versus 0.039). Six of the 23 experiments contributed to this peak based on our strict cutoff, but five others made smaller contributions. In comparison with the left STG peak, this right STG/STS peak was driven more by experiments using passive tasks and lower-level control stimuli. The main right temporal cluster extended farther anteriorly than





**Fig. 1.** Results of the main speech versus non-speech analysis (Aim 1). Significant ALE clusters for the speech versus non-speech analysis. Top: surface rendering of the left and right hemispheres. Bottom: coronal slices every 4 mm from Talairach Y of 10 to  $-50$ . Displayed results are significant at FDR  $q < .01$ , 3 or more experiments contributing, cluster size  $>100 \text{ mm}^3$ .

the left temporal cluster (to a Y of 10). A small right posterior MTG/STS cluster, driven mainly by experiments 1 and 9, also extended more posteriorly than the left STG cluster (to a Y of  $-52$ ). This does not imply that the expected extent of right temporal activity for a given imaging study is greater than the extent of left temporal activity. As demonstrated in the laterality analysis below (Aim 3), this greater antero-posterior extent of ALE values is due to greater variability in the location of activation peaks in the right temporal lobe compared to the left.

The only extra-temporal cluster of activation likelihood occurred in the middle frontal gyrus, near the junction of the left precentral and inferior frontal sulci. The ALE value here derived mainly from three studies using lower-level control conditions (Experiments 1, 18, and 21).

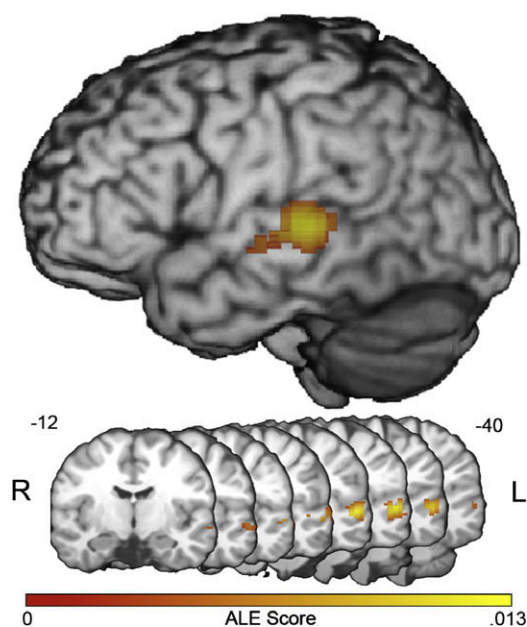
The significant voxels in this main speech versus non-speech analysis indicate the most consistent areas of activation in experiments comparing perception of speech with non-speech, but cannot be directly attributed to any specific cognitive process (e.g. acoustic, phonetic, phonological processing) because of the wide variety of stimulus types and tasks in this main analysis.

### 3.1.2. Localization of Phonetic and Phonologic Processing (Aim 2)

We performed two sub-analyses of meta-analysis 1 dataset to localize phonetic and phonological processing more specifically (Phonetic/Phonological Analyses 1 and 2). In *Phonetic/Phonological Analysis 1*, we performed an ALE analysis on a subset of the dataset including the six experiments using speech-analog control stimuli

which closely match speech on spectral and temporal features, and should result in relatively pure phonetic processing activity (Table 1, total 41 foci). Some phonological activity is expected here as well, to the degree that explicit tasks requiring access to phonological representations contribute to the significant ALE values. Experiment 6 was excluded from this analysis because it used both speech analog and low-level control conditions.

*Phonetic/Phonological Analysis 1* yielded a single cluster of activation likelihood in the left STG/STS, posterior and ventral to the center of the left STG cluster in the main ALE analysis (Fig. 2, Table 2). Based on the LPBA40 atlas, the cluster was centered on the upper lip of the STS and followed the course of the STS, but extended into the STG and MTG somewhat. The ALE peak in this analysis was located at a Talairach Y of  $-30$ , and significant ALE values ranged from a Y of  $-12$  to  $-40$ , spanning mainly the middle and posterior portions of the STS. The effect was driven primarily by the experiments using explicit decision tasks, although two of the experiments using passive tasks contributed to a lesser degree. The peak ALE value was 0.015, which is similar in magnitude to the peak right STG/STS voxel that was derived from the main analysis using 23 experiments and 122 foci. In light of the relatively small number of studies and foci included in this analysis, the ALE value represents a striking degree of concordance. We inspected the unthresholded *Phonetic/Phonological Analysis 1* ALE map to ensure that the absence of activation likelihood in the right hemisphere was not due to a threshold effect. There were no right hemisphere ALE peaks with contributions from more than one experiment.



**Fig. 2.** Results of *Phonetic/Phonological Analysis 1*. The significant left STG/STS phonetic processing ALE cluster is shown. Top: surface rendering of the left hemisphere. Bottom: coronal slices every 4 mm from Talairach Y of  $-12$  to  $-40$ . Slices are stacked anterior to posterior. Displayed results are significant at FDR  $q < .01$ , 3 or more experiments contributing, cluster size  $>100$  mm<sup>3</sup>.

Next, we reasoned that tasks requiring explicit attention to sub-lexical phonology would drive phonological processing areas to a greater degree than passive tasks. We performed *Phonetic/Phonological Analysis 2* in which a subtraction between the experiments using explicit decision tasks and those using passive or monitoring tasks was performed. The first group included 13 experiments reporting 54 foci, and the second group included 10 experiments reporting 68 foci. The six experiments from *Phonetic/Phonological Analysis 1* were split equally between these two groups.

This subtraction demonstrated that tasks requiring explicit attention to phonology resulted in higher activation likelihoods in the left posterior STS compared with passive tasks (Fig. 3, Table 2). As in *Phonetic/Phonological Analysis 1*, this effect was driven mainly by the experiments using speech-analog control stimuli and synthetic speech stimuli. The ALE peak was at a Y of  $-40$  in this analysis, and significant values extended from Y of  $-36$  to  $-44$ , overlapping with the cluster identified in *Phonetic/Phonological Analysis 1*. In contrast, passive tasks resulted in greater activation likelihood in the bilateral dorsolateral STG and right STS compared with tasks that require explicit attention to phonology. The maximum ALE value was essentially equal in the left and right STG. No significant differences between these groups were found in extra-temporal regions.

### 3.1.3. Laterality of temporal lobe activity (Aim 3)

The speech versus non-speech ALE analysis revealed qualitative differences between the left and right temporal lobes in magnitude and extent of activation likelihood. Next, we determined whether these perceived temporal lobe laterality effects were statistically significant. We multiplied the  $x$ -coordinates of the 122 speech versus non-speech foci by  $-1$ , effectively left–right reversing them. We then performed a subtraction ALE analysis between the original foci and the left–right reversed foci, and used a mask to specifically probe the temporal lobes. This method allows a direct voxelwise statistical comparison between left hemisphere and right temporal activations in the literature. The two temporal lobes are not perfectly symmetric anatomically, which may cause some

error in this analysis. The relatively low spatial resolution of ALE, especially using a 12 mm Gaussian model as we have, mitigates this concern considerably.

The laterality analysis confirmed that the previously described peak ALE value hemispheric asymmetry was statistically significant (Fig. 4, Table 2). As the peak ALE values were localized to nearly homotopic regions of the STG, this left lateralization cannot be accounted for by a difference in peak location. Nor could anatomical asymmetries between the temporal lobes account for the large difference in peak ALE values between the hemispheres.

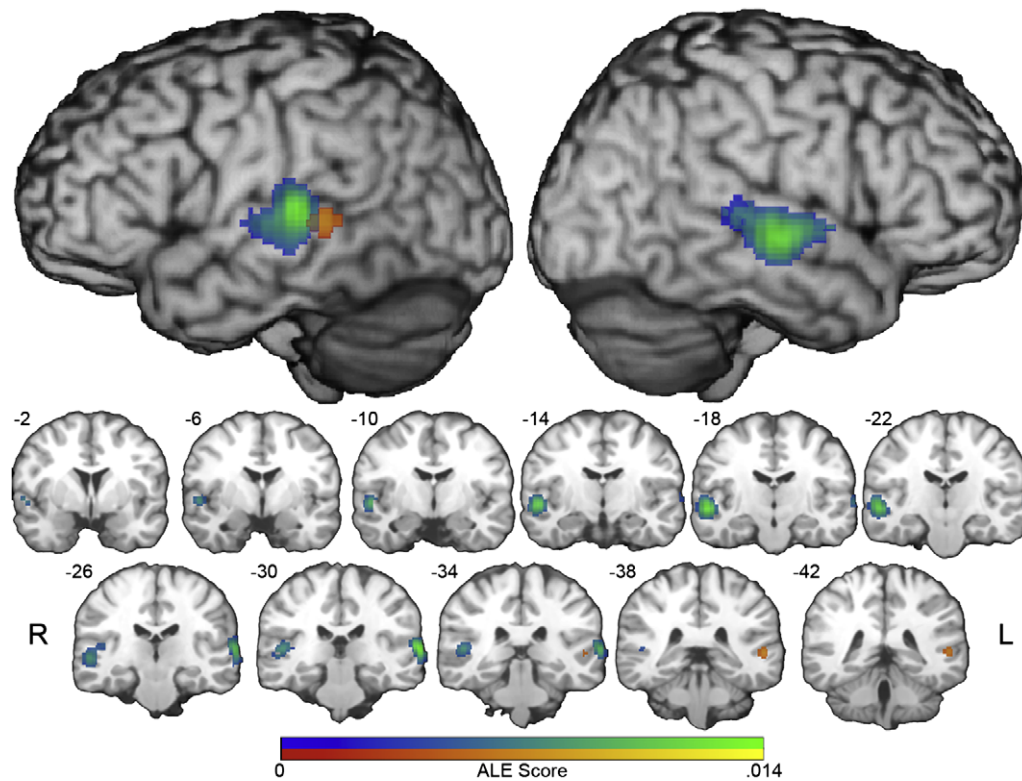
Although the peak activation likelihood in STG is left lateralized, Fig. 4 demonstrates that right lateralized clusters flank this STG peak region. This pattern of a central cluster of left lateralization surrounded by peripheral clusters of right lateralization implies that the left temporal foci from the literature are more densely clustered near the left STG peak in the ALE map, while the locations of right temporal foci are more variable or that the processing in the right hemisphere is more distributed. More variable localization of foci in the right temporal lobe could also explain why the anterior–posterior extent of the right temporal ALE clusters was greater than that of the left in the main speech versus non-speech analysis (Aim 1).

To test this hypothesis, we identified all the foci within the temporal lobe based on the nearest gray matter using the Talairach Client (Lancaster et al., 2000). This yielded 47 left temporal foci from 22 experiments, and 34 right temporal foci from 16 experiments ( $p_{(\text{binomial})} = .09$ ). Next, we calculated the Euclidean distance of each focus to the peak left or right STG voxel in the speech versus non-speech ALE map (from Aim 1). The left temporal foci were more closely clustered around the STG peak than the right temporal foci were, and their distance to the peak was less variable (mean distance to peak – left: 12.5 mm, SD 6.7 mm; right: 17.9 mm, SD 10.8 mm;  $T(51) = 2.57$ ,  $p = .01$ ). This confirms that right temporal lobe foci are less consistent in their localization than left temporal lobe foci in speech versus non-speech comparisons.

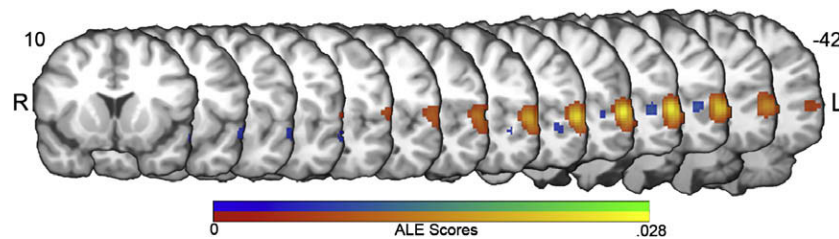
The ALE method does not consider the magnitude or the extent of the activity reported in the literature because the units and scales of these measures vary considerably across studies. Still, these values could reveal important information regarding lateralization of activity. To evaluate this possibility, we simply noted whether the magnitude and extent of reported peak activity was greater in the left temporal lobe or right temporal lobe for each experiment. If activity was only reported in one hemisphere, both the magnitude and extent of activity were deemed greater for that hemisphere. Studies not reporting these values were excluded. Fifteen of 21 experiments reported greater magnitude of activity in the left temporal lobe ( $p_{(\text{binomial})} = .04$ ). Thirteen of 15 experiments reported greater extent of activity in the left temporal lobe ( $p_{(\text{binomial})} = .004$ ). Again, this demonstrates that longer anterior–posterior extent of activation likelihoods in the right temporal lobe compared to the left in the main speech versus non-speech analysis resulted from variable localization of right temporal activity across studies, and does not predict a greater extent of right temporal activity in a given imaging experiment.

### 3.2. Meta-analysis 2 (Aim 4)

We identified eight experiments that met our inclusion and exclusion criteria for the categorical phoneme perception analysis (Table 3). These experiments were published in eight different papers, and included 123 subjects. They reported a total of 51 foci of activity. All experiments used either block design or event related fMRI, with a variety of experimental paradigms. The source of activation contrast varied considerably between experiments, including same-different phoneme comparisons which directly contrast between- and within-category phoneme pairs; release of habituation



**Fig. 3.** Results of *Phonetic/Phonological Analysis 2*. Areas of greater activation likelihood for experiments using decision tasks than experiments using passive or monitoring tasks are shown in flame scale. Areas of greater activation for passive or monitoring tasks than decision tasks are shown in the blue–green cool scale. Top: surface renderings of the right and left hemispheres. Bottom: coronal slices every 4 mm from Talairach Y of  $-2$  to  $-42$ . Displayed results are significant at  $FDR q < .01$ , 3 or more experiments contributing to the positive side of the subtraction, cluster size  $>100 \text{ mm}^3$ .



**Fig. 4.** Results of *Laterality of Sublexical Speech Processing Analysis (Aim 3)*. Significant clusters from the subtraction ALE analysis contrasting the left and right temporal lobes in the speech versus non-speech dataset overlaid on the left hemisphere. Regions in which activation likelihoods were greater in the left hemisphere than right are shown in red–yellow flame scale. Regions in which activation likelihoods were less in the left hemisphere than the right are shown in blue–green cool scale. The right hemisphere is not shown here because it would appear as a perfect mirror image of the left hemisphere. Coronal slices are shown every 4 mm from Talairach Y of 10 to  $-42$ . Slices are stacked anterior to posterior. Displayed results are significant at  $FDR q < .01$ , cluster size  $>100 \text{ mm}^3$ .

designs, which induce neural habituation to within-category phoneme sequences and then assess activity induced by releasing that habituation using a phoneme from a new category; multidimensional pattern classification, which assesses differences in patterns of activation for within-category and between-category phonemes; and neural amplification/correlation with behavioral measures, which identifies brain areas in which activity to phoneme sequences either within- or between-category correlate with phoneme discrimination scores measured outside the scanner. While these approaches vary considerably, each aims to identify brain activity that is greater for between-category phoneme comparisons than within-category comparisons. One additional experiment met our selection criteria but was not included because it did not identify any significant activity (Hutchison, Blumstein, & Myers, 2008). Another experiment could not be included because it did not report stereotactic coordinates for results, and the authors were not available

to provide them upon request (Formisano, De Martino, Bonte, & Goebel, 2008).

ALE analysis of the eight experiments yielded two significant clusters of activation likelihood: the left SMG (peak  $-48, -34, 26$ ; BA 40; extent  $1624 \text{ mm}^3$ , ALE value 0.011), and the left AG (peak  $-38 -60 42$ ; BA 39/7; extent  $568 \text{ mm}^3$ ; ALE value 0.006; Fig. 5). Three out of eight experiments contributed to each of the two clusters (Table 3). Neither cluster derived solely from experiments using the same source of imaging contrast, thus demonstrating consistent localization despite varying experimental designs. However, two studies (Raizada & Poldrack, 2007; Zevin & McCandliss, 2005) contributed the majority of the ALE value to the peak of each cluster. These two studies reported more foci than the other six studies combined. Thus, while these results are statistically significant even using our conservative thresholding methods, they may not demonstrate robust coherence across the

**Table 3**  
Meta-analysis 2 experiments.

ID#	Reference	N	Language	stimulus type	Source of contrast signal	Task	Foci		Percent of total ALE value contributed by each experiment
							SMG peak	AG peak	
24	Celsis et al. (1999)	6	French	Natural CV – sequence of 4	Habituation – last item discrepant versus repeated	Listening	1		0.3
25	Husain et al. (2006)	12	American English	Synthetic V and CV – pairs	Different versus same pairs	Phoneme discrimination	7		15.3
26	Dehaene-Lambertz et al. (2005)	19	French	Synthetic CV – sequence of 4	Habituation – last item discrepant versus repeated	Phoneme discrimination	2	21.7	2.7
27	Desai et al. (2008)	28	American English	Synthetic CV – triads	Correlation with “categorical perception index”	Phoneme Discrimination	2	0.3	
28	Myers, Blumstein, Walsh, and Eliassen (2009)	18	American English	Synthetic CV – sequence of 5	Habituation – last item discrepant versus repeated	Monitoring for high pitched target	1		
29	Zevin and McCandliss (2005)	8	American English	Natural CV – sequence of 4	Habituation – last item discrepant V5 repeated	Listening	22	46.9	25.8
30	Raizada, Tsao, Liu, and Kuhl (2009)	20	10 American English, 10 Japanese	Synthetic CV	Multidimensional pattern classification – American versus Japanese /ra/-/la/ identification	Phoneme identification	3		
31	Raizada and Poldrack (2007)	12	American English	Synthetic CV – pairs	Neural amplification (correlation with discrimination score)	Monitoring for low volume target	13	31.1	55.9

body of literature on categorical phoneme perception. There was no spatial overlap of significant ALE values between meta-analyses 1 and 2.

#### 4. Discussion

We used advanced neuroimaging meta-analysis techniques to examine the localization of sublexical speech perception in normal adults. The main goal of this research was to quantitatively localize sublexical speech processes based on previously published neuroimaging studies. We listed four aims for this research, and will discuss them in turn below.

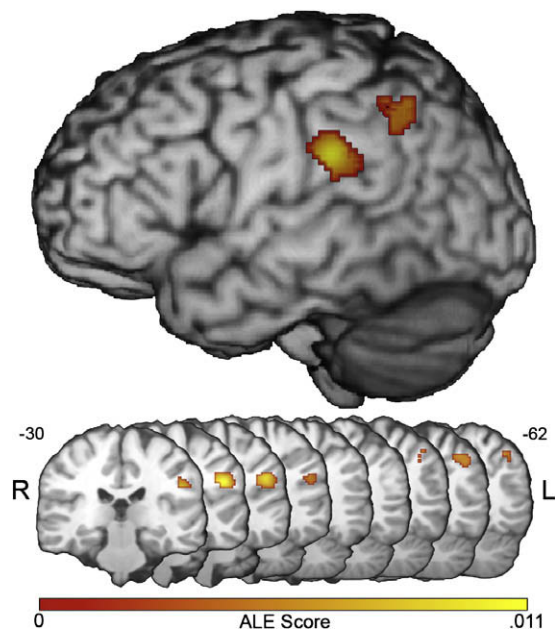
##### 4.1. Aim 1: specifying brain areas more responsive to sublexical speech sounds than non-speech sounds

This aim was addressed by the main speech versus non-speech analysis in meta-analysis 1, which identified significant activation likelihoods primarily within the bilateral STG, extending into the STS. This result provides quantitative support for current models of speech processing (Binder et al., 2000; Hickok & Poeppel, 2007; Obleser & Eisner, 2009; Rauschecker & Scott, 2009; Zatorre et al., 2002). Furthermore, by providing precise specification of the anatomic locus of this effect, the results may be useful for a variety of purposes, including targeting regions for non-invasive brain stimulation studies on speech perception, and defining regions-of-interest to address specific hypotheses in future imaging studies. These significant clusters of activation likelihoods represent the most consistent locations of activation peaks among neuroimaging studies on sublexical speech perception. They could be used to predict the most likely areas of peak localization in a new imaging study, but do not necessarily represent the likely extent of activations for any given imaging study.

Areas responding more to speech sounds than other auditory stimuli could perform a number of functions, including acoustic analysis, phonetic processing, phonological processing, lexical access, semantic access, or articulatory preparation. However, in this analysis we intentionally limited recruitment of lexical and semantic systems by only examining perception of sublexical speech sounds. Thus, cortical areas identified in Aim 1 could perform acoustic analysis of complex auditory stimuli, phonetic or

phonological processing, or articulatory preparation. Because of the relatively low spatial resolution of ALE, specific peaks of ALE values related to each of these cognitive processes are not separable in this main analysis. Thus the peak locations in these main ALE clusters represent the mostly likely location of reported activity in sublexical speech versus non-speech experiments, but should not be viewed as the locus for any one specific cognitive function related to speech perception. However, careful interpretation of which experiments contribute to specific areas within these large clusters can help delineate some finer structure-function relationships.

Most speech models assign the bilateral dorsal STG a role in spectrotemporal (i.e. acoustic) analysis of speech and other complex auditory stimuli (Binder et al., 2000; Hickok & Poeppel, 2007; Obleser & Eisner, 2009; Rauschecker & Scott, 2009; Scott & Johnsrude, 2003; Zatorre et al., 2002). Dorsal superior temporal lobe areas, including Heschl's gyrus and planum temporale, are sensitive to spectral and temporal features of non-speech auditory stimuli (Boemio et al., 2005; Griffiths & Warren, 2002; Overath, Kumar, von Kriegstein, & Griffiths, 2008). Our results in this analysis are consistent with this organization. The significant clusters of activation likelihood in the main speech versus non-speech analysis encompassed the bilateral dorsal STG and also the STS reflecting the wide range of cognitive constructs assayed in these studies, as mentioned above. In contrast, when we attempted to eliminate acoustic processes by only examining those studies using acoustically well-matched control stimuli in *Phonetic/Phonological Analysis 1*, the dorsal STG activation likelihood disappeared, leaving only left posterior STS and ventrolateral STG activation likelihoods. The bilateral dorsolateral STG activation likelihoods observed for passive tasks in *Phonetic/Phonological Analysis 2* compared with active tasks might also support this arrangement if directing attention to phonology using explicit decision tasks draws attention away from lower level acoustic aspects of speech perception. Some studies we included in this meta-analysis also reported dorsal superior temporal activity for non-speech stimuli in comparison to baseline control conditions (Heinrich, Carlyon, Davis, & Johnsrude, 2008; Jancke, Wustenberg, Scheich, & Heinze, 2002; Rimol, Specht, Weis, Savoy, & Hugdahl, 2005; Vouloumanos, Kiehl, Werker, & Liddle, 2001). Thus dorsal STG areas within the ALE clusters we identified in the main speech versus non-speech analysis may relate to acoustic processing not specific to speech. Right STG/STS activation likelihoods in particular



**Fig. 5.** Results of meta-analysis 2 (Aim 4). Significant clusters from the meta-analysis of experiments examining categorical discrimination of phonemes are shown in flame scale. Top: surface rendering of the left hemisphere. Bottom: coronal slices are shown every 4 mm from Talairach Y of  $-30$  to  $-62$ . Slices are stacked anterior to posterior. Displayed results are significant at  $FDR q < .01$ , 3 or more experiments contributing, cluster size  $>100 \text{ mm}^3$ .

derived mainly from experiments using passive tasks and lower-level control stimuli, suggesting a primary role in acoustic analysis.

Outside the temporal lobes, only a small region of the left MFG near the precentral and inferior frontal sulci demonstrated significant activation likelihood for the comparison of sublexical speech versus non-speech. This cluster localized just anterior to motor areas for aloud reading and syllable singing, at the same dorsal-ventral level (Brown et al., 2009; Turkeltaub et al., 2002). Thus, this cluster may correspond to premotor cortex for the vocal apparatus. Such activity could occur due to direct involvement of motor planning areas in sublexical speech perception, or due to silent articulation of speech stimuli in the scanner despite task instructions. The latter seems less likely, given that no primary motor activation likelihood was identified. The activation likelihood in this cluster derived from studies comparing perception of syllables to music, noise, and hums using monitoring and decision tasks. Its specific cognitive function cannot be determined by this meta-analysis.

The involvement of motor systems in speech perception was suggested over 40 years ago (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), and is still a matter of ongoing research and debate (Devlin & Aydelott, 2009). Transcranial magnetic stimulation studies have recently demonstrated excitation of primary motor mouth areas during speech perception, and modulation of phoneme discrimination after TMS to primary motor cortex (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Mottonen & Watkins, 2009; Roy, Craighero, Fabbri-Destro, & Fadiga, 2008; Watkins, Strafella, & Paus, 2003). Two other studies excluded from this analysis also found premotor activity during sublexical speech perception (Pulvermuller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004), and TMS studies have demonstrated alteration of phoneme discrimination tasks by premotor stimulation (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Sato, Tremblay, & Gracco, 2009). All of these studies, however, implicated a premotor area superior and posterior to the cluster identified here. Further studies could specifically probe the location identified in this analysis

to assess whether it is indeed a premotor area for the mouth, and whether it plays any causal role in speech perception.

#### 4.2. Aim 2: localizing phonetic and phonological processes

The models of speech perception we outlined in Section 1 indicate that the STS operates in phonetic and phonological processing of speech. Some models specify a mid-posterior localization of this processing unit (Belin & Zatorre, 2000; Hickok & Poeppel, 2007). Others suggest a mid-anterior localization (Rauschecker & Scott, 2009; Scott & Johnsrude, 2003), and some models imply a longer anterior-posterior extent, centered ventral to primary auditory cortex (Binder et al., 2000; Oleser & Eisner, 2009). Most specify left lateralization of STS phonetic and phonological processes, but Hickok and Poeppel (2007) suggest some degree of bilaterality. *Phonetic/Phonological Sub-analyses 1 and 2* were intended to quantitatively confirm that the STS processes phonetic and phonological aspects of speech, and to specify the localization in terms of laterality and anterior-posterior localization. *Phonetic/Phonological Analysis 1* should primarily address phonetic processes because studies were selected based on their phonetic content, not on whether access to phonological representations were required by the tasks. In contrast, *Phonetic/Phonological Analysis 2* should primarily address phonological processes because experiments that require attention to phonology were contrasted with those that do not. Unfortunately, phonetic and phonological processes are not easily separable in speech versus non-speech experiments and both functions likely contribute somewhat to each of these analyses, especially since the same studies using closely matched control stimuli and explicit tasks contributed significantly to both results. These analyses yielded activation likelihoods in the left mid-posterior STS, at a Talairach Y of  $-30$  for *Phonetic/Phonological Analysis 1* and a Y of  $-40$  for *Phonetic/Phonological Analysis 2*. There was no ambiguity in this lateralization, as there was virtually no consistent right hemisphere activity in these experiments. These analyses thus confirm speech models that propose a mid-posterior site in the left STS that processes phonetic and phonological aspects of speech. The relative location of the peaks of these two analyses could imply a more posterior localization of phonological processes relative to phonetic processes, but should not be over interpreted given the caveats above.

Since this meta-analysis excluded experiments using word level speech units to avoid contamination by lexical and semantic content, it remains possible that STS sites for phonetic and phonological processing of lexical speech extend farther anteriorly than the area we identified here. Models proposing a greater anterior extent of phonetic processes frequently cite neuroimaging studies using lexical or sentence level auditory speech (Binder, 2000; Binder et al., 2000; Scott & Johnsrude, 2003; Scott et al., 2000, 2006). An additional meta-analysis aiming to identify phonetic and phonological processing areas in lexical speech perception would help to confirm that these effects are consistent *en mass*. If more anterior areas of STS do participate in phonological analysis of lexical speech specifically, further imaging, stimulation, or patient studies could assess whether semantic content or lexicality primarily drives this anterior extension.

A limitation of these meta-analyses is the cognitive imprecision of any speech versus non-speech comparison. For example, articulatory preparation may occur subconsciously even though articulation is not explicitly required in speech perception tasks. More importantly, even when closely matched speech analogs are used account for acoustic differences between task and control stimuli, enough acoustic difference must remain for the subjects to perceive task stimuli as speech rather than non-speech. Regions of the brain performing spectral and temporal acoustic analysis regardless of speech status might then respond preferentially to

speech partly because it is acoustically different from non-speech sounds. Two experiments in our analysis did match the acoustic content of their stimuli perfectly by using sine wave speech analogs that are initially perceived as non-speech, but then switch to speech perception after multiple exposures or brief training (Dehaene-Lambertz et al., 2005; Desai et al., 2008). Comparing stimulus-induced activity after the perceptual shift to that before the shift should eliminate the possibility that resulting activity is based on acoustic differences. In this comparison, both studies activated a posterior region of the left STS ( $y = -40$ ,  $y = -43$ ), and also the left middle to mid-posterior STG ( $y = -24$ ,  $y = -33$ ). Thus, in two studies, this acoustically well-controlled comparison yielded only left hemisphere activity, in accordance with our phonetic and phonological processing analyses. Indeed, these two studies made significant contributions to the results of both Phonetic/Phonological Analyses, which could be considered a strength of these analyses given the cognitive precision of these experiments, and a caveat, given that sine wave speech differs substantially from natural speech in its spectrotemporal properties. These two experiments also identified activity in the left STG region more typically ascribed to acoustic analysis. The top-down mechanisms responsible for the perceptual shift from non-speech to speech perception may serve to tune regions responsive to spectral and temporal features of the stimuli, resulting in differential activity in acoustic processing areas. Still, this finding should be noted as a caveat to the attribution of purely acoustic functions to the dorsal STG.

#### 4.3. Aim 3: lateralization of sublexical speech perception

This meta-analysis provides quantitative specification of the degree and nature of lateralization of temporal lobe sublexical speech processing. We did not find a significant difference between the hemispheres in the frequency of activation, although there was a trend toward left lateralization. We did find that left temporal activity during sublexical speech perception is of greater magnitude and extent than right temporal activity and is more consistently localized to the mid STG. The main bilateral STG peaks of activation likelihood are nearly homotopic, but right STG and STS activation likelihoods derive mainly from experiments using passive tasks and lower-level control conditions, like tones, noise, or hums. In contrast, the nearly homotopic left STG area is activated across all varieties of experimental designs. When tasks requiring explicit attention to phonology are removed (as in *Phonetic/Phonological Analysis 2*), the activation likelihood of the STG is essentially equal in magnitude in each hemisphere, but the extent is greater in the right. These findings are consistent with a bilateral superior temporal acoustic processing system used for sublexical speech sounds. Overall activity is clearly left lateralized by multiple measures. The right STG function is not likely specific to speech, and is mainly acoustic in nature. The left STG is operant in acoustic processes but phonological demands drive activity here also. Lateralization of STG activity results mainly from this effect. As discussed above, phonetic and phonological processes are clearly left lateralized in the mid-posterior STS, and we found no evidence for any consistent phonetic activity in the right temporal lobe in *Phonetic/Phonological Analysis 1*. As discussed below, the results of meta-analysis 2 in the present study, and prior electrophysiological evidence also support left hemisphere dominance for some aspects of categorical phoneme perception (Dehaene-Lambertz, 1997; Naatenen et al., 1997; Schofield et al., 2009).

Because these findings are derived purely from neuroimaging data in healthy subjects, no assumptions can be made regarding the necessity of either temporal lobe for sublexical speech perception. The lesion data on sublexical speech perception is informative in this matter, but far from definitive. Interruption of either hemisphere during a Wada procedure on a patient with epilepsy did not

impair phonemic discrimination (Boatman et al., 1998), indicating bilateral processing. The seizure semiology in this case, however, suggested an ictal focus in the left temporal lobe, which could have resulted in significant reorganization of speech networks. Left internal carotid WADA induced mainly semantic errors on an auditory comprehension task in 20 epilepsy patients with left lateralized language (Hickok et al., 2008). Still, the absolute number of phonological errors was greater after left carotid anesthesia compared to right. Patients with fluent and non-fluent aphasias, who generally have left hemisphere lesions, clearly have deficits in word and sentence comprehension, but their sublexical speech perception deficits are more nuanced and do not predict their comprehension (Basso, Casati, & Vignolo, 1977; Blumstein, Baker, & Goodglass, 1977). Generally, aphasic patients perform poorly on phoneme identification tasks (Baum, 2001; Blumstein, Burton, Baum, Waldstein, & Katz, 1994; Blumstein, Cooper, Zurif, & Caramazza, 1977; Boyczuk & Baum, 1999; Gow & Caplan, 1996; Ravizza, 2003). In contrast, phoneme discrimination performance, although not normal, is relatively preserved (Blumstein et al., 1977; Gow & Caplan, 1996). This pattern suggests that access to phonological representations, phoneme labeling, or categorization is impaired, but non-categorical acoustic or phonetic processing is relatively preserved. In contrast, pure word deafness disrupts both phoneme discrimination and identification (Saffran, Marin, & Yeni-Komshian, 1976). Dichotic listening experiments have suggested right hemisphere language processing in these patients, indicating injury or disconnection of left temporal cortex (Geschwind, 1965; Saffran et al., 1976). Cases since the widespread use of CT and MRI have more often been associated with bilateral lesions to auditory association cortex (Coslett, Brashear, & Heilman, 1984; von Stockert, 1982), although left hemisphere and bilateral subcortical lesions involving the auditory radiations have also been reported (Hayashi & Hayashi, 2007). The lack of high-resolution structural imaging in the aphasia reports, and the paucity of pure word deafness cases limit interpretation of laterality based on the lesion literature. Still, these reports are generally consistent with our meta-analysis findings, reflecting a bilateral organization of acoustic perception of complex auditory signals, with left lateralization of phonological and categorical phonemic processes. Further studies using modern imaging and analysis techniques could clarify some of the ambiguity in the lesion literature (Bates et al., 2003).

#### 4.4. Aim 4: perceptual discrimination between phoneme categories

Meta-analysis 2 investigated the neural basis of categorical phoneme perception, specifically the perception of differences between phoneme categories. We found significant activation likelihoods in both the left SMG and AG. However, two experiments that reported the majority of foci in our analysis drove this effect. Thus, while the results are statistically significant, they do not indicate robust coherence across the literature on categorical phoneme perception. Still, our findings agree with the synthesis by Obleser and Eisner (2009), at least with regard to the SMG. One of the four studies from their analysis was excluded from ours (Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003), and another did not contribute significantly to our SMG cluster (Celsis et al., 1999). One study not cited by Obleser and Eisner did contribute to the SMG cluster (Zevin & McCandliss, 2005). Thus, our findings, although not particularly robust, do provide some quantitative confirmation and extension of Obleser and Eisner's synthesis. Further confirmation with additional studies would be useful.

It is not clear what role the left SMG or AG might play in categorical perception of speech sounds. One possibility is that they are active in phoneme discrimination tasks because they play a role in phonological working memory (Baddeley, 2003; Jacquemot & Scott, 2006), which is required for these tasks (Obleser & Eisner, 2009). However, the two experiments in meta-analysis 2 that

contributed most of the ALE value to the AG and SMG clusters used passive tasks that should not place a heavy load on phonological working memory (Raizada & Poldrack, 2007; Zevin & McCandliss, 2005). Alternatively, the left SMG and AG might serve as a “sensorimotor sketchpad” for conveying predictive modeling information between motor and perceptual areas during speech perception and production (Rauschecker & Scott, 2009). Phoneme discrimination could be performed via comparison between phonological features and models of the articulatory plan needed to produce a given speech sound.

The left inferior parietal lobe might also play a domain general role in either consolidating continuous features of percepts or concepts into categories or in comparison between stimuli during discrimination tasks. The AG is sometimes viewed as an amodal integrative processing hub because it receives input from associational areas but not primary sensory cortices, and lesions to it produce a wide variety of deficits in multiple domains (Binder, Desai, Graves, & Conant, 2009; Geschwind, 1965). In addition to phoneme discrimination, the left SMG has been implicated in processing categorical spatial relations (Amorapanth, Widick, & Chatterjee, 2009; Kemmerer, 2006; Noordzij, Neggers, Ramsey, & Postma, 2008; Tranel & Kemmerer, 2004), and in social stereotyping, which requires conceptual categorization of people (Quadflieg et al., 2009). Categorical discrimination of colors does not involve the left SMG, but is left lateralized, possibly due to linguistic labeling (Franklin et al., 2008; Roberson & Hanley, 2007; Ting Siok et al., 2009). Perhaps then the left hemisphere is involved in categorization because of category labeling, and the SMG is involved in categorization specifically for dorsal stream domains. This hypothesis is speculative, but could be systematically tested in the future.

Regardless of the specific roles of the left SMG and AG in categorical perception of phonemes, it is notable that neither of these areas were implicated in sublexical speech perception by meta-analysis 1. The lack of spatial overlap between these two analyses could suggest that categorical perception of phonemes is not involved in sublexical speech perception, at least as each are operationalized in neuroimaging studies. The lack of overlap between these two analyses should not be over interpreted though, in light of the limitations discussed below. Further research is warranted specifically probing the causal role of left parietal cortex in speech perception.

#### 4.5. Limitations

Because meta-analysis requires a degree of abstraction and averaging of the original research under consideration, our interpretation is limited to a relatively low “cognitive resolution.” That is to say, it is difficult to precisely identify a single cognitive process performed by an area of significant activation likelihood derived from multiple different experiments with different parameters. For this reason, we used strict criteria to select relatively homogeneous groups of experiments, which improved the precision of the analyses. Where possible, we also examined the contributions of single experiments to significant ALE clusters in order to maximize our precision, but this retrospective approach is still limited.

Our strict inclusion/exclusion criteria improved the specificity of the analyses, but limited the number of studies included, particularly in *Phonetic/Phonological Analysis 1* and *meta-analysis 2*. Small Ns in ALE analyses can lead to false positive results generated by a few foci clustering purely due to chance. To guard against false positives, we used a strict critical thresholding scheme designed to ensure that all significant results represent true convergence across multiple independent experiments. The striking degree of concordance we observed across experiments further mitigates

concern over false positive results in *Phonetic/Phonological Analysis 1*, but this issue is still a concern regarding *meta-analysis 2*.

Whereas our strict thresholding approach is likely to enhance the reliability of our significant ALE findings, our methods could make the analyses vulnerable to false negative results. For this reason, when claiming that we found no evidence of phonetic or phonological processing in the right hemisphere, we carefully examined the unthresholded ALE map to ensure that we were not missing subthreshold activation likelihoods. Still, a degree of caution is warranted any time one interprets a negative result.

## 5. Summary

Collectively, our results are consistent with a sublexical speech processing network in which the bilateral dorsal superior temporal cortex performs early acoustic analysis of speech and other auditory stimuli, whereas the left mid-posterior STS performs phonetic and phonological processing of speech. The left inferior parietal lobule, including the SMG and AG, is operative in categorical phoneme perception. This organization clarifies current models of speech processing, by specifying that phonetic and phonological STS processing is left lateralized and located in the mid-posterior portion, and that both the SMG and AG may be involved in phoneme discrimination. Furthermore, we have specified the central location and extent of consistent activation peaks within the superior temporal cortex regions involved in sublexical speech perception. Further investigation using advanced neuroimaging analysis techniques, non-invasive brain stimulation, and lesion analysis, in parallel with animal research, will help to further clarify this network.

## Acknowledgments

We thank Jose Maisog for providing C++ programs necessary for this analysis, and David Poeppel for his advice. This work was supported by the American Academy of Neurology Foundation (Clinical Research Training Fellowship to P.E.T.) and The National Institute for Child Health and Human Development (R24 HD050836). The authors have no conflicts of interest to declare.

## References

- Amorapanth, P. X., Widick, P., & Chatterjee, A. (2009). The neural basis for spatial relations. *Journal of Cognitive Neuroscience*.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839.
- Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 13(1), 85–95.
- Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*, 6(5), 448–450.
- Baum, S. R. (2001). Contextual influences on phonetic identification in aphasia: The effects of speaking rate and semantic bias. *Brain and Language*, 76(3), 266–281.
- Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience*, 3(10), 965–966.
- Benson, R. R., Richardson, M., Whalen, D. H., & Lai, S. (2006). Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *NeuroImage*, 31(1), 342–353.
- Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, 78(3), 364–396.
- Binder, J. (2000). The new neuroanatomy of speech perception. *Brain: A Journal of Neurology*, 123(Pt 12), 2371–2372.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(9), 1991.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10(5), 512–528.
- Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, 15(1), 19–30.

- Blumstein, S. E., Burton, M., Baum, S., Waldstein, R., & Katz, D. (1994). The role of lexical status on the phonetic categorization of speech in aphasia. *Brain and Language*, 46(2), 181–197.
- Blumstein, S. E., Cooper, W. E., Zurif, E. G., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, 15(3), 371–383.
- Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17(9), 1353–1366.
- Boatman, D., Hart, J., Jr., Lesser, R. P., Honeycutt, N., Anderson, N. B., et al. (1998). Right hemisphere speech perception revealed by amobarbital injection and electrical interference. *Neurology*, 51(2), 458–464.
- Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8(3), 389–395.
- Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S., & Sprenger-Charolles, L. (2008). Discrimination of speech sounds by children with dyslexia: Comparisons with chronological age and reading level controls. *Journal of Experimental Child Psychology*, 101(2), 137–155.
- Boyczuk, J. P., & Baum, S. R. (1999). The influence of neighborhood density on phonetic categorization in aphasia. *Brain and Language*, 67(1), 46–70.
- Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain and Cognition*, 70(1), 31–41.
- Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., et al. (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(1), 135–144.
- Coslett, H. B., Brashear, H. R., & Heilman, K. M. (1984). Pure word deafness after bilateral primary auditory cortex infarcts. *Neurology*, 34(3), 347–352.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8, 919–924.
- 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(1), 21–33.
- Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience*, 20(7), 1174–1188.
- Devin, J. T., & Aydelott, J. (2009). Speech perception: Motoric contributions versus the motor theory. *Current Biology*, 19(5), R198–R200.
- Drucker, D. M., Kerr, W. T., & Aguirre, G. K. (2009). Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. *Journal of Neurophysiology*, 101(6), 3310–3324.
- Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *The European Journal of Neuroscience*, 15(2), 399–402.
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*, 322(5903), 970–973.
- Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 105(47), 18221–18225.
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Sathannuwong, N., Tong, Y., et al. (2002a). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience*, 14(7), 1076–1087.
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Sathannuwong, N., Tong, Y., et al. (2002b). Neural circuitry underlying perception of duration depends on language experience. *Brain and Language*, 83(2), 268–290.
- Geschwind, N. (1965). Disconnection syndromes in animals and man. I. *Brain: A Journal of Neurology*, 88(2), 237–294.
- Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*, 21(2), 494–506.
- Gow, D. W., Jr., & Caplan, D. (1996). An examination of impaired acoustic–phonetic processing in aphasia. *Brain and Language*, 52(2), 386–407.
- Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, 25(7), 348–353.
- Hayashi, K., & Hayashi, R. (2007). Pure word deafness due to left subcortical lesion: Neurophysiological studies of two patients. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 118(4), 863–868.
- Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 20(10), 1737–1752.
- Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., et al. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language*, 107(3), 179–184.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Husain, F. T., Fromm, S. J., Pursley, R. H., Hosey, L. A., Braun, A. R., & Horwitz, B. (2006). Neural bases of categorization of simple speech and nonspeech sounds. *Human Brain Mapping*, 27(8), 636–651.
- Hutchison, E. R., Blumstein, S. E., & Myers, E. B. (2008). An event-related fMRI investigation of voice-onset time discrimination. *NeuroImage*, 40(1), 342–352.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(29), 9541–9546.
- Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing. *Trends in Cognitive Sciences*, 10(11), 480–486.
- Jamison, H. L., Watkins, K. E., Bishop, D. V., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex*, 16(9), 1266–1275.
- Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *NeuroImage*, 15(4), 733–746.
- Joanisse, M. F., & Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage*, 19(1), 64–79.
- Kemmerer, D. (2006). The semantics of space: Integrating linguistic typology and cognitive neuroscience. *Neuropsychologia*, 44(9), 1607–1621.
- Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., et al. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping*, 25(1), 155–164.
- Laird, A. R., McMillan, K. M., Lancaster, J. L., Kochunov, P., Turkeltaub, P. E., Pardo, J. V., et al. (2005). A comparison of label-based review and ALE meta-analysis in the stroop task. *Human Brain Mapping*, 25(1), 6–21.
- Lancaster, J. L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., et al. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, 28(11), 1194–1205.
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, 10(3), 120–131.
- Leonard, C. M., Puranik, C., Kuldau, J. M., & Lombardino, L. J. (1998). Normal variation in the frequency and location of human auditory cortex landmarks Heschl’s gyrus: Where is it. *Cerebral Cortex*, 8(5), 397–406.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, 15(10), 1621–1631.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19), 1692–1696.
- Mottronen, R., Calvert, G. A., Jaaskelainen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, 30(2), 563–569.
- Mottronen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(31), 9819–9825.
- Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science: A Journal of the American Psychological Society/APS*, 20(7), 895–903.
- Naatunen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotolainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic responses. *Nature*, 385, 432–434.
- Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Kobayashi, M., et al. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca’s area: An open-protocol study. *Brain and Language*, 93(1), 95–105.
- Noordzij, M. L., Neggens, S. F., Ramsey, N. F., & Postma, A. (2008). Neural correlates of locative prepositions. *Neuropsychologia*, 46(5), 1576–1580.
- Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roetlinger, M., et al. (2006). Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping*, 27(7), 562–571.
- Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, 13(1), 14–19.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 28(32), 8116–8123.
- Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex*, 17(10), 2251–2257.
- Overath, T., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 28(49), 13268–13273.
- Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science*, 25, 711–731.
- Price, C. J. (2000). The anatomy of language: Contributions from functional neuroimaging. *Journal of Anatomy*, 197(Pt 3), 335–359.
- Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), 7865–7870.
- Quadflieg, S., Turk, D. J., Waiter, G. D., Mitchell, J. P., Jenkins, A. C., & Macrae, C. N. (2009). Exploring the neural correlates of social stereotyping. *Journal of Cognitive Neuroscience*, 21(8), 1560–1570.

- Raizada, R. D., Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2009). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. *Cerebral Cortex*.
- Raizada, R. D., & Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, 56(4), 726–740.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Ravizza, S. M. (2003). Dissociating the performance of cortical and subcortical patients on phonemic tasks. *Brain and Cognition*, 53(2), 301–310.
- Rimol, L. M., Specht, K., Weis, S., Savoy, R., & Hugdahl, K. (2005). Processing of sub-syllabic speech units in the posterior temporal lobe: An fMRI study. *NeuroImage*, 26(4), 1059–1067.
- Roberson, D., & Hanley, J. R. (2007). Color vision: Color categories vary with language after all. *Current Biology*, 17(15), R605–R607.
- Roy, A. C., Craighero, L., Fabbri-Destro, M., & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Physiology, Paris*, 102(1–3), 101–105.
- Saffran, E. M., Marin, O. S., & Yeni-Komshian, G. H. (1976). An analysis of speech perception in word deafness. *Brain and Language*, 3(2), 209–228.
- Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, 111(1), 1–7.
- Schofield, T. M., Iverson, P., Kiebel, S. J., Stephan, K. E., Kilner, J. M., Friston, K. J., et al. (2009). Changing meaning causes coupling changes within higher levels of the cortical hierarchy. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), 11765–11770.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain: A Journal of Neurology*, 123(Pt 12), 2400–2406.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26(2), 100–107.
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech – A positron emission tomography study. *The Journal of the Acoustical Society of America*, 120(2), 1075–1083.
- Serniclaes, W., Sprenger-Charolles, L., Carre, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research: JSLHR*, 44(2), 384–399.
- Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*, 39(3), 1064–1080.
- Szenkovits, G., & Ramus, F. (2005). Exploring dyslexics' phonological deficit I: Lexical vs sub-lexical and input vs output processes. *Dyslexia*, 11(4), 253–268.
- Ting Siok, W., Kay, P., Wang, W. S., Chan, A. H., Chen, L., Luke, K., et al. (2009). Language regions of brain are operative in color perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(20), 8140–8145.
- Tranel, D., & Kemmerer, D. (2004). Neuroanatomical correlates of locative prepositions. *Cognitive Neuropsychology*, 21(7), 719–749.
- Turkeltaub, P., Eden, G., Jones, K., & Zeffiro, T. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, 16(3 Pt 1), 765–780.
- Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech-sound processing in human temporal cortex. *NeuroImage*, 31(3), 1284–1296.
- von Stockert, T. R. (1982). On the structure of word deafness and mechanisms underlying the fluctuation of disturbances of higher cortical functions. *Brain and Language*, 16(1), 133–146.
- Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, 13(7), 994–1005.
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989–994.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702.
- Zaehle, T., Geiser, E., Alter, K., Jancke, L., & Meyer, M. (2008). Segmental processing in the human auditory dorsal stream. *Brain Research*, 1220, 179–190.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46.
- Zevin, J. D., & McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioral and Brain Functions*, 1(1), 4.