Check for updates

METHOD ARTICLE

Comparison of auditory fMRI protocols for a voice localizer

[version 1; peer review: 1 approved, 1 approved with reservations]

Régis Trapeau[1], Julien Sein[1], Manon Obliger-Debouche[1], Pascal Belin[1,2]

[1]Aix-Marseille Université, UMR 7289 CNRS, Institut de Neurosciences de la Timone, Marseille, Provence-Alpes-Côte d'Azur, 13005, France
[2]Aix-Marseille Université, Institute of Language Communication and the Brain, Aix-en-Provence, Provence-Alpes-Côte d'Azur, 13100, France

v1 **First published:** 11 Dec 2024, **4**:261
https://doi.org/10.12688/openreseurope.18899.1
Latest published: 11 Dec 2024, **4**:261
https://doi.org/10.12688/openreseurope.18899.1

Abstract

Auditory functional magnetic resonance imaging (fMRI) presents unique challenges due to scanner noise interference, which can limit the detection of stimulus-related brain activity. This study systematically evaluates five different fMRI protocols—continuous, sparse, fast sparse, clustered sparse, and interleaved silent steady state (ISSS)—to determine their effectiveness in capturing auditory and voice-related brain activity under identical scanning conditions. Participants passively listened to vocal and non-vocal sounds during fMRI protocols of the same duration, and the ability of each protocol to detect auditory and voice-specific activation was evaluated. Results showed that continuous imaging produced the largest and highest auditory activation, followed closely by clustered sparse sampling. Both sparse and fast sparse sampling yielded intermediate results, with fast sparse sampling performing better at detecting voice-specific activation. ISSS had the lowest activation sensitivity. The results highlight that continuous imaging is optimal when participants are well protected from scanner noise, while clustered sparse sequences offer the best alternative when stimuli are to be presented in silence.

Plain language summary

This study compared different methods for brain imaging to find the best way to study how the brain processes sounds, particularly voices, using functional magnetic resonance imaging (fMRI). Five different techniques that are commonly used in auditory fMRI were tested, focusing on how well they could capture brain activity during the same amount of time and without being affected by the loud noise the MRI scanner makes. Results show that, for humans well protected from the scanner noise, continuously acquiring images while sounds were presented was optimal in terms of showing where in the brain the voice-related activity was happening. However, a method called "clustered sparse sampling," which allows for short silent periods

Open Peer Review

Approval Status ? ✓

	1	2
version 1 11 Dec 2024	? view	✓ view

1. **Xi-Nian Zuo**, Beijing Normal University, Beijing, China
National Basic Science Data Center, China, China

2. **Federico De Martino**, Maastricht University, Maastricht, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

during sound presentation, also performed very well and was nearly as effective. This method is particularly useful when it is important to present sounds in a silent environment to avoid interference from the scanner's noise. The study highlights that while continuous scanning is effective when subjects have good noise protection, clustered sparse sampling can be a better choice when presenting sounds in silence is necessary, like in certain research conditions.

Keywords
fMRI, auditory, cortical, voice, neuroimaging

This article is included in the European Research Council (ERC) gateway.

This article is included in the Horizon 2020 gateway.

Corresponding authors: Régis Trapeau (regis.trapeau@univ-amu.fr), Pascal Belin (pascal.belin@univ-amu.fr)

Author roles: Trapeau R: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Sein J**: Investigation, Methodology, Writing – Review & Editing; **Obliger-Debouche M**: Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Belin P**: Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (COVOPRIM) (grant agreement no. 788240).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Trapeau R *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Trapeau R, Sein J, Obliger-Debouche M and Belin P. **Comparison of auditory fMRI protocols for a voice localizer [version 1; peer review: 1 approved, 1 approved with reservations]** Open Research Europe 2024, **4**:261 https://doi.org/10.12688/openreseurope.18899.1

First published: 11 Dec 2024, **4**:261 https://doi.org/10.12688/openreseurope.18899.1

Introduction

Functional magnetic resonance imaging (fMRI) has established itself as a key tool in the field of auditory neuroscience. However, designing auditory fMRI experiments poses unique challenges due to the substantial noise generated by the rapid switching of gradient coils during echo planar imaging (EPI), which can reach up to 120 dB SPL. In a typical continuous EPI sequence, scans are continuously collected while stimuli are presented alongside the noise generated by the scanner, which offers the advantage of maximizing data collection and preserving fMRI temporal resolution. However, the scanner noise interferes with the stimulus and partially saturates the response of the auditory cortex, limiting the dynamic range available for detecting stimulus-related brain activity (Edmister *et al.*, 1999). Moreover, in the context of speech-related tasks, the decreased acoustic clarity caused by the scanner noise increases cognitive load, significantly impacting fMRI sensitivity and data interpretation (Peelle, 2014).

The typical approach to protecting human participants from scanner noise is to use passive sound attenuation, e.g. using over-ear hearing protection in conjunction with foam earplugs, insert earphones, and/ or active noise control. This results in a maximum total noise reduction of approximately 40 dB (Hall *et al.*, 2009; Ravicz & Melcher, 2001). Despite such protective measures, a substantial amount of noise is still transmitted to the cochlea, either through the tympani or by bone conduction. Consequently, the stimulus is perceived differently when presented along with the scanner noise or in quiet conditions. This may be acceptable depending on the stimuli and the brain regions being examined. However, presenting stimuli in silence can be crucial when frequency ranges of the stimuli and EPI noise significantly overlap (Langers *et al.*, 2014); when small alterations in the stimuli spectra result in large perceptual changes (Trapeau & Schönwiesner, 2018); when comparisons across experimental conditions require fine-grained acoustic discrimination; or, more critically, when the experimental setup does not allow for optimal acoustic protection of the subjects, as is often the case in animal research.

The classical method to presenting auditory stimuli that are not masked by the scanner noise, sparse sampling, involves separating volume acquisitions by a long gap (10 to 16 s) that allows for the presentation of sound stimuli in silence, and exploits the fact that the peak of the hemodynamic response occurs several seconds after the stimulus (Belin *et al.*, 1999; Hall *et al.*, 1999). Furthermore, the long silent intervals avoid contamination by the hemodynamic response to the scanner noise and allow for complete recovery of the magnetization from one volume acquisition to the next. Sparse sampling has been extensively used when studying primary auditory cortex (A1) and subcortical areas (Fuglsang *et al.*, 2022; Oya *et al.*, 2018; Trapeau & Schönwiesner, 2015; Wolak *et al.*, 2017), as well as higher-level areas in both humans and animal studies (Bach *et al.*, 2013; Belin & Zatorre, 2003; Joly *et al.*, 2012; Ortiz-Rios *et al.*, 2015; Petkov *et al.*, 2008; Schönwiesner *et al.*, 2015). Comparisons of sparse and continuous sequences comprising the same number of volumes emphasized the detrimental effect of scanner background noise. Identical sound stimulation sequences evoked higher signal in primary and higher-order auditory regions (Gaab *et al.*, 2007), as well as in other intrinsic networks (Yakunina *et al.*, 2015), when data were acquired with sparse sampling compared to a subset of the scan data acquired during continuous scanning that exactly matched the data sampling during sparse acquisition. Sparse sampling also elicits greater response in rhesus macaque auditory areas compared to a continuous sequence with the same number of volumes but a shorter duration (Petkov *et al.*, 2009). In contrast, duration-matched comparisons of both protocols yielded divergent results. Continuous paradigms exhibited a significantly larger extent of activation when compared to duration-matched versions of the sparse paradigm (Dewey *et al.*, 2021; Hall *et al.*, 2014; Peelle *et al.*, 2010). These results can be attributed to the primary limitation of sparse sampling, namely the relatively low number of images acquired per time unit. Furthermore, sparse sampling only provides a single measure per stimulus, which precludes the assessment of the profile of the hemodynamic response without scanning each stimulus multiple times with jittered intervals.

Clustered sparse sampling is a hybrid method that combines the advantages of sparse sampling with more efficient data acquisition. In this protocol, multiple fMRI volumes are acquired in rapid succession (a “cluster”) after a silent period, allowing for greater temporal coverage and statistical power than classical sparse sampling while still preserving relatively long moments of silence for stimulus presentation. Duration-matched comparisons between sparse and clustered sparse protocols, using clusters of three volumes, showed significantly enhanced functional responses along the superior temporal plane (Schmidt *et al.*, 2008; Zaehle *et al.*, 2007) for the clustered sparse protocol. However, a drawback of this method is that longitudinal magnetization does not reach a steady state within the acquisition of a single cluster, resulting in a T1 signal decay over the course of the cluster acquisition. To address the issue of T1-decay, various corrective approaches have been employed. These include the incorporation of regressors to model the decay (Schmidt *et al.*, 2008; Zaehle *et al.*, 2007), the averaging of the volumes within each cluster (Riecke *et al.*, 2017; Riecke *et al.*, 2018), and the analysis of each volume within a cluster as a distinct session (Keil *et al.*, 2020; Stewart *et al.*, 2022; Zaehle *et al.*, 2004)

To entirely prevent T1-decay during clustered sparse data acquisition, longitudinal magnetization can be maintained in a steady state during pseudo-quiet periods, by maintaining silent slice-selective excitation pulses, while disabling the noisier readout and phase encoding gradients (Schwarzbauer *et al.*, 2006). While reducing acoustic noise of approximately 15 dB during the quiet periods (Ress & Chandrasekaran, 2013), this imaging method, termed Interleaved Silent Steady State (ISSS), still generates significant scanner noise during volume acquisition. In comparison to sparse sampling, ISSS demonstrates a reduced spatial extent of activation (Schwarzbauer *et al.*, 2006), yet exhibits higher sensitivity in detecting subtle activity in subcortical brain regions (Mueller *et al.*, 2011). Furthermore,

it provides time course information that is missing in sparse imaging data.

An alternative method for overcoming the limitations of sparse sampling is to reduce the repetition time (TR) of the sequence (Perrachione & Ghosh, 2013). This approach, often referred to as fast sparse sampling, allows for more frequent data acquisition and enhanced temporal resolution in comparison to sparse sampling. Although the hemodynamic response to the sound is contaminated by the response to the frequent scanner noise, the fast sparse sampling technique has the advantage that sound perception is unaltered by the scanner noise as sounds are presented during the short silent periods between acquisitions. Fast sparse sampling has been employed in both human and non-human primate studies using a TR between 2s and 3.5 s, an acquisition time (TA) between 0.25 and 2 s, leaving time for a silent period between 1 and 2.5 s (Erb *et al.*, 2019; Jafari *et al.*, 2023; Norman-Haignere *et al.*, 2015; Norman-Haignere *et al.*, 2019; Rogers & Davis, 2017; Sadagopan *et al.*, 2015).

The objective of the present study is to evaluate and compare the efficacy of these five auditory fMRI protocols—continuous, sparse, clustered sparse, ISSS, and fast sparse—under identical scanning conditions (i.e., same scanner, same auditory stimulation setup, same stimulus sequence, same run duration, same participants). The aim is to determine which protocol provides the optimal level and extent of activation for a voice localizer within a specified scanning time. Furthermore, five variations of the clustered sparse protocol were tested, each with a different sound stimulation and data acquisition duration.

Methods

Participants

Three native French human speakers were scanned (two females in June 2021 (including author M.O.D.) and one male in February 2024 (author P.B.); 26–57 years old). Participants gave written informed consent and were not paid for their participation. The experiment was granted ethical approval by Aix-Marseille Université "Comité Protection des Personnes"—"Committee for the Protection of Individuals": CPP 2017-A03614–49, February 13, 2019. The experimental procedures conformed to the World Medical Association's Declaration of Helsinki.

Procedure

Participants were scanned for nine functional runs of the same duration of about 7 min 30, each corresponding to a specific imaging protocol. One participant was scanned using only four protocols: continuous, sparse, clustered sparse, and ISSS.

During each run, participants passively listened to sounds from a voice localizer (Belin *et al.*, 2000; Capilla *et al.*, 2013). Stimuli were presented in a block design. Each block consisted of a certain number of stimuli (depending on the functional imaging sequence, see below) belonging to the same category. Silent blocks were interleaved with sound blocks and occurred at the same frequency as the other two main categories.

Stimuli & audio apparatus

Stimuli were 500-ms long and comprised two main categories: human vocalizations (n = 252), consisting of either speech or non-speech sounds (e.g. yawn); and other sounds (n = 252), consisting of sounds from nature (e.g. wind, waterfall), animals (e.g. vocalizations, gallops), the industrialized human environment (e.g. cars, telephones, airplanes), and musical instruments (e.g. bells, harp, instrumental orchestra).

Stimuli were digitally processed using a TDT System 3 RM1 (Tucker Davis Technologies, Alachua, FL, USA) and delivered through a pair of S15 insert earphones (Sensimetrics, Maiden, MA, USA) at an approximate sound pressure level of 85 dB. Over-ear foam pieces were placed along each participant's head to minimize head movement and provide additional hearing protection beyond that provided by the insert earphones.

MRI data acquisition

Images were acquired at the MRI Center of the Institut de Neurosciences de la Timone (Marseille, France) using a 3-Tesla PRISMA scanner (Siemens, Erlangen, Germany) with a Head-Neck 64-channels receive coil. High-resolution T1 & T2-weighted anatomical volumes were first acquired for coregistration and normalization purposes (T1w: MPRAGE sequence, TR = 2.3 s, TE = 2.9 ms, TI = 900 ms flip angle: 9°, matrix size = 192 × 240 × 256, resolution 1 × 1 x 1 mm^3; T2w: SPACE sequence, TR = 3.2 s, TE = 563 ms,, matrix size = 192 × 240 × 256, resolution 1 × 1 x 1 mm^3). Functional volumes were acquired with the same coil using EPI sequences using these common parameters for all sequences: multiband acceleration factor = 4, TA = 0.945 s, TE = 30 ms, matrix size = 84 × 84 × 56, resolution 2.5 x 2.5 x 2.5 mm^3. Functional acquisition's TR and flip angle varied depending on the imaging sequence, see below. All sequences were obtained from the Center for Magnetic Resonance Research (CMRR, University of Minnesota) multiband EPI sequences, version R016a.

Functional imaging sequences

Imaging sequences and their associated hemodynamic response time courses are depicted in Figure 1. A summary of each sequence specifications, in terms of number of volumes acquired, number of stimuli presented, number of blocks of stimuli presented and sequence durations, is represented in Figure 2.

Continuous. During the continuous sequence, scans were continuously collected (TR = TA, flip angle = 65°) while blocks of 15 stimuli (block duration = 8.2 s, inter-stimulus interval (ISI) = 50 ms) were presented alongside volume acquisitions. The time interval between two consecutive

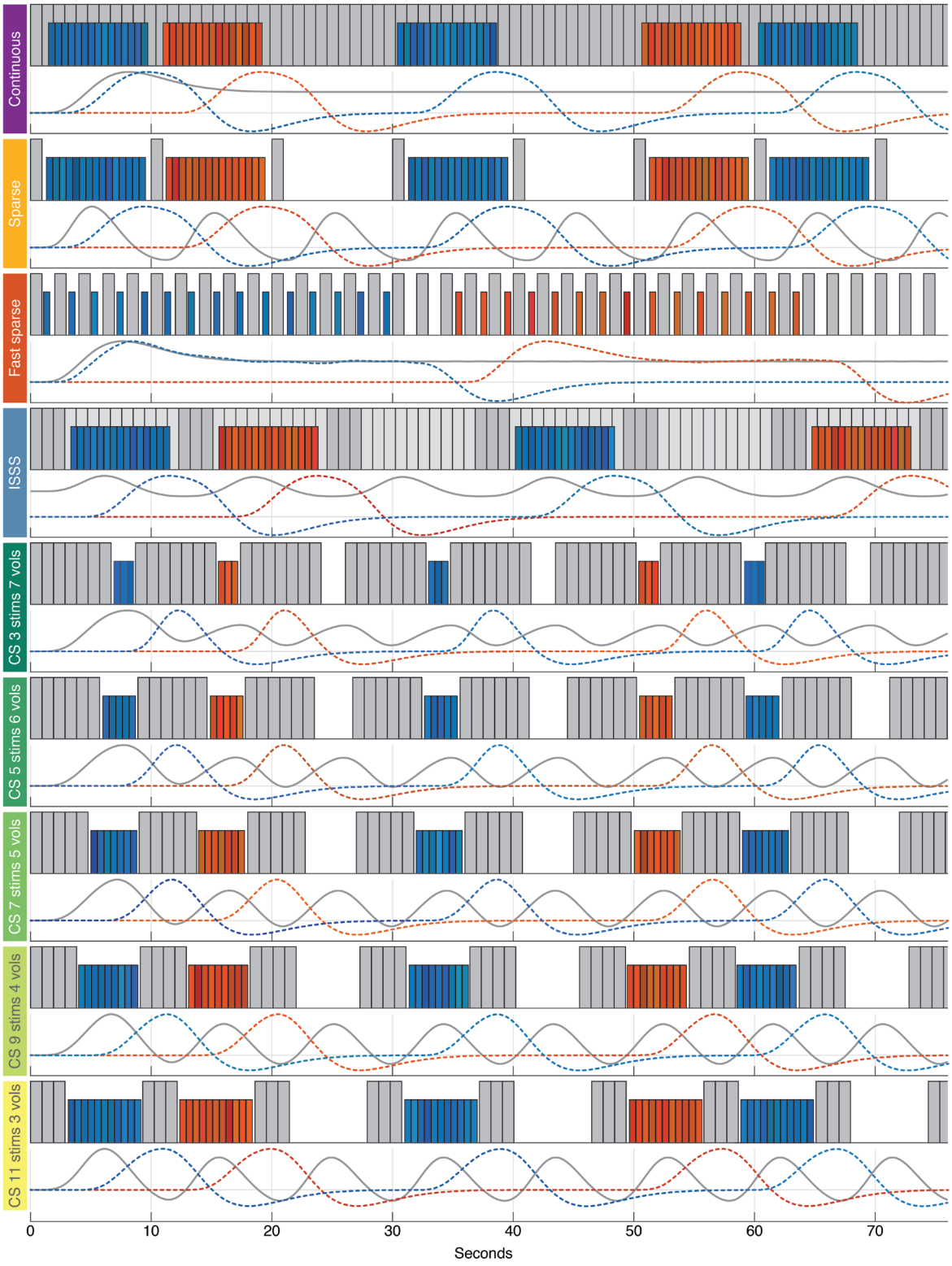

Figure 1. Schematic representation of all sequences and their associated hemodynamic response time courses. [Each gray bar represents an acquisition volume (light gray bars are ISSS silent slice-selective excitation pulses). Each colored bar represents a sound stimulus (blue for one category, red for the other). Evoked hemodynamic responses to scanner noise and to sound stimuli are depicted at the bottom of each sequence, using the same color codes. CS, Clustered Sparse. For each clustered sparse sequence, the specific number of stimuli within a block and the number of acquisition volumes per cluster are specified.].

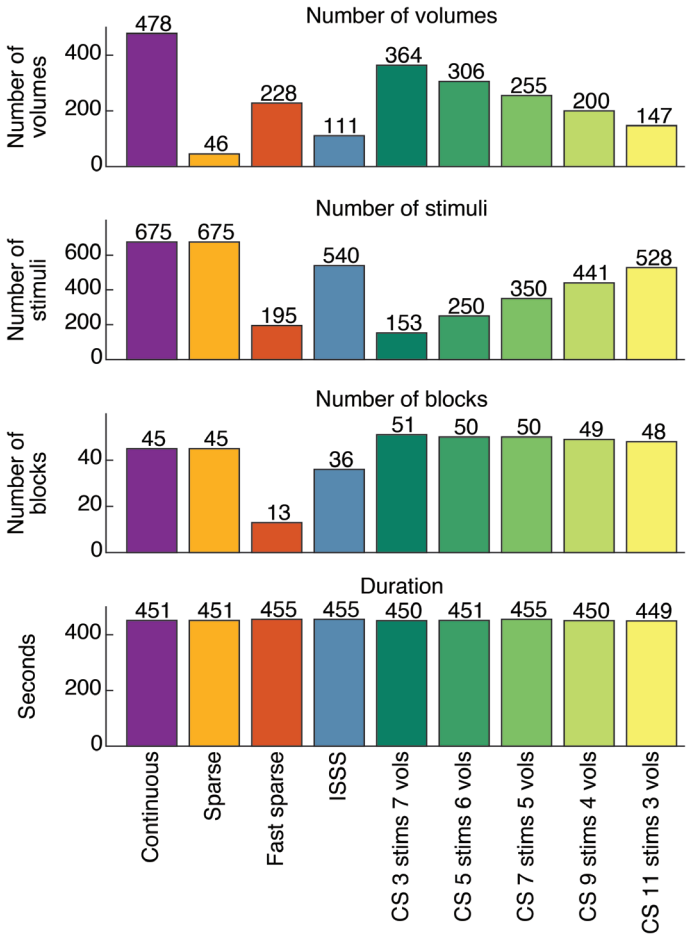

Figure 2. Sequence specifications. Comparison of the number of volumes acquired, the number of stimuli presented, the number of blocks of stimuli presented and the total duration of each sequence (CS, Clustered Sparse).

blocks was randomly jittered from 1 s to 2.5 s. This sequence allowed the acquisition of 478 volumes in 451 s and the presentation of 675 stimuli in 45 blocks.

Sparse. During sparse sampling, blocks of 15 stimuli (block duration = 8.2 s, ISI = 50 ms) were presented in between each volume acquisition (TR = 10 s, flip angle = 90°, silent period = 9.055 s). The gap between the end of a volume acquisition and the onset of sound stimulation was randomly jittered from 200 ms to 500 ms. This sequence allowed the acquisition of 46 volumes in 451 s and the presentation of 675 stimuli in 45 blocks.

Fast sparse. The fast sparse sampling sequence had a TR of 2 s (flip angle = 90°), leaving 1.055 s of silence between two acquisitions, during which a stimulus could be presented. The time interval between the end of a volume acquisition and the onset of a stimulus was randomly jittered from 100 ms to 400 ms. Thus, a block of 15 stimuli was approximately 15 TRs long (~30 s). Blocks of stimuli were randomly separated by either 2 or 3 TRs. This sequence allowed the acquisition of 228 volumes in 455 s and the presentation of 195 stimuli in 13 blocks.

ISSS. During the ISSS sequence, silent slice-selective excitation pulses were applied continuously for 10 TAs (quiet period = 9.45 s), followed by the acquisition of a cluster of 3 volumes (TR = TA, flip angle = 65°). The gap between the end of an acquisition cluster and the beginning of a block of stimuli (block duration = 8.2 s, ISI = 50 ms) was randomly jittered from 200 ms to 500 ms. This sequence allowed the acquisition of 111 volumes in 455 s and the presentation of 540 stimuli in 36 blocks.

Clustered sparse. Five variations of the clustered sparse sequence, with different block sizes and cluster sizes, were tested. For all variations, the following parameters were kept constant: ISI = 50 ms; the time interval between the end of an acquisition cluster and the start of a block of stimuli was randomly jittered from 100 ms to 300 ms; the TR within a cluster was

955 ms (10 ms more than the TA, to allow the scanner to listen to the volume acquisition trigger sent by the TDT RM1; flip angle = 60°).

Variations in the number of stimuli within a block were 3, 5, 7, 9 or 11, followed by an acquisition cluster of 7, 6, 5, 4 or 3 volumes, respectively. These sequences allowed the acquisition of 364, 306, 255, 200 or 147 volumes in 450, 451, 455, 450 or 449 s and the presentation of 153, 250, 350, 441 or 528 stimuli in 51, 50, 50, 49 or 48 blocks, respectively.

fMRI data analysis

Tissue segmentation and brain extraction were performed on the structural images using SPM's default segmentation procedure. Additionally, white matter and cerebrospinal fluid masks were eroded using a 3 mm cubic kernel in order to provide a mask that contains no region of interest to a principal component analysis of the functional data (see next paragraph). Transformation matrices between functional, anatomical and template volumes were computed using non-linear registration (ANTS, SyN). These transformation matrices were used to register the tissues segmentation to the functional data as well as to register the functional results to individual high-resolution anatomical scans and to the MNI152 ICBM 2009c Nonlinear Asymmetric template (Fonov *et al.*, 2011). Functional data pre-processing included motion correction and spatial smoothing with a full-width half-maximum 3-dimensional Gaussian kernel that was twice the size of the functional voxels (i.e., 5 mm).

General linear model (GLM) estimates of responses to all sounds versus silence ("Sound vs. Silence" contrast) and to human voices against other sounds ("Human vs. Others" contrast) were computed with fMRISTAT (Worsley *et al.*, 2002) using a hemodynamic response function modeled as the difference of two gamma density functions (Glover, 1999) with a peak at 5 s and an undershoot at 10 s. For all the sequences, the GLM included motion parameters (translations and rotations) as covariates of no interest. For the clustered sparse sequences, the first two principal components of a principal component analysis performed on the eroded mask of the functional voxels identified as containing white matter and cerebrospinal fluid were added as covariates of no interest to the GLM. Voxel significance was assessed by thresholding t-maps at $p < 0.05$, corrected for multiple comparisons using Gaussian Random Field Theory (Worsley *et al.*, 1996). Spatial extent of significant voxels was computed for each sequence and contrast.

Maximum t-value is reported for three sets of ROIs: primary auditory cortex (A1), temporal voice areas (TVAs) and frontal voice areas (FVAs). In each subject and hemisphere, the A1 ROI contained the voxels of the probabilistic map (non-linearly registered to each subject functional space) of Heschl's gyri provided with the MNI152 template (Penhune *et al.*, 1996) that were above a probability of 0.3. TVAs and FVAs were sphere of 5 mm radius around the average ROI coordinates provided by (Aglieri *et al.*, 2018). ROI locations are represented on Figure 3.

Fixed-effects group t-maps of both contrasts were computed using the template registered data from the two participants who were tested with all the protocols.

Results

Group results

Group results showed sound-related activation in response to all sounds of the voice localizer ("Sound vs. Silence" contrast) in all imaging protocols (Figure 4A). Auditory activation was found bilaterally in Heschl's gyrus, planum temporale and wider regions along the superior temporal gyrus and sulcus (STG and STS). Substantial activation was also found in the prefrontal cortex around the FVAs.

Activations in response to human voices against other sounds ("Human vs. Others" contrast) were observed at typical TVA locations, along the STS and STG, for all imaging protocols except ISSS, and were also observed in the FVAs, for the continuous and clustered sparse sequences presenting five or more stimuli per trial (Figure 4B).

For both contrasts, the loci of maximal activation were very consistent from one imaging sequence to the other.

"Sound vs. Silence" individual results

Individual results for spatial extent and ROI maximum t-value are shown in Figure 5. On average, the continuous sequence showed larger spatial extent and greater maximum t-value than the other imaging sequences for the "Sound vs. Silence" contrast (Figure 5A). Continuous imaging performed better than sparse, fast sparse and ISSS for all three participants in both spatial extent and maximum t-value.

The clustered sparse sequence, set to present seven stimuli per trial and acquire five volumes per cluster, showed the highest and largest auditory activation among the protocols that presented stimuli in silence (sparse, fast sparse, clustered sparse) or during reduced scanner noise (ISSS). Its average results approach those obtained with the continuous sequence (Figure 5A), and it shared best results with continuous scanning for the two participants who were tested with all the protocols.

Using ISSS, auditory activation barely reached significance in all three participants and maximum t-values were about one-sixth of those obtained with the continuous sequence (Figure 5A). ISSS always performed the worst, except for the FVAs maximum t-value in one participant.

Average spatial extents and maximum t-values obtained with the sparse and fast sparse sequences are about half of those obtained with the continuous sequence (Figure 5A). Results for two sequences were consistently in between those of continuous and ISSS for all three participants.

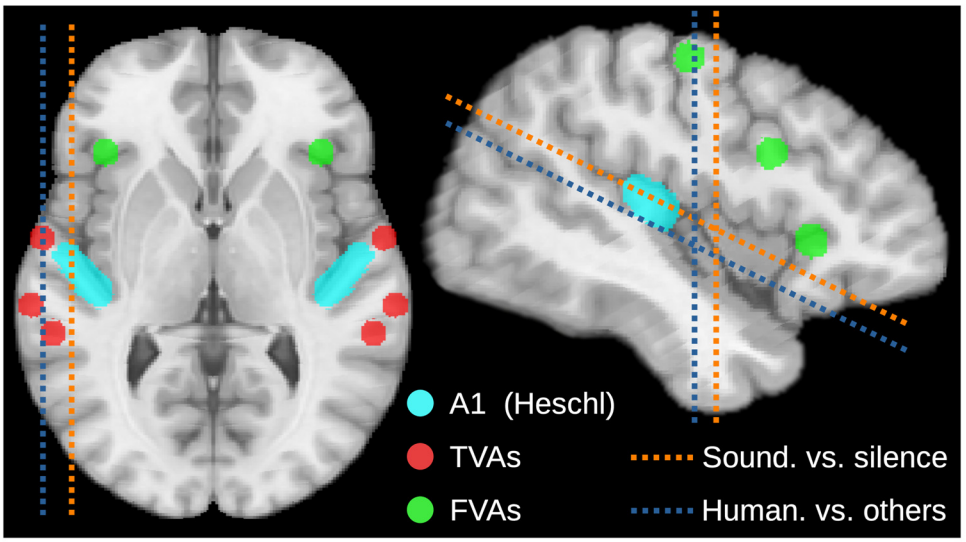

Figure 3. ROI and slice locations. Location and angle of axial, sagittal and coronal slices shown on Figure 4 (dotted lines). Location of the ROIs at which max-t value is extracted and shown on Figure 5 (colored areas).

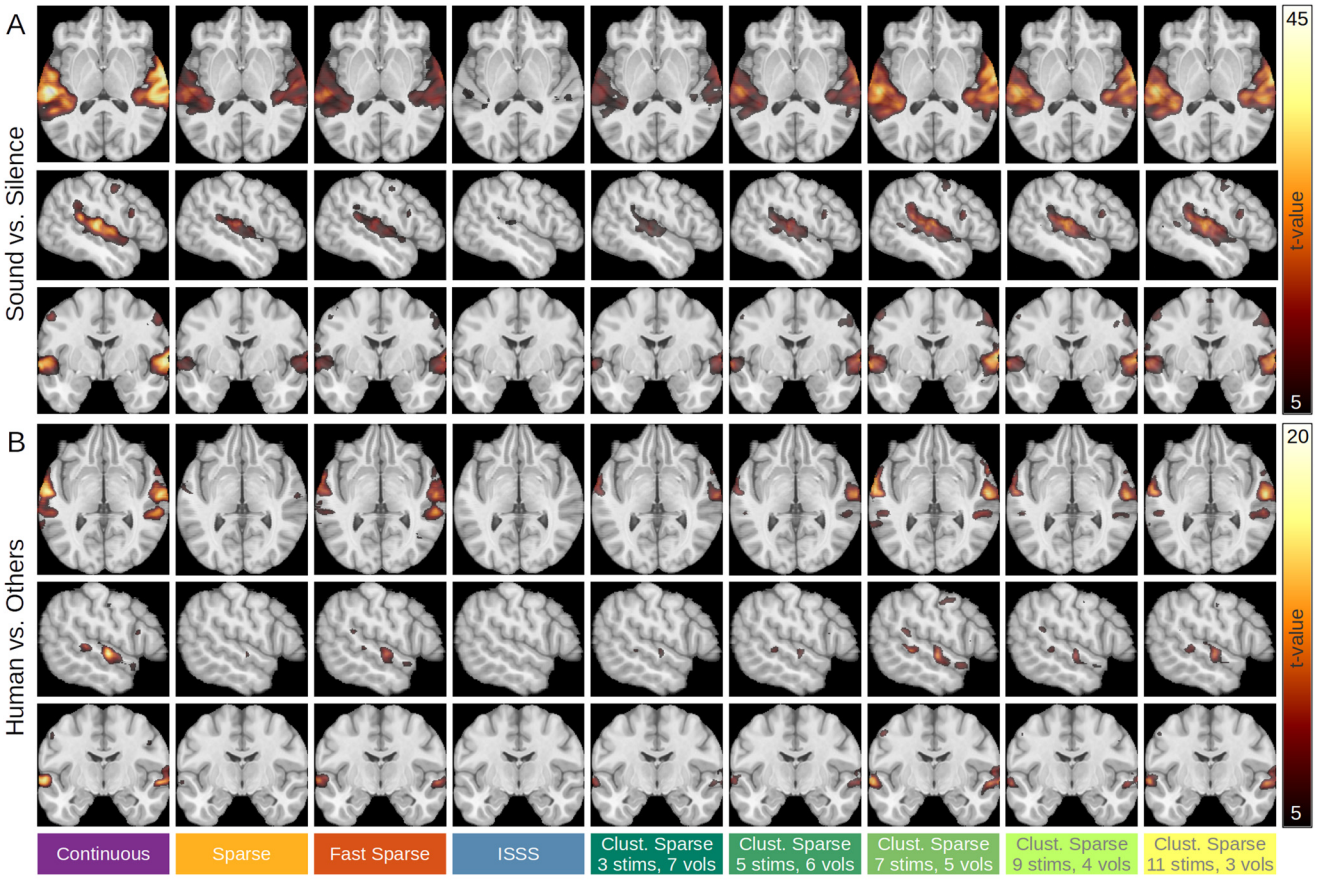

Figure 4. Group-averaged brain activations for all imaging sequences. (**A**) "Sound vs. Silence" contrast. (**B**) "Human vs. Others" contrast. Slice locations for each contrast are represented on Figure 3.

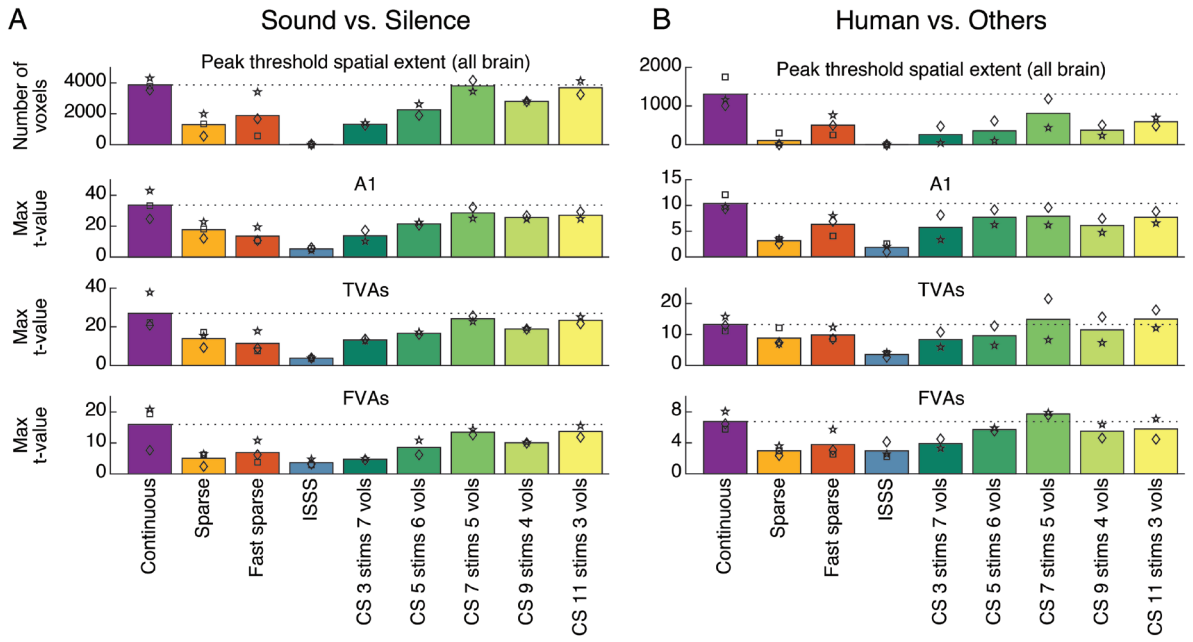

Figure 5. Spatial extent and ROI max t-values. For the "Sound vs. Silence" (**A**) and the "Human vs. Others" (**B**) contrasts. Each bar represents the spatial extent or maximum t-value observed for a particular sequence and averaged across participants. Individual values are represented by symbols (square, diamond or star). To facilitate the comparison with the other sequences, the horizontal dotted line indicates the average value obtained using the continuous scanning sequence. CS, Clustered Sparse. ROI locations are represented on Figure 3.

"Human vs. Others" individual results

The "Human vs. Others" contrast produced comparable but slightly different patterns (Figure 5B). In particular, the fast sparse sequence consistently outperformed the sparse sequence for this contrast, especially in terms of average spatial extent which was markedly higher. In TVAs and FVAs, the average maximum t-value was slightly higher for the clustered sparse sequence, which was set to present seven stimuli per trial and acquire five volumes per cluster, than for the continuous sequence. Both sequences yielded the best results for the two participants who were tested with all the protocols.

Discussion

We compared the performance of five different fMRI protocols of the same duration at detecting auditory and voice-related activation in human participants adequately protected from scanner noise. Under these conditions, continuous imaging was found to provide the best results, closely followed by a clustered sparse sequence in which stimulus and acquisition durations were both set to be around 4 s. Sparse and fast sparse sampling provided intermediate results, while ISSS showed the poorest performance.

The superior results of continuous imaging align with previous duration-matched comparisons between continuous and sparse sampling protocols, and can be explained by the combination of its established strength in maximizing data acquisition and the good protection from scanner noise provided by contemporary equipment. However, in setups where optimal protection from the scanner noise cannot be achieved, or where the magnetic field is higher, resulting in increased scanner noise (Akbar *et al.*, 2023), the observed superiority of continuous imaging performance may be compromised, while that of the protocols presenting stimuli during silence may not. In such setups, or when the experimental design necessitates the presentation of stimuli in silence, a clustered sparse sequence with the appropriate parameters may be the optimal choice. Prior work using clustered sparse imaging has demonstrated that it can outperform continuous scanning in certain contexts (Schmidt *et al.*, 2008), and the present study likewise showed that its performance approached that of continuous scanning. Additionally, we show that T1 signal decay can be readily corrected by adding the first two principal components of a PCA performed on the white matter tissue as nuisance regressors to the linear model.

Fast sparse sampling, with the advantage of constant TRs and longitudinal magnetization maintained at a steady state, also showed promising results, especially for the "Human vs. Others" contrast. For this sequence, we presented the same number of stimuli per block as for the continuous and sparse sequences, resulting in much longer blocks in comparison. In this particular design, fewer blocks could be presented in the fixed duration of a run, to the detriment of the sequence. Comparing variations of the fast sparse sequence, as we did for the clustered sparse sequence, would shed light on the optimal

parameters for block duration and improve the performance of this sequence. However, because the stimuli are quickly interleaved with volume acquisitions, fast sparse sampling is only suitable for short stimuli, limiting its generalizability to tasks involving longer sounds.

The original method that allows stimuli to be presented in silence, sparse sampling, produced results that were about half those of clustered sparse and continuous imaging. This highlights the main limitation of this method, its low data sampling per unit of time. However, If sparse does not seem to be the optimal sequence choice for a voice localizer, it may be crucial for certain study designs to separate acquisitions by a long enough interval to ensure that the hemodynamic response to the previous acquisition noise has passed. In such cases, our clustered sparse results suggest that it may be appropriate to cluster multiple volumes instead of acquiring volumes one after the other.

The results obtained with ISSS gave the lowest activation results of all the acquisition schemes. We have reservations about these results as they are very different from previous comparisons of ISSS with continuous or sparse sequences. As a side note, participants perceived the pseudo-silent periods of ISSS as moderately less loud than the acquisition periods. The ISSS sequence used in this study was obtained from the CMRR multiband sequence, version R016a. In the current version (R017), the implementation of ISSS is described as providing quieter "silent" periods.

A key limitation of this study is the small sample size of only three participants, which restricts the generalizability of the findings. However, clear trends were observed and sequences were ranked in a consistent manner across participants, suggesting that the overall picture appears to be valid. Furthermore, this study employed a block design, and it remains an open question as to which protocol would perform best in an event-related design.

In summary, continuous imaging is a powerful tool for auditory fMRI, especially in designs in which scanner noise masking does not significantly alter categorical perception and when human subjects are well protected from scanner noise. However, when stimuli must be presented in silence, clustered sparse sampling, with its specific balance between stimulation and acquisition time, emerges as a more effective protocol.

Ethics and consent

Participants gave written informed consent and were not paid for their participation. The experiment was granted ethical approval by Aix-Marseille Université "Comité Protection des Personnes"—"Committee for the Protection of Individuals": CPP 2017-A03614–49, February 13, 2019. The experimental procedures conformed to the World Medical Association's Declaration of Helsinki.

Data and software availability statement

SPM and fMRISTAT are compatible with GNU Octave.

Underlying data

Zenodo: Human fMRI data - Comparison of Auditory fMRI Protocols for a Voice Localizer, https://doi.org/10.5281/zenodo.13982532 (Trapeau *et al.*, 2024)

Data are available under the terms of the Creative Commons Attribution 4.0 International

Extended data

Zenodo: Audio stimuli - Voice localizer short, https://doi.org/10.5281/zenodo.14169736

Data are available under the terms of the Creative Commons Attribution 4.0 International

Author contributions

Conceptualization, R.T., and P.B.; Methodology, R.T., J.S., and P.B.; Data Curation, R.T. ; Investigation, R.T., J.S., and M.O.D.; Formal Analysis, R.T.; Software, R.T.; Visualization, R.T.; Funding Acquisition, P.B.; Supervision, P.B.; Writing - Original Draft Preparation, R.T.; Writing - Review & Editing, R.T., J.S., M.O.D. and P.B.

References

Aglieri V, Chaminade T, Takerkart S, *et al.*: **Functional connectivity within the voice perception network and its behavioural relevance.** *Neuroimage.* 2018; **183**: 356–365.
PubMed Abstract | Publisher Full Text | Free Full Text

Akbar AF, Sayyid ZN, Roberts DC, *et al.*: **Acoustic noise levels in high-field Magnetic Resonance Imaging scanners.** *OTO Open.* 2023; **7**(3): e79.
PubMed Abstract | Publisher Full Text | Free Full Text

Bach JP, Lüpke M, Dziallas P, *et al.*: **Functional Magnetic Resonance Imaging of the ascending stages of the auditory system in dogs.** *BMC Vet Res.* 2013; **9**(1): 210.
PubMed Abstract | Publisher Full Text | Free Full Text

Belin P, Zatorre RJ: **Adaptation to speaker's voice in right anterior temporal lobe.** *Neuroreport.* 2003; **14**(16): 2105–2109.
PubMed Abstract | Publisher Full Text

Belin P, Zatorre RJ, Hoge R, *et al.*: **Event-Related fMRI of the auditory cortex.** *Neuroimage.* 1999; **10**(4): 417–429.
PubMed Abstract | Publisher Full Text

Belin P, Zatorre RJ, Lafaille P, *et al.*: **Voice-selective areas in human auditory cortex.** *Nature.* 2000; **403**(6767): 309–312.
PubMed Abstract | Publisher Full Text

Capilla A, Belin P, Gross J: **The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG.** *Cereb Cortex.* 2013; **23**(6): 1388–1395.
PubMed Abstract | Publisher Full Text

Dewey RS, Hall DA, Plack CJ, *et al.*: **Comparison of continuous sampling with active noise cancelation and sparse sampling for cortical and subcortical auditory functional MRI.** *Magn Reson Med.* 2021; **86**(5): 2577–2588.
PubMed Abstract | Publisher Full Text

Edmister WB, Talavage TM, Ledden PJ, *et al.*: **Improved auditory cortex imaging using clustered volume acquisitions.** *Hum Brain Mapp.* 1999; **7**(2): 89–97.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Erb J, Armendariz M, De Martino F, *et al.*: **Homology and specificity of natural sound-encoding in human and monkey auditory cortex.** *Cereb Cortex.* 2019; **29**(9): 3636–3650.
PubMed Abstract | **Publisher Full Text**

Fonov V, Evans AC, Botteron K, *et al.*: **Unbiased average age-appropriate atlases for pediatric studies.** *Neuroimage.* 2011; **54**(1): 313–27.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Fuglsang SA, Madsen KH, Puonti O, *et al.*: **Mapping cortico-subcortical sensitivity to 4 Hz Amplitude Modulation depth in human auditory system with functional MRI.** *Neuroimage.* 2022; **246**: 118745.
PubMed Abstract | **Publisher Full Text**

Gaab N, Gabrieli JDE, Glover GH: **Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise.** *Hum Brain Mapp.* 2007; **28**(8): 703–720.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Glover GH: **Deconvolution of impulse response in event-related BOLD fMRI.** *Neuroimage.* 1999; **9**(4): 416–429.
PubMed Abstract | **Publisher Full Text**

Hall AJ, Brown TA, Grahn JA, *et al.*: **There's more than one way to scan a cat: imaging cat auditory cortex with high-field fMRI using continuous or sparse sampling.** *J Neurosci Methods.* 2014; **224**: 96–106.
PubMed Abstract | **Publisher Full Text**

Hall DA, Chambers J, Akeroyd MA, *et al.*: **Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional Magnetic Resonance Imaging.** *J Acoust Soc Am.* 2009; **125**(1): 347–359.
PubMed Abstract | **Publisher Full Text**

Hall DA, Haggard MP, Akeroyd MA, *et al.*: **Sparse temporal sampling in auditory fMRI.** *Hum Brain Mapp.* 1999; **7**(3): 213–223.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jafari A, Dureux A, Zanini A, *et al.*: **A vocalization-processing network in marmosets.** *Cell Rep.* 2023; **42**(5): 112526.
PubMed Abstract | **Publisher Full Text**

Joly O, Pallier C, Ramus F, *et al.*: **Processing of vocalizations in humans and monkeys: a comparative fMRI study.** *Neuroimage.* 2012; **62**(3): 1376–1389.
PubMed Abstract | **Publisher Full Text**

Keil P, Nettekoven C, Weiss K, *et al.*: **Accelerated clustered sparse acquisition to improve functional MRI for mapping language functions.** *J Neurol Surg A Cent Eur Neurosurg.* 2020; **81**(2): 95–104.
PubMed Abstract | **Publisher Full Text**

Langers DRM, Sanchez-Panchuelo RM, Francis ST, *et al.*: **Neuroimaging paradigms for tonotopic mapping (II): the influence of acquisition protocol.** *Neuroimage.* 2014; **100**: 663–675.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Mueller K, Mildner T, Fritz T, *et al.*: **Investigating brain response to music: a comparison of different fMRI acquisition schemes.** *Neuroimage.* 2011; **54**(1): 337–343.
PubMed Abstract | **Publisher Full Text**

Norman-Haignere S, Kanwisher NG, McDermott JH: **Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition.** *Neuron.* 2015; **88**(6): 1281–1296.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Norman-Haignere SV, Kanwisher N, McDermott JH, *et al.*: **Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones.** *Nature Neurosci.* 2019; **22**(7): 1057–1060.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ortiz-Rios M, Kuśmierek P, DeWitt I, *et al.*: **Functional MRI of the vocalization-processing network in the macaque brain.** *Front Neurosci.* 2015; **9**: 113.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Oya H, Gander PE, Petkov CI, *et al.*: **Neural phase locking predicts BOLD response in human auditory cortex.** *Neuroimage.* 2018; **169**: 286–301.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Peelle JE: **Methodological challenges and solutions in auditory functional magnetic resonance imaging.** *Front Neurosci.* 2014; **8**: 253.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Peelle JE, Eason RJ, Schmitter S, *et al.*: **Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing.** *Neuroimage.* 2010; **52**(4): 1410–1419.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Penhune VB, Zatorre RJ, MacDonald JD, *et al.*: **Interhemispheric anatomical differences in human Primary Auditory Cortex: probabilistic mapping and volume measurement from magnetic resonance scans.** *Cereb Cortex.* 1996; **6**(5): 661–672.
PubMed Abstract | **Publisher Full Text**

Perrachione TK, Ghosh SS: **Optimized design and analysis of sparse-sampling fMRI experiments.** *Front Neurosci.* 2013; **7**: 55.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Petkov CI, Kayser C, Augath M, *et al.*: **Optimizing the imaging of the monkey auditory cortex: sparse vs. continuous fMRI.** *Magn Reson Imaging.* 2009; **27**(8): 1065–1073.
PubMed Abstract | **Publisher Full Text**

Petkov CI, Kayser C, Steudel T, *et al.*: **A voice region in the monkey brain.** *Nat Neurosci.* 2008; **11**(3): 367–374.
PubMed Abstract | **Publisher Full Text**

Ravicz ME, Melcher JR: **Isolating the auditory system from acoustic noise during functional Magnetic Resonance Imaging: examination of noise conduction through the ear canal, head, and body.** *J Acoust Soc Am.* 2001; **109**(1): 216–231.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ress D, Chandrasekaran B: **Tonotopic organization in the depth of human Inferior Colliculus.** *Front Hum Neurosci.* 2013; **7**: 586.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Riecke L, Peters JC, Valente G, *et al.*: **Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex.** *Cereb Cortex.* 2017; **27**(5): 3002–3014.
PubMed Abstract | **Publisher Full Text**

Riecke L, Peters JC, Valente G, *et al.*: **Frequency-specific attentional modulation in human Primary Auditory Cortex and midbrain.** *Neuroimage.* 2018; **174**: 274–287.
PubMed Abstract | **Publisher Full Text**

Rogers JC, Davis MH: **Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds.** *J Cogn Neurosci.* 2017; **29**(5): 919–936.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Sadagopan S, Temiz-Karayol NZ, Voss HU: **High-field functional Magnetic Resonance Imaging of vocalization processing in marmosets.** *Sci Rep.* 2015; **5**(1): 10950.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schmidt CF, Zaehle T, Meyer M, *et al.*: **Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task.** *Hum Brain Mapp.* 2008; **29**(1): 46–56.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schönwiesner M, Dechent P, Voit D, *et al.*: **Parcellation of human and monkey core auditory cortex with fmri pattern classification and objective detection of tonotopic gradient reversals.** *Cereb Cortex.* 2015; **25**(10): 3278–3289.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schwarzbauer C, Davis MH, Rodd JM, *et al.*: **Interleaved Silent Steady State (ISSS) imaging: a new sparse imaging method applied to auditory fMRI.** *Neuroimage.* 2006; **29**(3): 774–782.
PubMed Abstract | **Publisher Full Text**

Stewart HJ, Cash EK, Hunter LL, *et al.*: **Speech cortical activation and connectivity in typically developing children and those with Listening Difficulties.** *Neuroimage Clin.* 2022; **36**: 103172.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Trapeau R, Schönwiesner M: **Adaptation to shifted interaural time differences changes encoding of sound location in human auditory cortex.** *Neuroimage.* 2015; **118**: 26–38.
PubMed Abstract | **Publisher Full Text**

Trapeau R, Schönwiesner M: **The encoding of sound source elevation in the human auditory cortex.** *J Neurosci.* 2018; **38**(13): 3252–3264.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Trapeau R, Sein J, Obliger-Debouche M, *et al.*: **Human fMRI data—Comparison of Auditory fMRI Protocols for a Voice Localizer.** [Dataset]. Zenodo. 2024.
http://www.doi.org/10.5281/zenodo.13982532

Wolak T, Cieśla K, Lorens A, *et al.*: **Tonotopic organisation of the auditory cortex in sloping sensorineural hearing loss.** *Hear Res.* 2017; **355**: 81–96.
PubMed Abstract | **Publisher Full Text**

Worsley KJ, Liao CH, Aston J, *et al.*: **A general statistical analysis for fMRI data.** *Neuroimage.* 2002; **15**(1): 1–15.
PubMed Abstract | **Publisher Full Text**

Worsley KJ, Marrett S, Neelin P, *et al.*: **A unified statistical approach for determining significant signals in images of cerebral activation.** *Hum Brain Mapp.* 1996; **4**(1): 58–73.
PubMed Abstract | **Publisher Full Text**

Yakunina N, Kang EK, Kim TS, *et al.*: **Effects of scanner acoustic noise on intrinsic brain activity during auditory stimulation.** *Neuroradiology.* 2015; **57**(10): 1063–1073.
PubMed Abstract | **Publisher Full Text**

Zaehle T, Schmidt CF, Meyer M, *et al.*: **Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks.** *Neuroimage.* 2007; **37**(4): 1195–1204.
PubMed Abstract | **Publisher Full Text**

Zaehle T, Wüstenberg T, Meyer M, *et al.*: **Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study.** *Eur J Neurosci.* 2004; **20**(9): 2447–2456.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ? ✓

Version 1

Reviewer Report 25 January 2025

https://doi.org/10.21956/openreseurope.20450.r49441

© **2025 De Martino F.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ **Federico De Martino**
Maastricht University, Maastricht, The Netherlands

The study compares acquisition protocols for auditory fMRI in the context of the localisation of voice responsive cortical regions. The study is methodologically sound and provides useful additional information in a literature that has performed similar comparisons in different contexts (e.g. tonotopic mapping) or using different settings (e.g. not matching the total acquisition time). Despite shown in a very limited number of volunteers, the results are interesting albeit perhaps not surprising: the increased statistical power afforded by continuous sampling overcomes limitations introduced by the saturation introduced by the scanner noise.
There are a few issues that the authors could perhaps clarify:

1. Was there any correction for serial autocorrelation applied when the GLM was performed?
2. Was there any way that the variable TR in the clustered acquisition was accounted for during the analysis – especially in connection to point 1, if serial autocorrelation was corrected how did the authors take into consideration the variable TR.
3. It is unclear why only for the clustered acquisition additional predictors are added to the GLM. I understand this allows to account for T1 saturation but the issue is that the reduced degrees of freedom renders some of the comparisons unfair (e.g. those based on t-values and p-values).
4. How do the responses across acquisition types compare in terms of test re-test reliability? It may be interesting to assess if e.g. spatial patterns are more stable across multiple data splits for one acquisition type compared to another.
5. While I find the discussion overall balanced, I wonder if the author would think that continuous acquisitions may have an effect of the response to specific stimuli also in their context. For example, the responses elicited by voice stimuli whose main frequency component is closer to the scanner noise may nevertheless be more affected than others. In this respect, I wonder if the authors would translate their conclusions from a block design oriented at categorical distinctions to designs that are optimized to estimate single stimulus responses and consequently perform model based “encoding” type of analyses.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: fMRI, cognitive neuroscience, auditory neuroscience, auditory fMRI, computational modeling, high field fMRI

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 24 January 2025

https://doi.org/10.21956/openreseurope.20450.r49448

© 2025 Zuo X. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Xi-Nian Zuo
[1] Developmental Population Neuroscience Research Center, Beijing Normal University, Beijing, China
[2] National Basic Science Data Center, China, China

This is a well deserved work to be index for promoting the utility of auditory fMRI. As claimed by the authors already, one limitation of the present work is the very small size of the samples (only three participants). I do think this would not change the trends observed on the brain activations significantly, however, would be better to validate the auditory fMRI methods from a more comprehensive way. Beyond the classical activation methodology, it is highly valuable to include connectional methods, e.g., the functional gradient of whole brain connectivity [1]. This would provide more evidence that the auditory fMRI can offer novel findings from a systems neuroscience perspective.

References
1. Margulies DS, Ghosh SS, Goulas A, Falkiewicz M, et al.: Situating the default-mode network along a principal gradient of macroscale cortical organization.*Proc Natl Acad Sci U S A*. 2016; **113** (44): 12574-12579 PubMed Abstract | Publisher Full Text

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: human brain mapping

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.