NeuroImage 117 (2015) 319–326

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Mapping cortical responses to speech using high-density diffuse optical tomography

Mahlega S. Hassanpour [a,*], Adam T. Eggebrecht [b], Joseph P. Culver [a,b], Jonathan E. Peelle [c]

[a] Department of Physics, Washington University in Saint Louis, USA
[b] Department of Radiology, Washington University in Saint Louis, USA
[c] Department of Otolaryngology, Washington University in Saint Louis, USA

ARTICLE INFO

Article history:
Received 27 January 2015
Accepted 20 May 2015
Available online 28 May 2015

ABSTRACT

The functional neuroanatomy of speech processing has been investigated using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) for more than 20 years. However, these approaches have relatively poor temporal resolution and/or challenges of acoustic contamination due to the constraints of echoplanar fMRI. Furthermore, these methods are contraindicated because of safety concerns in longitudinal studies and research with children (PET) or in studies of patients with metal implants (fMRI). High-density diffuse optical tomography (HD-DOT) permits presenting speech in a quiet acoustic environment, has excellent temporal resolution relative to the hemodynamic response, and provides noninvasive and metal-compatible imaging. However, the performance of HD-DOT in imaging the brain regions involved in speech processing is not fully established. In the current study, we use an auditory sentence comprehension task to evaluate the ability of HD-DOT to map the cortical networks supporting speech processing. Using sentences with two levels of linguistic complexity, along with a control condition consisting of unintelligible noise-vocoded speech, we recovered a hierarchically organized speech network that matches the results of previous fMRI studies. Specifically, hearing intelligible speech resulted in increased activity in bilateral temporal cortex and left frontal cortex, with syntactically complex speech leading to additional activity in left posterior temporal cortex and left inferior frontal gyrus. These results demonstrate the feasibility of using HD-DOT to map spatially distributed brain networks supporting higher-order cognitive faculties such as spoken language.

Introduction

Cognitive neuroscientists who study how the brain perceives spoken language desire a quiet imaging technique that can record brain function noninvasively and provide reliable results. Such measurements have proven challenging to collect using functional magnetic resonance imaging (fMRI) due to the substantial acoustic noise associated with echoplanar imaging (Foster et al., 2000; McJury and Shellock, 2000; Moelker and Pattynama, 2003; Price et al., 2001; Ravicz et al., 2000). Background noise can interfere with the presentation of auditory stimuli and adds additional perceptual and cognitive demands to the experimental task (Peelle, 2014). Such auditory task demands are likely to differentially affect participants with hearing impairment or reduced cognitive capacity (Caldwell and Nittrouer, 2013; Grimault et al., 2001; Peelle et al., 2011). In addition, high magnetic fields generated by the scanner pose a critical limitation on the studies of patients with metal implants who cannot receive MRIs. Although electroencephalography (EEG), magnetoencephalography (MEG), and positron emission tomography (PET) provide quiet imaging settings, each of these modalities has limitations. For example, anatomical localization can be challenging with EEG and MEG (Baumgartner, 2004; He, 1999), and PET uses ionizing radiation and has relatively low temporal resolution (Cabeza and Nyberg, 1997).

In theory, optical neuroimaging offers an appealing alternative. Optical methods use a quiet, safe, and noninvasive technique based on near infrared spectroscopy (NIRS) to record hemodynamic activity from the brain. However, traditional functional NIRS (fNIRS) imaging suffers from low spatial resolution (sparse source–detector arrangements) and signal contamination from superficial tissues. More recently, the development of high-density diffuse optical tomography (HD-DOT) instrumentation has dramatically improved the spatial resolution

Abbreviations: FOV, field of view; PET, positron emission tomography; MRI, magnetic resonance imaging; HD-DOT, high-density diffuse optical tomography; EEG, electroencephalography; MEG, magnetoencephalography; NIRS, near infrared spectroscopy; SR, subject-relative; OR, object-relative; LCD, liquid-crystal-display; ISI, interstimulus interval; HbO, oxygenated hemoglobin; HbR, deoxygenated hemoglobin; HbT, total hemoglobin; GLM, general linear model; HRF, hemodynamic response function; ROIs, regions of interest; STD, standard deviation; SD, source–detector; MFG, middle frontal gyrus; IIFG, inferior frontal gyrus; STG, superior temporal gyrus; STG, superior temporal gyrus; MTG, middle temporal gyrus; aSTG, anterior superior temporal gyrus; vPFC, ventral prefrontal cortex; dPFC, dorsal prefrontal cortex; MPFC, middle prefrontal cortex; STC, superior temporal cortex; MTC, middle temporal cortex; aSTC, anterior superior temporal cortex.

* Corresponding author.
E-mail address: hassanpourm@mir.wustl.edu (M.S. Hassanpour).

http://dx.doi.org/10.1016/j.neuroimage.2015.05.058
1053-8119/

and brain specificity of optical neuroimaging (Gregg et al., 2010; Joseph et al., 2006; Koch et al., 2010; Saager and Berger, 2008; White and Culver, 2010; Zeff et al., 2007). In addition, algorithms incorporating realistic forward light models, spatial normalization methods, and advanced statistical tools have significantly improved overall image quality, coregistration to anatomy, and reliability (Custo et al., 2010; Eggebrecht et al., 2012; Ferradal et al., 2014; Hassanpour et al., 2014; Okamoto and Dan, 2005).

Early HD-DOT studies covered about ~1/8 of the head, limiting imaging to small select regions of the brain. Recently we reported a large field of view HD-DOT system that covers approximately 50% of the head surface and is capable of mapping distributed brain functions and networks (Eggebrecht et al., 2014). We validated the performance of this system for functional imaging of distributed cognitive processes and networks through quantitative comparisons to coregistered fMRI, and were able to map the neuroanatomical organization of single-word processing (i.e., distinct cortical regions for hearing, reading, speaking, and subvocally generating single words). However, the ability of HD-DOT to capture the neural responses to connected speech has not yet been established. Connected speech comprehension is more complex than single word perception, incorporating syntactic structure and richer conceptual representations (Price, 2012). Processing these relationships requires a larger network of cortical regions that must be imaged simultaneously. Thus, being able to capture a range of cortical responses during speech processing is critical to studies that aim to understand how the healthy brain processes speech, and to understand the impact of auditory noise, hearing loss, or cognitive deficits that modulate the brain's strategy for speech comprehension.

To evaluate the performance of our HD-DOT system in imaging speech comprehension we presented listeners with spoken sentences that varied in their syntactic complexity. We chose this manipulation because in other neuroimaging modalities there are consistent differences in neural activity based on syntactic complexity, and because syntactic information is processed in a highly distributed and hierarchical fashion throughout the cortex at different cortical depths (e.g., sulci and gyri) (Bornkessel et al., 2005; Caplan et al., 2008; Friederici et al., 2003; Griffiths et al., 2013; Stromswold et al., 1996; Tyler et al., 2010). Using an event-related sentence comprehension task, we tested whether HD-DOT would be able to detect the effect of syntactic complexity caused by a word-order manipulation. Our imaging results show that HD-DOT is capable of mapping a hierarchical organization of the language system, and the spatial location of the functional maps are in good agreement with previous sentence comprehension studies using MRI and PET. Being able to detect the subtle changes in cortical activation induced by increased processing demand demonstrates that HD-DOT has the sensitivity and spatial specificity to serve as a general tool for cognitive neuroscience.

Materials and methods

Participants

We scanned 10 healthy, right-handed, native English speakers (6 female) between the ages of 20 and 32 years (mean = 27.6, STD = 3.3). All had normal hearing by self-report and no history of neurological or psychiatric disorders. Written informed consent was obtained for all subjects as approved by the Human Research Protection Office by Washington University School of Medicine.

Subject-specific light models were generated using each subject's own structural T1- and T2-weighted MRI images obtained from a previous study.

Materials

Auditory stimuli consisted of sentences and unintelligible noise (as a control condition). Sentences were constructed to contain a subject-relative (SR) or object-relative (OR) center-embedded clause. Sentences with object-relative clauses are reliably found to be more difficult to comprehend compared to subject-relative clauses (Gibson, 1998; Traxler et al., 2002), resulting in longer response times, more errors, or equivalent performance but through increased neural activity. These sentences were selected from a list of 60 meaningful 6-word sentences, each with a subject-relative embedded clause, used in previous studies (Peelle et al., 2004, 2010b; Wingfield et al., 2003). In half of the sentences the character performing the action was a male (e.g., king, brother) and in the other half the actor was a female (e.g., queen, sister). The 60 original sentences were then re-worded to vary syntactic complexity (turning subject-relative into object-relative construction) and whether a male or female was performing the action, as shown in the following examples:

1. Subject-relative clause, male agent: “Men that assist women are helpful.”
2. Object-relative clause, male agent: “Women that men assist are helpful.”
3. Subject-relative clause, female agent: “Women that assist men are helpful.”
4. Object-relative clause, female agent: “Men that women assist are helpful.”

These rearrangements resulted in 240 total sentences, each of which was presented a single time during the experiment (120 subject-relative sentences and 120 object-relative sentences). During the experiment, subjects were asked to indicate the gender of the character performing the action (male or female) using a button-press response.

In addition to intelligible sentences, we included unintelligible speech trials (“noise”) as a control condition. The noise stimuli consisted of one channel noise-vocoded speech, created by modulating white noise (lowpass filtered at 8 kHz) with the amplitude envelope of the sentence (lowpass filtered at 30 Hz); the vocoded sentences were a subset of the intelligible sentences used in the study. Noise vocoding removes the spectral detail from the sentence while retaining its temporal amplitude envelope (Shannon et al., 1995).

The mean length of auditory stimuli (sentences or noise) was 1.76 ± 0.05 s (range: 1.32–1.89 s).

HD-DOT system

Full details on our HD-DOT system are reported by Eggebrecht et al. (2014). Briefly, our HD-DOT array contains 96 sources and 92 detectors that are coupled with fiber optic bundles to a flexible imaging cap. Source locations are illuminated by continuous-wave light emitting diodes at two wavelengths (750 nm and 850 nm) that enable hemoglobin spectroscopy. Light is detected by avalanche photodiodes (Hamamatsu C5460-01) and digitized by dedicated 24-bit analog-to-digital converters (MOTU HD192) (Zeff et al., 2007), which enable high dynamic range ($>10^6$) and low crosstalk ($<10^{-6}$). The dynamic range allows the detection of light from multiple source detector distances (e.g., first through fourth nearest neighbors are 13, 30, 39 and 47 mm apart) (Supplementary Figs. 1A and B). This array provides more than 1200 usable source–detector measurements at a 10 Hz full-field frame rate.

Procedure

Subjects were seated in an adjustable chair in a sound-isolated room facing a 19 in. LCD screen (located 1 m from subjects and at approximately eye level), and two stereo speakers (each located at 1.5 m from subjects at approximately ear level). Subjects held a keyboard on their lap. The HD-DOT cap was put on the subject's head covering portions of occipital, temporal, motor, and frontal cortices (Fig. 1A). Once the cap was placed comfortably with good signal-to-noise ratio (Supplementary Fig. 1C), the placement of the cap with respect to

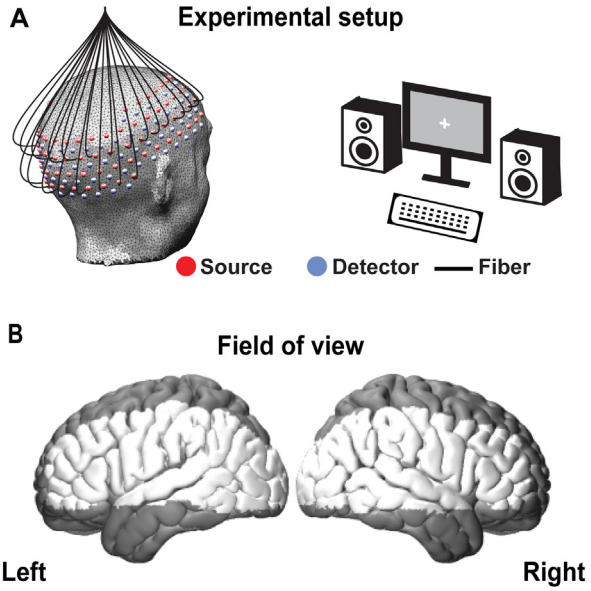


Fig. 1. (A) Schematic view of the HD-DOT experimental set-up, subject position and imaging cap structure (a subset of optical fibers is shown for clarity). (B) Group field of view on the cortical surface of an MNI atlas.

anatomical landmarks on the head and face of the subject (e.g., the nasion) was noted to locate the cap relative to the subject's head, later used to generate a subject-specific light model.

We presented stimuli using Psychophysics Toolbox 3 (Brainard, 1997), sending audio to the speakers via an external audio interface (M-Audio Fast Track Pro). We set the sound level at a comfortable listening level that did not change over the course of a session. Stimuli were presented in four separate runs, each of which contained 30 subject-relative (easy) sentences, 30 object-relative (complex) sentences, and 10 noise trials. These stimuli were presented in a pseudorandom order, with the order of conditions varied between runs but constant across subjects. Following each sentence, subjects were instructed to press a key with their left index finger if the person performing action was female and a separate key with their right index finger if the person performing the action was male. A central fixation cross was displayed at the center of a gray screen; after each key press the cross was changed to an 'x' to inform the subject that a response was received. This sign was on the screen during part of interstimulus interval (ISI). ISIs were pseudorandomly distributed between 2 and 10 s; the subject's reaction time on each trail was considered as a part of the ISI of that trial. If the reaction time was longer than the predetermined ISI for a given trial, the next stimulus was presented immediately following the subject's key press. One second prior to the stimulus trial the 'x' was changed back to a cross to prepare the subject for listening to the next stimulus.

Prior to the experiment subjects were given a short practice session containing 24 trials (8 trials per condition) to explain the instructions and ensure they were performing the task correctly. None of these sentences appeared in the actual experiment.

Data preprocessing

A flow chart outlining data preprocessing is shown in Supplementary Fig. 3. Raw detector data (sampling rate: 10 Hz) were decoded to source–detector pair data, and converted to log-ratio. The data were then bandpass filtered (0.02–0.5 Hz) to remove low-frequency trends and pulse artifacts. We averaged all signals from the first nearest neighbor channels to create a measure of superficial hemodynamics; we used linear regression to remove this nuisance signal from all channels. To create a realistic forward light model, we used subject-specific T1- and T2-weighted structural MRI images. After bias field correction using SPM8 software (Wellcome Trust Centre for Neuroimaging, London, UK), we used an in-house script to segment an individual head into five different tissue types (scalp/skin, skull, CSF, white matter, and gray matter) (Supplementary Fig. 2A). We used the segmented images to create finite element head meshes using NIRview software (version 1.10, http://www.dartmouth.edu/~nir/nirfast/) (Jermyn et al., 2013). The light propagation inside the mesh was modeled using the diffusion approximation and a sensitivity matrix was generated using NIRFAST software (Dehghani et al., 2009a) (Supplementary Fig. 2B). The sensitivity matrix was inverted, smoothed with a Gaussian kernel ($\sigma =$ 2.4 mm), and used to reconstruct absorption coefficient changes for each wavelength (Eggebrecht et al., 2012). Relative changes in the concentrations of oxygenated hemoglobin (ΔHbO), deoxygenated hemoglobin (ΔHbR), and total hemoglobin (ΔHbT) were obtained from the absorption coefficient changes by the spectral decomposition of the extinction coefficients of HbO and HbR at the two wavelengths. Additionally, data were downsampled to 1 Hz. For group analysis, we registered all data to the Montreal Neurological Institute (MNI) 152 atlas using in-house developed linear affine transformation code and concurrently resampled data to a voxel size of $3 \times 3 \times 3$ mm following Eggebrecht et al. (2014). Due to the cap fitting on a variety of head sizes and shapes, the field of view (FOV) measured within each subject varied across the group. For the current study, we included only voxels sampled with acceptable sensitivity in all subjects in the group FOV (displayed in white in Fig. 1B). To find these voxels, we calculated a flat field reconstruction (Dehghani et al., 2009b) and considered voxels with a reconstructed value within two orders of magnitude of the maximum value to have acceptable sensitivity. The group FOV contains approximately 700 cm^3 of head volume, covering occipital and parts of parietal, temporal, motor and frontal cortices and spans up to 2 cm into the brain tissue.

For the cortical surface representation of results, we mapped volumetric results onto the mid-thickness surface of MNI152 atlas extracted using FreeSurfer software (version 5.1.0, Martinos Center for Biomedical Imaging, Massachusetts General Hospital) (Dale et al., 1999). Volumetric activations are overlaid on the T1 images of the MNI152 atlas.

Timeseries analysis

We used custom HD-DOT SPM code for statistical analyses (Hassanpour et al., 2014), outlined in Supplementary Fig. 4. Five conditions were included in the general linear model (GLM) design matrix: subject-relative sentences, object-relative sentences, noise trials, and left and right button presses. All trials were included, regardless of behavioral accuracy. Auditory stimuli were modeled as events with 2 s duration and button presses as events with 0 s duration. Events were convolved with a canonical hemodynamic response function (HRF) to model hemodynamic responses to the predicted neural activity. We constructed the canonical HRF using a double-gamma function matched to the general properties of the hemodynamic response in primary auditory cortex averaged over all data (e.g., delay time of 2 s, time to peak of 7 s and undershoot at 17 s).

For each subject, we combined data for all four runs using a fixed effects analysis and generated linear contrast maps. We then assessed group-level activity using random effects analyses of these contrast maps and calculated statistical z-value maps for each contrast. We calculated voxelwise degrees of freedom and spatial smoothness from estimates of temporal and spatial autocorrelation structures of GLM residuals, respectively. Unless otherwise specified, all statistical maps are thresholded at $p < 0.001$ (voxelwise, uncorrected) and corrected for multiple comparisons using a nonstationary cluster analysis technique at $p < 0.05$ (Hassanpour et al., 2014; Hayasaka et al., 2004; Worsley et al., 1998).

In the main text we focus on maps of ΔHbO as we have found ΔHbO signal to exhibit a higher contrast-to-noise ratio compared to ΔHbR or ΔHbT (Eggebrecht et al., 2014; Hassanpour et al., 2014). Results from

other hemoglobin contrasts are reported in Supplementary Figs. 5–7 and are generally consistent with ΔHbO results.

Finally, we also estimated the temporal profile of hemodynamic activity for each stimulus type using the GLM un-mixing method (also known as a finite impulse response, or FIR model) (Glover, 1999; Hassanpour et al., 2014; Miezin et al., 2000). This procedure allowed us to evaluate the timecourse of the evoked responses to ensure they were physiologically plausible.

Results

Behavioral data

We collected behavioral measures including accuracy (percentage of correct responses) and response time (measured from the stimulus start time to key press time). The mean accuracy for the subject-relative (easy) sentences was 97.7% (STD = 3.13), and for the object-relative (complex) sentences was 97.6% (STD = 3.29). Accuracy was equivalent between these two conditions, $t(78) = 0.13$. The mean response time for the subject-relative sentences was 2.0 s (STD = 0.36), and for the object-relative sentences 2.1 s (STD = 0.37), which also did not significantly differ, $t(78) = 0.37$.

Hierarchical processing of spoken language

Hearing unintelligible vocoded speech ("noise") caused an increase in oxy-hemoglobin (HbO) concentration in parts of the temporal cortex bilaterally (Fig. 2A). Beyond these regions, hearing intelligible sentences resulted in widespread activity in bilateral temporal cortex along with additional activity in the left frontal cortex (Figs. 2B and C). Similar results were obtained from other hemoglobin contrasts, as shown in Supplementary Fig. 5.

We next statistically compared the responses to each sentence type with responses to noise, shown in Figs. 3A and B, with maxima listed in Supplementary Tables 1 and 2. The comparison between activity in response to sentences from that to noise helps differentiate higher-level speech processing regions from general auditory processing regions. Compared to noise, both subject-relative and object-relative sentences led to significant increases in activation in large portions of the left hemisphere including frontal cortex, lateral superior and middle temporal cortex, and ventral premotor cortex. In addition, we found a significant response to both types of sentences relative to noise in anterior parts of the right superior and middle temporal cortex.

We then identified regions showing more activity for the complex object-relative sentences than the easier subject-relative sentences, shown in Fig. 3C and listed in Supplementary Table 3. We found that the increased processing load resulted in a significant increase in the response in bilateral ventral parts of posterior prefrontal cortex, left lateral middle and superior temporal gyri (MTG and STG) and posterior parts of bilateral temporal cortex.

Fig. 4 illustrates the hierarchical organization of the speech network by overlaying maps for speech intelligibility (all sentences > noise) and syntactic complexity (object-relative > subject-relative sentences) on the cortical surface. The overlap between these two contrasts (shown in white) includes left lateral superior temporal gyrus and ventral inferior frontal gyrus, regions that have been previously associated with processing sentences containing center-embedded relative clauses (Caplan et al., 2008; Friederici et al., 2003; Peelle et al., 2010b; Stromswold et al., 1996). Regions in yellow, including posterior parts of temporal cortex and more anterior region of ventral prefrontal cortex, are not recruited for easier intelligible speech, but increase activity when the processing load increases. These two patterns highlight a "core" speech processing network that is active for more basic auditory sentence processing (Davis and Johnsrude, 2003; Peelle et al., 2010a; Rauschecker and Scott, 2009), and an expanded associative network that is differentially engaged as linguistic demands increase (Peelle, 2012; Wingfield and Grossman, 2006).

To examine the temporal profile of subjects' responses we extracted the hemodynamic response to sentences in several regions of interest (ROIs), listed in Table 1 and shown in Fig. 5. The ROIs were defined as a cube of 3-voxels per side centered on the peak voxel of the clusters that passed the significance threshold in either the noise > baseline contrast or the sentences > noise contrast. Results show that the hemodynamic response starts with a delay (up to 3 s) and peaks approximately 4–8 s after the stimulus (e.g., Figs. 5A and D). These responses serve as a quality control check and verify that our localized responses are consistent with evoked hemodynamic activity (Aguirre et al., 1998).

Finally, in order to ascertain how reliably we were able to detect hierarchical responses to speech comprehension in individual subjects and individual runs of data, we created single-subject and single run renderings of the main contrasts (sentences > noise and complex > easy). The single subject maps of the response to intelligibility were highly consistent across subjects (Supplementary Fig. 8A). While the contrast to noise ratio drops significantly at single run level, these also show qualitatively similar features (Supplementary Fig. 8B).

Discussion

The goal of the present study was to evaluate the use of HD-DOT for imaging speech comprehension. We assessed the ability of HD-DOT to

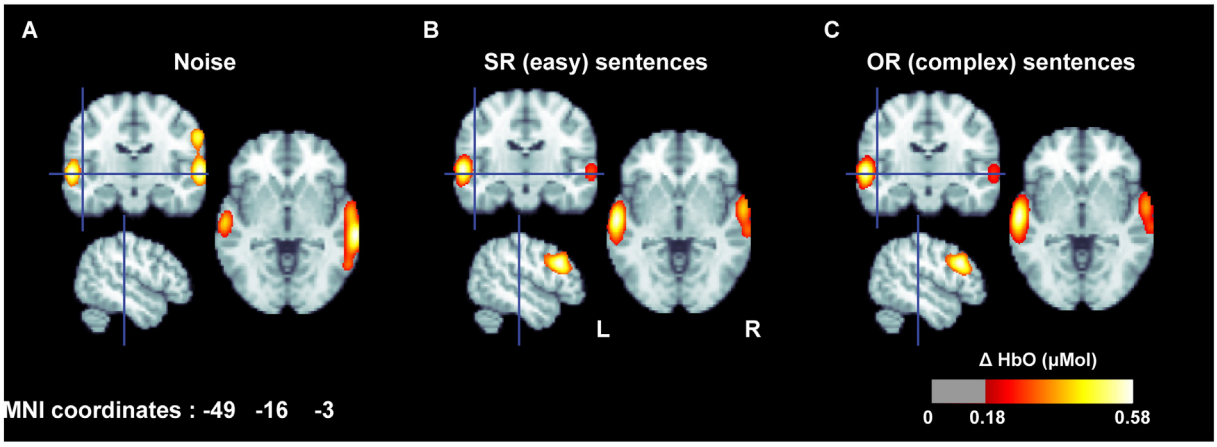


Fig. 2. Oxy-hemoglobin increase in response to (A) noise, (B) subject-relative sentences and (C) object-relative sentences. Individual data were spatially normalized to MNI152 space and group averaged. The volumetric activations are overlaid on T1 images of MNI152 atlas, and are shown in parasagittal, coronal and axial views ($x = -49$, $y = -16$ and $z = -3$). Images are thresholded at 0.18 μmol.

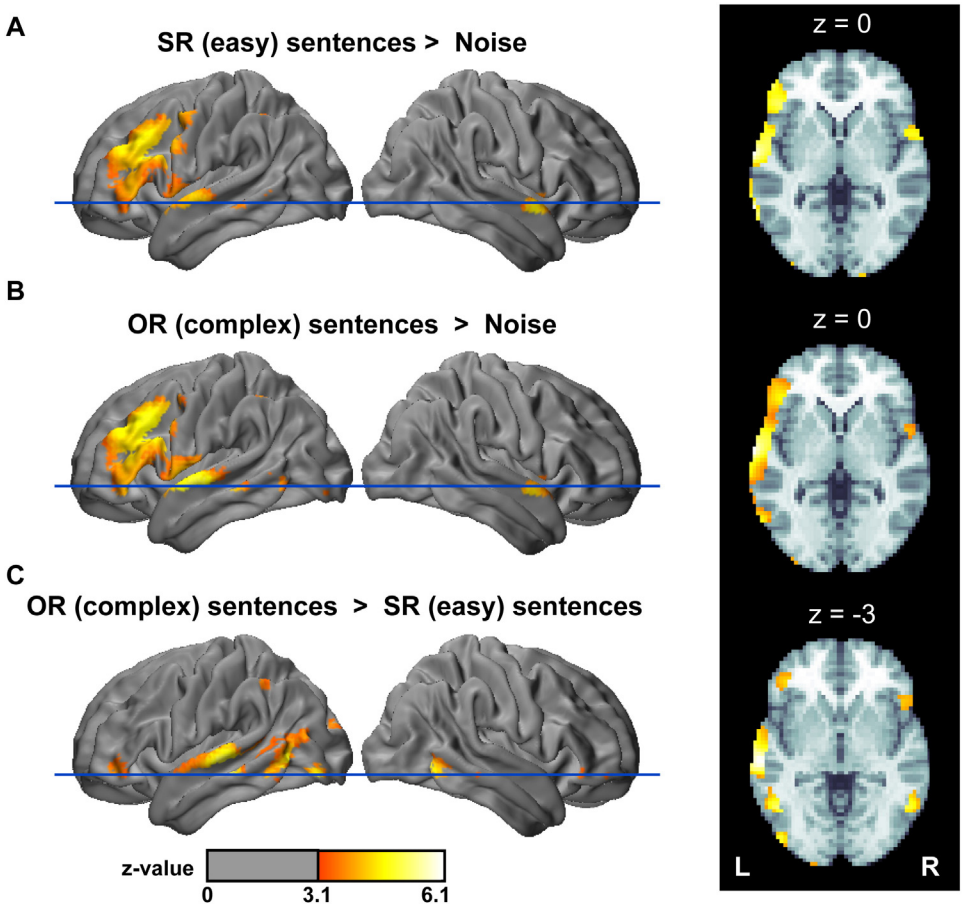
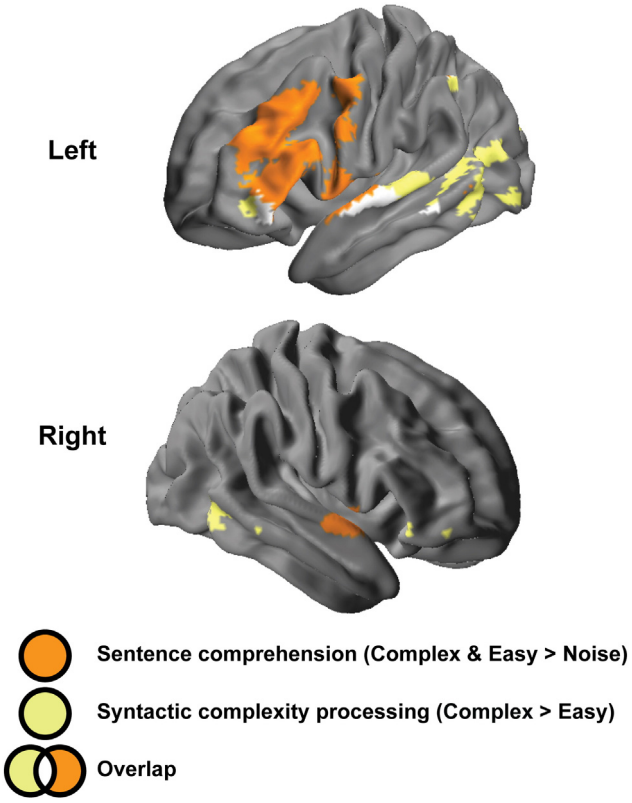


Fig. 3. Cortical processing for spoken language. (A) Differential activation by subject-relative (SR) sentences > noise highlights brain areas involved in intelligible speech processing. (B) A similar map was obtained for object-relative (OR) sentences > noise. (C) Directly comparing object-relative sentences to subject-relative sentences shows the effect of syntactic complexity. z-Maps are thresholded at voxelwise $p < 0.001$ ($z = 3.1$) and (corrected) cluster significance threshold of $p < 0.05$. These are results obtained from the oxy-hemoglobin signal.



measure brain activity during multiple levels of speech processing by presenting listeners with spoken sentences that varied in intelligibility and linguistic complexity. Overall, our results highlight a network of neural areas that are described in the literature for supporting speech comprehension, and map the hierarchical organization of spoken language processing. Below we assess our findings in more detail and discuss advantages and drawbacks of using HD-DOT in neurocognitive studies.

Cortical responses to intelligibility and syntactic complexity

Our results show a distributed network of cortical activity associated with the processing of spoken language. This network includes regions located in the temporal, parietal and frontal cortices of both hemispheres. However, speech related processes activated up to six times larger cortical volume of the left hemisphere compared to its contralateral side. The largest significant cluster spans several left frontotemporal sub-regions, classical language processing centers. While both ventral and dorsal parts of the left hemisphere are involved in speech comprehension, right hemisphere activity is mainly limited to temporal cortex. These results are largely consistent with previous fMRI and PET studies (Davis and Johnsrude, 2003; Hickok and Poeppel, 2007; Osnes et al., 2011; Pulvermüller et al., 2006; Tyler and Marslen-Wilson, 2008; Wilson et al., 2008).

Fig. 4. Hierarchy in speech processing. Regions with significantly increased activity in response to syntactic complexity (yellow) are overlaid on the regions with significantly increased activity in response to intelligibility (orange). Regions that overlap (white) are recruited by both types of sentences (easy and complex), but are recruited to greater degree by syntactically complex sentences, implying a role in syntax processing. Yellow regions are recruited only to support increased syntactic processing load.

Table 1
Coordinates of the centers of regions of interest.

Region	Center coordinate			Z score	
	x	y	z	Noise	Speech
Left:					
Lateral middle temporal cortex[a,b]	−61.5	−21	0	3.81	3.80
Lateral superior temporal cortex[a,b]	−61.5	−30	−9	3.81	3.13
Anterior superior temporal cortex[b]	−61.5	6	−9	2.16	3.21
Middle prefrontal cortex[b]	−52.5	14	23	1.82	4.18
Dorsal prefrontal cortex[b]	−46.5	28	30	0.94	4.30
Ventral prefrontal cortex [b]	−43.5	39	−6	−0.72	3.46
Right:					
Anterior superior temporal cortex[b]	55.5	6	−9	2.55	4.20
Ventral prefrontal cortex[b]	52.5	27	−9	−0.70	3.12

[a] ROI is within a cluster from the noise > baseline comparison.
[b] ROI is within a cluster from the sentences > noise comparison.

Sentences with object-relative construction are more complex than those with subject-relative clauses due to both memory and linguistic integration costs (Cooke et al., 2002; Fiebach et al., 2001). Although the behavioral performance of participants is sometimes poorer for object-relative sentences (Caplan et al., 2008; Wingfield et al., 2003), in our study differential processing was only revealed in the imaging results. One potential reason for this could be the age of our subjects, as younger adults are less impacted by syntactic complexity. However, processing differences were indeed apparent in the patterns of neural activation. During successful comprehension of object-relative sentences, we found brain regions showing significantly stronger activity that overlapped the core speech processing network, as well as complementary regions not seen in response to the simpler subject-relative sentences. Overall, HD-DOT revealed an enlargement of the speech processing network when processing load increased.

We designed this study to focus on the group level results, and therefore acquired approximately 30 min of data per subject. Nonetheless, individuals' maps of the main effect of intelligibility were consistent across subjects. This includes a significant increase in the activity of the left temporal and prefrontal cortices (core speech processing centers) during intelligible speech processing compared to listening to unintelligible noise (in all subjects). In our previous study we have shown that the HD-DOT performance at the single subject level is statistically comparable to fMRI (Eggebrecht et al., 2014). Further studies with larger number of samples per subject may shed more light whether the inter-subject variance is due to a) low number of samples per subject for detecting the subtle effect of a word-order change or b) individual differences in the brain recruitment for processing syntactically complex speech.

Using HD-DOT to measure neural responses to speech

Although the language subsystems of human brain have been studied for many years using both lesion studies and functional neuroimaging, HD-DOT has notable advantages compared to other methods. HD-DOT provides non-invasive, radiation-free and metal-compatible tomographic imaging of cortical hemodynamic activity in a noise-free environment with relatively good spatial and temporal resolution. These features provide an advantage in many settings including studies of cognitive development in children, studies of noise-degraded speech, and studies of subjects with implanted metal devices such as cochlear implants.

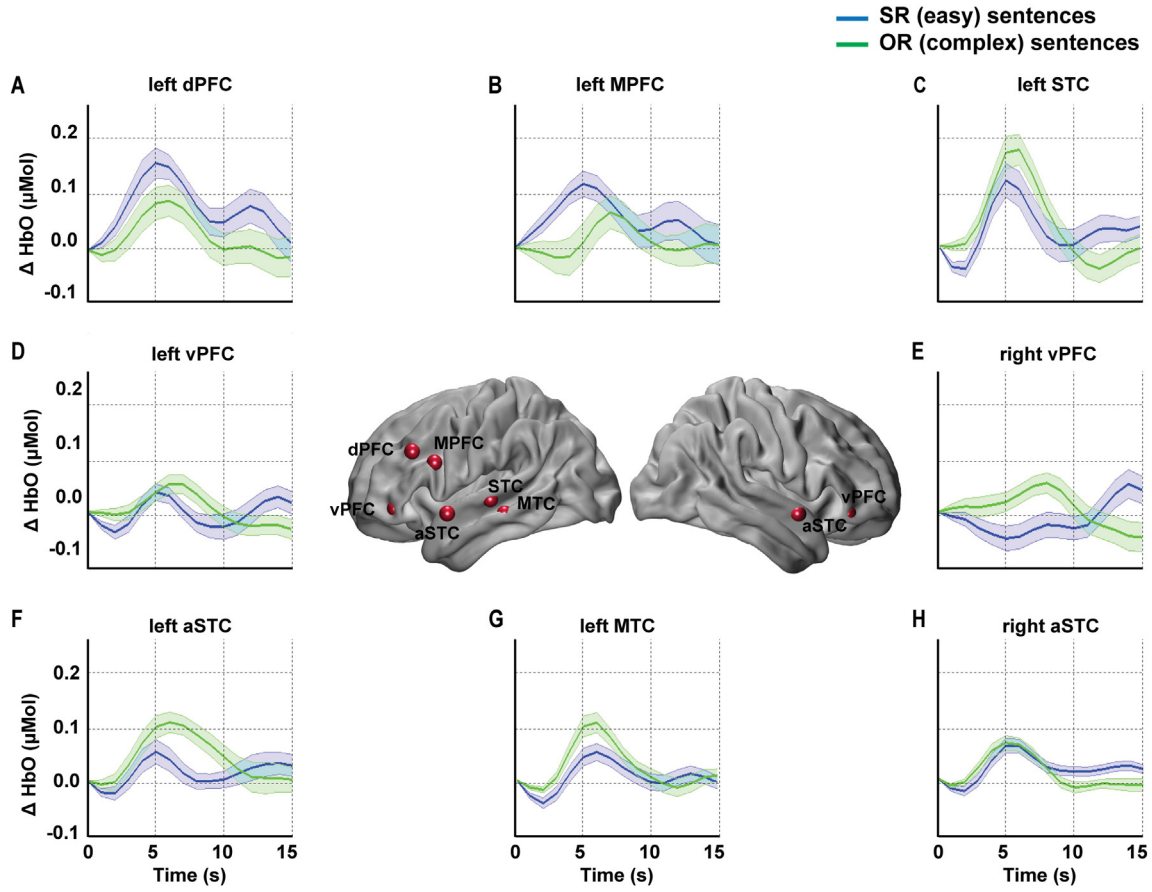


Fig. 5. Temporal profile of hemodynamic response: (A–H) show oxy-hemoglobin changes in response to subject-relative (blue) and object-relative (green) sentences at different ROIs. Shaded areas show (±) standard error.

In particular, the quietness of the HD-DOT system is an important advantage compared to fMRI in speech studies, as the acoustic noise associated with fMRI scanning can interfere with the normal auditory language processing in numerous ways. Loud sounds can affect the hearing threshold by causing a stapedius muscle reflex (Olsen, 1999; Ulmer et al., 1998), and affect the perception of stimuli by acoustic-spectral masking (Shah et al., 1999). At the physiological level, acoustic noise can saturate neuronal populations in the auditory cortex (Bandettini et al., 1998; Gaab et al., 2007). At a cognitive level, speech comprehension in the midst of scanner noise may require additional functional responses in extra-auditory frontal areas (Peelle, 2014; Schmidt et al., 2008; Skouras et al., 2013), making it difficult to separate executive processes required for language processing from those required to deal with the background noise. A common approach for reducing acoustic contamination in auditory fMRI studies is to use sparse imaging, which allows the presentation of stimuli in quiet by collecting fewer MRI volumes (Hall et al., 1999). However, sparse imaging reduces temporal resolution and thus the ability to efficiently detect the shape of hemodynamic responses. Thus, HD-DOT provides an acoustically superior alternative to fMRI for auditory neuroscience.

One potential drawback of using HD-DOT to assess language processing is that, unlike fMRI, HD-DOT imaging is limited to superficial cortex (i.e., ~1–2 cm into the brain) and cannot access deep cortical structures (e.g., insula or operculum) or subcortical brain structures (e.g., striatum or thalamus). Fortunately, a number of regions critical for speech processing are relatively near to the cortical surface, including large portions of frontal and temporal cortex frequently highlighted in anatomically-constrained models of language processing. A related limitation of the current study is that the HD-DOT system we used, while having a relatively large field of view, does not provide full head coverage—a limitation shared by all existing DOT systems. For the purposes of establishing the regional sensitivity to syntax processing the current system was sufficient. However, for more comprehensive mapping of the neural response to speech, particularly in frontal cortex, an HD-DOT system with greater coverage will be required. As with any modality, a full picture of network function can only be obtained through the use of converging evidence from multiple techniques.

Conclusion

Our results demonstrate the feasibility of imaging hierarchical cognitive processes during speech comprehension with HD-DOT. Our findings are in general agreement with previous fMRI studies in demonstrating that increased processing demand for syntactically complex sentences results in greater activation in left temporal and prefrontal cortex. With the advantages of being acoustically quiet, ability to image subjects with electronic implants, accurate anatomical localization, and relatively high spatial resolution, HD-DOT is well-suited for studying cortical responses to spoken language.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.05.058.

Acknowledgments

This work was supported in part by NIH grants R01EB009223 (J.P.C.), R01NS090874 (J.P.C.), R01NS046424 (S.E.P.), and R01AG038490 (J.E.P.), the McDonnell Center for Systems Neuroscience #220020256, (S.E.P.) K01 MH103594 (A.T.E.) and an Autism Speaks Postdoctoral Translational Research Fellowship #7962 (A.T.E.). The funding sources had no involvement in the study design, collection, analysis, interpretation of the data, writing of the paper, or decision to submit the paper for publication. J.P.C. and Washington University have financial interests in Cephalogics LLC based on a license of related optical imaging technology by the University to Cephalogics LLC.

References

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. NeuroImage 8, 360–369.
Bandettini, P.A., Jesmanowicz, A., Van Kylen, J., Birn, R.M., Hyde, J.S., 1998. Functional MRI of brain activation induced by scanner acoustic noise. Magn. Reson. Med. 39, 410–416.
Baumgartner, C., 2004. Controversies in clinical neurophysiology. MEG is superior to EEG in the localization of interictal epileptiform activity: con. Clin. Neurophysiol. 115, 1010–1020.
Bornkessel, I., Zysset, S., Friederici, A.D., von Cramon, D.Y., Schlesewsky, M., 2005. Who did what to whom? The neural basis of argument hierarchies during language comprehension. NeuroImage 26, 221–233.
Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.
Cabeza, R., Nyberg, L., 1997. Imaging cognition: an empirical review of PET studies with normal subjects. J. Cogn. Neurosci. 9, 1–26.
Caldwell, A., Nittrouer, S., 2013. Speech perception in noise by children with cochlear implants. J. Speech Lang. Hear. Res. 56, 13–30.
Caplan, D., Stanczak, L., Waters, G., 2008. Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. J. Cogn. Neurosci. 20, 643–656.
Cooke, A., Zurif, E.B., DeVita, C., Alsop, D., Koenig, P., Detre, J., Gee, J., Pinango, M., Balogh, J., Grossman, M., 2002. Neural basis for sentence comprehension: grammatical and short-term memory components. Hum. Brain Mapp. 15, 80–94.
Custo, A., Boas, D.A., Tsuzuki, D., Dan, I., Mesquita, R., Fischl, B., Grimson, W.E.L., Wells, W., 2010. Anatomical atlas-guided diffuse optical tomography of brain activation. NeuroImage 49, 561–567.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis — I. Segmentation and surface reconstruction. NeuroImage 9, 179–194.
Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431.
Dehghani, H., Eames, M.E., Yalavarthy, P.K., Davis, S.C., Srinivasan, S., Carpenter, C.M., Pogue, B.W., Paulsen, K.D., 2009a. Near infrared optical tomography using NIRFAST: algorithm for numerical model and image reconstruction. Commun. Numer. Methods Eng. 25, 711–732.
Dehghani, H., White, B.R., Zeff, B.W., Tizzard, A., Culver, J.P., 2009b. Depth sensitivity and image reconstruction analysis of dense imaging arrays for mapping brain function with diffuse optical tomography. Appl. Opt. 48, D137–D143.
Eggebrecht, A.T., White, B.R., Ferradal, S.L., Chen, C., Zhan, Y., Snyder, A.Z., Dehghani, H., Culver, J.P., 2012. A quantitative spatial comparison of high-density diffuse optical tomography and fMRI cortical mapping. NeuroImage 61, 1120–1128.
Eggebrecht, A.T., Ferradal, S.L., Robichaux-Viehoever, A., Hassanpour, M.S., Dehghani, H., Snyder, A.Z., Hershey, T., Culver, J.P., 2014. Mapping distributed brain function and networks with diffuse optical tomography. Nat. Photonics 8, 448–454.
Ferradal, S.L., Eggebrecht, A.T., Hassanpour, M., Snyder, A.Z., Culver, J.P., 2014. Atlas-based head modeling and spatial normalization for high-density diffuse optical tomography: in vivo validation against fMRI. NeuroImage 85, 117–126.
Fiebach, C.J., Schlesewsky, M., Friederici, A.D., 2001. Syntactic working memory and the establishment of filler-gap dependencies: insights from ERPs and fMRI. J. Psycholinguist. Res. 30, 321–338.
Foster, J.R., Hall, D.A., Summerfield, A.Q., Palmer, A.R., Bowtell, R.W., 2000. Sound-level measurements and calculations of safe noise dosage during EPI at 3 T. J. Magn. Reson. Imaging 12, 157–163.
Friederici, A.D., Ruschemeyer, S.A., Hahne, A., Fiebach, C.J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb. Cortex 13, 170–177.
Gaab, N., Gabrieli, J.D.E., Glover, G.H., 2007. Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. Hum. Brain Mapp. 28, 721–732.
Gibson, E., 1998. Linguistic complexity: locality of syntactic dependencies. Cognition 68, 1–76.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.
Gregg, N.M., White, B.R., Zeff, B.W., Berger, A.J., Culver, J.P., 2010. Brain specificity of diffuse optical imaging: improvements from superficial signal regression and tomography. Front. Neuroenerg. 2, 14.
Griffiths, J.D., Marslen-Wilson, W.D., Stamatakis, E.A., Tyler, L.K., 2013. Functional organization of the neural language system: dorsal and ventral pathways are critical for syntax. Cereb. Cortex 23, 139–147.
Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., Collet, L., 2001. Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. Br. J. Audiol. 35, 173–182.
Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.
Hassanpour, M.S., White, B.R., Eggebrecht, A.T., Ferradal, S.L., Snyder, A.Z., Culver, J.P., 2014. Statistical analysis of high density diffuse optical tomography. NeuroImage 85 (Pt 1), 104–116.
Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J., Nichols, T.E., 2004. Nonstationary cluster-size inference with random field and permutation methods. NeuroImage 22, 676–687.
He, B., 1999. Brain electric source imaging: scalp Laplacian mapping and cortical imaging. Crit. Rev. Biomed. Eng. 27, 149–188.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.
Jermyn, M., Ghadyani, H., Mastanduno, M.A., Turner, W., Davis, S.C., Dehghani, H., Pogue, B.W., 2013. Fast segmentation and high-quality three-dimensional volume mesh creation from medical images for diffuse optical tomography. J. Biomed. Opt. 18, 86007.
Joseph, D.K., Huppert, T.J., Franceschini, M.A., Boas, D.A., 2006. Diffuse optical tomography system to image brain activation with improved spatial resolution and validation with functional magnetic resonance imaging. Appl. Opt. 45, 8142–8151.
Koch, S.P., Habermehl, C., Mehnert, J., Schmitz, C.H., Holtze, S., Villringer, A., Steinbrink, J., Obrig, H., 2010. High-resolution optical functional mapping of the human somatosensory cortex. Front. Neuroenerg. 2, 12.
McJury, M., Shellock, F.G., 2000. Auditory noise associated with MR procedures: a review. J. Magn. Reson. Imaging 12, 37–45.
Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., Buckner, R.L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. NeuroImage 11, 735–759.
Moelker, A., Pattynama, P.M., 2003. Acoustic noise concerns in functional magnetic resonance imaging. Hum. Brain Mapp. 20, 123–141.
Okamoto, M., Dan, I., 2005. Automated cortical projection of head-surface locations for transcranial functional brain mapping. NeuroImage 26, 18–28.
Olsen, S.O., 1999. The relationship between the uncomfortable loudness level and the acoustic reflex threshold for pure tones in normally-hearing and impaired listeners — a meta-analysis. Audiology 38, 61–68.
Osnes, B., Hugdahl, K., Specht, K., 2011. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. NeuroImage 54, 2437–2445.
Peelle, J.E., 2012. The hemispheric lateralization of speech processing depends on what "speech" is: a hierarchical perspective. Front. Hum. Neurosci. 6, 309.
Peelle, J.E., 2014. Methodological challenges and solutions in auditory functional magnetic resonance imaging. Front. Neurosci. 8, 253.
Peelle, J.E., McMillan, C., Moore, P., Grossman, M., Wingfield, A., 2004. Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI. Brain Lang. 91, 315–325.
Peelle, J.E., Johnsrude, I.S., Davis, M.H., 2010a. Hierarchical processing for speech in human auditory cortex and beyond. Front. Hum. Neurosci. 4, 51.
Peelle, J.E., Troiani, V., Wingfield, A., Grossman, M., 2010b. Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity. Cereb. Cortex 20, 773–782.
Peelle, J.E., Troiani, V., Grossman, M., Wingfield, A., 2011. Hearing loss in older adults affects neural systems supporting speech comprehension. J. Neurosci. 31, 12638–12643.
Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. NeuroImage 62, 816–847.
Price, D.L., De Wilde, J.P., Papadaki, A.M., Curran, J.S., Kitney, R.I., 2001. Investigation of acoustic noise on 15 MRI scanners from 0.2 T to 3 T. J. Magn. Reson. Imaging 13, 288–293.
Pulvermüller, F., Huss, M., Kherif, F., Martin, F.M.D.P., Hauk, O., Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. Proc. Natl. Acad. Sci. U. S. A. 103, 7865–7870.
Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724.
Ravicz, M.E., Melcher, J.R., Kiang, N.Y.S., 2000. Acoustic noise during functional magnetic resonance imaging. J. Acoust. Soc. Am. 108, 1683–1696.
Saager, R., Berger, A., 2008. Measurement of layer-like hemodynamic trends in scalp and cortex: implications for physiological baseline suppression in functional near-infrared spectroscopy. J. Biomed. Opt. 13, 034017.
Schmidt, C.F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., Jancke, L., 2008. Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. Hum. Brain Mapp. 29, 46–56.
Shah, N.J., Jancke, L., Grosse-Ruyken, M.L., Muller-Gartner, H.W., 1999. Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. J. Magn. Reson. Imaging 9, 19–25.
Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.
Skouras, S., Gray, M., Critchley, H., Koelsch, S., 2013. fMRI scanner noise interaction with affective neural processes. PLoS One 8, e80564.
Stromswold, K., Caplan, D., Alpert, N., Rauch, S., 1996. Localization of syntactic comprehension by positron emission tomography. Brain Lang. 52, 452–473.
Traxler, M.J., Morris, R.K., Seely, R.E., 2002. Processing subject and object relative clauses: evidence from eye movements. J. Mem. Lang. 47, 69–90.
Tyler, L.K., Marslen-Wilson, W., 2008. Fronto-temporal brain systems supporting spoken language comprehension. Philos. Trans. R. Soc., B 363, 1037–1054.
Tyler, L.K., Wright, P., Randall, B., Marslen-Wilson, W.D., Stamatakis, E.A., 2010. Reorganization of syntactic processing following left-hemisphere brain damage: does right-hemisphere activity preserve function? Brain 133, 3396–3408.
Ulmer, J.L., Biswal, B.B., Mark, L.P., Mathews, V.P., Prost, R.W., Millen, S.J., Garman, J.N., Horzewski, D., 1998. Acoustic echoplanar scanner noise and pure tone hearing thresholds: the effects of sequence repetition times and acoustic noise rates. J. Comput. Assist. Tomogr. 22, 480–486.
White, B.R., Culver, J.P., 2010. Quantitative evaluation of high-density diffuse optical tomography: in vivo resolution and mapping performance. J. Biomed. Opt. 15, 026006.
Wilson, S.M., Molnar-Szakacs, I., Iacoboni, M., 2008. Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. Cereb. Cortex 18, 230–242.
Wingfield, A., Grossman, M., 2006. Language and the aging brain: patterns of neural compensation revealed by functional brain imaging. J. Neurophysiol. 96, 2830–2839.
Wingfield, A., Peelle, J.E., Grossman, M., 2003. Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. Aging Neuropsychol. Cognit. 10, 310–322.
Worsley, K.J., Cao, J., Paus, T., Petrides, M., Evans, A.C., 1998. Applications of random field theory to functional connectivity. Hum. Brain Mapp. 6, 364–367.
Zeff, B.W., White, B.R., Dehghani, H., Schlaggar, B.L., Culver, J.P., 2007. Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. Proc. Natl. Acad. Sci. U. S. A. 104, 12169–12174.