

Published in final edited form as:

Neuroimage. 2014 October 15; 100: 663–675. doi:10.1016/j.neuroimage.2014.07.042.

Neuroimaging paradigms for tonotopic mapping (II): The influence of acquisition protocol

Dave R.M. Langers^{a,b,*}, Rosa M. Sanchez-Panchuelo^c, Susan T. Francis^c, Katrin Krumbholz^d, and Deborah A. Hall^{a,b}

^aNational Institute for Health Research (NIHR) Nottingham Hearing Biomedical Research Unit, University of Nottingham, Nottingham, UK

^bOtology and Hearing Group, Division of Clinical Neuroscience, School of Medicine, University of Nottingham, Nottingham, UK

^cSir Peter Mansfield Magnetic Resonance Centre, School of Physics and Astronomy, University of Nottingham, Nottingham, UK

^dMRC Institute for Hearing Research, Nottingham, UK

Abstract

Numerous studies on the tonotopic organisation of auditory cortex in humans have employed a wide range of neuroimaging protocols to assess cortical frequency tuning. In the present functional magnetic resonance imaging (fMRI) study, we made a systematic comparison between acquisition protocols with variable levels of interference from acoustic scanner noise. Using sweep stimuli to evoke travelling waves of activation, we measured sound-evoked response signals using *sparse*, *clustered*, and *continuous* imaging protocols that were characterised by inter-scan intervals of 8.8, 2.2, or 0.0 s, respectively. With regard to sensitivity to sound-evoked activation, the sparse and clustered protocols performed similarly, and both detected more activation than the continuous method. Qualitatively, tonotopic maps in activated areas proved highly similar, in the sense that the overall pattern of tonotopic gradients was reproducible across all three protocols. However, quantitatively, we observed substantial reductions in response amplitudes to moderately low stimulus frequencies that coincided with regions of strong energy in the scanner noise spectrum for the clustered and continuous protocols compared to the sparse protocol. At the same time, extreme frequencies became over-represented for these two protocols, and high best frequencies became relatively more abundant. Our results indicate that although all three scanning protocols are suitable to determine the layout of tonotopic fields, an exact quantitative assessment of the representation of various sound frequencies is substantially confounded by the presence of scanner noise. In addition, we noticed anomalous signal dynamics in response to our travelling wave paradigm that suggest that the assessment of frequency-dependent tuning is non-trivially influenced by time-dependent (hemo)dynamics when using sweep stimuli.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

*Corresponding author at: NIHR Nottingham Hearing Biomedical Research Unit, Ropewalk House, 113 The Ropewalk, Nottingham NG1 5DU, UK. Fax: +44 115 82 32615. davey.langers@nottingham.ac.uk (D.R.M. Langers).

Keywords

Human auditory cortex; Tonotopy; Sound; Functional magnetic resonance imaging (fMRI); Acoustic scanner noise

Introduction

In the context of functional magnetic resonance imaging (fMRI), the loud acoustic scanner noise (ASN) that is emitted by the read-out gradient switches of echo-planar imaging sequences is a major confounding factor that becomes increasingly serious as available magnetic field strengths get higher (Foster et al., 2000; Moelker and Pattynama, 2003). Apart from evoking activation in the auditory cortex (Bandettini et al., 1998; Elliott et al., 1999; Hall et al., 2000), ASN has also been shown to influence task-related activity in non-auditory brain regions related to vision (Zhang et al., 2005), motion (Fuchino et al., 2006), imagery (Mazard et al., 2002), nociception (Boyle et al., 2006), emotion (Skouras et al., 2013), attention (Novitski et al., 2001), working memory (Novitski et al., 2003; Haller et al., 2005; Tomasi et al., 2005), and the default mode network (Gaab et al., 2008). Similarly, various intrinsic brain networks were shown to be affected during so-called resting-state experiments (Langers and van Dijk, 2011; Rondinoni et al., 2013). Although often still underappreciated, ASN therefore is a ubiquitous factor that forms an important consideration in the design of fMRI experiments in general (Amaro et al., 2002).

For experiments that involve sound presentations in particular, ASN may influence outcomes at an acoustic or neural level in the form of a direct masking of the delivered stimuli and interference with task performance, or at a metabolic or vascular level in the form of a sustained elevation of baseline activity and reduced activation due to non-linear ceiling effects (Talavage and Edmister, 2004). Various solutions have been proposed to overcome these detrimental effects (Okada and Nakai, 2003). Several strategies aim to make the acquisition sequence quieter, for instance by optimising the design of coil geometries (Bowtell and Mansfield, 1995), excitation pulses (Schmitter and Bock, 2010), readout gradient shapes (Loenneker et al., 2001; Zapp et al., 2012), switching frequencies (Chapman et al., 2003; Segbers et al., 2010), slew rates (de Zwart et al., 2002), and *k*-space trajectories (Oesterle et al., 2001). Other approaches are based on reduction of the produced ASN by means of passive noise attenuation (Ravicz and Melcher, 2001; Nordell et al., 2009; Li and Mechefske, 2010) or active noise cancellation (McJury et al., 1997; Hall et al., 2009; Blackman and Hall, 2011; Kannan et al., 2011). Although these approaches allow ASN to be managed to some degree, fMRI sequences tend to remain far from quiet, such that additional measures remain necessary in practice.

A popular approach to deal with the problem of ASN is to employ sparse scanning protocols (Eden et al., 1999; Edmister et al., 1999; Hall et al., 1999). By spacing short fMRI acquisitions by extended periods of scanner inactivity, stimuli can be delivered in a nearly silent environment and a sustained elevation of the hemodynamic baseline in sound-responsive brain areas can be avoided. Ideally, the duration of acquisitions should not exceed approximately 2 s and the inter-scan interval should be of the order of 20 s in order for ASN

effects to become negligible (Talavage et al., 1999; Olulade et al., 2011). Because this would lead to a very low data acquisition rate, various authors have investigated the benefit of clustering multiple image acquisitions together (Schmithorst and Holland, 2004; Langers et al., 2007; Zaehle et al., 2007), sometimes accompanied by additional quiet radio-frequency excitation pulses in the inter-scan interval to retain magnetisation equilibrium (Schwarzbauer et al., 2006; Mueller et al., 2011). At the same time, it has been argued that an optimal sensitivity to sound-evoked activation would be obtained for moderate inter-scan intervals of 2 to 8 s (Liem et al., 2012; Perrachione and Ghosh, 2013). It has even been suggested that continuous scanning might be preferable because it does not involve sudden sound onsets (Seifritz et al., 2006). The discrepancies between findings may have arisen because the level of interference depends not only on the characteristics of ASN itself, but also on those of the stimulus and task paradigm. For instance, interactions may depend on stimulus attributes like sound intensity, frequency content, spectrotemporal dynamics, and behavioural significance, or task-related factors like instruction and attention. In summary, there is ample evidence that the effects of ASN can be reduced by implementing appropriate image acquisition protocols, but which protocol is optimal likely depends upon the study goals and design.

In this report, we focus on the effects of ASN in relation to tonotopic mapping. Whereas the first detailed cortical frequency tuning maps that were non-invasively produced a decade ago were inconsistent and therefore difficult to interpret (Schönwiesner et al., 2002; Formisano et al., 2003; Talavage et al., 2004), a rising number of studies have since resulted in a better consensus regarding the large-scale tonotopic organisation in human auditory cortex. As a result, tonotopic maps are gradually becoming a suitable research target to investigate in the context of studies on normal and disordered hearing (Langers, 2014). Homologous to the organisation in non-human primates (Hackett et al., 2001; Baumann et al., 2013), multiple abutting tonotopic gradients in core auditory cortex are found to fold at an angle across Heschl's gyrus (HG) in humans (reviewed by Saenz and Langers, 2014). Yet, despite the current agreement in outcomes across studies, researchers have employed very diverse imaging protocols. In particular, in relation to ASN, widely diverging scanning “duty cycles” have been employed: some authors employed continuous acquisitions (Da Costa et al., 2011; Dick et al., 2012), whereas others used sparse scanning (Humphries et al., 2010; Langers and van Dijk, 2011), and yet others settled on an intermediate design involving brief inter-scan intervals (Moerel et al., 2012; Norman-Haignere et al., 2013); for a representative overview of recent studies, see Table 1 in our accompanying paper (Langers et al., 2014-in this issue).

The question how ASN interacts with tonotopic maps is a highly relevant one because it has been shown that interference from ASN is frequency-dependent. In an elegant study, Scarff et al. (2004) showed not only that the perceived loudness of tones that best coincided with the frequency content of ASN decreased, but these tones also induced less detectable fMRI activity than tones at other frequencies. When the imaging sequence was modified in such a way that the dominant spectral content of the ASN shifted to higher frequencies, it was found that the interference with stimulus audibility and the corresponding dip in detectable activation also shifted towards those higher frequencies. Thus, the observed effects could be causally linked to ASN. At least two other studies also reported frequency-specific

interference of ASN with experimental stimuli (Langers et al., 2005; Novitski et al., 2006). Because studies on tonotopy rely upon the assessment of frequency-dependent response characteristics in order to extract a map of best frequencies, it is therefore highly plausible that tonotopic mapping outcomes are affected by the presence of ASN.

In the present study, we aimed to compare tonotopic maps derived using several distinct imaging protocols. We designed three acquisition sequences that we deemed reasonable alternatives based on current evidence and existing practices: (1) a “sparse” protocol with moderate-to-long silent intervals between pairs of contiguous acquisitions; (2) a “clustered” protocol with gaps between successive acquisitions that were long enough to avoid direct acoustic masking but still short compared to hemodynamic timescales; and (3) a “continuous” protocol in which acquisitions were carried out without any intermittent silent periods. All other imaging parameters, including the total measurement time, were kept constant across the three protocols. The stimuli were also the same: like many previous studies, we used an iterated sweep stimulus, the activation to which travels across the tonotopic map in periodic fashion (Talavage et al., 2004). Our subsequent analysis focussed on quantitative differences in the frequency-dependent sound-evoked responses, and qualitative differences in the layout of the extracted tonotopic maps.

Methods

Seven healthy volunteers (gender: 3♂, 4♀; age [years]: 33.6 ± 5.5 , range 26–42) participated in this study after having given written informed consent. The procedures accorded with the Declaration of Helsinki and were approved by the Research Ethics Committee of the School of Medicine at the University of Nottingham. Subjects had no history of neurological or hearing-related disease.

Imaging session

Subjects were positioned supinely in the bore of a 3.0-T MR system (Philips Achieva, Best, the Netherlands) that was equipped with a 32-channel receive head coil. The scanner coolant pump was turned off during all measurements and subjects wore foam ear plugs and MR-compatible electrostatic headphones (NordicNeuroLab AudioSystem, Bergen, Norway) to diminish ambient noise levels. After the acquisition of an anatomical reference scan, subjects performed an automated audiometric test in situ while no scanning took place to determine their hearing thresholds. Results were used to calibrate the stimulus delivery in subsequent functional runs. For the remainder of the session, subjects watched a silent video (a nature documentary).

The functional imaging session comprised six 6-minute runs, each consisting of a sequence of high-resolution T_2^* -sensitive 2-D gradient-echo echo-planar imaging (EPI) volume acquisitions (2.2-s acquisition time; 40-ms echo time; 90° flip angle; $128 \times 126 \times 25$ matrix; $192 \times 168 \times 37.5\text{-mm}^3$ field of view; $1.5 \times 1.5 \times 1.5\text{-mm}^3$ reconstructed voxel size; 0-mm slice gap; EPI-factor 53). The acquisition volume was positioned in an oblique axial orientation, tilted forward parallel to the Sylvian fissure, and approximately centred on the superior temporal plane. Regional saturation slabs were added to null the signal from the eyes. Three different acquisition protocols were employed that differed with respect to their

repetition times (see Fig. 1). During *sparse* scanning, 26 pairs of EPI images (labelled “a” and “b” in Fig. 1) were acquired separated by 8.8-s periods of scanner inactivity (alternatingly 2.2-s or 11.0-s repetition time). During *clustered* scanning, 78 EPI images were acquired separated by 2.2-s gaps of scanner inactivity (4.4-s repetition time). During *continuous* scanning, 156 EPI images were acquired contiguously (2.2-s repetition time). Preparation scans were acquired to achieve stable image contrast and to trigger the start of stimulus delivery. These were excluded from the analysis.

The stimulus paradigm presented in each run consisted of a sequence of upward or downward tone sweeps. Each sweep comprised a sequence of 242 pure tones, presented at a rate of 10 per second (i.e., total duration equivalent to 11 contiguous acquisitions). Each tone lasted 75 ms, with 5-ms cosine ramps. At any given time, the tone frequencies were randomly selected from an interval that dynamically swept from 125–177 to 5657–8000 Hz for the upward sweep, or in the reverse direction for the downward sweep. In other words, tone frequencies spanned a total range of 6 octaves (125–8000 Hz) and were jittered over 1/2 octave while the frequency contour progressed continuously at a rate of 1/2 octave per 2.2 s. In order to obtain constant loudness, each tone was presented diotically at a sensation level (i.e., level above the individually determined detection threshold) corresponding to a loudness level of 60 phon, as defined in the international ISO-226 standard (Suzuki et al., 2003). To reduce startle due to sudden sweep on- and offsets, levels were faded in and out during the first and last 1.1 s of the sweep. Finally, each sweep was padded with 2.2 s of silence on each end, resulting in a stimulus repetition period of 28.6 s. The stimulus was repeated twelve times per run (see Fig. 1).

Thus, six runs were obtained, combining three acquisition protocols (*sparse*, *clustered* and *continuous* imaging) with each of two stimulus protocols (*upward* and *downward* sweeps). The ordering of these runs was randomised across subjects.

Acoustics measurements

The spectrum and intensity level of the ASN were measured within the scanner bore (i.e. unattenuated by headphones or earplugs) using a type-4143 microphone and type-2669 pre-amplifier (Brüel & Kjær, Nærum, Denmark). These were placed next to a phantom in the head coil, near where the ear of a participant would be. The microphone output was calibrated using a type-4230 calibrator (Brüel & Kjær, Nærum, Denmark). Signals were recorded and analysed using a 51.2-kHz sampling rate and 32-bit precision by means of an Apollo-box sound level meter (Sinus Messtechnik GmbH, Leipzig, Germany).

Sound waveforms of approximately 30-s duration were recorded for the sparse, clustered, and continuous imaging protocols. Minimum, maximum, and equivalent sound intensity levels were determined using linear and A weightings with an integration time of 1.0 s.

Data analysis

Data were preprocessed using the SPM12 software package (Wellcome Department of Imaging Neuroscience, <http://www.fil.ion.ucl.ac.uk/spm/>) (Friston et al., 2007). Functional imaging volumes were corrected for motion effects using rigid body transformations and coregistered to the subject's anatomical image. A logarithmic transformation was carried out

in order to express all derived voxel signal measures in fMRI units of percentage signal change relative to the mean, and images were moderately smoothed by convolution with a 5-mm full-width at half-maximum (FWHM) Gaussian kernel. The two images of each pair in the sparse runs (“a” and “b”, see above) were averaged using an equal weighting. This was determined to optimise the signal-to-noise ratio while accounting for differences in image contrast and activation level (see the Results section). The resulting pairwise-averaged images were treated as if they were obtained from a sequence with fixed 13.2-s repetition time. The anatomical images were segmented and all images were normalised and resampled at 1-mm resolution inside a bounding box of $x = -75 \dots +75$, $y = -60 \dots +40$, $z = -20 \dots +30$ in Montreal Neurological Institute (MNI) stereotaxic space. Cortical surface meshes were generated from the anatomical images using the standard processing pipeline of the FreeSurfer v5.1.0 software package (Martinos Center for Biomedical Imaging, <http://surfer.nmr.mgh.harvard.edu/>) (Dale et al., 1999).

Linear regression models were formulated for each of the three acquisition protocols. Individual subject models as well as fixed-effects group models were evaluated. All models comprised three regressors for each of the runs. These included a constant offset to model the baseline signal, plus a cosine and a sine function to model a sinusoidal response to the sweep stimuli with arbitrary amplitude and phase (Fig. 1). The functions were discretised by temporal sampling at the middle of each acquisition for the clustered and continuous protocols, and the middle of each acquisition pair for the sparse protocol. For the runs with downward sweeps, the sine regressor's sign was flipped compared to that of those with upward sweeps in order to reverse the modelled phase. The cosine and sine regression coefficients, averaged across runs, were converted to polar coordinates to determine the amplitude (i.e. modulus $A = \sqrt{(\beta_{cos}^2 + \beta_{sin}^2)}$) and phase (i.e. argument $\varphi = \arctan(\beta_{sin} / \beta_{cos})$) of the response, which are held to reflect the activation level and frequency tuning, respectively.

Results

Acoustic measurements

For the continuous acquisition protocol, the intensity level of the ASN in the scanner bore was 101.8 dB SPL on average, and ranged between 101.1 and 104.1 dB SPL at different time points. The A-weighted intensity level equalled 99.1 dB(A). The sound spectrum is shown in Fig. 2. The fundamental frequency of the ASN, determined by the switching frequency of the gradient coils, was equal to 423 Hz. The spectrum was highly dominated by the fundamental component, which contained approximately half of the entire sound energy. The second largest component was the third harmonic, followed by the second and fifth harmonics. In addition, sub-harmonics were observed corresponding with the 11.4 Hz rate of slice readouts (i.e. 25 slices in 2.2 s). Approximately two-thirds of the sound energy were contained within a frequency range of 400–1000 Hz. Negligible energy occurred above the fifth harmonic (~2.1 kHz).

For the clustered and sparse protocols, sound spectra were identical with regard to the periods when the scanner was active (not shown).

Activation and tuning maps

Fig. 3a shows maximum intensity projections of the significant group-level activations according to each of the three acquisition protocols, thresholded at a voxelwise confidence level $p < 0.001$ and cluster size $k_E > 1.0 \text{ cm}^3$. For the sparse protocol, data were analysed on the basis of only the first image of each pair (labelled “a”), only the second image (labelled “b”), or on the basis of the pairwise average. Both the first and second images yielded activation confined to the bilateral auditory cortices. The first image proved moderately less sensitive to sound-evoked activation than the second: the total volume of supra-threshold activation amounted to 16.0 and 19.4 cm^3 for “a” and “b”, respectively. The pairwise averaging improved sensitivity further (23.2 cm^3). The clustered protocol yielded similarly extensive activation (23.4 cm^3) as the sparse protocol, whereas the continuous protocol yielded the least extensive activation (17.9 cm^3). A single ROI was subsequently defined by merging the pairwise-averaged sparse, the clustered, and the continuous activation clusters (27.4 cm^3).

In order to assess the sensitivity of the different acquisition protocols, we averaged the voxel-wise activation levels, A (in percent signal change), and the Fisher–Snedecor F -values across the ROI. The first image in the sparse protocol resulted in $A = 0.11\%$ and $F = 12.7$; for the second image, $A = 0.17\%$ and $F = 19.6$. This implies that an optimal linear combination of the images would be achieved if the images were weighted as “a”:“b” = 0.49:0.51 (see the Appendix A), which is very close to the equal weighting that we employed in our preprocessing. Using that equal weighting, the pairwise-averaged image yielded $A = 0.14\%$ and $F = 26.0$ for the sparse protocol, compared to $A = 0.12\%$ and $F = 22.0$ for the clustered protocol and $A = 0.10\%$ and $F = 16.2$ for the continuous protocol. Henceforth, all reported outcomes for the sparse protocol will be based on the pairwise-averaged data.

Fig. 3b compares the subject-level activations for the sparse, clustered, and continuous protocols. For these individual activation maps, the cluster extent threshold was lowered to $k_E > 0.1 \text{ cm}^3$. Unsurprisingly, activation extents varied across subjects, but this variability tended to be similar across the three acquisition protocols; that is, strongly responsive subjects for one protocol remained strongly responsive for other protocols as well. Individual activations tended to be more extensive and less specifically confined to auditory cortex for the sparse protocol than for the other two protocols.

The regression coefficients of the cosine and sine regressors were converted to response amplitudes, A , and response phases, φ , which are mapped in Fig. 4 for the group-level analyses, and in Fig. 5 for the subject-level analyses. Although the spatial pattern of response amplitudes (Figs. 4a and 5a) was similar between the three acquisition protocols, responses were typically largest for the sparse protocol, intermediate for the clustered protocol, and smallest for the continuous protocol. The group-level activation peaked in lateral Heschl's gyrus (HG), but individual subjects often showed multiple elongated activation clusters running more or less parallel along the banks of HG. The fact that these patterns were reproducible across the three protocols suggests that they reflect true neural or vascular physiological characteristics.

At the group level, the spatial distributions of the response phases, indicating tonotopic tuning, were qualitatively highly similar across protocols (Fig. 4b). Small phase values (i.e., low-frequency preferences; 90° coincides with ~ 0.3 kHz) were found extensively in the lateral auditory cortex. This was especially the case for lateral HG, but additional phase minima appeared to exist on lateral planum polare (PP) and lateral planum temporale (PT) near the rostrolateral and caudolateral borders of the ROI, respectively. Conversely, large phase values (i.e., high-frequency preferences; 270° coincides with ~ 3.4 kHz) occurred in a more confined region on medial HG. Two distinct phase maxima were observed on the rostromedial and caudomedial aspects of HG, separated by a region with lower phases on HG's axis. Thus, up to three lateral low-frequency regions and two medial high-frequency regions appeared to form an interdigitated pattern, with iso-frequency contours as well as tonotopic progressions zig-zagging in between. Although individual subjects showed more detailed features and substantial small-scale variability (Fig. 5b), similar overall patterns could be identified in most subjects (with the exception of the phase map of subject S1 according to the sparse protocol).

Comparisons across acquisition protocols

The distribution of the group-level responses is detailed further in Fig. 6. The amplitudes and phases of all voxels in the ROI are plotted in a polar plot; equivalently, the coefficients of the cosine and sine regressors correspond with the Cartesian coordinates in these points. For all protocols, the low-frequency voxels (in red) tended to attain larger amplitudes than the high-frequency voxels (in blue). Still, some noteworthy differences existed between the distributions for the three protocols. For moderately low frequencies (~ 0.7 kHz, corresponding with $\sim 150^\circ$), response amplitudes were markedly larger according to the sparse protocol than according to either the clustered or the continuous protocols. Similar but smaller differences extended up to ~ 1.6 kHz ($\sim 220^\circ$). Conversely, the response amplitudes at extreme tuning frequencies below ~ 0.5 kHz ($\sim 120^\circ$) or above ~ 2.1 kHz ($\sim 240^\circ$) were found to be largest in the clustered protocol, also large in the continuous protocol, but smaller in the sparse protocol.

A more direct comparison of the voxel-wise response amplitudes and phases across the three acquisition protocols is made in Fig. 7. The graphs in the lower left of panels (a) and (b) show the amplitude and phase histograms, respectively, for the three protocols. The scatter plots show pairwise comparisons of the amplitudes and phases between each two protocols. With regard to the response amplitude (Fig. 7a), all three protocols yielded results that were highly correlated. Nevertheless, the correlation between the clustered and continuous protocols was higher ($R > 0.9$) than both correlations involving the sparse protocol ($R < 0.9$). Furthermore, the sparse protocol typically showed the strongest responses, followed by the clustered protocol, and finally the continuous protocol. With regard to the phases (Fig. 7b), a more complicated relationship was observed. In the histogram for the sparse protocol, a majority of voxels was tuned to low frequencies, followed by an extended, almost uniform tail of higher-frequency preference. For the clustered and continuous protocol, the low-frequency mode in the distribution progressively diminished and shifted towards lower frequencies, while a secondary peak arose at high frequencies. The pairwise comparisons again showed a high correlation between the different protocols, but a non-linear

relationship was apparent: in particular, the data points at intermediate frequencies (i.e., in the middle of the cloud) were shifted towards higher frequencies for the clustered and continuous compared to the sparse protocol, and also for the continuous compared to the clustered protocol. Again, the clustered vs. continuous correlation ($R > 0.9$) exceeded both of those involving the sparse protocol ($R < 0.9$), suggesting that out of all three protocols the sparse protocol behaved most differently from the other two. These observations indicate that, compared to the sparse protocol, tonotopic representations of moderately low frequencies were diminished in favour of high frequencies in the clustered protocol, and even more so in the continuous protocol.

Effects of sweep direction

In an effort to study the response phases in more detail, we determined phase maps based on the upward sweeps alone, or based on the downward sweeps alone. Fig. 8a shows the upward and downward phases plotted against one another for each of the three acquisition protocols. The diagonal grid quantifies the average phases and the phase differences between the two sweep directions. The upward and downward phases were found to be highly correlated. Yet, the upward sweep was found to result in a systematically delayed phase compared to the downward sweep. The average phase difference between the upward and downward sweep directions equalled $\varphi = 104^\circ$, 96° , and 108° for the sparse, clustered, and continuous protocols, respectively. This can be accounted for by assuming hemodynamic response delays of $\tau = 4.1$, 3.8 , and 4.3 s.

Finally, we visualised the underlying signal dynamics in relation to the presented sweep stimuli. The previously defined ROI was subdivided into six sub-ROIs corresponding to regions of different frequency tuning. For that purpose, we averaged the group-level phase maps for all three protocols (shown in Fig. 4b) and classified all voxels into six 30° -bins between 90° and 270° ; remaining voxels with mean phases of $0-90^\circ$ were included in the $90-120^\circ$ -bin, and voxels with mean phases of $270-360^\circ$ were included in the $240-270^\circ$ -bin. Fig. 8b shows the hemodynamic signals of each of these six sub-ROIs, normalised to unit variance to facilitate their comparison and averaged across all subjects in the group and across all sweeps in a run, but separated according to the direction of the sweep and according to the acquisition protocol. Because the sweep duration (13×2.2 s) and the period of the acquisition protocols (6×2.2 s) were chosen to be incommensurate multiples of the acquisition time (see Methods), all signals were effectively sampled at 2.2-s temporal resolution when folded into a single period.

The response dynamics for the sparse protocol appeared to contain more short-term oscillations than for the clustered or continuous protocols, which can likely be attributed to the lower sampling frequency and therefore noisier averages and larger aliasing effects. Disregarding this difference, the protocols resulted in comparable signals. As expected, the hemodynamic response in low-frequency voxels (red curves) started to grow almost immediately after the onset of the upward sweep, reaching a maximum after approximately 9 s, and then slowly declined back to baseline in a quasi-monotonic fashion. In response to the downward sweep, these voxels showed a reversed response, starting with a slow build-up to a maximum, followed by a sharp drop back to baseline after the end of the sweep. High-

frequency voxels (blue curves) displayed a similar sharp growth after the onset of the downward sweep, and a similar fast decline after the offset of the upward sweep. However, these voxels did not show the same gradual decline or growth over the course of the sweep; instead they appeared to show a small secondary peak near the onset of the upward sweep and towards the end of the downward sweep, creating a response minimum near the midpoint of the sweep. This behaviour was clearly visible in the voxels tuned to moderate frequencies as well (green curves). Their hemodynamic response did not peak near the middle of the sweep, as might have been expected from the fact that this coincided with the presentation of moderate tone frequencies. Instead, they showed a bimodal response with more or less equal peaks after the onset and near the offset of the sweep. This behaviour is perhaps most dramatically displayed in the continuously acquired data obtained from the upward sweeps, but evidence for such behaviour occurred for all three acquisition protocols and either sweep direction. In summary, voxels did not show a single response peak that was positioned along the sweep according to their best frequency, but their responses tended to consist of two peaks for which the relative amplitudes varied according to their assigned frequency tuning.

Discussion

Sound-evoked activation

We compared three different acquisition protocols to study the influence of ASN on the outcomes of a tonotopic mapping experiment. Continuous scanning is the default protocol in experiments that do not involve sound stimulation. In the context of studies on audition, it may be used when stimuli can be presented at such loud levels that the ASN does not substantially affect their audibility, in particular when the focus of the study is not towards primary auditory cortex or subcortical auditory nuclei. Compared to the continuous scanning protocol, a clustered protocol offers the advantage that at least during the short silent periods between scans auditory stimuli can be perceived with minimal direct acoustic interference. This allows tasks to be performed and responses to be assessed in relation to stimuli presented at normal loudness levels. However, because the silent periods are short compared to the hemodynamic response duration, a sustained elevation of the fMRI baseline signal due to previous scans should be expected. Insofar as response saturation due to hemodynamic non-linearities plays a role, this will decrease the measurable amplitude of stimulus-evoked activation and reduce the dynamic range for discerning the difference between various contrasting conditions. Finally, sparse scanning paradigms employ intermittent silent periods that are so long that responses to previous scans have largely decayed back to baseline. These allow the “cleanest” brain responses to be measured, since they avoid interactions at the hemodynamic level. However, this comes at the cost of acquiring substantially reduced amounts of data per unit time.

In an attempt to partially overcome the limitations of sparse scanning, we employed moderately long (8.8-s) intervals of scanner inactivity and acquired pairs of clustered acquisitions in this study. Although longer silent intervals of up to 20 s have been advocated in order for ASN effects to become negligible (Talavage et al., 1999; Olulade et al., 2011), intervals in the order of 10 s appear to be in common usage. The acquisition of multiple

image volumes in a cluster is a less common practice, although it has been proposed and successfully employed by multiple authors (Schmithorst and Holland, 2004; Langers et al., 2007; Zaehle et al., 2007). In our case, the acquisition of a second image (labelled “b” in Fig. 1) came at a small cost, by extending the periodicity of the acquisition paradigm from 11.0 to 13.2 s. At the same time, a more traditional data analysis could still have been carried out by considering only the first images (labelled “a”) that have similar image contrast as images from a standard sparse sequence due to their long repetition time. By averaging the image pairs, we were able to avoid irregular serial correlations due to the variable repetition time. The observation that the obtained mean F -values in the ROI substantially increased from 12.7 and 19.6 for the individual images to $F = 26.0$ for the pairwise averages indicates that the approach proved beneficial.

With regard to the sensitivity to sound-evoked brain activation, we found that the continuous protocol as well as the sparse protocol based on single images (i.e. only the first or second in a pair) performed worse than the clustered protocol and the sparse protocol based on both images in the pair. For the former protocols, the group-level activation extent was between 15 and 20 cm³, whereas the latter protocols detected larger volumes of between 20 and 25 cm³ of significantly responsive brain tissue. Our results are in agreement with the finding in literature that responses to auditory stimuli are significantly diminished in the presence of ASN (Talavage and Edmister, 2004). They indicate that the insertion of silent periods improves the sensitivity of auditory fMRI experiments. The duration of the silent periods did not prove critical, given that the 2.2-s periods in the clustered protocol provided similar sensitivity as the 8.8-s periods in the sparse protocol. However, our data do not support previous suggestions that continuous scanning might be preferable due to the absence of sudden ASN onsets (Seifritz et al., 2006), at least for the types of stimuli and intensity levels in combination with the scanner and imaging sequence employed in the present experiment (60-phon tones with a 3-T EPI sequence).

Tonotopic maps

To further investigate whether the tonotopic organisation appears distorted in the presence of ASN due to frequency-dependent interactions, we generated tonotopic maps. In contrast to characteristic-frequency maps that are generated based on invasive neurophysiological measurements and that reflect exact quantitative measures of tuning in individual neurons or small neuronal assemblies, the best-frequency maps that are derived from non-invasive fMRI data should be regarded as qualitative correlates of tuning in mesoscopic neural populations. In our experience, although voxels differ from each other in their tuning characteristics, most voxels tend to show a rather broad response to a large range of frequencies, certainly for fMRI at or below 3 T. This may partly be due to the fact that a single voxel comprises numerous cortical columns, or to the fact that fMRI is sensitive to hemodynamic effects that originate from an extensive patch of cortex (Turner, 2002). Consequently, even voxels that are primarily tuned to – say – the highest frequency of 8 kHz in our experiment still tend to show substantially non-zero responses to the lowest frequency of 125 Hz as well. When fitting the response to a frequency sweep, the effective frequency that is assigned to such a voxel will therefore tend to be shifted towards the moderate frequency values that correspond with the middle of the sweep. As a result, extreme frequency tuning was not

notably found in our results (e.g. the 0–90° and 270–0° quadrants in Fig. 6, or the equivalent ranges in the histogram in Fig. 7b), whereas neurons tuned to the extremest frequencies represented in our sweep stimuli should still be expected to be abundant. Hence, we caution against a direct interpretation of the observed Fourier phase in terms of an exact frequency scale. In particular, the outcome likely depends on the characteristics of the sweep stimuli (e.g. frequency range, duration, sweep rate, silent periods). Although this complicates comparisons across studies with different stimuli, it nevertheless seems justified to compare detailed results across paradigms with identical stimuli, like we do in our study.

Qualitatively, the extracted maps of preferred frequencies were highly similar for the three scanning protocols. In particular, all protocols revealed low-frequency representations on the lateral aspect of the superior surface of the temporal lobe, and high-frequency representations more medially deep within the Sylvian fissure. Multiple extrema could be distinguished along the anterior-to-posterior direction, resulting in a typical zig-zag pattern of reversing tonotopic gradients. This pattern agrees well with previous literature (Saenz and Langers, 2014). This observation suggests that, with regard to the delineation of tonotopic gradients and corresponding fields in auditory cortex, some interference from ASN is acceptable.

For various applications, more exact quantitative measures may be of interest. With regard to cortical reorganisation, for instance, one does not necessarily expect the number or relative location of cortical fields to change. However, one might compare the extent of cortical representations of particular sound frequencies across distinct subjects groups, for instance the edge-frequency in subjects with high-frequency hearing loss or the frequency corresponding with the tinnitus pitch in tinnitus patients. Our results show that such measures are notably distorted by interference due to ASN. Therefore, the use of additional measures to avoid detrimental influences from ASN seems essential in such contexts.

We found that the response amplitude to moderate frequencies was decreased for the clustered acquisition protocol compared to the sparse protocol, and progressively more so for the continuous protocol. In contrast, the responses in voxels that were found to be tuned to extreme (low or high) frequencies were simultaneously increased. The various employed scanning protocols differed in the repetition time of the acquisitions. This has an effect on the tissue contrast and signal-to-noise of the obtained images, but it is difficult to envisage how this would result in an effect that depends upon the frequency of the stimulus presentations. In our view, the only plausible explanation for the observed differences is related to ASN. The moderate stimulus frequencies near the middle of a sweep coincided with the regions in the sound spectrum of the ASN that contained high spectral power and that were dominated by strong harmonic peaks. If neurons tuned to such moderate frequencies adapt to the ongoing presence of sound, or if hemodynamic signals in tonotopic regions corresponding with those frequencies saturate, then evoked responses to the sound stimuli of interest will decrease. At the same time, neurons may become relatively more sensitive to frequencies that are not present in the ASN due to an absence of adaptation at those frequencies, possibly supplemented by adaptation to inhibitory input from side-bands that do coincide with ongoing ASN. Additionally, hemodynamic baseline levels may conceivably decrease in the lowest- and highest-frequency tonotopic endpoints due to blood

stealing effects, which might further increase the available dynamic response range at extreme frequencies. These neural and vascular mechanisms may explain the apparent increase in responses to stimulus frequencies that were weakly represented in the ASN.

Overall, we conclude that although ASN has only minor influence on the qualitative layout of tonotopic maps, it does interfere with quantitative measures of frequency tuning. Out of the several protocols that we compared, we advocate the use of sparse scanning with multiple contiguous acquisitions because it provides similar sensitivity as the clustered protocol, more sensitivity than the continuous protocol, and it best avoids the obvious frequency-specific interference from ASN. Moreover, sparse scanning has the advantage that it allows the use of softer stimuli than those employed in this study, which is likely to improve the stimulus-specificity of the measured responses (Langers and van Dijk, 2012).

Travelling wave responses

Finally, we discuss an unexpected observation regarding the temporal dynamics of the responses to the employed tone sweeps. Travelling wave stimuli have been used for vision (Engel et al., 1997), touch (Sanchez-Panchuelo et al., 2010), and audition (Talavage et al., 2004). They allow the best-frequency tuning to be encoded as the phase of the response, which can easily be extracted by correlation, regression, or Fourier analysis. By averaging results over both sweep directions, hemodynamic delays can be cancelled without explicit measurement of the hemodynamic response function. As far as we can oversee, our outcomes using this protocol are generally consistent with previous results: not only did the extracted tonotopic maps agree with present literature (Humphries et al., 2010; Da Costa et al., 2011; Striem-Amit et al., 2011; Herdener et al., 2013; Langers, 2014), the inferred 4-s hemodynamic delay also well matches other reports concerning auditory cortex (Backes and van Dijk, 2002; Inan et al., 2004; Langers et al., 2005; Olulade et al., 2011).

Nevertheless, we found that the hemodynamic signals did not behave as expected. In particular, we found that responses in voxels that were assigned a moderate best frequency did not peak in the middle of the sweeps, but instead tended to show two peaks, one in the first half and one in the second half of the sweep, often accompanied by an apparent dip coinciding with the presentation of moderate frequencies. It is arguable whether this behaviour should be interpreted as “moderate tuning”, as the phase of the fitted sinusoid suggests. Overall, the range of behaviours was better summarised by an occurrence of two peaks, one corresponding to low frequencies and another corresponding to high frequencies; the assignment of a best frequency appeared to depend mostly on the relative strength of these peaks. To our knowledge, such bimodal response behaviour has not previously been reported in the context of travelling wave designs. We are unable to ascertain whether our data diverge from those in other studies in this respect, or whether other authors simply have not investigated the underlying signal dynamics.

At first sight, the observed dip may seem to be another effect of ASN. The dominant frequencies in the ASN corresponded with those presented in the middle of the sweep, so it might have been expected that ASN would particularly diminish response levels at that instance in time. However, a similar effect was observed irrespective of the scanning protocol. In particular, for the sparse scanning protocol, moderately-tuned voxels did not

show a unimodal peak near the middle of the sweep either. For that reason, we find it unlikely that the observed behaviour is fully accounted for by ASN.

One might conceivably interpret these bimodal peaks to be related to stimulus onset and offset responses. Indeed, it has been shown that the response at a cortical level is dominated by transients, rather than by sustained responses as they occur in subcortical nuclei (Harms and Melcher, 2002, 2003). However, this explanation is not appropriate in our case, given that the second peak in the response started well before the end of the sweep. Moreover, the sweep tones were gradually faded in and out, and in addition to the onset and offset of the sweep as a whole individual tone onsets and offsets occurred throughout the stimulus, rendering this explanation less persuasive. Yet, it is likely that the strongest onset responses occurred at the start of a sweep, following the period of silence between sweeps. Conceivably, that onset activity may have depleted the brain tissue to some extent such that neurons that are tuned to moderate frequencies were less responsive at the moment that their preferred frequencies were presented, whereas neurons tuned to frequencies that were presented towards the end of the sweep had more time to recover. Such a slow recovery behaviour was reported by Harms and Melcher (2003) in the form of a ramp function. However, given the long duration of the sweeps (24.2 s), all neurons should have had sufficient time to recover from sweep onset activity, including those that were presented with their preferred frequencies after approximately 10 s in the middle of the sweep. Therefore, although we do not completely discard this explanation, it seems implausible.

A different straightforward interpretation would be that there are two distinct populations of neurons: one tuned to low frequencies, explaining one peak, and one tuned to high frequencies, explaining the other. Neurons tuned to moderate frequencies might simply be absent, or underrepresented, in such a scenario. Different regions of the auditory cortex might contain different mixtures of these two types of neurons, resulting in the observed dynamics. Alternatively, the spatial location of regions strictly composed of one subpopulation or the other might vary across subjects, similarly resulting in continuously variable contributions to the dynamics when considering voxel-wise averages at the group level. In fact, it has been previously suggested that apparent continuous tonotopic gradients might actually arise from regions with discretely differing frequency tuning (Schönwiesner et al., 2002). Although this explanation is compatible with our observations, such a model would be very difficult to reconcile with findings from numerous animal studies that report an abundance of neurons that are tuned to moderate frequencies in a variety of species (Merzenich et al., 1976; Hellweg et al., 1977; Stiebler et al., 1997; Capsius and Leppelsack, 1999; Kusmirek and Rauschecker, 2009; Scott et al., 2011). Yet, it could be that neurons that are tuned to lower frequencies than 125 Hz or higher frequencies than 8 kHz respond maximally at the edges of the sweep, or perhaps even shift their best frequencies centripetally towards more moderate frequencies due to the stimulus statistics in the experiment (Suga, 2012). Thus, population tuning could accumulate at the extreme frequencies. Although this makes the hypothesis that neurons can roughly be subdivided into two subpopulations more plausible, it remains hard to envisage that this would completely explain our observation given that we employed a rather large range of frequencies (six octaves) that should encompass the best frequency of a majority of neurons.

Interestingly, comparable effects have been reported in the context of vision. There, evidence suggests that phase encoding techniques give biased estimates of topographic maps near retinotopic boundaries related to masks in visual field stimuli (Haak et al., 2012). It has been forwarded that this bias arises in the analysis as a consequence of failing to model the absence of stimulation in parts of the visual field (Binda et al., 2013). Translating this to our auditory paradigm, it may well be that similar effects led to an overrepresentation of the responses at the edges of the sweep stimuli. Because we do not observe a similarly notable dip at moderate frequencies when employing stimuli presented in a block-design instead of sweeps (Langers et al., 2014-in this issue), the problem appears to be specific for the travelling wave paradigm. We therefore tentatively attribute the dip to response dynamics rather than frequency-tuning.

Despite the fact that the reasons for the observed bimodal response behaviour remain poorly understood, it seems safe to conclude that travelling wave stimuli in the auditory modality may suffer from some distortion in the derived tonotopic maps. Although we do not expect this effect to render our results regarding the effects of ASN to be invalidated, we wish to point out that sweep stimuli potentially confuse response characteristics as a function of frequency and hemodynamics as a function of time. In our view, this deserves further investigation, and we therefore recommend that authors who employ travelling wave stimuli report on the underlying evoked signal dynamics as well, instead of presenting tonotopic maps based on fitted response phases alone.

Acknowledgments

The authors would like to acknowledge Prof. Richard Bowtell, who contributed to the experimental design and provided valuable input during discussions, and Dr. Julien Besle who in addition to that performed the acoustic measurements.

Appendix A

In an analysis of variance (ANOVA), a measured signal mean is tested against the null-hypothesis by computing a Fisher-Snedecor F -value that equals the explained variance of interest (per degree of freedom) divided by the unexplained variance (per degree of freedom): $F = m^2 / s^2$, where m and s^2 are the signal's estimated mean and residual variance, respectively.

If two independent signals y_1 and y_2 have means m_1 and m_2 and variances s_1^2 and s_2^2 , then a weighted linear combination $y = b_1 \cdot y_1 + b_2 \cdot y_2$ will have mean $m = b_1 \cdot m_1 + b_2 \cdot m_2$ and variance $s^2 = b_1^2 \cdot s_1^2 + b_2^2 \cdot s_2^2$. By optimising the signal to noise ratio of the combined signal using the parameters b_1 and b_2 , it can be shown that a maximum equal to $F = F_1 + F_2$ can be obtained for weights b_1 and b_2 that relate as

$$b_1 : b_2 = \frac{F_1}{m_1} : \frac{F_2}{m_2}.$$

More generally, the vector of optimal weights \mathbf{b} can be obtained through $\mathbf{b} = \mathbf{S}^{-1} \mathbf{m}$, where \mathbf{m} is the vector of estimated means and \mathbf{S} equals the estimated covariance matrix of signal

noises. For this choice of weights, $F = \mathbf{m}^T \mathbf{S}^{-1} \mathbf{m}$ is achieved. This generalisation allows for an arbitrary number of signals and non-zero covariances between signals.

References

- Amaro E, Williams SCR, Shergill SS, Fu CHY, MacSweeney M, Picchioni MM, Brammer MJ, McGuire PK. Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. *J Magn Reson Imaging*. 2002; 16:497–510. [PubMed: 12412026]
- Backes WH, van Dijk P. Simultaneous sampling of event-related BOLD responses in auditory cortex and brainstem. *Magn Reson Med*. 2002; 47:90–96. [PubMed: 11754447]
- Bandettini PA, Jesmanowicz A, Van Kylen J, Birn RM, Hyde JS. Functional MRI of brain activation induced by scanner acoustic noise. *Magn Reson Med*. 1998; 39:410–416. [PubMed: 9498597]
- Baumann S, Petkov CI, Griffiths TD. A unified framework for the organization of the primate auditory cortex. *Front Syst Neurosci*. 2013; 7:11. [PubMed: 23641203]
- Binda P, Thomas JM, Boynton GM, Fine I. Minimizing biases in estimating the reorganization of human visual areas with BOLD retinotopic mapping. *J Vis*. 2013; 13:13.
- Blackman GA, Hall DA. Reducing the effects of background noise during auditory functional magnetic resonance imaging of speech processing: qualitative and quantitative comparisons between two image acquisition schemes and noise cancellation. *J Speech Lang Hear Res*. 2011; 54:693–704. [PubMed: 20844253]
- Bowtell RW, Mansfield P. Quiet transverse gradient coils: Lorentz force balanced designs using geometrical similitude. *Magn Reson Med*. 1995; 34:494–497. [PubMed: 7500892]
- Boyle Y, Bentley DE, Watson A, Jones AKP. Acoustic noise in functional magnetic resonance imaging reduces pain unpleasantness ratings. *Neuroimage*. 2006; 31:1278–1283. [PubMed: 16517183]
- Capsius B, Leppelsack H. Response patterns and their relationship to frequency analysis in auditory forebrain centers of a songbird. *Hear Res*. 1999; 136:91–99. [PubMed: 10511628]
- Chapman BLW, Haywood B, Mansfield P. Optimized gradient pulse for use with EPI employing active acoustic control. *Magn Reson Med*. 2003; 50:931–935. [PubMed: 14587003]
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M. Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci*. 2011; 31:14067–14075. [PubMed: 21976491]
- Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*. 1999; 9:179–194. [PubMed: 9931268]
- De Zwart JA, van Gelderen P, Kellman P, Duyn JH. Reduction of gradient acoustic noise in MRI using SENSE-EPI. *Neuroimage*. 2002; 16:1151–1155. [PubMed: 12202101]
- Dick F, Taylor Tierney A, Lutti A, Josephs O, Sereno MI, Weiskopf N. In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *J Neurosci*. 2012; 32:16095–16105. [PubMed: 23152594]
- Eden GF, Joseph JE, Brown HE, Brown CP, Zeffiro TA. Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. *Magn Reson Med*. 1999; 41:13–20. [PubMed: 10025606]
- Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM. Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp*. 1999; 7:89–97. [PubMed: 9950066]
- Elliott MR, Bowtell RW, Morris PG. The effect of scanner sound in visual, motor, and auditory functional MRI. *Magn Reson Med*. 1999; 41:1230–1235. [PubMed: 10371456]
- Engel SA, Glover GH, Wandell BA. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex*. 1997; 7:181–192. [PubMed: 9087826]
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*. 2003; 40:859–869. [PubMed: 14622588]
- Foster JR, Hall DA, Summerfield AQ, Palmer AR, Bowtell RW. Sound-level measurements and calculations of safe noise dosage during EPI at 3 T. *J Magn Reson Imaging*. 2000; 12:157–163. [PubMed: 10931575]

- Friston, KJ., Ashburner, JT., Kiebel, SJ., Nichols, TE., Penny, WD. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. 1st ed. Academic Press; 2007.
- Fuchino Y, Sato H, Maki A, Yamamoto Y, Katura T, Obata A, Koizumi H, Yoro T. Effect of fMRI acoustic noise on sensorimotor activation examined using optical topography. *Neuroimage*. 2006; 32:771–777. [PubMed: 16829140]
- Gaab N, Gabrieli JDE, Glover GH. Resting in peace or noise: scanner background noise suppresses default-mode network. *Hum Brain Mapp*. 2008; 29:858–867. [PubMed: 18454447]
- Haak KV, Cornelissen FW, Morland AB. Population receptive field dynamics in human visual cortex. *PLoS ONE*. 2012; 7:e37686. [PubMed: 22649551]
- Hackett TA, Preuss TM, Kaas JH. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol*. 2001; 441:197–222. [PubMed: 11745645]
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp*. 1999; 7:213–223. [PubMed: 10194620]
- Hall DA, Summerfield AQ, Gonçalves MS, Foster JR, Palmer AR, Bowtell RW. Time-course of the auditory BOLD response to scanner noise. *Magn Reson Med*. 2000; 43:601–606. [PubMed: 10748437]
- Hall DA, Chambers J, Akeroyd MA, Foster JR, Coxon R, Palmer AR. Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional magnetic resonance imaging. *J Acoust Soc Am*. 2009; 125:347–359. [PubMed: 19173422]
- Haller S, Bartsch AJ, Radue EW, Klarhöfer M, Seifritz E, Scheffler K. Effect of fMRI acoustic noise on non-auditory working memory task: comparison between continuous and pulsed sound emitting EPI. *MAGMA*. 2005; 18:263–271. [PubMed: 16320092]
- Harms MP, Melcher JR. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol*. 2002; 88:1433–1450. [PubMed: 12205164]
- Harms MP, Melcher JR. Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. *Hum Brain Mapp*. 2003; 20:168–183. [PubMed: 14601143]
- Hellweg FC, Koch R, Vollrath M. Representation of the cochlea in the neocortex of guinea pigs. *Exp Brain Res*. 1977; 29:467–474. [PubMed: 913526]
- Herdener M, Esposito F, Scheffler K, Schneider P, Logothetis NK, Uludag K, Kayser C. Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex*. 2013; 49:2822–2833. [PubMed: 23706955]
- Humphries C, Liebenthal E, Binder JR. Tonotopic organization of human auditory cortex. *Neuroimage*. 2010; 50:1202–1211. [PubMed: 20096790]
- Inan S, Mitchell T, Song A, Bizzell J, Belger A. Hemodynamic correlates of stimulus repetition in the visual and auditory cortices: an fMRI study. *Neuroimage*. 2004; 21:886–893. [PubMed: 15006655]
- Kannan G, Milani AA, Panahi IMS, Briggs RW. An efficient feedback active noise control algorithm based on reduced-order linear predictive modeling of FMRI acoustic noise. *IEEE Trans Biomed Eng*. 2011; 58:3303–3309. [PubMed: 21134809]
- Kusmieriek P, Rauschecker JP. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol*. 2009; 102:1606–1622. [PubMed: 19571201]
- Langers DRM. Assessment of tonotopically organised subdivisions in human auditory cortex using volumetric and surface-based cortical alignments. *Hum Brain Mapp*. 2014; 35:1544–1561. [PubMed: 23633425]
- Langers DRM, Krumbholz K, Bowtell RW, Hall DA. Neuroimaging paradigms for tonotopic mapping (I): The influence of sound stimulus type. *Neuroimage*. 2014; 100:650–662. (in this issue). [PubMed: 25069046]
- Langers DRM, van Dijk P. Robustness of intrinsic connectivity networks in the human brain to the presence of acoustic scanner noise. *Neuroimage*. 2011; 55:1617–1632. [PubMed: 21255663]
- Langers DRM, van Dijk P. Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex*. 2012; 22:2024–2038. [PubMed: 21980020]

- Langers DRM, Van Dijk P, Backes WH. Interactions between hemodynamic responses to scanner acoustic noise and auditory stimuli in functional magnetic resonance imaging. *Magn Reson Med*. 2005; 53:49–60. [PubMed: 15690502]
- Langers DRM, Backes WH, van Dijk P. Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *Neuroimage*. 2007; 34:264–273. [PubMed: 17049275]
- Li G, Mechefske CK. A comprehensive experimental study of micro-perforated panel acoustic absorbers in MRI scanners. *MAGMA*. 2010; 23:177–185. [PubMed: 20464444]
- Liem F, Lutz K, Luechinger R, Jäncke L, Meyer M. Reducing the interval between volume acquisitions improves “sparse” scanning protocols in event-related auditory fMRI. *Brain Topogr*. 2012; 25:182–193. [PubMed: 22015572]
- Loenneker T, Hennel F, Ludwig U, Hennig J. Silent BOLD imaging. *MAGMA*. 2001; 13:76–81. [PubMed: 11502421]
- Mazard A, Mazoyer B, Etard O, Tzourio-Mazoyer N, Kosslyn SM, Mellet E. Impact of fMRI acoustic noise on the functional anatomy of visual mental imagery. *J Cogn Neurosci*. 2002; 14:172–186. [PubMed: 11970784]
- McJury M, Stewart RW, Crawford D, Toma E. The use of active noise control (ANC) to reduce acoustic noise generated during MRI scanning: some initial results. *Magn Reson Imaging*. 1997; 15:319–322. [PubMed: 9201679]
- Merzenich MM, Kaas JH, Roth GL. Auditory cortex in the grey squirrel: tonotopic organization and architectonic fields. *J Comp Neurol*. 1976; 166:387–401. [PubMed: 1270613]
- Moelker A, Pattynama PMT. Acoustic noise concerns in functional magnetic resonance imaging. *Hum Brain Mapp*. 2003; 20:123–141. [PubMed: 14601139]
- Moerel M, De Martino F, Formisano E. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci*. 2012; 32:14205–14216. [PubMed: 23055490]
- Mueller K, Mildner T, Fritz T, Lepsien J, Schwarzbauer C, Schroeter ML, Möller HE. Investigating brain response to music: a comparison of different fMRI acquisition schemes. *Neuroimage*. 2011; 54:337–343. [PubMed: 20728550]
- Nordell A, Lundh M, Horsch S, Hallberg B, Aden U, Nordell B, Blennow M. The acoustic hood: a patient-independent device improving acoustic noise protection during neonatal magnetic resonance imaging. *Acta Paediatr*. 2009; 98:1278–1283. [PubMed: 19432831]
- Norman-Haignere S, Kanwisher N, McDermott JH. Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J Neurosci*. 2013; 33:19451–19469. [PubMed: 24336712]
- Novitski N, Alho K, Korzyukov O, Carlson S, Martinkauppi S, Escera C, Rinne T, Aronen HJ, Näätänen R. Effects of acoustic gradient noise from functional magnetic resonance imaging on auditory processing as reflected by event-related brain potentials. *Neuroimage*. 2001; 14:244–251. [PubMed: 11525334]
- Novitski N, Anourova I, Martinkauppi S, Aronen HJ, Näätänen R, Carlson S. Effects of noise from functional magnetic resonance imaging on auditory event-related potentials in working memory task. *Neuroimage*. 2003; 20:1320–1328. [PubMed: 14568500]
- Novitski N, Maess B, Tervaniemi M. Frequency specific impairment of automatic pitch change detection by fMRI acoustic noise: an MEG study. *J Neurosci Methods*. 2006; 155:149–159. [PubMed: 16530843]
- Oesterle C, Hennel F, Hennig J. Quiet imaging with interleaved spiral read-out. *Magn Reson Imaging*. 2001; 19:1333–1337. [PubMed: 11804761]
- Okada T, Nakai T. Silent fMRI acquisition methods for large acoustic noise during scan. *Magn Reson Med Sci*. 2003; 2:181–187. [PubMed: 16222112]
- Olulade O, Hu S, Gonzalez-Castillo J, Tamer GG Jr, Luh W-M, Ulmer JL, Talavage TM. Assessment of temporal state-dependent interactions between auditory fMRI responses to desired and undesired acoustic sources. *Hear Res*. 2011; 277:67–77. [PubMed: 21426929]
- Perrachione TK, Ghosh SS. Optimized design and analysis of sparse-sampling FMRI experiments. *Front Neurosci*. 2013; 7:55. [PubMed: 23616742]

- 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; 109:216–231. [PubMed: 11206150]
- Rondinoni C, Amaro E Jr, Cendes F, Santos ACD, Salmon CEG. Effect of scanner acoustic background noise on strict resting-state fMRI. *Braz J Med Biol Res*. 2013; 46:359–367. [PubMed: 23579634]
- Saenz M, Langers DRM. Tonotopic mapping of human auditory cortex. *Hear Res*. 2014; 307:42–52. [PubMed: 23916753]
- Sanchez-Panchuelo RM, Francis S, Bowtell R, Schluppeck D. Mapping human somatosensory cortex in individual subjects with 7 T functional MRI. *J Neurophysiol*. 2010; 103:2544–2556. [PubMed: 20164393]
- Scarff CJ, Dort JC, Eggermont JJ, Goodyear BG. The effect of MR scanner noise on auditory cortex activity using fMRI. *Hum Brain Mapp*. 2004; 22:341–349. [PubMed: 15202112]
- Schmithorst VJ, Holland SK. Event-related fMRI technique for auditory processing with hemodynamics unrelated to acoustic gradient noise. *Magn Reson Med*. 2004; 51:399–402. [PubMed: 14755667]
- Schmitter S, Bock M. Acoustic noise-optimized VERSE pulses. *Magn Reson Med*. 2010; 64:1446–1452. [PubMed: 20597118]
- Schönwiesner M, von Cramon DY, Rübsem R. Is it tonotopy after all? *Neuroimage*. 2002; 17:1144–1161. [PubMed: 12414256]
- Schwarzbauer C, Davis MH, Rodd JM, Johnsrude I. Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. *Neuroimage*. 2006; 29:774–782. [PubMed: 16226896]
- Scott BH, Malone BJ, Semple MN. Transformation of temporal processing across auditory cortex of awake macaques. *J Neurophysiol*. 2011; 105:712–730. [PubMed: 21106896]
- Segbers M, Rizzo Sierra CV, Duifhuis H, Hoogduin JM. Shaping and timing gradient pulses to reduce MRI acoustic noise. *Magn Reson Med*. 2010; 64:546–553. [PubMed: 20665798]
- Seifritz E, Di Salle F, Esposito F, Herdener M, Neuhoﬀ JG, Scheﬄer K. Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *Neuroimage*. 2006; 29:1013–1022. [PubMed: 16253522]
- Skouras S, Gray M, Critchley H, Koelsch S. FMRI scanner noise interaction with affective neural processes. *PLoS ONE*. 2013; 8:e80564. [PubMed: 24260420]
- Stiebler I, Neulist R, Fichtel I, Ehret G. The auditory cortex of the house mouse: left-right differences, tonotopic organization and quantitative analysis of frequency representation. *J Comp Physiol A*. 1997; 181:559–571. [PubMed: 9449817]
- Triem-Amit E, Hertz U, Amedi A. Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding FMRI. *PLoS ONE*. 2011; 6:e17832. [PubMed: 21448274]
- Suga N. Tuning shifts of the auditory system by corticocortical and corticofugal projections and conditioning. *Neurosci Biobehav Rev*. 2012; 36:969–988. [PubMed: 22155273]
- Suzuki, Y., Mellert, V., Richter, U., Møller, H., Nielsen, L. Precise and Full-Range Determination of Two-Dimensional Equal Loudness Contours. 2003. Available at: <http://www.nedo.go.jp/itd/grant-e/report/00pdf/is-01e.pdf>
- Talavage TM, Edmister WB. Nonlinearity of FMRI responses in human auditory cortex. *Hum Brain Mapp*. 2004; 22:216–228. [PubMed: 15195288]
- Talavage TM, Edmister WB, Ledden PJ, Weisskoff RM. Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Hum Brain Mapp*. 1999; 7:79–88. [PubMed: 9950065]
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM. Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol*. 2004; 91:1282–1296. [PubMed: 14614108]
- Tomasi D, Caparelli EC, Chang L, Ernst T. fMRI-acoustic noise alters brain activation during working memory tasks. *Neuroimage*. 2005; 27:377–386. [PubMed: 15893942]
- Turner R. How much cortex can a vein drain? Downstream dilution of activation-related cerebral blood oxygenation changes. *Neuroimage*. 2002; 16:1062–1067. [PubMed: 12202093]

- Zaehle T, Schmidt CF, Meyer M, Baumann S, Baltes C, Boesiger P, Jancke L. Comparison of “silent” clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. *Neuroimage*. 2007; 37:1195–1204. [PubMed: 17644001]
- Zapp J, Schmitter S, Schad LR. Sinusoidal echo-planar imaging with parallel acquisition technique for reduced acoustic noise in auditory fMRI. *J Magn Reson Imaging*. 2012; 36:581–588. [PubMed: 22585371]
- Zhang N, Zhu X-H, Chen W. Influence of gradient acoustic noise on fMRI response in the human visual cortex. *Magn Reson Med*. 2005; 54:258–263. [PubMed: 16032671]

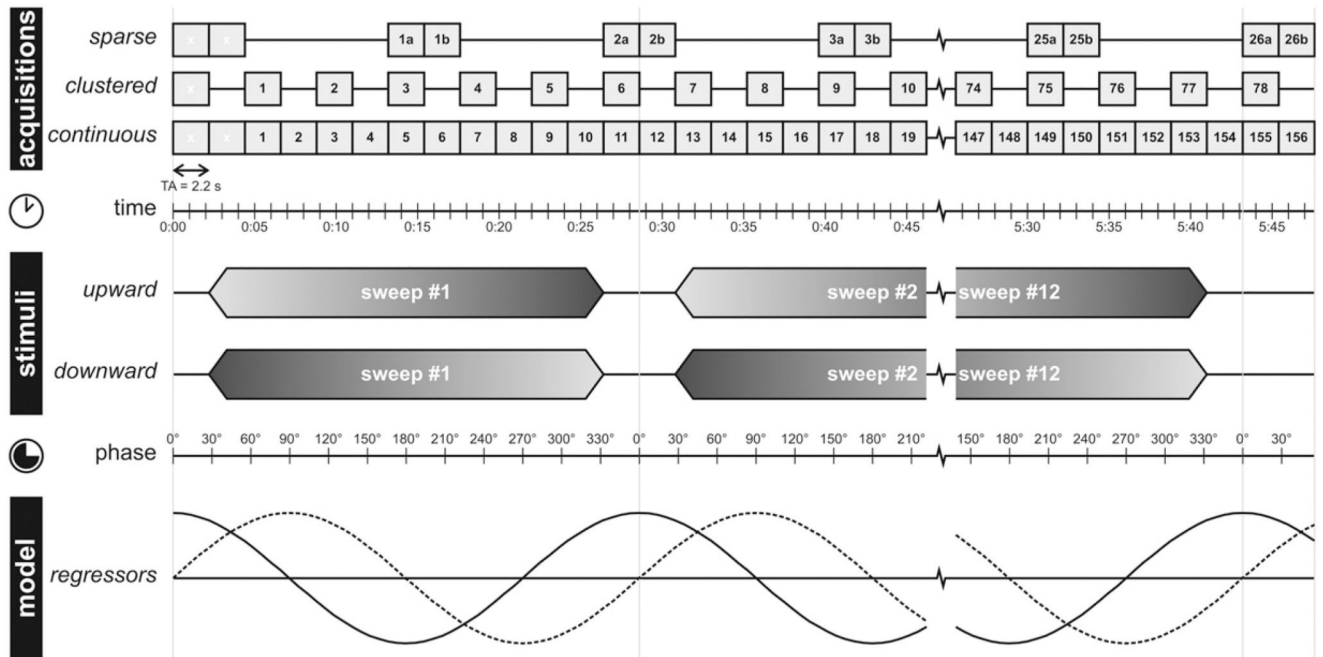


Fig. 1. Schematic illustration of the employed paradigm. Acquisitions were carried out using a sparse protocol (with long periods of scanner inactivity between pairs of images), a clustered protocol (with short periods of scanner inactivity between all images), or a continuous protocol (without interspersed periods of scanner inactivity). At the same time, jittered tone sequences were presented with frequency profiles that were slowly swept upward or downward. The regression model comprised sinusoidal regressors to model a periodic response with the same periodicity as the sweeps, but with arbitrary amplitude and phase (for the downward sweeps, the modelled phase was reversed compared to the upward sweeps by flipping the sign of the sine regressors).

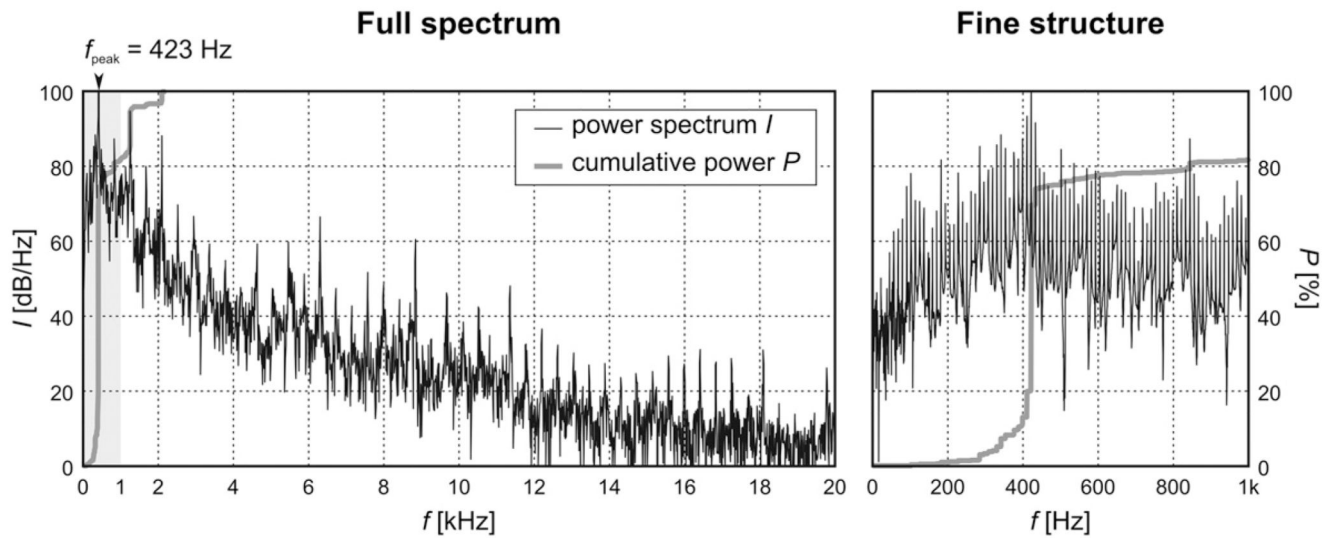


Fig. 2.

The spectrum power density I of the recorded acoustic scanner noise (ASN), shown in black against the left ordinate axis, displayed a large-scale harmonic structure corresponding with the gradient switches at a rate of $f = 423$ Hz (left panel) and an additional sub-harmonic fine-structure corresponding with the slice read-outs at a rate of $f = 11.4$ Hz (right panel). The bold grey curve plots the cumulative power P as a function of frequency against the right ordinate axis, scaled as a percentage relative to the total power in the sound wave.

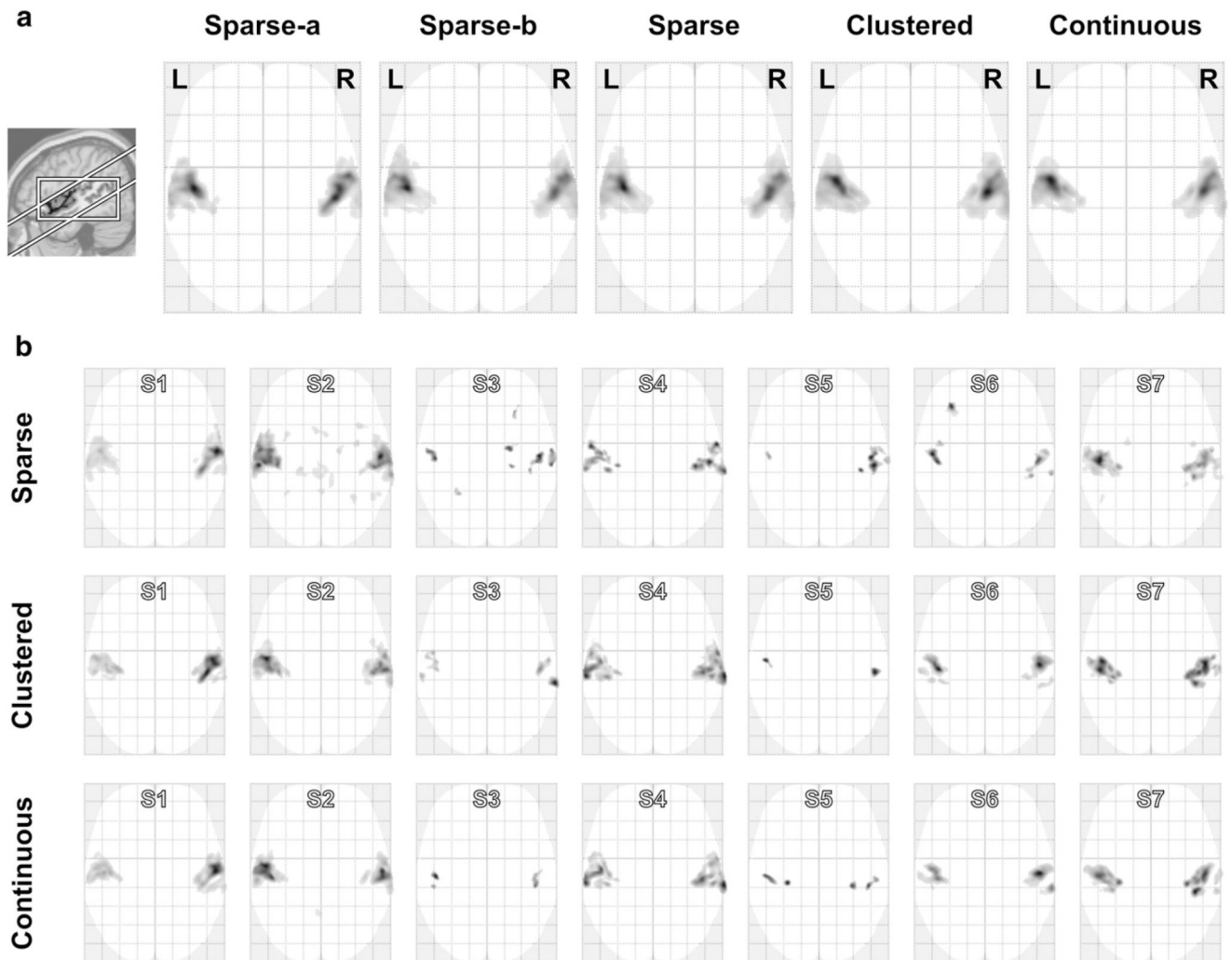


Fig. 3. Significant activation according to various models is displayed in an axial "glass-brain" projection, leniently thresholded at $p < 0.001$ (uncorrected for family-wise errors). Activation was observed in bilateral auditory cortex in all models, but the activation extent and the amount of apparent non-auditory activation differed between analyses. (a) Group models were evaluated for the sets of first and second images from the sparse acquisition protocol (labelled "a" and "b" in Fig. 1), the corresponding set of pairwise-averaged images, and for all images from the clustered and continuous protocols. A group-level cluster extent threshold $k_E > 1.0 \text{ cm}^3$ was additionally imposed. The inset shows the approximate orientation of the imaging volume and the employed analysis' bounding box. (b) Individual subjects models were evaluated for the (pairwise-averaged) sparse, clustered and continuous protocols. A subject-level cluster extent threshold $k_E > 0.1 \text{ cm}^3$ was additionally imposed.

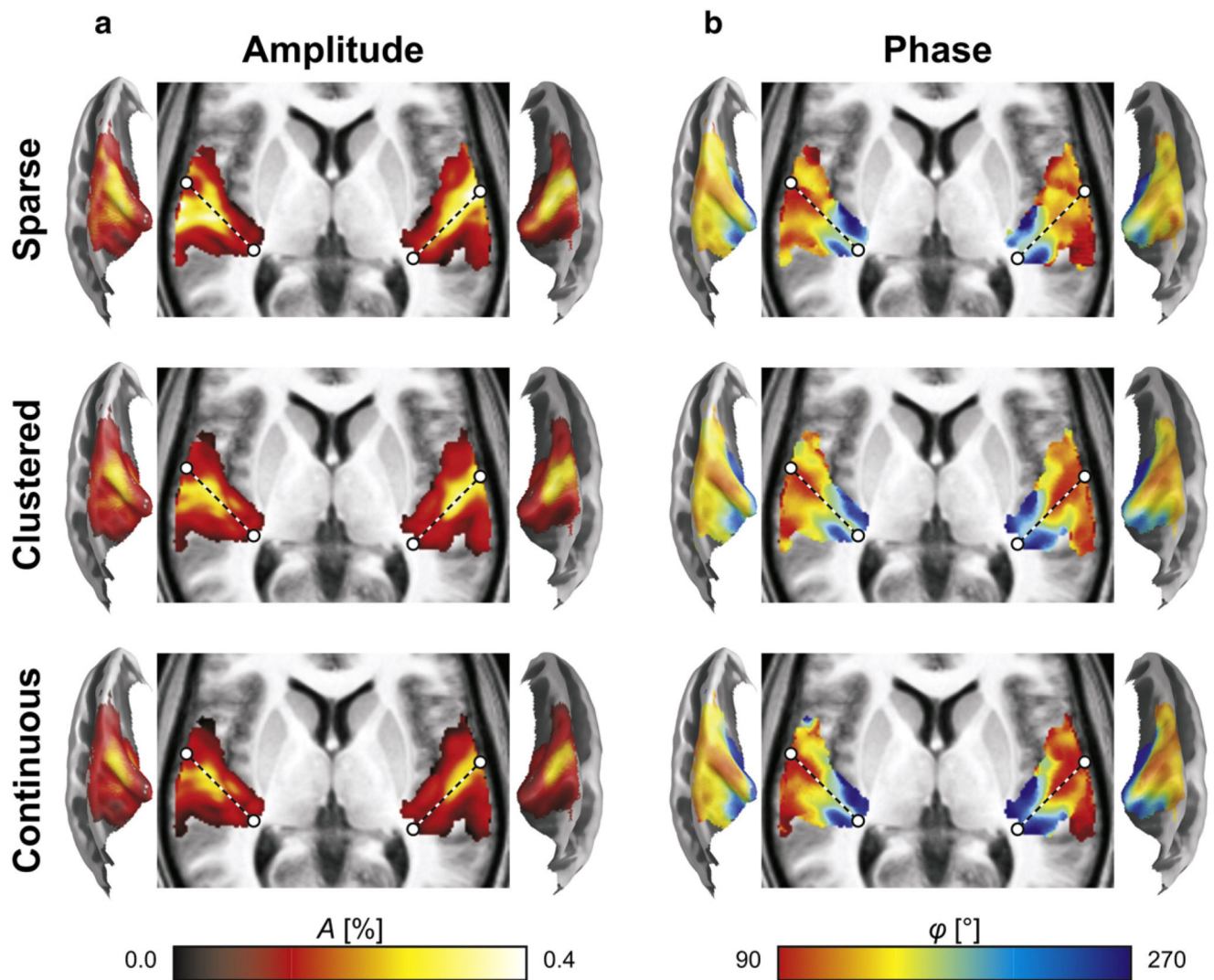


Fig. 4.

The group-level responses are projected on an axial anatomical background (in the middle) and cross-sectioned with reconstructed cortical surfaces of the temporal lobes (to the sides). The dashed line approximately outlines the axis of Heschl's gyrus. (a) Response amplitudes were calculated as the modulus of the pair of cosine and sine regression coefficients. The axial projections show maximal amplitudes across the activated voxels in the z -direction. Activation peaked in lateral Heschl's gyrus. (b) Response phases were calculated as the argument of the pair of cosine and sine regression coefficients. The axial projections show averaged phases across the activated voxels in the z -direction. Multiple phase minima and maxima (i.e. low- and high-frequency endpoints) were observed in lateral and medial regions of auditory cortex, respectively.

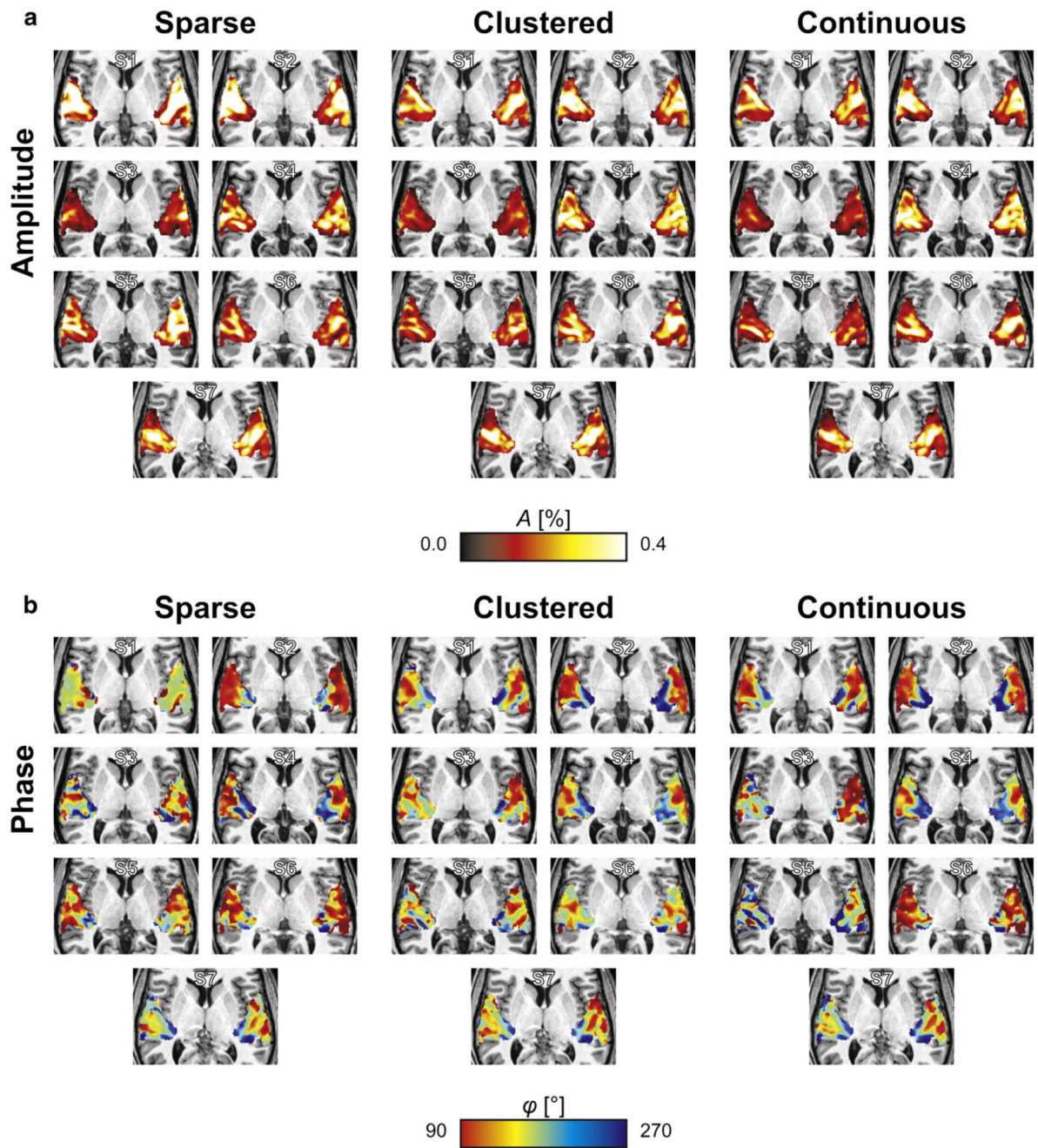


Fig. 5. The subject-level responses are projected on an axial anatomical background. As in Fig. 4, (a) maximal response amplitudes and (b) average response phases were projected across the z-direction. Despite small-scale variations, general patterns of activation and frequency tuning appeared consistent with the corresponding group-level results in most subjects.

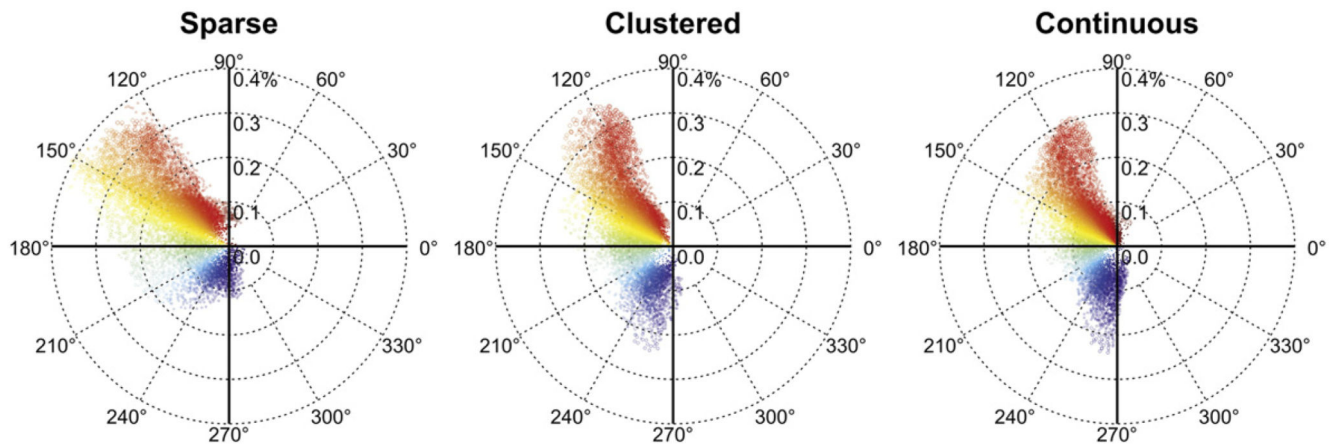
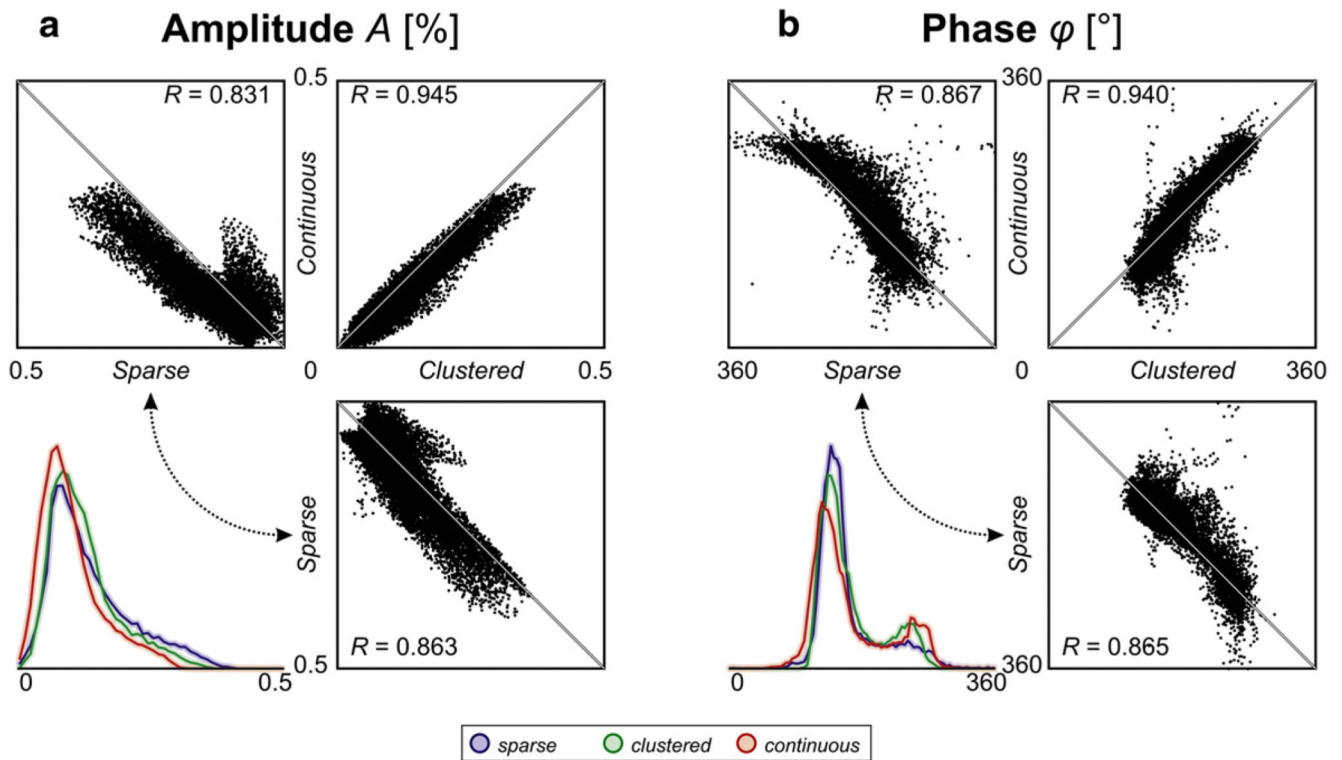


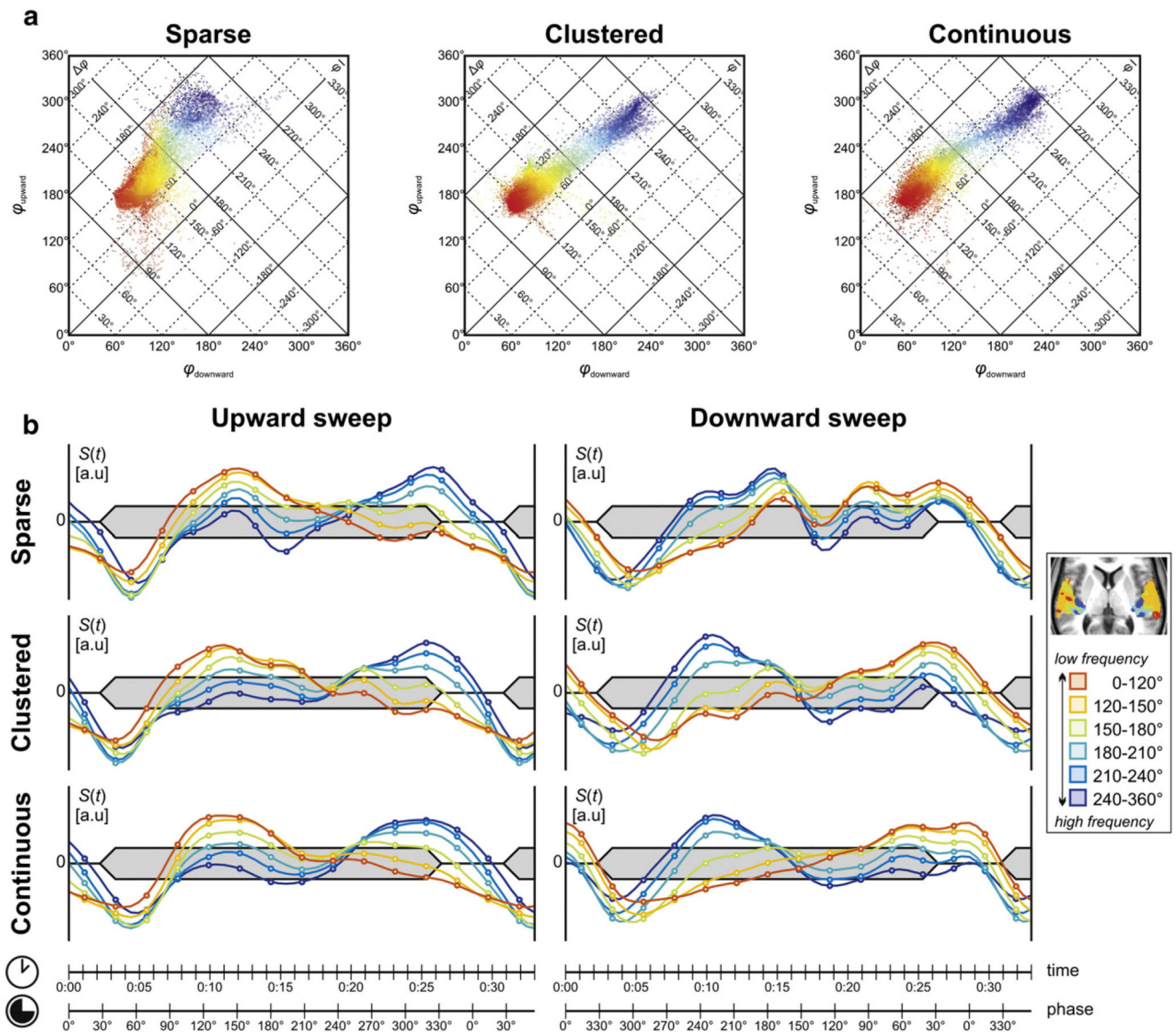
Fig. 6.

Scatter pots of the responses of all voxels are shown for the sparse, clustered, and continuous acquisition protocols. Each voxel's polar (A, φ)-coordinates correspond with its response amplitude and phase, encoding activation level and frequency tuning, respectively.

Equivalently, each voxel's cartesian (x, y)-coordinates correspond with its regression coefficients of the cosine and sine regressors (see Fig. 1). Each voxel contributes one data point; marker size increases with significance of activation, and marker colour maps best frequency.

**Fig. 7.**

The (a) response amplitude and (b) response phase of all voxels are compared across the acquisition protocols. The lower left plots display histograms for the sparse (blue), clustered (green), and continuous (red) protocols. The other three plots display pairwise scatter plots. The outcomes of all three protocols were generally well correlated; the listed Pearson correlation coefficients always exceeded $R > 0.8$. Still, the response amplitudes were largest according to the sparse protocol, intermediate according to the clustered protocol, and smallest according to the continuous protocol. At the same time, the best frequencies shifted to progressively more extreme frequencies and from the moderately low to the highest frequencies when comparing the clustered protocol and particularly the continuous protocol to the sparse protocol.

**Fig. 8.**

(a) Response phases, encoding frequency tuning, were calculated on the basis of the upward and downward sweeps alone, and plotted against each other for each of the three acquisition protocols. Each voxel contributes one data point; marker size increases with significance of activation, and marker colour maps best frequency (from the group model based on upward and downward sweeps combined). The extracted upward and downward phases were generally well correlated, but substantial phase differences occurred that are consistent with an approximately 4-s hemodynamic delay. (b) Hemodynamic response curves were averaged across subjects and across repetitions of the sweep stimuli, and displayed for the upward and downward sweep and for the three acquisition protocols separately. A 33-s fragment is shown (i.e. 15 acquisition times), which slightly exceeds one sweep period; therefore, signals overlap left and right. Measured data points (circles) are interpolated by means of

their Fourier series (lines). Voxels were categorised into six bins according to their response phase (red = low-frequency tuning; blue = high-frequency tuning).