Neuroimage: Reports 1 (2021) 100050

Contents lists available at ScienceDirect

Neuroimage: Reports

journal homepage: www.sciencedirect.com/journal/neuroimage-reports

Sparse and continuous sampling approaches to fMRI of overt vocalization tasks

Dawn L. Merrett [a,b,1,*], Chris Tailby [a,b], David F. Abbott [b,c], Graeme D. Jackson [b,c], Sarah J. Wilson [a]

[a] Melbourne School of Psychological Sciences, 12th Floor Redmond Barry Building, The University of Melbourne, Victoria, 3010, Australia
[b] Florey Institute of Neuroscience and Mental Health, Melbourne Brain Centre, 245 Burgundy Street, Heidelberg, Victoria, 3084, Australia
[c] Department of Medicine - Austin Health, The University of Melbourne, Victoria, 3084, Australia

ARTICLE INFO

Keywords:
fMRI
Sparse temporal sampling
Auditory imaging
Singing
Overt response

ABSTRACT

Sparse temporal sampling has become the dominant paradigm for functional magnetic resonance imaging studies of auditory stimuli or verbal responses, as it allows the presentation or production of stimuli during the relatively quiet periods when there is no gradient switching and ensures that task-related movements are not occurring during scan acquisitions. To date, however, there has been no direct comparison between sparse and continuous acquisition protocols for overt auditory-verbal studies (i.e., speaking or singing). The aim of this study was to determine whether sparse temporal sampling would reduce movement artefacts and show better network activation for overt singing compared to continuous imaging. Fifteen healthy adults performed the same overt singing task under both sparse and continuous scanning conditions. We noted significant variations in signal intensity between adjacent slices in our sparse acquisition, with (odd-numbered) slices acquired in the second half of each volume acquisition being of lower intensity and showing less reliable task-related activation, and thus requiring the removal of these slices prior to preprocessing. Edge artefacts, presumably due to movement, were observed in both acquisition types at a subthreshold level, although ventricular space artefacts were more apparent in the continuous data. However, statistical comparison revealed no significant differences in functional activation nor in motion correction parameters. Our results show that sparse imaging has the potential to introduce significant image artefacts affecting downstream analyses. While sparse sampling provides benefits that may be essential for certain studies (e.g., periods free from scanner noise), the technical requirements of such sequences should not be overlooked and inspection of raw data is essential. Our data also show that continuous imaging can be used for overt response auditory-verbal studies and may be of wider utility than previously appreciated.

1. Introduction

Sparse temporal sampling (STS) has become a dominant technique in auditory-related functional imaging studies since its introduction in the late 1990s (Talavage and Hall, 2012). Around that time, a number of research groups developed functional Magnetic Resonance Imaging (fMRI) protocols that did not use the continuous acquisition typical of echo planar imaging (EPI), but rather used imaging sequences in which the repetition time (TR) was substantially longer than the acquisition time (TA) (Belin et al., 1999; Eden et al., 1999; Edmister et al., 1999; Elliott et al., 1999; D. A. Hall et al., 1999; Scheffler et al., 1998; Shah et al., 1999). These protocols were given various names, such as sparse temporal sampling, clustered volume acquisition, behavior interleaved gradients technique, and stroboscopic event-related method, but all included temporal gaps between volume acquisitions intended to address the considerable noise generated by the rapidly switching magnetic gradients required for EPI. Thus, the term "sparse temporal sampling" used by Hall et al. (1998) applies more generally and has

* Corresponding author. BRAMS / UdeM – FAS – Department of Psychology, CP 6128, succ. Centre-ville, Montréal, QC H3C 3J7, Canada.
E-mail addresses: dawn.merrett@umontreal.ca (D.L. Merrett), chris.tailby@florey.edu.au (C. Tailby), david.abbott@florey.edu.au (D.F. Abbott), graeme.jackson@florey.edu.au (G.D. Jackson), sarahw@unimelb.edu.au (S.J. Wilson).
[1] Present address: Department of Psychology and International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Montréal, Quebec, Canada.

https://doi.org/10.1016/j.ynirp.2021.100050
Received 23 February 2021; Received in revised form 6 August 2021; Accepted 1 September 2021
Available online 21 September 2021

become the most common term in the literature to describe such protocols.

By including a temporal gap between image acquisitions, STS allows the presentation or production of auditory stimuli, such as speech or music, to occur in the absence of most scanner noise. The effects of scanner noise on auditory paradigms have been well-documented (reviewed in Amaro et al., 2002; Moelker and Pattynama, 2003; Peelle, 2014) and include issues such as masking of auditory stimuli, creating saturation or habituation effects in auditory activations, increasing task difficulty, altering cognitive processing, and obscuring auditory feedback of a participant's own production during overt tasks. The use of sparse sampling mostly eliminates these scanner noise effects, making it an attractive option for many fMRI studies involving sound. It is also thought useful for studies that require overt vocal production, as it may ameliorate some of the movement artefacts in overt studies by allowing scanning to take place during silent, movement-free periods (Eden et al., 1999; Gracco et al., 2005).

Despite the many advantages provided by STS, it has a number of distinct disadvantages that have not been given full consideration. First, as indicated by the term "sparse", less data is typically acquired than in standard continuous sequences. This may take the form of reduced brain coverage, where fewer slices allow for faster volume acquisitions and somewhat shorter TRs, and/or fewer volumes (and hence degrees of freedom) acquired for each condition. The latter may not reduce power substantially, as sparse sequences have longer signal recovery times between excitations and may therefore show larger changes in blood-oxygenation-level-dependent (BOLD) signal magnitude for each volume (D. A. Hall et al., 1999). However, it does mean that the number of events that contribute to any given result is typically small in sparse sampling when trying to avoid unreasonably long functional runs. Such results may then be more vulnerable to sources of deviance, such as missed responses due to inattention, cognitive lapses, or head motion.

A second disadvantage of STS is uncertainty around capturing the peak of the hemodynamic response. Sparse sampling is possible because it exploits the natural delay in the brain's hemodynamic response to a stimulus. It allows for scanning after, rather than during, the stimulus—hopefully near the peak of the response. The canonical hemodynamic response function (HRF) places the peak at approximately 5–6 s after stimulus onset. However, this timing can vary substantially between individuals and even between brain regions within the same individual (Handwerker, Ollinger, & D'Esposito, 2004). Age, brain injury, and rehabilitation training have also been shown to change the timing of the peak hemodynamic response (Bonakdarpour et al., 2007; Huettel et al., 2001; Peck et al., 2004). Given this high degree of variability, deciding when to scan after stimulus onset is challenging. A fairly common way of addressing this issue is to "jitter" the timing of the stimulus in relation to the scan. While this allows different portions of the BOLD timecourse to be sampled, the peak will likely be under-sampled unless the number of scans (and therefore the run length) is increased to accommodate the jittering. Errors in estimating the magnitude of the response in both jittered and unjittered paradigms are likely, due to either under-sampling or missing the peak of the response (Peelle, 2014). In the context of an under-sampled HRF, it is usually not possible to disambiguate variations in amplitude from variations in latency.

Variations to the STS protocol in which a number of volumes are acquired in succession for each event are another way to obtain more data and ensure that the peak response is fully captured. Two different methods have been proposed: clustered temporal acquisition (Schmidt et al., 2008; Yang et al., 2000; Zaehle et al., 2007) and interleaved silent steady state imaging (ISSS; Schwarzbauer et al., 2006). The primary difference between these methods is whether or not longitudinal magnetization is held at steady state during the non-scanning periods. The ISSS protocol uses acoustically near-silent slice-selective excitation pulses to maintain steady state longitudinal magnetization during periods of stimulus presentation or production, with a number of continuous volumes acquired after each event. The principal argument against holding the magnetization at steady state, as in the ISSS protocol, is the increase in signal to noise ratio when the full longitudinal magnetization is available for the first volume obtained, as occurs with the long (relative to T1) TRs typical of STS (Zaehle et al., 2007). However, simply adding regressors into the data analysis for clustered temporal acquisition protocols to account for T1 decay-related changes, as suggested in Schmidt et al. (2008), may not sufficiently resolve the issues created by differences between volumes in longitudinal magnetization. In addition, whether or not magnetization is held constant, the more volumes acquired for each event and the more of the hemodynamic response that is sampled, the more likely that subsequent volumes will include brain responses to scanner noise, given that auditory regions may show a peak response as early as 3–5 s after stimulus onset (Belin et al., 1999; D. A. Hall et al., 2000).

A third disadvantage of sparse sampling protocols is that standard functional imaging analysis pipelines cannot be used. Given the temporal gaps between scans, slice timing correction should not be applied. However, unless very short TRs are used, either slice timing correction or some other method of accounting for variation in the acquisition time of slices should be used if the BOLD response is to be appropriately modeled through convolution of the HRF with predicted neuronal activity. Given the challenges this presents for sparsely acquired data, common practice has been to ignore the HRF and use simple boxcar model functions, but the adequacy of this approach has been questioned (Perrachione and Ghosh, 2013). In the case of clustered and ISSS acquisitions, the temporal filtering typically applied to continuous time series (e.g., high pass filtering, autoregressive modeling) also cannot be applied. Regrettably, the analysis pipelines used in extant studies are often not reported in sufficient detail, and no consistent and principled analysis pipeline has emerged. A review of the current literature suggests that STS protocols are often used without adequately addressing the analysis issues that they present, likely leading to suboptimal imaging analysis practices in some studies.

Since STS presents both advantages and disadvantages, comparisons between sparse and continuous protocols are of value in making decisions about whether to use the STS technique. A descriptive, but not statistical, comparison between continuous and sparse protocols was included in one of the first papers on this topic by Hall et al. (1999). Their conclusion was that for their auditory task (listening to speech samples) the two methods were equivalent in their "ability to detect clusters of voxels" (p. 219), although they found a greater mean percentage signal change in sparse imaging than in continuous. Across the two techniques, they found about 80% concordance for activation of auditory regions, with the continuous imaging more often showing a region not seen in sparse imaging. At an individual level, there were fairly large differences between sparse and continuous sampling, with either one or the other showing more significant voxels. In contrast, Edmister et al. (1999) reported that their clustered volume acquisition sequence consistently produced auditory cortex activations with a greater spatial extent and statistical significance than the distributed (continuous) acquisition for a music listening task. Further investigations in both humans and animals have often supported the latter finding, with a number of studies suggesting that STS approaches are superior to continuous imaging in terms of statistical power and/or spatial localization within auditory regions (Gaab et al., 2007; Mueller et al., 2011; Petkov et al., 2009; Schmidt et al., 2008; Woods et al., 2009; Yang et al., 2000).

Nevertheless, the question about which scanning protocol provides better statistical power and reliability remains open. Many of the comparison studies investigated only auditory areas using a small number of slices, very basic perceptual tasks, and very small numbers of participants (often only 3–4 participants; for example, Edmister et al., 1999; D. A. Hall et al., 1999; Petkov et al., 2009; Yang et al., 2000). Even under these conditions, contradictory results in favor of continuous scanning have been published (A. J. Hall et al., 2014). For studies examining the

whole brain (taking a network level approach), larger responses have often been found for continuous imaging in some extra-temporal regions, particularly the insula (Gaab et al., 2007; Schmidt et al., 2008). Schmidt et al. interpreted these findings as showing that continuous imaging provokes more demanding cognitive processing because of required stimulus detection against a background of noise. The insula is thought to support auditory attention and speech comprehension under challenging conditions (Adank, 2012; Bamiou et al., 2003). Such an interpretation is supported by the slower in-scanner reaction times observed for continuous imaging when making grammatical judgments about speech samples used in their study. However, other possible explanations, such as potential differences in time-to-peak across various regions of the brain, that could differentially influence the sensitivity of sparse versus continuous protocols should be considered. As suggested by Gaab et al. (2007), numerous variables likely play a role in determining whether the benefits of using an STS design outweigh the costs.

Despite their recommended use for overt response studies, to our knowledge no direct comparisons between STS and continuous designs for overt-production studies have been conducted. The current study aims to advance the literature by comparing sparse and continuous protocols in the context of an overt singing study. STS has been used in a number of previous imaging studies of overt singing for two primary reasons: 1) singing in the scanner is likely to produce significant movement artefacts that may be reduced by scanning after the singing-related movement has already taken place, and 2) participants may need to be able to hear themselves and to integrate their own vocal feedback in order to obtain ecologically valid results for the singing network (Zarate and Zatorre, 2008). For these reasons, we hypothesized that, despite its drawbacks, the STS protocol implemented in our study would show fewer movement artefacts and would more clearly delineate the known singing network than a matched continuous protocol.

2. Methods

2.1. Participants

The participants were 15 healthy adults (12 female) with a mean age of 33 (Range = 19–56), recruited primarily through university and community referrals. All participants were right-handed, reported normal hearing, and spoke English as their first language. Participants were screened for neurological and psychiatric illness, as well as any contraindications for MRI. None had participated in more than three years of formal music education on any instrument. All participants were novice singers, with seven having limited previous singing experience in the form of mandatory school choirs (n = 4), musical theatre participation (n = 1), or brief formal training (n = 3).

2.2. General experimental procedures

The study received approval from the Human Research Ethics Committee at The University of Melbourne, and all participants gave written informed consent in accordance with the Declaration of Helsinki. In compliance with the ethics approvals that were obtained for this research, participants were assured raw individual data would remain confidential and would not be shared with third parties. The current data were acquired as part of a randomized, cross-over longitudinal study investigating the influence of singing training on functional brain activation for singing and speech. Before training, a neuroimaging protocol was administered that included two different acquisitions for the same overt singing task: continuous EPI and STS EPI, allowing direct comparison between the two protocols.

2.3. Functional activation paradigm

The novel song used for this overt singing task was approximately 12 s in length with four 3 s phrases. Simple lyrics and a simple tune with a range smaller than one octave allowed novice singers to learn and perform the song with relative ease. It was introduced in a standardized manner at a separate session the day prior to baseline imaging, at which time a mock scan was also conducted to familiarize the participants with the scanning equipment, scanner noise, and the sensation of singing in a supine position. Participants were instructed to sing with minimal movement of the face and jaw. A template of the song was played to participants during same-day out-of-scanner rehearsal and immediately prior to the relevant run in the scanner to promote a consistent rate of performance. Recordings of the in-scanner performances were obtained to ensure that each participant completed the task correctly.

The design was mixed, with singing events jittered within task blocks (see Fig. 1). One phrase of the song was produced during each singing event, with four singing events (the four phrases of the song) comprising one task block. Four task blocks were interspersed with four rest blocks, for a total of 16 singing events in each run. During rest blocks, participants were asked to breathe in and exhale (cued using the same timing as in the singing events) to keep physiological processes as consistent as possible across task and rest, but were otherwise asked to "relax and just let your mind go blank." Participants attended to a visual display containing graphics that indicated when to inhale (in preparation) and when to sing or exhale. The graphics were delivered to a screen viewed by a double mirror attached to the head coil using Presentation software (Presentation 15.0, Neurobehavioral Systems, Berkeley, CA, www.neurobs.com), which allowed the onset and offset of singing and exhaling to be precisely timed to image acquisition. The length of each functional run was 7 min 14 s, exclusive of time spent providing instructions and playing the recorded song template.

2.4. fMRI acquisition

Functional MRI of overt singing was performed using a 3-T Siemens Trio system (Siemens, Erlanger, Germany) with a 12-channel birdcage head coil. Foam wedges were used to stabilize each participant's head and thereby minimize motion during overt responses. Passive sound attenuation was provided through the Siemens Trio over-ear headphones. Images were acquired using a multi-slice EPI sequence (single shot gradient recalled echo) providing T2*-weighted BOLD contrast. Imaging parameters across both the sparse and continuous sampling conditions were 44 interleaved axial oblique slices; voxel size = 3.0 × 3.0 × 3.0 mm; no inter-slice gap; echo time (TE) = 30 ms; field of view = 216 mm × 216 mm; flip angle = 85°. The continuous protocol had a TR of 3 s. After allowing magnetization to reach steady-state, a total of 129 whole brain volumes were acquired. The STS protocol had an effective TR of 12 s (Fig. 1). The scanner was turned on at the beginning of each 12 s TR for one 3 s EPI acquisition, during which a single whole brain volume was acquired. A total of 33 volumes were acquired in the sparsely sampled functional run.

2.5. fMRI data analysis

The images were pre-processed and then statistically analyzed using a general linear model (GLM) with Statistical Parametric Mapping Software (SPM8 r6313; Welcome Department of Imaging Neuroscience, London, UK) and the iBrain™ analysis toolbox for SPM (Abbott et al., 2011). Inspection of the raw STS images revealed variations in signal intensity between adjacent axial slices, with reduced signal within the odd-numbered slices (obtained during the second half of the acquisition; Fig. 2B, Section 3.2). Additional scrutiny of the raw and normalized time courses in motor and auditory regions showed that the slices obtained in the second half of each volume showed considerable decreases in intensity and did not show reliable task-related activation. For this reason, the sparse data were investigated in two ways. In one, all of the slices were retained, while in the other, the odd-numbered slices (obtained in the second half of each interleaved acquisition) were removed from the analysis, essentially creating a 3-mm gap between the remaining slices

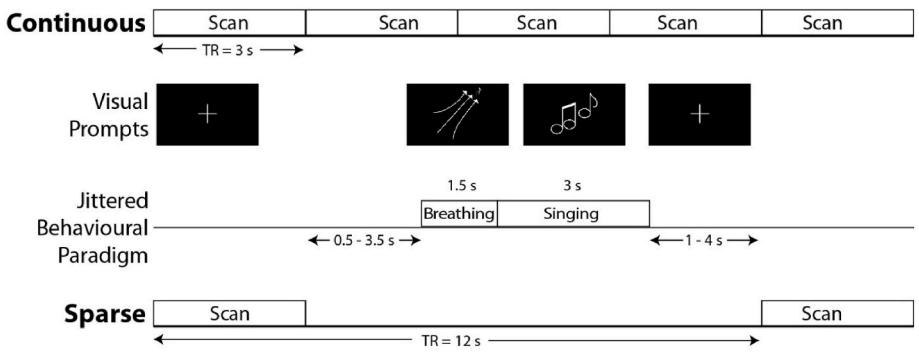

Fig. 1. A schematic of the jittered continuous and sparse temporal sampling imaging paradigms. For the sparse paradigm, the time between participant responses and the onset of scanning was varied by shifting the experimental tasks within the 12 s repetition time (TR). This jittering resulted in sets of axial images with delays of 4, 5, 6, and 7 s for the first image in the stack of axial slices in relation to task onset. The same behavioral timing (jittering) was also used when scanning continuously with a TR of 3 s.

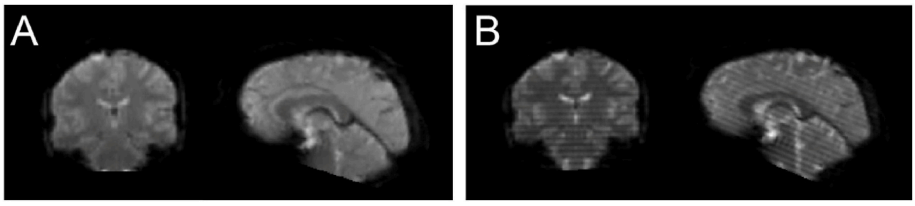

Fig. 2. Raw EPI image data from a single volume across the same representative participant in coronal and saggital planes, respectively, with (**A**) continuous sampling and (**B**) sparse sampling.

and a sequential acquisition. For comparison purposes, the continuous data were also analyzed both ways, with all slices included and with the alternate slices from the second half of each acquisition removed.

For both acquisition types and both investigations (all slices, even slices), pre-processing included 3D motion correction (using iBrain to select the optimal target acquisition time for realignment – the acquisition where the centre of intensity of within-brain voxels was closest to the median of the entire time series), normalization to the standard EPI template in SPM, resampling to a voxel size of 3 × 3 × 3 mm, and smoothing using a Gaussian kernel at 8 mm full width at half maximum (FWHM). For the continuous data, pre-processing also included slice timing correction. The design matrix for statistical analysis of the sparse data included the task regressor convolved with the canonical HRF, as well as six motion parameters (x, y, and z translations and pitch, yaw, and roll rotations) and a high-pass filter set at 128 s to capture low-frequency drift. The statistical analysis of the continuous data was the same, with the exception that it also included autoregression modeling (using an AR(1) process with alpha = 0.2, the default in SPM).

For sparse and continuous data both with and without alternate slices, sing-exhale contrasts were estimated for each participant, with these contrasts entered into a series of second-level random effects analyses. The threshold for statistical significance was set at $p < .001$, and subsequently, only voxels belonging to a cluster size large enough to survive a threshold of $p < .05$ corrected for topological false discovery rate (FDRc) were retained (Chumbley and Friston, 2009). One-sample t-tests were used to create second-level maps for each acquisition type for visual comparison. To assess the influence of removing alternate slices on group-level activations, paired t-tests were used to compare the analyses with all slices retained and with alternate slices removed (contrasts: *all-alt* and *alt-all*) for both the sparse and continuous data. To quantitatively assess our hypothesis that STS would better delineate the singing network, sparse and continuous acquisitions were compared at the second-level with paired t-tests (contrasts: *sparse-continuous* and *continuous-sparse*). Suprathreshold activations of second-level contrasts were assigned anatomical labels using the Harvard-Oxford probabilistic cortical and subcortical structural atlases (implemented in FSL; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases; Desikan et al., 2006). Thresholded (but not cluster-corrected) 2nd-level images were also binarized to calculate the Dice similarity coefficient, a widely-used cluster overlap method for assessing similarity or reliability (Bennett and Miller, 2010). Average cluster overlap in Bennett and Miller's 2010 review of fMRI reliability studies was found to be .45, which is typically interpreted as moderate reliability.

Model fit and motion parameters were also evaluated across acquisition types using SPSS (IBM SPSS Statistics 25 for Windows; https://www.ibm.com/products/spss-statistics/), using Wilcoxon Signed Rank Tests for nonparametric data. Bonferroni correction was used to account for multiple comparisons. The distributions of the variance of the mean contrast images for sparse and continuous acquisitions were compared using the two-sample Kolmogorov-Smirnov test in R (R Core Team, 2018; http://www.R-project.org/). Region-of-interest (ROI) analyses to assess activation within brain regions previously reported to show differential activation in sparse versus continuous scanning were conducted using the MarsBaR Toolbox for SPM (Brett et al., 2002) with anatomical ROIs created with the WFU PickAtlas tool (Maldjian et al., 2003).

3. Results

3.1. Comparable ranges of motion in sparse and continuous acquisitions

Movement in the scanner, as indexed by the motion parameters estimated during image preprocessing, for both sparse and continuous acquisitions appeared to be reasonably well-controlled. Given the positively skewed nature of the data, both the median and maximum movements were considered (Table 1). Translational movements (in the x, y, and z planes) were typically less than 0.1 mm and very rarely exceeded 0.5 mm, while rotational movements (pitch, roll, and yaw) were generally less than 0.002 radians, and few exceeded 0.015 radians.

There were no significant differences between acquisition types in the median or maximum rotations or translations in any direction (all p-values > .1), with the exception of greater maximum translations in the z-plane in continuous scanning ($Mdn = 0.364$) than in sparse scanning ($Mdn = 0.284$), $z = -2.90$, $p = .004$, $r = -0.53$.

Table 1
Motion in sparse and continuous fMRI acquisitions.

	Median Motion		Maximum Motion	
	Continuous	Sparse	Continuous	Sparse
Translations (mm)				
x	**0.026** (0.012)	**0.032** (0.025)	**0.090** (0.046)	**0.110** (0.060)
y	**0.095** (0.065)	**0.094** (0.089)	**0.361** (0.194)	**0.288** (0.181)
z	**0.082** (0.041)	**0.075** (0.023)	**0.364***(0.181)	**0.284***(0.150)
Rotations (radians)				
Pitch	**0.002** (0.001)	**0.002** (0.002)	**0.010** (0.013)	**0.010** (0.009)
Roll	**0.001** (0.001)	**0.001** (0.001)	**0.003** (0.001)	**0.003** (0.002)
Yaw	**0.001** (0.001)	**0.001** (0.001)	**0.004** (0.002)	**0.003** (0.002)

Given non-normal distributions, reported values are medians, with interquartile ranges in parentheses. *Significant difference between acquisition types after Bonferroni correction, with $p = .004$; all other p-values $> .1$.

3.2. Alternate slice signal intensity banding in STS

Visual inspection of the raw imaging data revealed large differences in signal intensity between adjacent slices in the STS acquisition (Fig. 2B) but not in the continuous acquisition (Fig. 2A), giving a banded appearance to the unprocessed STS images. Specifically, signal intensity was reduced in the odd numbered slices (those acquired in the second half of each volume in the interleaved acquisition). We consider the likely basis of this signal intensity variation below (see section 3.4).

3D motion correction, spatial normalization and smoothing effectively interpolate across these intensity bands, effectively washing out some of the post-processed signal. Slices obtained in the second half of each volume acquisition were therefore removed to address this issue (see Section 2.5 and follow-up data in Section 3.4).

For the sparse acquisition, this removal led to a visually-evident increase in the extent of suprathreshold activation relative to the analysis retaining all slices (Fig. 3C and D). However, second-level random effects analyses (subtractions) comparing the sparse data with all slices to the sparse data with alternate slices did not show statistically significant differences. In contrast, removal of alternate slices from the continuous acquisition decreased the extent of suprathreshold activation (Fig. 3A and B) and increased the suprathreshold deactivation in frontal, inferior temporal, and cerebellar regions. These same regions of deactivation showed significant differences in an *all-alt* second-level contrast. Given these findings, the best outcome obtained for each acquisition type – that is, using alternate slices for STS and all slices for continuous sampling, as indexed by extent of suprathreshold activation – will be used to compare the sparse and continuous protocols on the overt singing task.

3.3. Group-level comparisons show similar effects for sparse and continuous data

Although the binarization imposed by thresholding gives the suggestion of differences when visually comparing the group maps from the sparse and continuous acquisitions (Fig. 4A and B; Table 2), statistical comparisons using paired *t*-tests showed no significant cortical differences between the (alternate slices) sparse and (all slices) continuous data. Using another method of comparison, the Dice similarity coefficient of the binarized thresholded (at $p < .001$) images was .403 or moderate. It falls within the range (min = 0.314, max = 0.670) of Dice coefficients obtained across a number of reliability studies (Bennet and Miller, 2010) and is similar to the coefficients obtained in a more recent reliability study of language tasks (Wilson et al., 2017).

The contrast *sparse* minus *continuous* showed two clusters in the ventricular space and surrounding white matter (Fig. 4C), which appear to be artefact and reflect weak "deactivation" in the continuous scans but no signal change in the STS scans, while the contrast *continuous* minus *sparse* showed no suprathreshold clusters. There were also no significant differences in the model fits (median Root Mean Square Error) at the first level for the sparse ($Mdn = 0.708$) and continuous ($Mdn = 0.704$) acquisitions, $z = -1.306$, $p = .191$.

Inspection of the mean of the contrast images shows much greater concordance (Fig. 5A) than the statistical parametric maps (Fig. 4); however, increased variance (across participants) in the sparse contrast images was observed (Fig. 5B–D; Sparse $Mdn = 1.27$, Continuous $Mdn = 1.07$, Kolmogorov–Smirnov $D = 0.157$, $p < .001$).

To place these data in the context of our hypothesis that STS would better capture the singing network, Table 3 lists the regions of the putative network that were expected to be activated by our overt singing

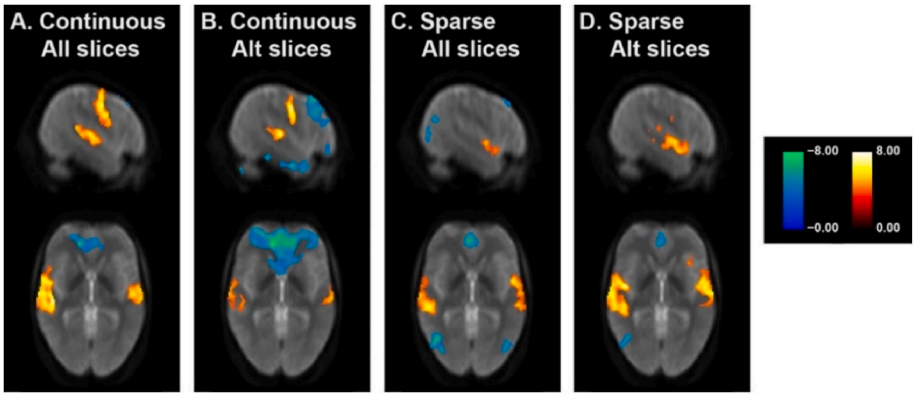

Fig. 3. Group mean activation maps (random effects analysis) for a novel singing task thresholded at $p < .001$ and cluster corrected at $p < .05$. (**A**) Continuous data with all slices. (**B**) Continuous data with alternate slices removed. (**C**) Sparse data with all slices. (**D**) Sparse data with alternate slices removed. Sagittal slices (top row) at $x = -50$; axial slices (bottom row) at $z = -4$.

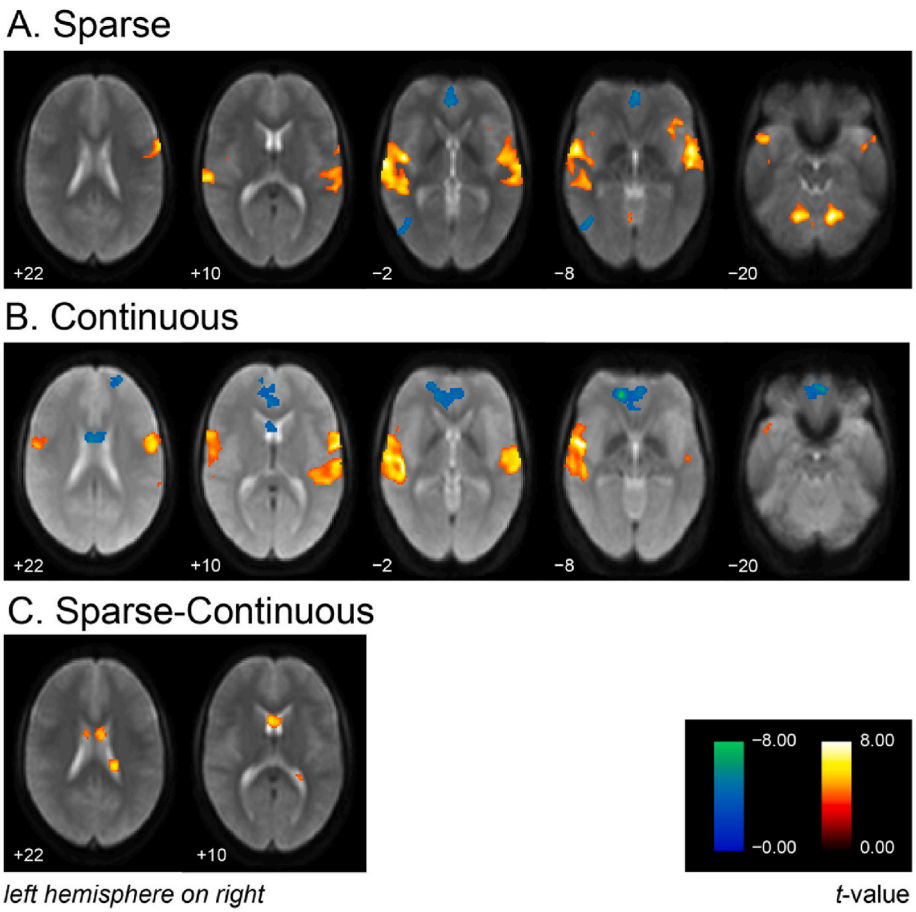

Fig. 4. Group mean activation maps (random effects analysis) for a novel singing task thresholded at $p < .001$ and cluster corrected at $p < .05$ from (A) the sparse temporal sampling acquisition and (B) the continuous sampling acquisition. (C) Subtraction of the continuous acquisition from the sparse sampling acquisition. There were no significant clusters for the continuous-sparse subtraction.

task based on the literature (for example, Brown et al., 2004; Kleber et al., 2007). Both acquisition types showed activation across most of this network; however, the activation that survived thresholding and FDRc did vary to some degree. The sparse data showed suprathreshold activation in the insula and cerebellum that was not above the corrected threshold in the continuous data, while the continuous data showed activation in the right precentral gyrus, right premotor cortex, and bilateral primary somatosensory cortex that was not above the corrected threshold in the sparse data.

Since the existing literature comparing sparse and continuous fMRI for auditory tasks reports greater activation in primary and secondary auditory areas for STS and greater activation in the insula for continuous sampling, region-of-interest (ROI) analyses were used to interrogate these specific structures. No significant differences in beta values were found between the two acquisition types for either ROI (p-values $> .25$).

3.4. Basis of signal loss in slices acquired later in each STS volume

To try to better understand the loss of signal intensity observed in the slices acquired later in each volume of the sparse acquisition we acquired a number of additional scans. We first considered whether slice cross-talk resulting from imperfections in the slice profiles, leading to excitation of adjacent slices given the use of contiguous slices (Edmister et al., 1999), could give rise to the effect. To this end, a series of post-hoc STS runs were obtained using the same sequence as for the sparse singing task acquisition but with distance factors/gaps of 0%, 10%, 30%, 50%, and 100% of the slice thickness (3 mm). Alternative slice intensity differences are still present even at the 100% gap width (Fig. 6), indicating that cross-talk cannot account for the effect.

We next looked at the spatial distribution of the effect in the unprocessed sparse images and noticed that the banding appeared to be most pronounced between adjacent slices at the bottom of the image (i.e. acquired earlier in the volume) and within white matter. We therefore quantified the degree of banding in grey and white matter separately, based on segmentation of the mean BOLD images (without motion correction or normalization) using the SPM12 segment routine. Specifically, we plotted the mean signal intensity in each axial slice separately for cortical grey matter voxels and white matter voxels, using a tissue class threshold of >0.85. This shows that in white matter, but not cortical grey matter, the mean signal intensity for the even (i.e. earlier acquired) slices is greater than that of their neighbouring odd (i.e. later acquired) slices (Fig. 6L-N;Q-S). The banding is also present when looking just at voxels with a lower grey matter tissue class threshold ($0.5 <$ threshold < 0.85), suggesting the effect permeates into partial volumed voxels that contain a mix of grey and white matter. The banding effect is not present with the continuous data (Fig. 6K, P). In other words, the signal intensity banding is only observed in the sparse acquisition, is present for all slice gaps/distance factors, and appears tissue specific, being strongest in white matter and weakest/absent in grey matter.

One sequence parameter that might produce such tissue specific effects is the fat saturation (*fat sat*) pulse, which was on by default in all our sparse acquisitions. The *fat sat* pulse seeks to null signal from fat by flipping the magnetization of fat-associated protons at the start of the

Table 2
Locations of peaks within activation clusters in sparse and continuous fMRI acquisitions.

	Coordinates (MNI space)			*t*-value	Cluster size
	x	*y*	*z*		
Sparse					
R superior temporal gyrus, posterior division	70	−26	6	9.48	2031
R superior temporal gyrus, anterior division	58	2	−10	8.91	
R planum polare	48	−8	−2	7.98	
L superior temporal gyrus, posterior division	−64	−28	2	9.33	1481
	−54	−12	−10	8.17	
L superior temporal gyrus, anterior division	−56	0	−10	7.78	
R cerebellum	16	−62	−22	9.00	427
	6	−60	−12	6.34	
	28	−58	−28	5.15	
L cerebellum	−14	−62	−20	7.85	489
L precentral gyrus	−64	2	24	6.40	279
	−44	−10	36	4.91	
	−64	0	12	4.71	
L frontal orbital cortex	−44	26	−10	6.13	139
L insular cortex	−34	20	−6	5.21	
Continuous					
L superior temporal gyrus, posterior division	−60	−26	4	11.16	2651
L precentral gyrus/central opercular cortex	−60	−4	10	7.36	
	−64	4	10	6.96	
R precentral gyrus	48	−6	34	9.87	2859
R superior temporal gyrus, posterior division	50	−24	−4	9.01	
R superior temporal gyrus, anterior division	56	−2	−8	8.48	

R = right; L = left.

imaging sequence (using an RF pulse tuned to the resonance frequency of fat) and using gradients to destroy this signal in the transverse plane; shortly thereafter signal from water-associated protons is imaged, before the fat-associated signal can regrow. We attempted to address this question by collecting corresponding data with *fat sat* switched off, but were unable to disable this feature for our standard sparse sampling protocol. We were, however, able to test this hypothesis by using an EPI pulse sequence obtained from the University of Minnesota Center for Magnetic Resonance Research (CMRR multi-band accelerated EPI version RO16a), which permitted disabling of *fat sat*. This sequence was primarily designed to support multiband acceleration; however, to maintain similarity with our earlier acquisitions, we disabled that

Table 3
Putative overt singing network regions (from Brown et al., 2004; Kleber et al., 2007) showing above-threshold activation ($p < .001$, FDRc = .05) for our overt singing task for sparse versus continuous image acquisitions. Activation that survived feature thresholding ($p < .001$) but not cluster correction (FDRc < .05) is shown in brackets ().

	Sparse	Continuous
Primary motor cortex	L, (R)	L, R
Premotor cortex (BA6)	L, (R)	L, R
Supplementary motor area	–	–
Inferior frontal gyrus	L*	(R)**
Anterior cingulate cortex	–	–
Anterior insula	L, R	(L, R)
Primary auditory cortex (Heschl's gyrus)	L, R	L, R
Planum temporale	L, R	L, R
Planum polare	L, R	L, R
Primary somatosensory cortex	(L, R)	L, R
Secondary somatosensory cortex	–	–
Inferior parietal cortex (supramarginal gyrus)	(L)	(L, R)
Dorsolateral prefrontal cortex	–	–
Putamen	–	–
Globus pallidus	–	(L, R)
Thalamus	–	–
Cerebellum	L, R	(L, R)
Brainstem	(L, R)	(L, R)

FDRc = correction for false discovery rate; L = left; R = right; *pars orbitalis; **pars opercularis.

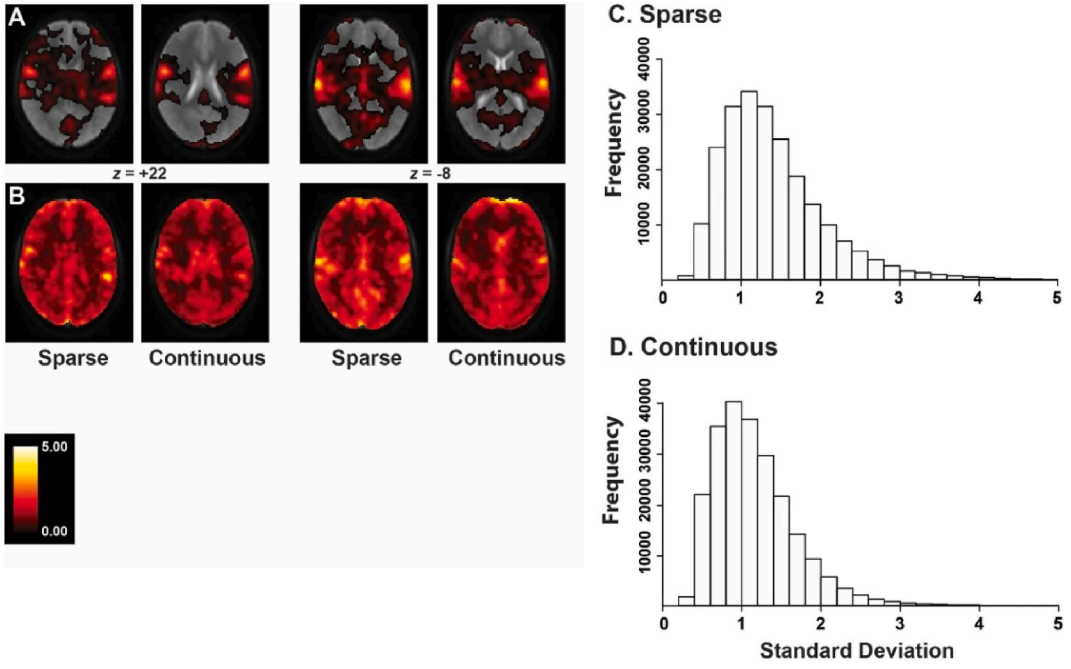

Fig. 5. Mean contrast images (row **A**) and standard deviations for those contrasts (row **B**) for sparse and continuous acquisitions. Axial slices at $z = +22$ and $z = -8$ are shown. **C** and **D** are histograms of the voxelwise standard deviations of the contrast images for sparse and continuous acquisitions, respectively.

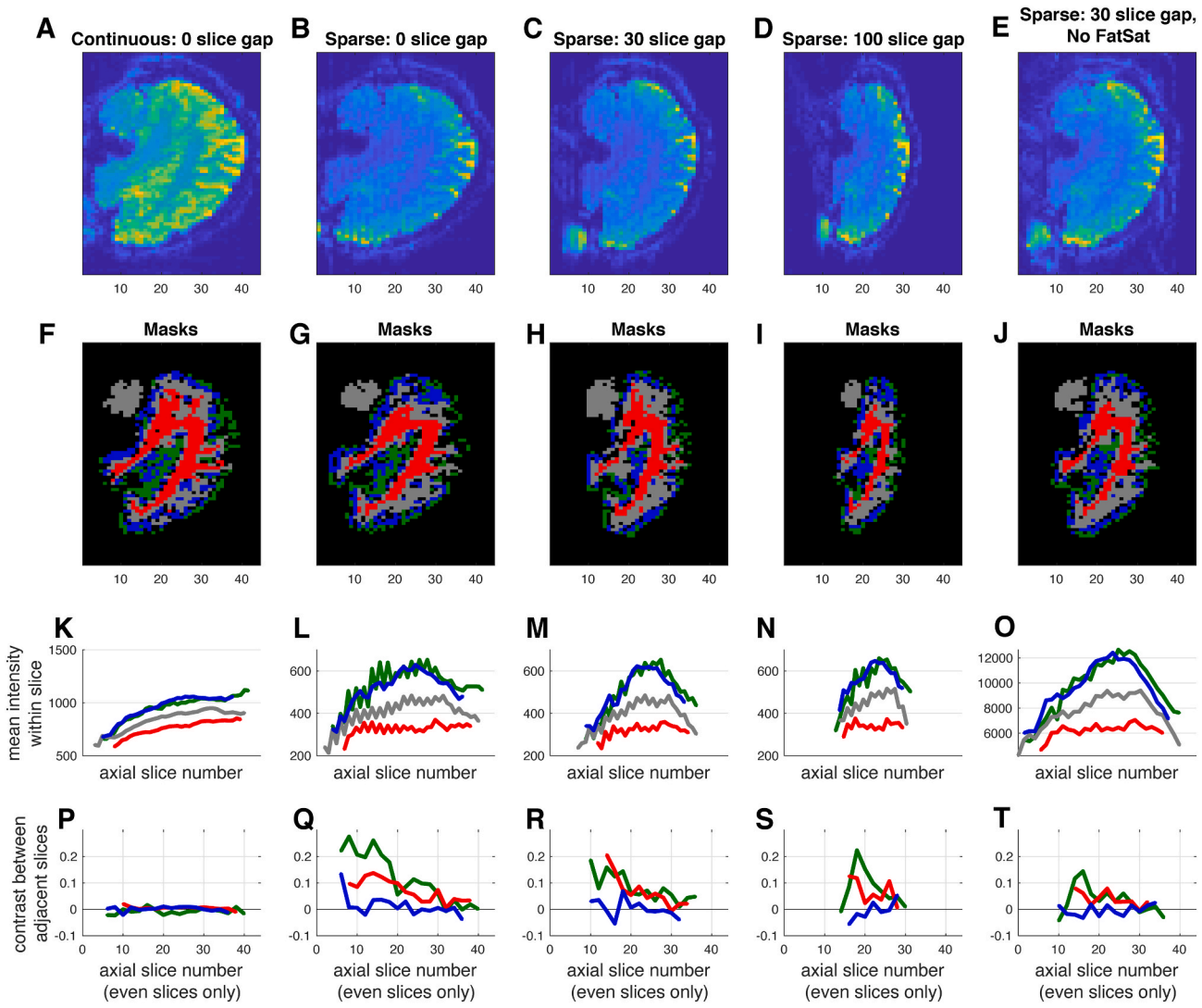

Fig. 6. **A-E**: mean functional images acquired using continuous data (**A**), sparse acquisitions with increasing distance factors/gaps between adjacent slices (**B**: 0% gap; **C**: 30% gap; **D**: 100% gap), and with a 30% gap but with *fat sat* disabled (**E**). The banding artefact is evident in all the sparse sampling acquisitions with *fat sat* enabled and is much reduced in **E** with *fat sat* disabled; however, an aliasing artefact is introduced with *fat sat* disabled. **F-J:** masks used to define voxels as high (blue voxels and blue curves in lower panels) or moderate (green) probability cortical grey matter, high probability white matter (red) or within brain (grey). Masks covered the full 3D volumes, shown here for a single parasagittal slice. **K-O**: mean signal intensity within each axial slice, plotted separately for each tissue type as defined in **F-J**. **P-T**: Contrast between signal intensity in a given *even* numbered slice relative to that in its neighbouring *odd* numbered slices, calculated separately for each *even* numbered slice as $[mean_{even} - (mean_{odd\text{-}below} + mean_{odd\text{-}above})/2]/[\ mean_{even} + mean_{odd\text{-}below} + mean_{odd\text{-}above})/3]$: 0 = comparable signal intensity among neighbouring slices, >0 = greater signal intensity in even slice relative to neighbouring odd slices. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

acceleration by setting the multiband factor to 1. The banding was significantly reduced with the elimination of *fat sat* pulses (Fig. 6, right column), suggesting that these repeated non-spatially selective *fat sat* pulses (one at the beginning of each slice) interact with the subsequently measured non-steady state EPI signal characteristic of sparse acquisitions to produce the temporal slice order-dependent intensity banding observed in our STS data (we confirmed that the artefact was present using this pulse sequence when *fat sat* was enabled, not shown). However, it should also be noted that a new aliasing artefact was introduced with the *fat sat* pulses disabled.

4. Discussion

Contrary to our hypothesis, there were no statistically significant differences between the sparse and continuous acquisitions in the fMRI activation of an overt singing task. Extensive and concordant activation in primary and secondary auditory cortices was evident across both acquisition types; however, above-threshold activation for motor and association regions varied, with some regions evident in sparse imaging but not continuous and vice versa. The reason for this discordance is not clear. Given the lack of statistically significant differences between the two datasets, it could simply be that these regions have an effect size near the limit of our detection power, and by chance only one or other of the methods happened to detect the regions in this particular study. It is also possible that significant activation seen in continuous but not sparse imaging could be related to the increased across-subject variance observed in the STS data. The presence of alternate slice signal intensity artefact in our implementation of the STS acquisition, and the processing steps undertaken to handle this, could also have influenced our findings. Our data nonetheless suggest that continuous acquisition protocols may be of wider utility in overt verbal response fMRI experiments than previously recognized.

Reduction of motion artefacts (both direct motion artefacts and indirect effects from movement-induced magnetic field inhomogeneities) is one the most often-cited reasons to use STS with overt tasks in the scanner, such as speaking or singing, since scanning can occur during periods when there is no task-related movement (Eden et al., 1999). In the current study, the extent of the motion correction required for translational and rotational movement between scans was roughly equivalent between the two types of imaging. The continuously acquired data showed possible motion artefacts in the ventricular space ("deactivation"), while both acquisition types showed edge artefacts in the unthresholded contrast images. As only the ventricular space artefact was above threshold, and other spurious activation was not evident, these results suggest that task-related movement did not significantly impact the continuously acquired data, at least in healthy participants who were trained to do the task with minimal head movement and were appropriately stabilized in the scanner. This could be more of an issue, however, in patients and research participants who are elderly or children and are known to move more in the scanner (Power et al., 2015). Further research investigating potential differences in overt task-related movement between STS and continuous acquisitions in clinical populations, such as those with speech and language disorders, is also warranted.

Another proposed reason to use STS for overt tasks is to ensure that participants can hear their own in-scanner responses. We felt that this might be particularly critical in a singing study, given the role of auditory feedback in vocal production (Zarate and Zatorre, 2008). However, during informal post-scanning debriefing, all participants reported that they could adequately hear themselves during the continuous scanning and did not find it harder to sing during the continuous acquisition than the sparse acquisition. A musically-trained listener (D.L.M.) did not observe any differences in production accuracy, corroborating participant reports that scanner noise did not seem to affect their singing; quantitative analyses of these data to confirm this finding are the subject of a separate investigation.

Previous whole-brain comparisons of sparse and continuous imaging during speech listening tasks found increased activation of primary and secondary auditory regions for STS and increased activation of insular cortex for continuous sampling (Gaab et al., 2007; Schmidt et al., 2008). We did not replicate these findings using an overt singing task; there were no significant differences in beta values in auditory or insular cortex. Reports of increased extra-temporal activation in continuous imaging, particularly in the insula, have been explained as increased cognitive load during processing of auditory material. In comparison to perceptual or discrimination tasks, the solo singing task used here may not have been more difficult to process in the presence of scanner noise, especially since bone conductance and kinesthetic information might provide additional feedback to complement the available auditory signal (Stenfelt, 2011; Kleber et al., 2013). Given that the insula has been recognized as part of the singing network (Riecker et al., 2000; Zarate, 2013), another possible explanation is that any potential differences in insula activation present as a result of increased task difficulty are too slight to be detected in the midst of singing task-related activity.

No significant differences between the sparse and continuous acquisitions in activation, model fit, or movement parameters were apparent in our data. However, a limitation of this study is the discrepancy between the sparse and continuous data that was created by removing alternate slices from the sparse data because of the *fat sat* artefact. Given that only half of the slices were retained in the sparse data for this comparison, it is possible that the sparse data were impoverished. On the other hand, the continuous data has possible optimizations that are not available in the sparse data, such as the inclusion of additional regressors to partial out activation related to externally-cued breathing in the preparatory period that would overlap with the singing task activation, which might also improve the model fit. In addition, the continuous acquisition protocol was designed to match the STS protocol and used the same limited number of events, although continuous imaging might allow greater detection power through the use of an efficient event-related design or a block design (which could be useful for certain types of overt responses) that included more singing. Thus, further research is required to determine whether a significant difference between acquisition protocols emerges when both are fully optimized.

In general, the collection and analysis of continuous imaging data is more straightforward than for STS. Continuous data can be analyzed using widely accepted and standardized protocols, whereas no principled and standardized analyses for STS data have yet been adopted in the field. The challenges surrounding slice timing correction, HRF convolution, and temporal filtering in STS data have not been fully resolved (Peelle, 2014; though we note that the need for slice timing correction may now be circumvented by using functional acquisitions with very brief TRs). Also, the potential influence of pulse sequence parameters on the resulting image properties in STS (e.g., Fig. 6) appears to be underappreciated, despite the description of many of these considerations in one of the earliest papers in the field (Edmister et al., 1999) and in a review of the development of STS methods (Talavage and Hall, 2012). The basis of the alternate slice intensity banding artefact in our own data appear to to be due to the interaction of fat suppression pulses (and their influence on longitudinal magnetization) with slice order in the long TR (12 s) sequence that we used. However, while disabling the *fat sat* pulses does address the slice intensity artefact, it introduced a new aliasing artefact which itself would need to be resolved before the sequence could be used in practice. It is clear that greater awareness of the technical requirements of STS acquisitions, improvements to the acquisition protocols themselves, inspection of the raw images (prior to steps like motion correction, warping and smoothing that will 'iron out' and thereby conceal inhomogeneities in the raw data) as a quality control step, and further refinement of image processing and modeling are all necessary to ensure that STS data are collected and analyzed in an appropriate fashion.

It is worth noting that use of the ISSS protocol mentioned previously could resolve the longitudinal magnetization issue, thereby potentially avoiding the signal intensity banding problem and allowing better characterization of the time course of activation and its peak. However, some of the unresolved analysis issues remain for these protocols, such as the temporal discontinuities in the data and the limited amount of data (number of events) that can be collected. Very recent attempts to address some of these issues in sparse scanning, such as novel sequences using "pre-saturated EPI" (Shrestha et al., 2021) or the application of acceleration techniques like multi-band imaging (Keil et al., 2020), offer potential solutions that should be investigated further. Despite these promising developments, it should not be discounted that continuous acquisitions can accommodate designs that may have better power to detect task-related activation. In the current data, where hearing and movement in the scanner did not pose a significant problem and whole brain coverage was required (cf. Schwarzbauer et al., 2006; Mueller et al., 2011), the ISSS protocol, an alternative sparse sequence that removes the alternate slice intensity banding artefact without introducing new image distortions, or even novel optimizations of STS may not have provided a significant benefit over optimized continuous imaging.

5. Conclusions

The use of STS is prevalent in auditory-related fMRI, yet under some conditions its advantages may be offset by its disadvantages. Considering the unresolved issues in STS analysis, the wide-spread adoption of STS protocols for all auditory and overt verbal studies may be unhelpful. Our study shows that unforeseen artefacts can be introduced when implementing STS from standard fMRI sequences. Furthermore, these artefacts can affect the results derived from standard preprocessing and analysis methods, but are 'smoothed out' and hence not readily apparent in the preprocessed images themselves; they are only apparent when

inspecting the raw images prior to preprocessing. Our results also suggest that concerns about movement and hearing in the scanner in the context of overt responses are not necessarily valid reasons to choose STS. These findings support the recommendations of Gaab et al. (2007) that for whole brain network-level exploration in healthy adults sparse sampling may not outperform continuous sampling. While it still may be important to use sparse sampling designs with certain populations (Davis et al., 2007) or for certain auditory tasks that require fine-grained analysis of auditory responses or auditory attention, thus necessitating further research to optimize this type of protocol, it appears that continuous protocols can be used for overt-response music and language fMRI tasks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions (CRediT statement)

Dawn Merrett: Conceptualization, Investigation, Formal analysis, Visualization, Writing – Original Draft. **Chris Tailby:** Software, Methodology, Formal analysis, Visualization, Writing – Review & Editing. **David Abbott:** Software, Methodology, Writing – Review & Editing. **Graeme Jackson:** Conceptualization, Resources, Supervision, Writing – Review & Editing. **Sarah Wilson:** Conceptualization, Supervision, Writing – Review & Editing.

Funding

The Florey Institute of Neuroscience and Mental Health acknowledges the strong support from the Victorian Government and in particular the funding from the Operational Infrastructure Support Grant of the State Government of Victoria, Australia. The authors also acknowledge the facilities and scientific and technical assistance of the National Imaging Facility (NIF), a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at The Florey Institute of Neuroscience and Mental Health. DLM was supported by a Melbourne International Research Scholarship and an Endeavor International Postgraduate Research Scholarship during the initial phases of this study. DFA is supported by fellowship funding from NIF. These funding/support sources had no involvement in the study.

Acknowledgements

We would like to thank Rob Smith, Eric Pierre, and Alan Connelly for helpful discussions about technical aspects of the study and also the anonymous reviewers for their helpful suggestions. We also acknowledge receipt of Software from the University of Minnesota Center for Magnetic Resonance Research, which we utilised to test the effect of disabling fat saturation pulses.

References

Abbott, D.F., Masterton, R.A.J., Waites, A.B., Bhaganagarapu, K., Pell, G.S., Harvey, M. R., et al., 2011. The iBrain(TM) analysis toolbox for SPM. In: *Proceedings for the 17th Annual Meeting of the Organization for Human Brain Mapping*, Quebec City, Canada, 364 WTh.

Adank, P., 2012. The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. Brain Lang. 122 (1), 42–54. https://doi.org/10.1016/j.bandl.2012.04.014.

Amaro, E., Williams, S.C.R., Shergill, S.S., Fu, C.H.Y., MacSweeney, M., Picchioni, M.M., McGuire, P.K., 2002. Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. J. Magn. Reson. Imag. 16 (5), 497–510. https://doi.org/10.1002/jmri.10186.

Bamiou, D.-E., Musiek, F.E., Luxon, L.M., 2003. The insula (Island of Reil) and its role in auditory processing: literature review. Brain Res. Rev. 42 (2), 143–154. https://doi.org/10.1016/S0165-0173(03)00172-3.

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. Neuroimage 10 (4), 417–429. https://doi.org/10.1006/nimg.1999.0480.

Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging? Ann. N. Y. Acad. Sci. 1191 (1), 133–155.

Bonakdarpour, B., Parrish, T.B., Thompson, C.K., 2007. Hemodynamic response function in patients with stroke-induced aphasia: implications for fMRI data analysis. Neuroimage 36 (2), 322–331. https://doi.org/10.1016/j.neuroimage.2007.02.035.

Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002, June. *Region of interest* analysis *using the MarsBar toolbox for SPM 99*. In: Presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.

Brown, S., Martinez, M.J., Hodges, D.A., Fox, P.T., Parsons, L.M., 2004. The song system of the human brain. Cognit. Brain Res. 20 (3), 363–375. https://doi.org/10.1016/j.cogbrainres.2004.03.016.

Chumbley, J.R., Friston, K.J., 2009. False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage 44 (1), 62–70. https://doi.org/10.1016/j.neuroimage.2008.05.021.

Davis, M.S., Fridriksson, J., Healy, E.W., Baylis, G.C., 2007. Effects of MRI scanner noise on language task performance in persons with aphasia. J. Med. Speech Lang. Pathol. 15 (2), 119–126.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021.

Eden, G., Joseph, J., Brown, H., Brown, C., Zeffiro, T., 1999. Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. Magn. Reson. Med. 41 (1), 13–20. https://doi.org/10.1002/(SICI)1522-2594(199901)41:1<13::AID-MRM4>3.0.CO;2-T.

Edmister, W.B., Talavage, T.M., Ledden, P.J., Weisskoff, R.M., 1999. Improved auditory cortex imaging using clustered volume acquisitions. Hum. Brain Mapp. *7* (2), 89–97. https://doi.org/10.1002/(SICI)1097-0193(1999)7:2<89::AID-HBM2>3.0.CO;2-N.

Elliott, M.R., Bowtell, R.W., Morris, P.G., 1999. The effect of scanner sound in visual, motor, and auditory functional MRI. Magn. Reson. Med. 41 (6), 1230–1235. https://doi.org/10.1002/(SICI)1522-2594(199906)41:6<1230::AID-MRM20>3.0.CO;2-1.

Gaab, N., Gabrieli, J.D., Glover, G.H., 2007. 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. 28 (8), 703–720. https://doi.org/10.1002/hbm.20298.

Gracco, V.L., Tremblay, P., Pike, B., 2005. Imaging speech production using fMRI. Neuroimage 26, 294–301. https://doi.org/10.1016/j.neuroimage.2005.01.033.

Hall, D.A., Elliott, M.R., Bowtell, R.W., Gurney, E., Haggard, M.P., 1998. "Sparse" temporal sampling in fMRI enhances detection of activation by sound for both magnetic and acoustical reasons. Neuroimage 7 (S551).

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., et al., 1999. "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7 (3), 213–223. https://doi.org/10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N.

Hall, D.A., Summerfield, A.Q., Gonçalves, M.S., Foster, J.R., Palmer, A.R., Bowtell, R.W., 2000. Time-course of the auditory BOLD response to scanner noise. Magn. Reson. Med. 43 (4), 601–606. https://doi.org/10.1002/(sici)1522-2594(200004)43:4<601::aid-mrm16>3.0.co;2-r.

Hall, A.J., Brown, T.A., Grahn, J.A., Gati, J.S., Nixon, P.L., Hughes, S.M., et al., 2014. 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 224, 96–106. https://doi.org/10.1016/j.jneumeth.2013.12.012.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21 (4), 1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029.

Huettel, S.A., Singerman, J.D., McCarthy, G., 2001. The effects of aging upon the hemodynamic response measured by functional MRI. Neuroimage 13 (1), 161–175. https://doi.org/10.1006/nimg.2000.0675.

Keil, P., Nettekoven, C., Weiss, K., Lichtenstein, T., Goldbrunner, R., Giese, D., Weiss Lucas, C., 2020. Accelerated clustered sparse acquisition to improve functional MRI for mapping language functions. J. Neurol. Surg. Cent. Eur. Neurosurg. 81 (2), 95–104. https://doi.org/10.1055/s-0039-1691821.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., Lotze, M., 2007. Overt and imagined singing of an Italian aria. Neuroimage 36 (3), 889–900. https://doi.org/10.1016/j.neuroimage.2007.02.053.

Kleber, B., Zeitouni, A.G., Friberg, A., Zatorre, R.J., 2013. Experience-dependent modulation of feedback integration during singing: role of the right anterior insula. J. Neurosci. 33 (14), 6070. https://doi.org/10.1523/JNEUROSCI.4418-12.2013.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19 (3), 1233–1239. https://doi.org/10.1016/S1053-8119(03)00169-1.

Moelker, A., Pattynama, P.M., 2003. Acoustic noise concerns in functional magnetic resonance imaging. Hum. Brain Mapp. 20 (3), 123–141. https://doi.org/10.1002/hbm.10134.

Mueller, K., Mildner, T., Fritz, T., Lepsien, J., Schwarzbauer, C., Schroeter, M.L., Möller, H.E., 2011. Investigating brain response to music: a comparison of different fMRI acquisition schemes. Neuroimage 54 (1), 337–343. https://doi.org/10.1016/j.neuroimage.2010.08.029.

Peck, K.K., Moore, A.B., Crosson, B.A., Gaiefsky, M., Gopinath, K.S., White, K., Briggs, R. W., 2004. Functional magnetic resonance imaging before and after aphasia therapy: shifts in hemodynamic time to peak during an overt language task. Stroke 35 (2), 554–559. https://doi.org/10.1161/01.STR.0000110983.50753.9D.

Peelle, J.E., 2014. Methodological challenges and solutions in auditory functional magnetic resonance imaging. Front. Neurosci. 8, 253. https://doi.org/10.1002/hbm.2037210.3389/fnins.2014.00253.

Perrachione, T.K., Ghosh, S.S., 2013. Optimized design and analysis of sparse-sampling fMRI experiments. Front. Neurosci. 7, 55. https://doi.org/10.1002/hbm.2037210.3389/fnins.2013.00055.

Petkov, C.I., Kayser, C., Augath, M., Logothetis, N.K., 2009. Optimizing the imaging of the monkey auditory cortex: sparse vs. continuous fMRI. Magn. Reson. Imag. 27 (8), 1065–1073. https://doi.org/10.1016/j.mri.2009.01.018.

Power, J.D., Schlaggar, B.L., Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105, 536–551. https://doi.org/10.1016/j.neuroimage.2014.10.044.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from. http://www.R-project.org/.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., Grodd, W., 2000. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. Neuroreport 11 (9), 1997–2000.

Scheffler, K., Bilecen, D., Schmid, N., Tschopp, K., Seelig, J., 1998. Auditory cortical responses in hearing subjects and unilateral deaf patients as detected by functional magnetic resonance imaging. Cerebr. Cortex 8 (2), 156–163. https://doi.org/10.1093/cercor/8.2.156.

Schmidt, C.F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., Jancke, L., 2008. Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. Hum. Brain Mapp. 29 (1), 46–56. https://doi.org/10.1002/hbm.20372.

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.

Shah, N.J., Jäncke, L., Grosse-Ruyken, M.-L., Mueller-Gaertner, H.W., 1999. Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. J. Magn. Reson. Imag. 9 (1), 19–25. https://doi.org/10.1002/(SICI)1522-2586(199901)9:1<19::AID-JMRI3>3.0.CO;2-K.

Shrestha, M., Lee, H.S., Nöth, U., Deichmann, R., 2021. A novel sequence to improve auditory functional MRI with variable silent delays. Magn. Reson. Med. 85 (2), 883–896. https://doi.org/10.1002/mrm.28479.

Stenfelt, S., 2011. Acoustic and physiologic aspects of bone conduction hearing. Adv. Oto-Rhino-Laryngol. 71, 10–21. https://doi.org/10.1159/000323574.

Talavage, T.M., Hall, D.A., 2012. How challenges in auditory fMRI led to general advancements for the field. Neuroimage 62 (2), 641–647. https://doi.org/10.1016/j.neuroimage.2012.01.006.

Wilson, S.M., Bautista, A., Yen, M., Lauderdale, S., Eriksson, D.K., 2017. Validity and reliability of four language mapping paradigms. Neuroimage: Clinical 16, 399–408.

Woods, D.L., Stecker, G.C., Rinne, T., Herron, T.J., Cate, A.D., Yund, E.W., et al., 2009. Functional maps of human auditory cortex: effects of acoustic features and attention. PloS One 4 (4). https://doi.org/10.1371/journal.pone.0005183 e5183-e5183.

Yang, Y., Engelien, A., Engelien, W., Xu, S., Stern, E., Silbersweig, D.A., 2000. A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. Magn. Reson. Med. 43 (2), 185–190. https://doi.org/10.1002/(sici)1522-2594(200002)43:2<185::aid-mrm4>3.0.co;2-3.

Zaehle, T., Schmidt, C.F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P., Jancke, L., 2007. Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. Neuroimage 37 (4), 1195–1204. https://doi.org/10.1016/j.neuroimage.2007.04.073.

Zarate, J.M., 2013. The neural control of singing. Front. Hum. Neurosci. 7, 237. https://doi.org/10.3389/fnhum.2013.00237.

Zarate, J.M., Zatorre, R.J., 2008. Experience-dependent neural substrates involved in vocal pitch regulation during singing. Neuroimage 40 (4), 1871–1887. https://doi.org/10.1016/j.neuroimage.2008.01.026.