Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com

Hybrid auditory fMRI: In pursuit of increasing data acquisition while decreasing the impact of scanner noise

Matthew Heard [a], Xiangrui Li [b], Yune S. Lee [a,*]

[a] School of Behavioral and Brain Sciences, University of Texas at Dallas, United States
[b] Center for Cognitive and Behavioral Brain Imaging, The Ohio State University, United States

ARTICLE INFO

Keywords:
Functional MRI
Auditory neuroimaging
Language
Speech perception
Silent imaging
Fast imaging
MR sequence

ABSTRACT

Background: Two challenges in auditory fMRI include the loud scanner noise during sound presentation and slow data acquisition. Here, we introduce a new auditory imaging protocol, termed "hybrid", that alleviates these obstacles.

New method: We designed a within-subject experiment (N = 14) wherein language-driven activity was measured by hybrid, interleaved silent (ISSS), and continuous multiband acquisition. To determine the advantage of noise attenuation during sound presentation, hybrid was compared to multiband. To identify the benefits of increased temporal resolution, hybrid was compared to ISSS. Data were evaluated by whole-brain univariate general linear modeling (GLM) and multivariate pattern analysis (MVPA).

Results: Comparison with existing methods:

- Hybrid vs. Multiband: in both GLM and MVPA, hybrid showed widespread activation throughout the language network including the left inferior frontal gyrus, bilateral superior temporal regions, thalamus, and inferior colliculus. By contrast, multiband showed activity mostly within the left frontotemporal cortices.
- Hybrid vs. ISSS: in MVPA, hybrid yielded more activation than ISSS throughout the language network. However, in GLM, hybrid detected less activation than ISSS. Despite the reduction of activation, hybrid more specifically detected activity in the canonical language network compared to ISSS.

Conclusions: Our data revealed that hybrid imaging restored neural activity in the canonical language network that was absent due to the loud noise or slow sampling in the conventional imaging protocols. With its noise-attenuated sound presentation windows and increased acquisition speed, the hybrid protocol is well-suited for auditory fMRI research tracking neural activity pertaining to fast, time-varying acoustic events.

1. Introduction

For fMRI researchers who wish to present sound stimuli, scanner noise may be a great concern. The high-frequency noise of fMRI generated by the gradient coils can reach up to 120 dB, which can easily mask out the sound-of-interest. Worse, it can cause instant hearing damage if precautions are not taken (Peelle, 2014; Price et al., 2001). To alleviate these issues, sound stimuli are often delivered via MRI-compatible ear plugs while subjects wear earmuffs to passively decrease background noise. Nevertheless, the presence of this attenuated scanner noise still competes with the auditory stimuli and potentially alters brain activity (Peelle, 2014). For example, it has been shown in monkeys as well as humans that blood oxygenation-level dependent (BOLD) activity in response to acoustic stimuli decreased within the primary auditory cortices, likely due to saturation by scanner noise (Gaab et al., 2007; Langers et al., 2005; Peelle, 2014; Petkov et al., 2009; Talavage and Edmister, 2004). Furthermore, the presence of scanner noise may recruit additional neural resources associated with effortful listening that may not be reflective of everyday listening tasks (Eckert et al., 2016; Peelle, 2014, 2018). From an analytic perspective, the subtraction-based approach (e.g., [sound A + noise] – [sound

* Corresponding author.
Email address: yunesang.lee@gmail.com (Y.S. Lee)

https://doi.org/10.1016/j.jneumeth.2021.109198
Received 21 October 2020; Received in revised form 28 March 2021; Accepted 16 April 2021
Available online xxx
0165-0270/

B + noise]) does not fully remove a potential interaction between the sound stimuli and scanner noise (Adank, 2012; Peelle, 2014).

Numerous attempts have been made to address these challenges. For example, a sparse imaging protocol was first introduced by Hall et al. (1999), which enabled sound presentation during a silent period. However, a critical drawback of sparse imaging lies in its lack of efficiency in data acquisition (Fig. 1A). This method not only hampers statistical power and provides no information regarding the temporal characteristics of the BOLD response, but it also limits application of standard modeling procedures such as convolution with the hemodynamic response function (HRF) due to the discontinuous nature of the dataset. More recently, the interleaved silent steady state (ISSS) protocol (Schwarzbauer et al., 2006) was developed, which allowed for consecutive data acquisition after stimulus presentation during the silent period (Fig. 1B). Hence, ISSS not only increases statistical power, it also allows for a more complete reconstruction of the BOLD signal (Lee et al., 2016, 2018).

However, typical ISSS protocol is conducted with a repetition time (TR) of 2–3 seconds; the slow temporal resolution is particularly unfavorable for capturing fast and time-varying acoustic events in auditory fMRI experiments. Indeed, attempts have been made to enhance temporal resolution in the field of fMRI physics. For example, multiband acceleration (Larkman et al., 2001; Moeller et al., 2010; Nunes et al., 2006; Xu et al., 2013) has enabled ultra-fast data acquisition by imaging multiple slices of tissue simultaneously (Fig. 1C). Unfortunately, the vast majority of multiband-accelerated protocols have been applied to non-auditory domains; there are only a few fMRI studies exploiting acceleration to accommodate fast auditory events such as phonemes (Correia et al., 2015; De Martino et al., 2015; Evans and McGettigan, 2017). For instance, De Martino et al. (2015) and Correia et al. (2015) both used a multiband-accelerated acoustic MRI protocol which collected single EPIs followed by brief silent gaps, during which various acoustic stimuli (e.g., voices, music, tool sounds, animals) were presented. Nevertheless, these auditory protocols are similar to sparse imaging in that they capture only a single image of the BOLD signal, which attenuates the maximum benefit of increased temporal resolution. Additionally, this protocol requires acoustic stimuli to be kept extremely brief.

In the present study, we developed a new auditory imaging protocol—termed *hybrid*—that integrates an interleaved silent scanning protocol with fast imaging acquisition (Fig. 1D). The hybrid protocol allowed us to present much longer auditory stimuli during the silent period and to collect consecutive images with a much faster sampling rate. To explore its practical benefits in auditory fMRI research, we designed a within-subject experiment in which language-driven activity was collected by hybrid, ISSS, and multiband imaging. We then analyzed these data using both mass univariate analysis with general linear modeling (GLM) and whole-brain MVPA searchlights (Kriegeskorte et al., 2006). The aims were twofold: first, to identify the impact of minimized scanner noise, the results were compared between hybrid and multiband. We hypothesized that hybrid would restore language-evoked activity that could have been susceptible to background noise during the continuous imaging protocol. Next, for the purpose of identifying the consequence of increased sampling rate in hybrid imaging, its results were compared to that of slower ISSS imaging. We hypothesized that hybrid, with its superior temporal resolution, would capture subtle differences in activation patterns that are invisible to ISSS, which can particularly benefit MVPA. However, hybrid and ISSS may exhibit comparable performance in GLM because increasing sampling rate would not likely increase the size of overall amplitude in BOLD activity (see Methods detailing the metric of summed positive area).

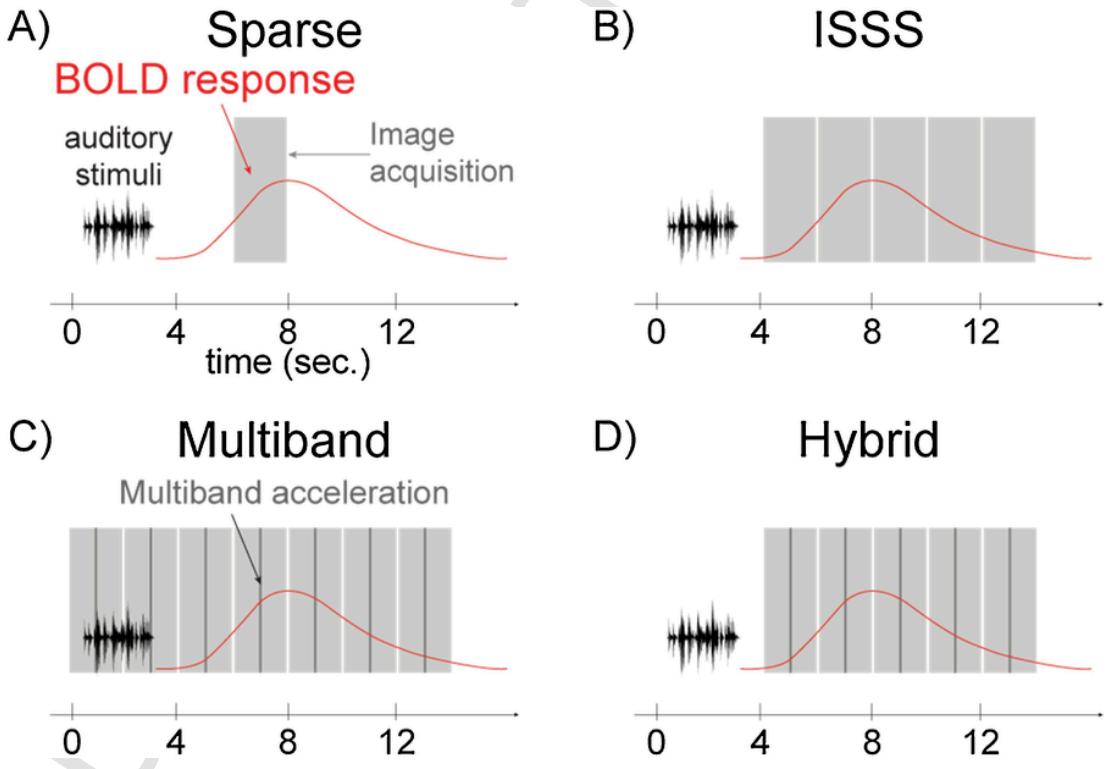

Fig. 1. A) Sparse imaging that acquires a single image after stimulus presentation; **(B)** Interleaved Silent Steady Scanning (ISSS) that acquires 5 images, 1 image per 2 s, after a 4 s silent window for sound presentation; **(C)** Multiband that continuously acquires 14 images with no silent window, 1 image per 1 s; **(D)** Hybrid that acquires 10 images, 1 image per 1 s, after a 4 s silent window reserved for sound presentation.

2. Methods

2.1. Participants

Nineteen participants were recruited from The Ohio State University student body. Each subject received either monetary reward or extra class credit for participation. Data from 5 subjects were removed from the final analyses, 3 of which were pilot subjects removed due to a severe tissue contrast artifact observed in both hybrid and ISSS imaging. This was due to steady-state magnetization being accidentally disabled during the silent period, which affected the tissue contrast of images acquired up to 6 s after the silent period. In addition, 2 subjects' data were discarded due to poor behavioral performance (at or below chance-level). This left a total of 14 intact data (7 females, mean age = 21.2 years, *SD* = 1.8 years; age range = 19–24 years, all right-handed). Participants were briefed on the experimental task before entering the scanner. Consent to participate in the experiment was obtained per Ohio State IRB protocols.

2.2. Stimuli and procedure

The sentence stimuli used were a subset of stimuli obtained from previous studies (Lee et al., 2016). Half of the sentences were embedded with a subject-relative (SR) clause and the other half were embedded with an object-relative (OR) clause. Both sentences were constructed by re-arranging the order of constituent words (e.g., SR: Boys that *help girls* are happy; OR: Boys that *girls help* are happy). Every sentence included two persons, one male and one female, and subjects indicated the gender of the agent in the sentence via button press. A total of 48 base sentences were used to construct SR and OR sentences and the gender was counterbalanced across sentences. Periodically, an unintelligible sentence (i.e., 1 ch vocoded sentence) was presented. During these 1 ch noise trials, subjects were instructed to randomly press either the male or female button, thus controlling for motor activity throughout trials. There were also silent events in which no stimuli were presented. All sound files were root-mean-square (RMS) normalized to match mean intensity. The MATLAB code used to vocode and normalize auditory stimuli is available at https://github.com/jpeelle/jp_matlab.

Before entering the scanner, subjects were given a brief practice session consisting of 4 trials in order to acclimate to the task. These sentences were not included in the main experiment. In each functional run, participants underwent a total of 16 trials that were evenly distributed across the 4 conditions (SR, OR, noise, silence). The sound stimuli (Mean duration = 1.74 s, SD = 0.10 s) were presented during the first 4 s of the silent window with +/-0.5 s of jittering on the onset. All sentences were presented to participants using the Siemens pneumatic sound system that is built into the MRI scanner, fitted with MRI-compatible earphones (Sensimetrics Corp., USA). Behavioral responses were tracked using a Response Time Box (Li et al., 2010), an MRI-compatible button box system (Current Designs, USA). Subjects were asked to maintain visual fixation on a cross during the task. The fixation screen consisted of a black cross on a gray background and was presented using a 3-chips DLP projector (Christie Digital Systems, USA). Participants viewed the screen projection using a coil-mounted mirror. Timing and stimuli presentation were managed by PsychToolbox version 3 (Kleiner et al., 2007) run on MATLAB 2014a (Mathworks, USA).

2.3. fMRI acquisition

Images were acquired at the Center for Cognitive and Behavioral Brain Imaging at The Ohio State University (Columbus, Ohio) using a 3 T Siemens PRISMA system (Erlangen, Germany). All images were acquired using a 32-channel phase array receiver head coil. Foam inserts were placed along each participant's head to ensure comfort, provide additional noise protection, and minimize head motion. Scanning began with structural imaging using a standard magnetization-prepared rapid acquisition gradient echo (MPRAGE) protocol (voxel size = 1 × 1 × 1 mm^3, TR = 1900 ms, TE = 4.44 ms, TI = 950 ms, 176 slices, matrix = 224 × 256 pixels). Next, B_0 field maps were acquired for unwarping of functional data (voxel size 3 × 3 × 3 mm^3, TR = 500 ms, TE1 = 5.17 ms, TE2 = 7.63 ms, 45 slices). Last, participants underwent 6 functional runs with echo planar imaging (EPI): 2 runs per each of the ISSS, multiband, and hybrid protocols. Across all acquisition paradigms, the following parameters were held constant: voxel size (3 × 3 × 3 mm^3), echo time (28 ms), number of slices acquired (36), slice thickness (3 mm, no gap between slices), acquisition (ascending), field of view (216 × 216 mm^2), and matrix size (72 × 72 pixels). The following parameters, by necessity, differed across protocols: multiband acceleration factor (1 in ISSS, 2 in hybrid and multiband), and repetition time (2000 ms in ISSS, 1000 ms in hybrid and multiband). Flip angle of ISSS (72°) was determined based upon previous studies (Lee et al., 2016, 2018) that reliably yielded language-evoked activity. For hybrid and multiband acquisition, a flip angle of 52° was chosen due to the shorter TR (Ernst and Anderson, 1966). The order of runs was counterbalanced across subjects.

2.4. Analysis

2.4.1. Data inspection

Upon completion of data collection, close inspection of the raw MR images revealed tissue contrast artifacts in the first 2 s of images after the silent period in ISSS and hybrid. Compared to other images acquired during these runs, the signal strength from ventricles and gray matter appeared higher while signal strength from white matter was reduced, akin to an anatomical T2 image. This artifact persisted even though steady state magnetization had been enabled for data collection. Consequently, these images were not included in the data analysis for both protocols. Although no such artifact was seen in the first two images of the multiband protocol, these images were also removed from data analysis for consistency. We confirmed the quality of these latter images spanning 2–10 seconds with visual inspection of the raw data as well as the temporal signal-to-noise ratio (tSNR) map. The tSNR map, drawn from the raw data, stored an index per voxel of image quality and was computed by dividing the mean intensity of a voxel by its standard deviation per each of 6 runs (Parrish et al., 2000). No artifacts caused by signal fluctuation, scanner malfunctions, or movement were detected by visual inspection of these maps. In addition, we compared mean tSNR maps (i.e., averaged across 2 runs and across voxels) for each of the 3 protocols via a repeated measures ANOVA in MATLAB (2017a). There were no significant differences in the tSNRs across the protocols [$F(2, 26) = 1.12$, $p = 0.34$], further assuring comparable data quality across all imaging protocols.

2.4.2. Preprocessing

Preprocessing was separately performed for each of the imaging protocols using SPM12 as follows. First, functional images were un-warped using a B_0 field map and realigned to the first image of the first run of each protocol using an affine transformation. Motion traces were visually inspected; one run of one subject demonstrated a spike of translational movement greater than 2 mm and was discarded from analysis. Then, a mean image generated by realignment was used in co-registration, to which the structural MPRAGE was aligned without re-slicing. Normalization was performed to transform all functional data from the native space to the MNI space using a transformation matrix generated by tissue segmentation on the co-registered structural image. Finally, two sets of spatially smoothed data were created, one

with an 8 mm full width at half maximum (FWHM) Gaussian kernel and another with a 3 mm FWHM. Slice timing correction (STC) was not included in the pipeline since it has minimal impact on the quality of images acquired using TR shorter than 1 s. Although STC is recommended for images acquired with a TR of 2 s by some studies (Parker et al., 2017; Parker and Razlighi, 2019; Power et al., 2017), our exploratory analysis of ISSS data with STC did not yield any better results (data not shown) than the data without it.

2.4.3. Univariate fMRI analysis

The 8-mm smoothed data were used for subsequent first level general linear modeling (GLM) analyses performed separately on each of the three protocols. We opted to use a finite impulse response (FIR) function to estimate activity for each discrete time point following stimulus presentation (hybrid and multiband: window length = 8 s; order = 8 (window length / TR); ISSS: window length = 8 s; order = 4). Thus, hybrid and multiband yielded 8 beta estimates per each condition (i.e., SR, OR, and 1 ch-vocoded) and ISSS yielded 4 beta estimates per each condition. We used the aCompCorr algorithm in the CONN toolbox (Behzadi et al., 2007; Whitfield-Gabrieli and Nieto-Castanon, 2012) to derive the first two principal components within cerebrospinal fluid and white matter respectively. Additional regressors were included in the models for motion and run effects. In total, 1st level models included 6 motion parameters, 2 cerebrospinal fluid regressors, 2 white matter regressors, and 2 run regressors per each protocol. High-pass filtering (SPM default of 128 s cutoff) was applied, and the first-order autoregressive AR(1) modeling for temporal autocorrelation was turned off due to the discontinuous nature of hybrid and ISSS. Because all images following the silent period were modeled discontinuously and without AR(1) in both hybrid and ISSS , the same parameter specification was applied to the multiband models. We note that our exploratory analysis of the continuous multiband time-series (14 s) with AR(1) did not show improved results (data not shown).

Subsequently, we computed the average summed-positive-area (SPA) underneath each series of beta estimates per condition at every voxel (Lee et al., 2016, 2018). With SPAs, we were able to measure the size of activation across time points (Fig. 2C). The contrast SPA maps (e.g., sentence vs. noise) were then submitted for 2nd level random effect analysis in SPM. Task accuracy was included in the model as a covariate to control for performance differences across subjects. The resulting t-maps were evaluated under a voxel-wise threshold of $p < .001$ (uncorrected) in combination with a cluster-wise threshold of $p < .05$ (corrected using family-wise error) based on Gaussian random field theory (Friston et al., 1994). The Brainnetome atlas (Fan et al., 2016) was used for anatomical labeling.

2.4.4. Multivariate fMRI data analyses

For multivariate analyses, we used another set of preprocessed data that were moderately smoothed (i.e. 3 mm FWHM; Hendriks et al., 2017). We performed a whole-brain searchlight analysis (Kriegeskorte et al., 2006) with 2-voxel radius sphere, wherein we performed a binary classification between sentences (OR + SR) and 1 ch vocoded stimuli using a Gaussian naïve Bayes (GNB) linear classifier (Raizada and Lee, 2013). We performed a 2-fold cross validation in which beta estimates from each of the 2 runs alternately served as a training and testing set. We excluded one subject who had only one run per protocol due to excessive head motion. To balance the amount of data between the two classes, we duplicated the input vectors of 1 ch vocoded speech. The chance-level accuracy was adjusted from 0.5 to 0 in the resulting individual searchlight maps in order to leverage the 2nd level random effect analysis (RFX) pipeline in SPM12 for one sample *t*-test. Behavioral accuracy of each participant was included as a covariate in order to remove the effects of task performance. The resulting t-maps were evaluated under a voxel-wise threshold of $p < .001$ (uncorrected) with a cluster-wise threshold of $p < .05$ (corrected using family-wise error). The Brainnetome atlas (Fan et al., 2016) was used for anatomical labeling.

All tables of activation foci report the three coordinates with the highest activation located within each of the significant clusters. Full activation foci for univariate and multivariate analyses can be found in the Supplementary tables. Un-thresholded T maps of all resulting images are available on Neurovault (https://identifiers.org/neurovault.collection:8563, Gorgolewski et al., 2015).

3. Results

3.1. Behavioral

Overall, participants performed relatively well on the task during fMRI scanning (M = 90.6 %; SD = 0.12). On accuracy data, we performed 3-by-2 repeated measures ANOVAs to test for main effects of scanning protocol (hybrid, ISSS, or multiband), syntax (object- or subject-relative), and interactions of the two. There were no significant main effects of syntax [$F(1, 13) = 3.43$, $p = 0.087$], scanning protocol [$F(2, 26) = 0.33$, $p = 0.72$], nor interaction of the two [$F(2, 26) = 0.19$, $p = 0.82$]. Likewise, in response time, there were no main effects of syntax [$F(1, 13) = 4.08$, $p = 0.064$], scanning protocol [$F(2, 26) = 0.162$, $p = 0.85$], nor interaction [$F(2, 26) = 0.534$, $p = 0.59$]. To explore whether the trending effect of syntax on accuracy and reaction time was due to lack of power, the unbiased effect size of the syntax manipulation was calculated using ESCI (Cumming and Calin-Jageman, 2016). Indeed, participants were more accurate in responding to subject-relative sentences (M = 93.1 %, SD = 0.10) compared to object-relative sentences (M = 88.1 %, SD = 0.14), separated by a medium effect size (d_{unb} = 0.407). Similar trends were observed in the reaction time data, with participants responding more quickly to subject-relative (M = 1.04 s, SD = 0.46) than object-relative sentences (M = 1.12, SD = 0.51), but the effect size was smaller (d_{unb} = 0.172).

3.2. fMRI (Hybrid versus Multiband)

We first evaluated the consequence of minimized background noise by comparing hybrid to multiband on the [sentence vs. 1 ch noise] contrast. In GLM, hybrid yielded widespread activity in the bilateral superior temporal gyri, anterior insula, superior frontal gyri/pre-SMA region, as well as in the left inferior frontal gyrus, left precentral gyrus/premotor cortex, and the right cerebellum (Fig. 2**A;** Table 1). In addition, subcortical activity was observed in the anterior thalamus and inferior colliculus (Fig. 2B). Unlike the hybrid protocol, multiband yielded activity predominantly in the left perisylvian network (Fig. 2**A;** Table 2): a majority of the significant cortical clusters were observed in the left superior temporal gyrus, precentral gyrus, insula, and inferior frontal gyrus. Midline activity was also seen in the superior frontal gyrus/pre-SMA. Although several clusters emerged in the right cerebellar hemisphere, no subcortical clusters appeared in the multiband data (Fig. 2B). In sum, hybrid yielded 10 significant clusters (a total of 39,447 mm^3) and multiband yielded 6 clusters (a total of 22,869 mm^3), all of which have previously been implicated in the neuroimaging of sentence processing literature as confirmed by Neurosynth (Yarkoni et al., 2011).

To further explore the subcortical activity in the anterior thalamus and inferior colliculus that appeared in hybrid, but not in multiband, we extracted and plotted beta values for both sentence and 1-ch noise conditions (Fig. 2C). We also compared the SPAs within these clusters; SPA values for the language condition were greater in hybrid than in multiband, while an opposite pattern was observed in 1-ch noise (Fig. 2D). Such interaction patterns indicate that hybrid selectively amplified language-related activity that may be suppressed by the scanner

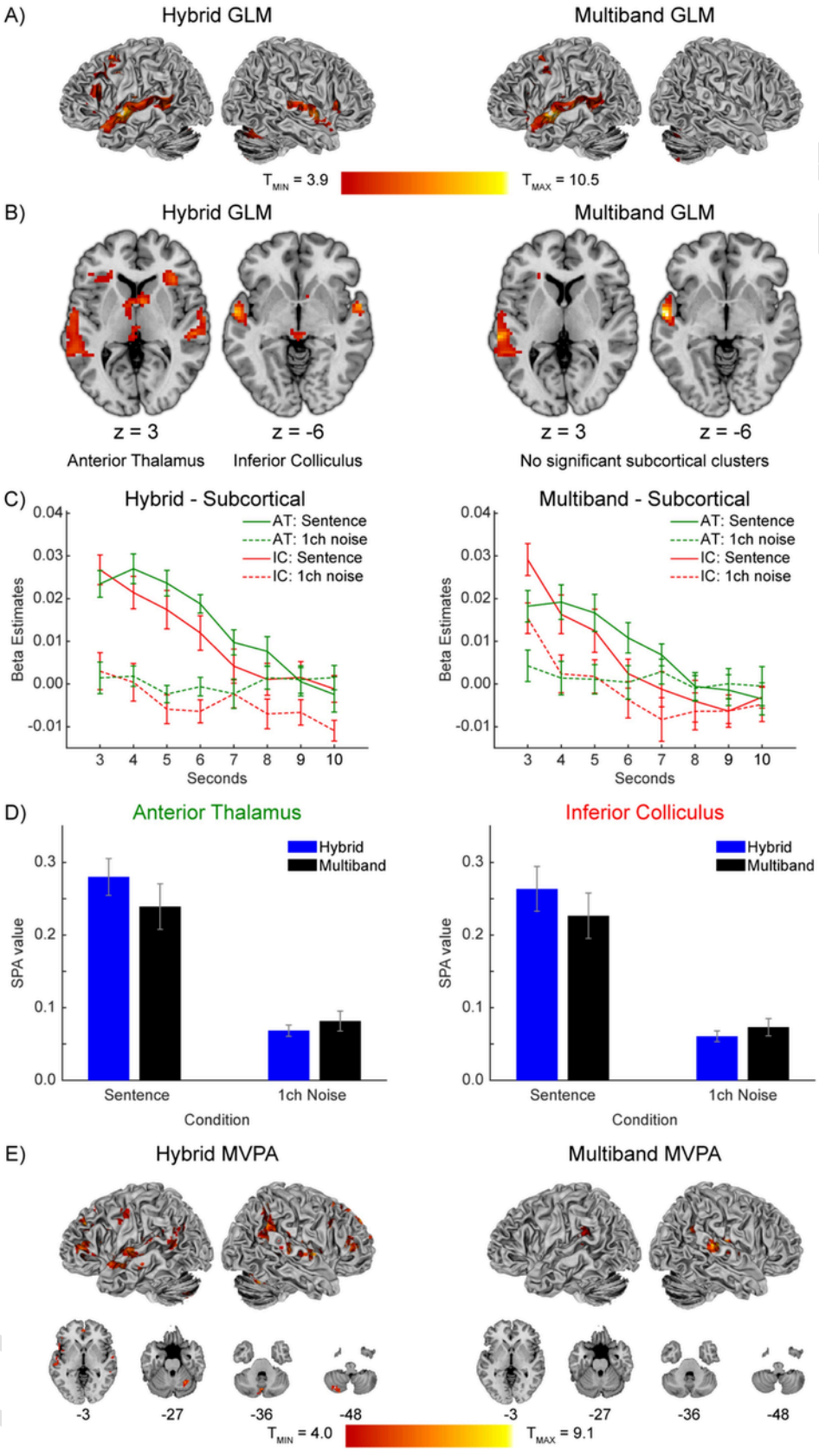

Fig. 2. **(A)** GLM results (sentence vs. 1 ch vocoded noise contrast) of hybrid and multiband. Both protocols show canonical language activation in the left hemisphere, but the hybrid protocol revealed additional activity in the right superior temporal gyrus and subcortical regions; **(B)** Representative horizontal images highlighting different GLM results in hybrid and multiband. Only hybrid reveals right superior temporal gyrus, right anterior insula and subcortical activation; **(C)** Beta estimate plots demonstrating the profile of activation within the

anterior thalamus (AT) and inferior colliculus (IC) that is significant during hybrid, but not multiband, acquisition; **(D)** SPA metrics of stimuli type across scanning protocols in the anterior thalamus and inferior colliculus. The interaction pattern between scanning protocol and stimuli type is consistently present across regions; **(E)** MVPA results of hybrid and multiband, with representative horizontal views below. Hybrid acquisition yielded widespread activity throughout the language network. By contrast, multiband acquisition yielded much less focal activity in the right posterior STG and bilateral inferior parietal lobules.

Table 1
Foci of activation for data acquired using hybrid acquisition and analyzed using univariate methods. Labels are from the Brainnetome atlas.

		MNI Coordinates			
Region name	T Score	X	Y	Z	Volume of cluster (mm^3)
L STG	10.00	−60	−10	−4	9369
L MTG	9.86	−66	−46	5	
L posterior STS	7.05	−54	−49	11	
R cerebellum	4.94	15	−70	−28	6048
R cerebellum	7.78	12	−79	−37	
R cerebellum	6.77	12	−73	−25	
L precentral gyrus	7.06	−48	5	47	4725
L precentral gyrus	6.34	−48	8	38	
L precentral gyrus	5.99	−42	−1	38	
L IFG	6.44	−51	20	14	3861
L IFG	6.29	−39	23	8	
L insula	5.36	−30	17	14	
R striatum	4.53	6	5	−1	3402
L thalamus	4.17	−9	−7	5	
L thalamus	3.54	−6	−13	11	
R STG	4.56	60	−1	−7	3294
R STG	4.26	57	−28	2	
R STG	3.98	63	−25	5	
R SFG	4.68	6	11	50	2376
L SFG	4.15	−6	11	53	
R cingulate gyrus	3.58	9	23	35	
R cerebellum	4.35	24	−64	−52	2187
R cerebellum	4.09	30	−67	−52	
R cerebellum	3.89	18	−67	−46	
R insula	4.19	36	26	2	2133
R IFG	3.32	33	23	14	
Midbrain	4.11	−3	−25	−1	2052
Midbrain	3.76	3	−31	−7	
Midbrain	3.60	−6	−25	−10	

STG: Superior temporal gyrus; MTG: Middle temporal gyrus; STS: Superior temporal sulcus; IFG: Inferior frontal gyrus; SFG: Superior frontal gyrus.

noise in multiband imaging. Accordingly, the difference in SPA between language and 1 ch noise became smaller in the multiband dataset, providing a possible explanation for the lack of significant subcortical activity in GLM.

Consistent with the univariate GLM, MVPA of hybrid data revealed widespread activity in the bilateral superior temporal gyri, anterior insula, middle frontal gyri/premotor cortex, cerebellum, inferior parietal lobules, and cingulate gyri, and left inferior frontal gyrus (Table 3). By contrast, MVPA of multiband data yielded a few sporadic activations in the upper bank of the right posterior superior temporal gyrus and the bilateral inferior parietal lobule (Table 4). In sum, hybrid detected 21 clusters consisting of 18,846 mm^3 and multiband found 3 clusters (a total of 2,322 mm^3) throughout the language network.

3.3. fMRI (Hybrid versus ISSS)

To characterize the impact of increased temporal resolution, the hybrid and ISSS protocols were compared as above. In GLM results, both protocols yielded activation in the expected language network including left inferior frontal and bilateral superior temporal areas (Fig. 3**A and B,** Table 5). The two protocols also showed deep brain activity in the anterior thalamus, posterior midbrain, and right cerebellum. However, ISSS yielded additional clusters outside of the canonical language network including the bilateral inferior parietal lobule and right inferior frontal gyrus. Together, univariate analysis of hybrid data yielded 10 significant clusters (a total of 39,447 mm^3) and ISSS yielded 9 clusters (a total of 61,317 mm^3). In MVPA (Fig. 3C, Table 6), hybrid detected activity throughout the established language network—the superior temporal gyri, inferior frontal gyri, middle frontal gyri/premotor cortex, cerebellum, inferior parietal lobules, and cingulate gyri as described above (Table 3). ISSS performed poorly in MVPA compared to hybrid although small clusters were found in the language network as well: the left inferior frontal gyrus, left premotor cortex, and left posterior superior temporal gyrus (Table 6). Together, hybrid detected 21 clusters consisting of 18,846 mm^3 throughout the language network and ISSS found 3 clusters (a total of 891 mm^3) in the left frontotemporal region.

Table 2
Foci of activation for data acquired using multiband acquisition and analyzed using univariate methods. Labels are from the Brainnetome atlas.

		MNI Coordinates			
Region name	T Score	X	Y	Z	Volume of cluster (mm^3)
L MTG	10.50	−57	−10	−7	10,449
L MTG	8.82	−63	−31	2	
L STG	6.37	−51	11	−13	
R cerebellum	5.82	18	−67	−25	3402
R cerebellum	5.43	21	−58	−22	
R cerebellum	5.09	12	−70	−34	
L precentral gyrus	10.50	−39	2	35	3078
L precentral gyrus	5.25	−45	5	50	
L precentral gyrus	4.42	−54	−1	38	
R cerebellum	8.39	24	−64	−52	2133
R cerebellum	4.40	39	−67	−55	
R SFG	7.42	6	11	50	2025
L SFG	5.75	−6	8	53	
L insula	4.74	−30	20	14	1782
L IFG	4.55	−27	32	8	
L orbital gyrus	4.48	−27	23	−1	

MTG: Middle temporal gyrus; STG: Superior temporal gyrus; SFG: Superior frontal gyrus; IFG: Inferior frontal gyrus.

4. Discussion

In the present study, we evaluated a new hybrid auditory fMRI protocol featuring increased temporal resolution and interleaved silent windows—an approach introduced by this article for the first time. The advantage of this new protocol was hypothesized to be two-fold; first, the introduction of silent windows would reveal language areas that experience interference from scanner background noise. Second, the high temporal resolution would increase sensitivity (i.e., statistical power) to detecting language-driven brain activity. Our data confirmed the hypotheses, in that hybrid uncovered activation that was not observed in the two existing protocols—ISSS and multiband. Moreover, hybrid was less prone to the false positives than ISSS which yielded additional activity in the areas outside of the typical language network. In MVPA, hybrid yielded activity throughout the language network that was undetected by ISSS indicating a drastic improvement in statistical power in hybrid. Together, hybrid can offer an excellent solution for addressing loud scanner noise and slow temporal resolution—two long-standing obstacles of auditory fMRI. Below, we elaborate these findings in more detail.

Table 3
Foci of activation for data acquired using hybrid acquisition and analyzed using multivariate methods. Labels are from the Brainnetome atlas.

		MNI Coordinates			
Region name	T Score	X	Y	Z	Volume of cluster (mm^3)
L STG	8.40	−45	2	−10	3483
L MTG	7.41	−60	−25	−10	
L STG	6.85	−57	−4	−10	
R STG	7.58	63	−10	−1	2025
R STG	7.17	60	−10	5	
R STG	7.04	63	−4	−1	
R IPL	6.98	63	−49	26	1998
R IPL	5.81	51	−55	20	
R IPL	5.66	51	−52	14	
L MFG	7.45	−39	32	29	1782
L MFG	5.07	−27	23	50	
L SFG	4.86	−21	26	47	
R IFG	6.29	51	32	8	1107
R IFG	5.75	48	38	5	
R IFG	5.43	57	26	8	
L IFG	5.29	−36	17	11	891
L precentral gyrus	4.55	−42	−1	17	
L IFG	6.65	−51	35	2	864
L IFG	5.87	−51	32	8	
L IFG	5.09	−45	41	5	
L cerebellum	9.13	−33	−73	−49	783
L cerebellum	5.18	−24	−82	−49	
R cerebellum	7.08	30	−61	−31	729
R cerebellum	5.07	24	−67	−25	
L cingulate gyrus	6.04	0	47	−4	702
L orbital gyrus	5.77	−3	38	−16	
L cingulate gyrus	4.87	−3	44	−7	
R MFG	5.57	30	17	44	621
R MFG	5.20	33	17	53	
R MFG	4.72	36	20	32	
R SFG	5.14	15	65	11	459
R SFG	5.08	9	65	14	
R MFG	4.80	18	62	17	
L IPL	5.08	−51	−55	23	459
L IPL	4.90	−51	−61	23	
L IPL	4.25	−60	−55	20	
R SFG	5.87	24	35	50	432
R SFG	4.78	18	35	50	
R precentral gyrus	5.23	57	11	26	405
R IFG	4.34	51	14	29	
R precentral gyrus	4.23	48	5	32	
L MTG	5.18	−57	−55	8	405
L STG	5.12	−66	−37	17	378
L STG	4.95	−63	−40	20	
L STG	4.91	−60	−31	14	
Medial cerebellum	6.42	0	−76	−37	351
R MFG	7.02	24	50	35	324
R MFG	5.21	21	53	32	
R MFG	4.77	24	44	38	
R IPL	5.72	51	−49	47	324
R IPL	5.30	54	−52	41	
R IPL	4.33	54	−52	35	
L postcentral gyrus	4.62	−57	−10	38	324
L postcentral gyrus	4.51	−60	−4	29	
L postcentral gyrus	4.38	−54	−13	41	

STG: Superior temporal gyrus; IPL: Inferior parietal lobule; MTG: Middle temporal gyrus; MFG: Middle frontal gyrus; SFG: Superior frontal gyrus; IFG: Inferior frontal gyrus.

4.1. Impact of diminished scanner noise on language-evoked activity

Background noise produced by MRI scanners introduces additional confounds during listening tasks (Peelle, 2014; Wild et al., 2012). That is, the loud scanner noise hampers the faithful encoding of acoustic signals in bottom-up regions (e.g., inferior colliculus, Heschl's gyrus) and also forces participants to recruit top-down areas (e.g., the cingulo-opercular and fronto-parietal network) due to added demands on attention and executive functions. The hybrid and ISSS protocols eliminated the loud scanner noise during sound presentation, which had a noticeable impact on the data we collected; the silent protocols revealed several additional clusters that the continuous multiband paradigm failed to detect, particularly within the inferior colliculus, anterior thalamus, and right STG.

Table 4
: Foci of activation for data acquired using multiband acquisition and analyzed using multivariate methods. Labels are from the Brainnetome atlas.

		MNI Coordinates			
Region name	T Score	X	Y	Z	Volume of cluster (mm^3)
R STG	7.94	66	−28	8	1512
R STG	6.89	60	−19	2	
R STG	6.65	63	−19	11	
L IPL	4.72	−51	−37	23	459
L STG	4.45	−45	−40	20	
L STS	4.16	−54	−37	11	
R IPL	5.51	54	−52	23	351
R IPL	4.99	51	−49	20	

STG: Superior temporal gyrus; IPL: Inferior parietal lobule; STS: Superior temporal sulcus.

The inferior colliculus is an important hub for many auditory functions (Andics et al., 2014; Ortiz-Rios et al., 2015; Portfors et al., 2009; Šuta et al., 2003; Warrier et al., 2011), including audiovisual sensory integration (Champoux et al., 2006; Gruters and Groh, 2012; Thompson et al., 2006), selective auditory attention (Rinne et al., 2008), and interaural time difference processing (Wagner et al., 2007). Beyond general auditory perception, the inferior colliculi are critical for speech processing and language learning (Chandrasekaran et al., 2011). As predicted by the sparse coding model (Carlson et al., 2012), the inferior colliculus was tuned to core features of speech such as fundamental frequency (Peng et al., 2018; Ranasinghe et al., 2013) and periodicity (Y. Xu et al., 2017). Interestingly, speech evoked activity within the inferior colliculus is modulated by age (Parthasarathy et al., 2019) and acoustic trauma (Heeringa and van Dijk, 2019). We found that this region's activity, which was evoked by meaningful sentence stimuli, was attenuated by the background scanner noise and was accordingly undetected by the multiband protocol. The subcortical activity, however, was restored by using silent protocols (hybrid and ISSS). This result is consistent with a previous fMRI study demonstrating that the inferior colliculi were vulnerable to the degree of background noise present in speech stimuli (Davis et al., 2011).

The thalamus has been implicated in past studies of language processing. Several lesion studies have shown that damage to this region resulted in language impairment (Grönholm et al., 2016; Hebb and Ojemann, 2013; Nishio et al., 2014) including semantic aphasia (Crosson, 2019). Also, connectivity studies have reported that the left anterior nucleus of thalamus showed a robust connection to the left inferior frontal gyrus (Barbas et al., 2013; Bohsali et al., 2015). Despite ample clinical and neurophysiological evidence, the anterior thalamus has been implicated by few fMRI studies of language processing (Ketteler et al., 2008). With hybrid and ISSS, we were able to delineate sizable clusters in the anterior part of the thalamus that multiband failed to detect.

It has been debated whether or not language network is predominantly mediated by the left-lateralized (Friederici et al., 2017; C. J. Price, 2012; Rauschecker and Scott, 2009) or the bilateral persisylvian network (Evans and McGettigan, 2017; Hickok et al., 2011; Hickok and Poeppel, 2007; McGettigan and Scott, 2012; Okada et al., 2010; Scott and McGettigan, 2013). Such discrepancy could be, in part, due to differences in participants (young vs. old) and the nature of particular

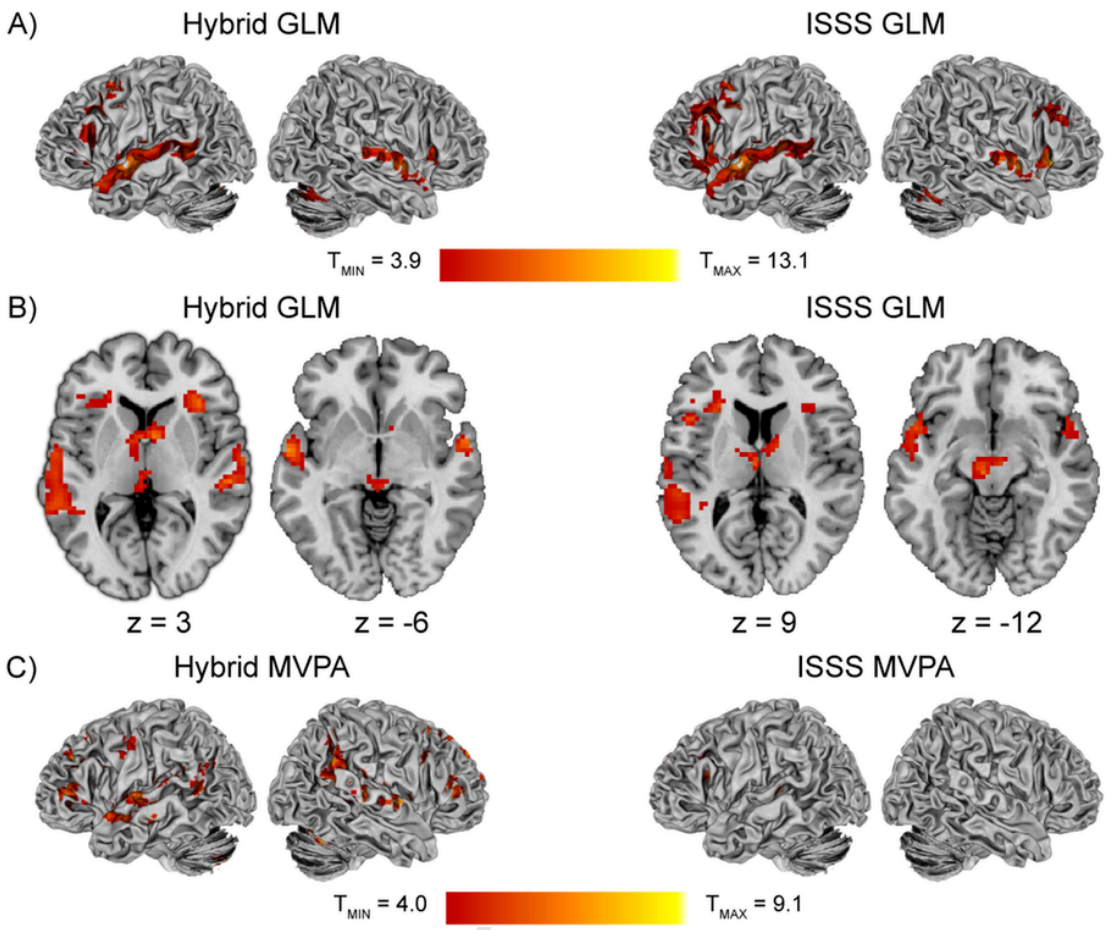

Fig. 3. **(A)** GLM results of hybrid and ISSS protocols. Both protocols detected significant clusters in the left inferior frontal and bilateral superior temporal regions; **(B)** Representative horizontal views demonstrating the different localization of similar subcortical activation in the anterior thalamus and midbrain; **(C)** MVPA results of hybrid and ISSS protocols. Hybrid detected more patterns of activity throughout the language network than ISSS.

language tasks. The present study employing a within-subject design on an identical task revealed a stark contrast in the degree of laterality between the silent and loud scanning protocols, an effect which has been greatly overlooked in the literature. Our data suggest that scanner noise can obscure language-related activity in the non-dominant right hemisphere. Together, the current data shed light on the consequence of background scanner noise by demonstrating differences between the silent hybrid and ISSS versus the loud multiband protocols.

4.2. Advantage of higher temporal resolution on auditory fMRI

Inherently, the characteristics of auditory stimuli including music, speech, and everyday sounds manifest in the temporal domain. Although the spatial resolution of fMRI has substantially increased over the last decades, fMRI still suffers from low temporal resolution (i.e., sampling rate), especially compared to electro- and magnetoencephalography. Numerous efforts have been devoted to overcoming this drawback. For example, multiband acceleration has been used to shorten acquisition time for resting state data (Smith et al., 2013; Smitha et al., 2018), to improve temporal signal-to-noise ratio (Todd et al., 2016), to create silent gaps (De Martino et al., 2015), and to maximize data acquisition (Vu et al., 2016). However, super-fast acceleration schemes suffer from signal autocorrelation, signal leakage, as well as vulnerability to physiological and motion-induced noise (Chen et al., 2019; Demetriou et al., 2018; Sahib et al., 2016; Todd et al., 2016). Attempts have been made to mitigate some of these drawbacks. For example, combining multiband acceleration with additional multi-echo acquisition and generalized auto-calibrating partial parallel acquisition (GRAPPA) reduced signal leakage (Boyacioğlu et al., 2015; Todd et al., 2016). In addition, improved autoregression techniques can be applied to highly accelerated data acquisition (Sahib et al., 2016).

In the present study, a moderate multiband acceleration factor was combined with ISSS imaging in order to enhance data acquisition following the silent stimulus presentation window. Enhanced data acquisition in hybrid was most evident in MVPA data. The hybrid protocol revealed a larger extent of activation throughout the brain than ISSS, which nearly failed to detect activity in the language network at a matched threshold. We stress that the amount data we acquired (i.e., 2 runs, each consisting of 16 trials) was not ideal for a sophisticated classification analysis with robust leave one run out cross-validation, which likely hampered the ISSS data. Nevertheless, by doubling the acquisition power within the same two runs of data collection, hybrid still yielded activity in the canonical language network without lowering statistical thresholds. Conversely, it is reasonable to conjecture that ISSS could have yielded comparable results if the acquisition time was doubled. However, increasing scan time leaves practical challenges to both participants and researchers due to fatigue and financial burden respectively. It is remarkable to observe that discriminable patterns emerged in the expected language network by increasing the temporal resolution from 2 s to 1 s repetition time during the limited scan time (2 runs lasting less than 4.5 min in total).

It is important to note that the performance of hybrid acquisition could have been improved further had it not been constrained by necessity in the present study. That is, with an increased acceleration fac-

Table 5
Foci of activation for data acquired using ISSS acquisition and analyzed using univariate methods. Labels are from the Brainnetome atlas.

		MNI Coordinates			
Region name	T Score	X	Y	Z	Volume of cluster (mm^3)
L STG	12.10	−57	−7	−4	31,401
L STG	8.25	−54	−31	2	
L precentral gyrus	8.07	−42	−1	35	
R orbital gyrus	10.85	33	23	−4	12,447
R STG	9.88	57	2	−7	
R MFG	6.97	45	26	26	
R cerebellum	6.46	36	−64	−28	4590
R cerebellum	6.02	39	−67	−31	
R cerebellum	5.80	27	−61	−28	
L SFG	8.50	0	14	53	4563
L SFG	5.44	−12	23	32	
R SFG	5.42	9	26	41	
L SPL	7.07	−27	−52	44	2916
L IPL	7.05	−39	−43	44	
L thalamus	7.87	−3	−13	11	1674
R thalamus	6.12	9	−7	11	
R thalamus	5.88	12	−4	8	
Midbrain	7.78	−6	−22	−13	1539
Midbrain	6.43	12	−19	−7	
Midbrain	6.06	6	−16	−13	
R SPL	5.22	33	−49	41	1188
R cerebellum	5.55	30	−67	−52	999
R cerebellum	4.80	24	−64	−49	

STG: Superior temporal gyrus; MFG: Middle frontal gyrus; SFG: Superior frontal gyrus; SPL: Superior parietal lobule; IPL; Inferior parietal lobule.

Table 6
: Foci of activation for data acquired using ISSS acquisition and analyzed using multivariate methods. Labels are from the Brainnetome atlas.

		MNI Coordinates			
Region name	T Score	X	Y	Z	Volume of cluster (mm^3)
L IFG	5.60	−51	17	20	432
L IFG	5.43	−48	23	20	
L IFG	5.36	−48	20	26	
L SFG	5.19	−12	53	29	243
L SFG	4.47	−15	50	35	
L posterior STG	5.43	−51	−34	8	216
L STG	4.93	−60	−31	5	
L STG	4.34	−57	−34	14	

IFG: Inferior frontal gyrus; SFG: Superior frontal gyrus; STG: Superior temporal gyrus.

tor (e.g., < 1 s TR), which would permit better slice coverage and lower in-plane resolution—both of which were deliberately matched to those of the ISSS protocol—hybrid can be further optimized. Bearing these constraints in mind, the true benefits of an optimized multiband-accelerated paradigm in auditory neuroimaging may be larger than demonstrated in the present study, which deserves further evaluation by future studies.

Nevertheless, increased temporal resolution in hybrid did not benefit the univariate results. As noted in the introduction, we expected comparable GLM results between hybrid and ISSS because increasing sampling would not affect the overall size of hemodynamic envelope, thereby yielding similar SPAs. Contrary to our hypothesis, however, ISSS yielded more voxels throughout the brain than hybrid acquisition. Similar to our finding, Demetriou et al. (2018) showed that highly accelerated data yielded reduced task-related activation across a wide battery of tasks (but see Boyacioğlu et al., 2017 for contradicting reports). Although this may seemingly suggest that ISSS acquisition has higher sensitivity over hybrid, this conclusion would only be valid if the activation had occurred in the established language network. Indeed, a close inspection of the data in reference to existing literature and meta-analysis archives suggested that many clusters in ISSS were located outside of the typical language network including the bilateral inferior parietal lobules and right premotor cortex. In contrast, hybrid yielded clusters, albeit smaller, more specifically in the language network, suggesting that hybrid may be less susceptible to false positives than ISSS. Further research is warranted to investigate the sensitivity and specificity of hybrid over existing silent protocols.

4.3. Other considerations and limitations

Our data is limited by a relatively small sample size and complex experimental design. We had to discard 26 % of the total data (5/19) due to an unexpected artifact in the new protocol and other issues (see Methods for more detail). This small sample size reduced the statistical power of MRI analysis, especially when considering data reported by ISSS as discussed above. It also impacted our ability to run direct comparisons between protocols (i.e. paired t-tests), as these comparisons were underpowered. Moreover, the group level MVPA suffered more than that of GLM due to further reduced sample size from 14 to 13 (see Methods for more detail), which may account for generally weaker results in MVPA than GLM across all comparisons. Our use of a two-fold cross validation could have led to a risk of overfitting. Although we found no evidence of overfitting in the classification outputs, it is desirable to increase data collection.

The small sample size may have decreased the statistical power in behavioral data as well. Previous studies using the same syntactic manipulation have led to significant differences in accuracy and reaction time, especially in children (Lee et al., 2020). However, the syntactic effect was only trending in the young adults; although this allowed us to legitimately combine the two syntactic conditions to be compared with 1 ch noise in fMRI data analyses, we were not able to evaluate performance of the protocol on delineating the more specific syntax network.

The sensorimotor activation observed in the present analysis—within the left premotor cortex and pre-supplementary motor area—may be due to an interaction between the cognitive and motor demands of the sentence comprehension task. Sensorimotor contributions to language comprehension have been observed in experiments that do not require motor responses (Buchsbaum and D'Esposito, 2019; Rong et al., 2018), and a review of previous work suggests that precentral gyrus contributions to language processing are impacted by both auditory noise and preparatory motor processing (Schomers and Pulvermüller, 2016). Activation was observed throughout all protocols, suggesting that precentral gyrus activity is less sensitive to acoustic noise, but the presence of a motor task prevents firm conclusions from being drawn with regards to language processing by the precentral gyrus and pre-supplementary motor area.

5. Conclusion

To alleviate the slow temporal resolution endemic to traditional sparse imaging and the detrimental effect of background noise that results from continuous imaging, we developed a new hybrid imaging method for auditory neuroimaging by adopting advantages of the quiet ISSS and the fast multiband-accelerated protocols. Our findings demonstrated that hybrid imaging is a viable option for auditory fMRI experiments with increased detecting power while minimizing false-positives. Further efforts are under way to improve the quality of the hybrid protocol for auditory neuroimaging.

Uncited reference

Worsley et al. (1992)

CRediT authorship contribution statement

Matthew Heard: Data curation, Formal analysis, Investigation, Project administration, Software, Validation, Visualization, Writing - original draft. **Xiangrui Li:** Methodology, Investigation, Resources. **Yune S. Lee:** Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

None of the authors have any conflicts of interest to disclose.

Acknowledgements

The authors would like to thank the Center for Cognitive and Behavioral Brain Imaging (CCBBI) at The Ohio State University for providing pilot scan time to develop these sequences, Essa Yacoub for technical support in implementing ISSS at the CCBBI, and Hyun-Woong Kim for providing feedback on the manuscript. We are especially grateful to the editor and 5 anonymous reviewers who provided valuable feedback on the study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jneumeth.2021.109198.

References

Adank, P., 2012. Design choices in imaging speech comprehension: an Activation Likelihood Estimation (ALE) meta-analysis. NeuroImage 63 (3), 1601–1613. https://doi.org/10.1016/j.neuroimage.2012.07.027.

Andics, A., Gácsi, M., Faragó, T., Kis, A., Miklósi, , 2014. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr. Biol. 24 (5), 574–578. https://doi.org/10.1016/j.cub.2014.01.058.

Barbas, H., García-Cabezas, , Zikopoulos, B., 2013. Frontal-thalamic circuits associated with language. Brain Lang. 126 (1), 49–61. https://doi.org/10.1016/j.bandl.2012.10.001.

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage 37 (1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Bohsali, A.A., Triplett, W., Sudhyadhom, A., Gullett, J.M., McGregor, K., FitzGerald, D.B., Mareci, T., White, K., Crosson, B., 2015. Broca's area – thalamic connectivity. Brain Lang. 141, 80–88. https://doi.org/10.1016/j.bandl.2014.12.001.

Boyacioğlu, R., Schulz, J., Koopmans, P.J., Barth, M., Norris, D.G., 2015. Improved sensitivity and specificity for resting state and task fMRI with multiband multi-echo EPI compared to multi-echo EPI at 7T. NeuroImage 119, 352–361. https://doi.org/10.1016/j.neuroimage.2015.06.089.

Buchsbaum, B.R., D'Esposito, M., 2019. A sensorimotor view of verbal working memory. Cortex 112, 134–148. https://doi.org/10.1016/j.cortex.2018.11.010.

Carlson, N.L., Ming, V.L., DeWeese, M.R., 2012. Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. PLoS Comput. Biol. 8 (7)https://doi.org/10.1371/journal.pcbi.1002594.

Champoux, F., Tremblay, C., Mercier, C., Lassonde, M., Lepore, F., Gagné, J.-P., Théoret, H., 2006. A role for the inferior colliculus in multisensory speech integration. NeuroReport 17 (15), 1607–1610. https://doi.org/10.1097/01.wnr.0000236856.93586.94.

Chandrasekaran, B., Kraus, N., Wong, P.C.M., 2011. Human inferior colliculus activity relates to individual differences in spoken language learning. J. Neurophysiol. 107 (5), 1325–1336. https://doi.org/10.1152/jn.00923.2011.

Chen, J.E., Polimeni, J.R., Bollmann, S., Glover, G.H., 2019. On the analysis of rapidly sampled fMRI data. NeuroImage 188, 807–820. https://doi.org/10.1016/j.neuroimage.2019.02.008.

Correia, J.M., Jansma, B.M.B., Bonte, M., 2015. Decoding articulatory features from fMRI responses in dorsal speech regions. J. Neurosci. 35 (45), 15015–15025. https://doi.org/10.1523/JNEUROSCI.0977-15.2015.

Crosson, B., 2019. The role of cortico-thalamo-Cortical circuits in language: recurrent circuits revisited. Neuropsychol. Rev. https://doi.org/10.1007/s11065-019-09421-8.

Cumming, G., Calin-Jageman, R., 2016. Introduction to the New Statistics: Estimation, Open Science, and Beyond. Routledge.

Davis, M.H., Ford, M.A., Kherif, F., Johnsrude, I.S., 2011. Does semantic context benefit speech understanding through "Top–Down" processes? Evidence from time-resolved sparse fMRI. J. Cogn. Neurosci. 23 (12), 3914–3932. https://doi.org/10.1162/jocn_a_00084.

De Martino, F., Moerel, M., Ugurbil, K., Formisano, E., Yacoub, E., 2015. Less noise, more activation: multiband acquisition schemes for auditory functional MRI. Magn. Reson. Med. 74 (2), 462–467. https://doi.org/10.1002/mrm.25408.

Demetriou, L., Kowalczyk, O.S., Tyson, G., Bello, T., Newbould, R.D., Wall, M.B., 2018. A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI. NeuroImage 176, 404–416. https://doi.org/10.1016/j.neuroimage.2018.05.011.

Eckert, M., Teubner-Rhodes, S., Vaden, K., 2016. Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. Ear Hear. 37, https://doi.org/10.1097/AUD.0000000000000300.

Ernst, R.R., Anderson, W.A., 1966. Application of fourier transform spectroscopy to magnetic resonance. Rev. Sci. Instrum. 37 (1), 93–102.

Evans, S., McGettigan, C., 2017. Comprehending auditory speech: previous and potential contributions of functional MRI. Lang. Cogn. Neurosci. 32 (7), 829–846. https://doi.org/10.1080/23273798.2016.1272703.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2016. The human brainnetome atlas: a new brain atlas based on connectional architecture. Cereb. Cortex 26 (8), 3508–3526. https://doi.org/10.1093/cercor/bhw157.

Friederici, A.D., Chomsky, N., Berwick, R.C., Moro, A., Bolhuis, J.J., 2017. Language, mind and brain. Nat. Hum. Behav. 1 (10), 713–722. https://doi.org/10.1038/s41562-017-0184-4.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C., 1994. Assessing the significance of focal activations using their spatial extent. Hum. Brain Mapp. 1 (3), 210–220. https://doi.org/10.1002/hbm.460010306.

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 (8), 721–732. https://doi.org/10.1002/hbm.20299.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.-B., Yarkoni, T., Margulies, D.S., 2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinform. 9, https://doi.org/10.3389/fninf.2015.00008.

Grönholm, E.O., Roll, M.C., Horne, M.A., Sundgren, P.C., Lindgren, A.G., 2016. Predominance of caudate nucleus lesions in acute ischaemic stroke patients with impairment in language and speech. Eur. J. Neurol. 23 (1), 148–153. https://doi.org/10.1111/ene.12822.

Gruters, K.G., Groh, J.M., 2012. Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. Front. Neural Circuits 6, https://doi.org/10.3389/fncir.2012.00096.

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 (3), 213–223.

Hebb, A.O., Ojemann, G.A., 2013. The thalamus and language revisited. Brain Lang. 126 (1), 99–108. https://doi.org/10.1016/j.bandl.2012.06.010.

Heeringa, A.N., van Dijk, P., 2019. Neural coding of the sound envelope is changed in the inferior colliculus immediately following acoustic trauma. Eur. J. Neurosci. 49 (10), 1220–1232. https://doi.org/10.1111/ejn.14299.

Hendriks, M.H.A., Daniels, N., Pegado, F., Op de Beeck, H.P., 2017. The effect of spatial smoothing on representational similarity in a simple motor paradigm. Front. Neurol. 8, https://doi.org/10.3389/fneur.2017.00222.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402. https://doi.org/10.1038/nrn2113.

Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69 (3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019.

Ketteler, D., Kastrau, F., Vohn, R., Huber, W., 2008. The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain; an fMRI study. NeuroImage 39 (4), 2002–2009. https://doi.org/10.1016/j.neuroimage.2007.10.023.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., Cornelissen, F., 2007. What's new in Psychtoolbox-3?. Perception 36 (14), 89.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U.S.A. 103 (10), 3863–3868. https://doi.org/10.1073/pnas.0600244103.

Langers, D.R.M., Dijk, P.V., Backes, W.H., 2005. Interactions between hemodynamic responses to scanner acoustic noise and auditory stimuli in functional magnetic resonance imaging. Magn. Reson. Med. 53 (1), 49–60. https://doi.org/10.1002/mrm.20315.

Larkman, D.J., Hajnal, J.V., Herlihy, A.H., Coutts, G.A., Young, I.R., Ehnholm, G., 2001. Use of multicoil arrays for separation of signal from multiple slices simultaneously excited. J. Magn. Reson. Imaging 13 (2), 313–317. https://doi.org/10.1002/1522-2586(200102)13:2<313::AID-JMRI1045>3.0.CO;2-W.

Lee, Y.S., Min, N.E., Wingfield, A., Grossman, M., Peelle, J.E., 2016. Acoustic richness modulates the neural networks supporting intelligible speech processing. Hear. Res. 333, 108–117. https://doi.org/10.1016/j.heares.2015.12.008.

Lee, Y.S., Wingfield, A., Min, N.E., Kotloff, E., Grossman, M., Peelle, J.E., 2018. Differences in hearing acuity among "Normal-Hearing" young adults modulate the neural

basis for speech comprehension. Eneuro 5 (3)https://doi.org/10.1523/ENEURO.0263-17.2018, ENEURO.0263-17.2018.
Lee, Y.S., Ahn, S., Holt, R.F., Schellenberg, E.G., 2020. Rhythm and syntax processing in school-age children. Dev. Psychol. 56 (9), 1632–1641. https://doi.org/10.1037/dev0000969.
Li, X., Liang, Z., Kleiner, M., Lu, Z.-L., 2010. RTbox: a device for highly accurate response time measurements. Behav. Res. Methods 42 (1), 212–225. https://doi.org/10.3758/BRM.42.1.212.
McGettigan, C., Scott, S.K., 2012. Cortical asymmetries in speech perception: What's wrong, what's right and what's left?. Trends Cogn. Sci. (Regul. Ed.) 16 (5), 269–276. https://doi.org/10.1016/j.tics.2012.04.006.
Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63 (5), 1144–1153. https://doi.org/10.1002/mrm.22361.
Nishio, Y., Hashimoto, M., Ishii, K., Ito, D., Mugikura, S., Takahashi, S., Mori, E., 2014. Multiple thalamo-cortical disconnections in anterior thalamic infarction: implications for thalamic mechanisms of memory and language. Neuropsychologia 53, 264–273. https://doi.org/10.1016/j.neuropsychologia.2013.11.025.
Nunes, R.G., Hajnal, J.V., Golay, X., Larkman, D.J., 2006. Simultaneous slice excitation and reconstruction for single shot EPI. Proc. Int. Soc. Magn. Reson. Med. 14, 1.
Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cerebral Cortex (New York, N.Y.: 1991) 20 (10), 2486–2495. https://doi.org/10.1093/cercor/bhp318.
Ortiz-Rios, M., Kuśmierek, P., DeWitt, I., Archakov, D., Azevedo, F.A.C., Sams, M., Jääskeläinen, I.P., Keliris, G.A., Rauschecker, J.P., 2015. Functional MRI of the vocalization-processing network in the macaque brain. Front. Neurosci. 9, https://doi.org/10.3389/fnins.2015.00113.
Parker, D., Razlighi, Q.R., 2019. The benefit of slice timing correction in common fMRI preprocessing pipelines. Front. Neurosci. 13, https://doi.org/10.3389/fnins.2019.00821.
Parker, D., Liu, X., Razlighi, Q.R., 2017. Optimal slice timing correction and its interaction with fMRI parameters and artifacts. Med. Image Anal. 35, 434–445. https://doi.org/10.1016/j.media.2016.08.006.
Parrish, T.B., Gitelman, D.R., LaBar, K.S., Mesulam, M.-M., 2000. Impact of signal-to-noise on functional MRI. Magn. Reson. Med. 44 (6), 925–932. https://doi.org/10.1002/1522-2594(200012)44:6<925::AID-MRM14>3.0.CO;2-M.
Parthasarathy, A., Herrmann, B., Bartlett, E.L., 2019. Aging alters envelope representations of speech-like sounds in the inferior colliculus. Neurobiol. Aging 73, 30–40. https://doi.org/10.1016/j.neurobiolaging.2018.08.023.
Peelle, J.E., 2014. Methodological challenges and solutions in auditory functional magnetic resonance imaging. Front. Neurosci. 8, 253. https://doi.org/10.3389/fnins.2014.00253.
Peelle, J.E., 2018. Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. Ear Hear. 39 (2), 204–214. https://doi.org/10.1097/AUD.0000000000000494.
Peng, F., Innes-Brown, H., McKay, C.M., Fallon, J.B., Zhou, Y., Wang, X., Hu, N., Hou, W., 2018. Temporal coding of voice pitch contours in mandarin tones. Front. Neural Circuits 12, 55. https://doi.org/10.3389/fncir.2018.00055.
Petkov, C., Kayser, C., Augath, M., Logothetis, N., 2009. Optimizing the imaging of the monkey auditory cortex: sparse vs. continuous fMRI. Magn. Reson. Imaging 27 (8), 1065–1073. https://doi.org/10.1016/j.mri.2009.01.018.
Portfors, C.V., Roberts, P.D., Jonson, K., 2009. Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. Neuroscience 162 (2), 486–500. https://doi.org/10.1016/j.neuroscience.2009.04.056.
Power, J.D., Plitt, M., Kundu, P., Bandettini, P.A., Martin, A., 2017. Temporal interpolation alters motion in fMRI scans: magnitudes and consequences for artifact detection. PLoS One 12 (9), e0182939https://doi.org/10.1371/journal.pone.0182939.
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 (2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062.
Price, D.L., Wilde, J.P.D., 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 (2), 288–293. https://doi.org/10.1002/1522-2586(200102)13:2<288::AID-JMRI1041>3.0.CO;2-P.
Raizada, R.D.S., Lee, Y.-S., 2013. Smoothness without smoothing: why gaussian naive bayes is not naive for multi-subject searchlight studies. PLoS One 8 (7), e69566https://doi.org/10.1371/journal.pone.0069566.
Ranasinghe, K.G., Vrana, W.A., Matney, C.J., Kilgard, M.P., 2013. Increasing diversity of neural responses to speech sounds across the central auditory pathway. Neuroscience 252, 80–97. https://doi.org/10.1016/j.neuroscience.2013.08.005.
Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12 (6), 718–724. https://doi.org/10.1038/nn.2331.
Rinne, T., Balk, M.H., Koistinen, S., Autti, T., Alho, K., Sams, M., 2008. Auditory selective attention modulates activation of human inferior colliculus. J. Neurophysiol. 100 (6), 3323–3327. https://doi.org/10.1152/jn.90607.2008.
Rong, F., Isenberg, A.L., Sun, E., Hickok, G., 2018. The neuroanatomy of speech sequencing at the syllable level. PLoS One 13 (10), e0196381https://doi.org/10.1371/journal.pone.0196381.
Sahib, A.K., Mathiak, K., Erb, M., Elshahabi, A., Klamer, S., Scheffler, K., Focke, N.K., Ethofer, T., 2016. Effect of temporal resolution and serial autocorrelations in event-related functional MRI. Magn. Reson. Med. 76 (6), 1805–1813. https://doi.org/10.1002/mrm.26073.
Schomers, M.R., Pulvermüller, F., 2016. Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. Front. Hum. Neurosci. 10, https://doi.org/10.3389/fnhum.2016.00435.
Schwarzbauer, C., Davis, M.H., Rodd, J.M., Johnsrude, I., 2006. Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. NeuroImage 29 (3), 774–782. https://doi.org/10.1016/j.neuroimage.2005.08.025.
Scott, S.K., McGettigan, C., 2013. Do temporal processes underlie left hemisphere dominance in speech perception?. Brain Lang. 127 (1), 36–45. https://doi.org/10.1016/j.bandl.2013.07.006.
Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Glasser, M.F., 2013. Resting-state fMRI in the human connectome project. NeuroImage 80, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039.
Smitha, K.A., Arun, K.M., Rajesh, P.G., Joel, S.E., Venkatesan, R., Thomas, B., Kesavadas, C., 2018. Multiband fMRI as a plausible, time-saving technique for resting-state data acquisition: study on functional connectivity mapping using graph theoretical measures. Magn. Reson. Imaging 53, 1–6. https://doi.org/10.1016/j.mri.2018.06.013.
Šuta, D., Kvašňák, E., Popelář, J., Syka, J., 2003. Representation of species-specific vocalizations in the inferior colliculus of the Guinea Pig. J. Neurophysiol. 90 (6), 3794–3808. https://doi.org/10.1152/jn.01175.2002.
Talavage, T.M., Edmister, W.B., 2004. Nonlinearity of FMRI responses in human auditory cortex. Hum. Brain Mapp. 22 (3), 216–228. https://doi.org/10.1002/hbm.20029.
Thompson, S.K., Kriegstein, Kvon, Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T.D., McAlpine, D., 2006. Representation of interaural time delay in the human auditory midbrain. Nat. Neurosci. 9 (9), 1096–1099. https://doi.org/10.1038/nn1755.
Todd, N., Moeller, S., Auerbach, E.J., Yacoub, E., Flandin, G., Weiskopf, N., 2016. Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: sensitivity and slice leakage artifacts. NeuroImage 124, 32–42. https://doi.org/10.1016/j.neuroimage.2015.08.056.
Vu, A.T., Phillips, J.S., Kay, K., Phillips, M.E., Johnson, M.R., Shinkareva, S.V., Tubridy, S., Millin, R., Grossman, M., Gureckis, T., Bhattacharyya, R., Yacoub, E., 2016. Using precise word timing information improves decoding accuracy in a multiband-accelerated multimodal reading experiment. Cogn. Neuropsychol. 33 (3–4), 265–275. https://doi.org/10.1080/02643294.2016.1195343.
Wagner, H., Asadollahi, A., Bremen, P., Endler, F., Vonderschen, K., Campenhausen, Mvon., 2007. Distribution of interaural time difference in the barn owl's inferior colliculus in the low- and high-frequency ranges. J. Neurosci. 27 (15), 4191–4200. https://doi.org/10.1523/JNEUROSCI.5250-06.2007.
Warrier, C.M., Abrams, D.A., Nicol, T.G., Kraus, N., 2011. Inferior colliculus contributions to phase encoding of stop consonants in an animal model. Hear. Res. 282 (1), 108–118. https://doi.org/10.1016/j.heares.2011.09.001.
Whitfield-Gabrieli, S., Nieto-Castanon, A., 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect. 2 (3), 125–141. https://doi.org/10.1089/brain.2012.0073.
Wild, C.J., Yusuf, A., Wilson, D.E., Peelle, J.E., Davis, M.H., Johnsrude, I.S., 2012. Effortful listening: the processing of degraded speech depends critically on attention. J. Neurosci. 32 (40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012.
, .
Xu, J., Moeller, S., Auerbach, E.J., Strupp, J., Smith, S.M., Feinberg, D.A., Yacoub, E., Uğurbil, K., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3T. NeuroImage 83, 991–1001. https://doi.org/10.1016/j.neuroimage.2013.07.055.
Xu, Y., Chen, M., LaFaire, P., Tan, X., Richter, C.-P., 2017. Distorting temporal fine structure by phase shifting and its effects on speech intelligibility and neural phase locking. Sci. Rep. 7, https://doi.org/10.1038/s41598-017-12975-3.
Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8 (8), 665–670. https://doi.org/10.1038/nmeth.1635.