Encoding and decoding models of functional MRI responses in human visual and auditory cortex.

Jessica M. Thomas

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2015

Reading Committee:
Dr. Ione Fine, Chair
Dr. Geoffrey M. Boynton
Dr. Chantel Prat

Program Authorized to Offer Degree:
Psychology Department

Acknowledgments

My thanks to all those who have taught and mentored me throughout my PhD training and graduate career, in particular my advisors and academic parents, Ione Fine and Geoffrey M. Boynton. Thanks also to all of members and fellow trainees on the Auditory Neuroscience Training Grant for sharing your knowledge and expertise. I am also immensely grateful for both past and present members of the Vision and Cognition lab, especially Scott Murray, Bjorn Hubert-Wallander, Fang Jiang, Libby Huber, Paola Binda, and Andrew Bock. My work was heavily influenced and improved by each of these wonderful individuals. Finally, thanks to my subjects who kindly volunteered their time as part of this research.

This work was supported by the National Institute of Health (2T32DC005361-11 grant awarded to J.M.T, a NEI EY-014645 grant awarded to I.F., and a NEI EY-12925 awarded to G.M.B.

Dedication

This dissertation is dedicated to my amazing support system, including all of those that have provided me with encouragement and guidance over the years. To my amazing parents and family members, for always believing I could achieve my goals. To all the teachers and mentors throughout my undergraduate and graduate career who pushed me to aim higher. To all my friends who have laughed, cried, and supported me along this journey.

And to my strong and steady husband Jeremy, who always has been and always will be, the solid earth that holds me together. And finally, to my precocious Maddie Moo, whose curiosity and glee rejuvenates me every day. Mommy loves you.

University of Washington

Abstract

Encoding and decoding models of functional MRI responses
in human visual and auditory cortex.

Jessica M. Thomas

Chair of Supervisory Committee:

Associate Professor Ione Fine
Psychology Department

Computational models of brain activity serve to not only uncover the organizational principles of the brain, but also quantify the neural processes underlying our perceptual experiences. This dissertation investigates the neural encoding of sensory information through computational models of blood oxygen level-dependent (BOLD) responses obtained with functional magnetic resonance imaging (fMRI). Based upon a technique known as population receptive field (pRF) modeling, voxel-based encoding and decoding models were used to predict and decode neural responses to various visual and auditory stimuli. The studies presented herein demonstrate the utility of this technique for examining the functional organization of the visual and auditory cortex, as well as provide methods for verifying model generalizability across different stimulus types.

Table of Contents

Table of Figures

Table of Supplementary Figures

Table of Tables

Table of Supplementary Tables

Introduction

Understanding how sensory information is encoded by our brains is an essential step in elucidating the neural mechanisms underlying our perceptual experiences. Recent functional magnetic resonance imaging (fMRI) studies have begun to explore neural encoding through the development of computational models used to predict and decode blood oxygen level-dependent (BOLD) responses to various sensory stimuli. Using this approach, researchers have begun to reveal not only where different features of sensory stimuli are represented within the human brain, but also the manner in which these neural representations are functionally organized.

This approach began with the development of computational models to predict fMRI responses to a wide range of presented stimuli. Initially, these models were used to establish whether fMRI activity within a particular brain area represented some amount of information about a specific sensory feature. For example, studies performed in the visual domain have demonstrated how orientation-selective responses could be classified from the pattern of BOLD activity across early areas of visual cortex (Kamitani and Tong, 2005). Decoding studies such as these have often applied a technique known as multi-voxel pattern analysis (MVPA) which uses various linear classification algorithms, such as support vector machine (SVM) or logistic regression, that are trained to discriminate stimuli according to their elicited pattern of activity across fMRI voxels. These conventional fMRI decoding methods have proven to be very successful in identifying and classifying perceived stimuli. Furthermore, because these models generate predictions about stimulus discriminability, they can serve as a method for linking neural activity to behavioral performance (Boynton et al., 1999; Buracas et al., 2005) However, conventional decoding studies often do not provide information about the feature space or manner in which these stimuli are functionally organized within the brain. Additionally, these

studies are limited to the candidate sets of stimuli that were used to train the classification algorithm and therefore cannot be used to identify novel stimuli.

An alternative approach to these decoding models, are voxel-based *encoding* models of neural activity. Instead of using trained linear classifiers to discriminate between voxel-activity patterns, encoding models aim to provide a more explicit description of how sensory information is represented within an individual voxel (Naselaris et al., 2011). For example, though the study by Kamitani and Tong (2005) found that orientation-selectivity is differentially encoded across voxels in early visual cortex, it did not establish how this stimulus dimension is represented across those same voxels. Vision researchers have therefore begun to apply an encoding approach by developing models to better characterize the response preferences of each voxel in terms of retinotopic location, spatial frequency, and orientation (Kay et al., 2008). These encoded models were then used to predict BOLD responses to novel image stimuli. Moreover, when combined with parametric decoding, these models were able to not only identify (Kay et al., 2008) but reconstruct novel images according to their visual features (Thirion et al., 2006; Miyawaki et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011).

The goal of this computational approach is to generate a model of fMRI responses that is capable of providing a complete functional description of a specific brain area. Once this model has been accurately generated, and all stimulus features have been mapped to cortex, it is then possible to decode what stimulus had evoked a given neural response. This allows researchers to directly link cortical representations with measures of behavioral performance and provides numerous insights about the role of a specific brain area in the context of our overall perceptual experiences.

This dissertation presents results from a number of experiments implementing both encoding and decoding models of fMRI data to investigate the functional organization of the human visual and auditory cortex. One commonality across these experiments is the use of a voxel-based encoding approach known as population receptive field (pRF) modeling, a technique originally developed as a method for visual field, or retinotopic, mapping of the visual cortex (Dumoulin and Wandell, 2008). This method models the response properties for the population of neurons within a given fMRI voxel as either a one or two dimensional Gaussian. A model time course is then generated for each voxel by calculating the linear sum of the overlap between the time course of the pRF model and the hemodynamically blurred input stimulus. This is mathematically equivalent to the matrix multiplication between the stimulus after hemodynamic blurring and the pRF model (**Figure 0-1**). Model fits for each voxel are then obtained by using an algorithm to iterate towards the model parameters that maximize the correlation between the voxel's pRF model time course and the acquired fMRI response time course.

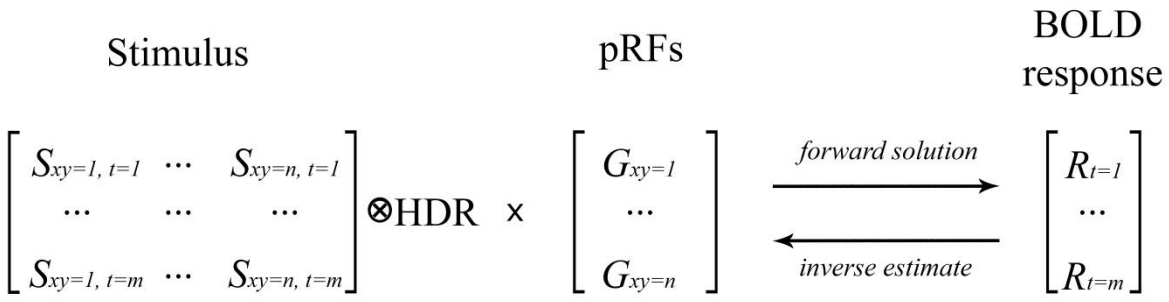


Figure 0-1. Components of the visual pRF linear model

Adapted from Binda et al. 2013. The input stimulus and the pRF are represented as matrices, where vertical and horizontal space are collapsed into a single dimension and time is the other dimension).

Chapter 1 describes work performed to develop and validate pRF modeling analysis for obtaining unbiased measurements of cortical re-organization as a consequence of vision loss (Binda et al., 2013). Using a simulated foveal scotoma, these studies demonstrate methods for

quantifying bottom-up changes in pRFs that are independent from the effects of spatio-temporal BOLD nonlinearities and various top-down influences including the effects of neural ‘fill-in.’

Chapter 2 presents a methodology for extending pRF modeling to the auditory domain to investigate functional representations of auditory frequency, or tonotopic maps within the auditory cortex (Thomas et al., 2015). Chapter 3 presents a method for investigating the generalizability of the auditory pRF models through a combined encoding/decoding approach that is used to identify and reconstruct perceived auditory stimuli over time.

Ultimately, this dissertation aims to demonstrate the utility of modeling the perceptual experiences of an individual using non-invasive measurements of brain activity; a technique which should serve as a powerful method for studying the neurological basis of both enhancements and deficits in sensory perception.

Chapter 1: Unbiased pRF estimates in human visual cortex.

Within the field of vision science, there has been substantial interest in the use of functional MRI and pRF modeling to examine how retinotopic maps are affected by long-term visual deprivation. Several high-profile studies have investigated the degree of cortical reorganization that occurs as a result of lesions or diseases affecting either the retina (Morland et al., 2001; Baseler et al., 2002; Sunness et al., 2004; Baker et al., 2005; Baker et al., 2008; Masuda et al., 2008; Dilks et al., 2009; Masuda et al., 2010; Baseler et al., 2011), retino-geniculate fibers (Hoffmann et al., 2012), or geniculo-cortical pathways (Baseler et al., 1999; Morland et al., 2001; Dilks et al., 2007).

As a control condition, several studies compared the retinotopic maps of subjects with long-term vision loss to those of normally-sighted subjects under conditions designed to simulate a visual impairment. Unexpectedly, evidence for cortical reorganization was observed within the normally-sighted subjects using either a simulated scotoma condition (Baseler et al., 2011; Haak et al., 2012), or scotopic illumination levels used to reduce the sensitivity within the central visual field (Barton & Brewer, 2011). For both of these cases, the pRFs of voxels representing the scotoma area were shifted and changed in size, mimicking the pattern of cortical reorganization previously observed in some cases of long-term visual impairment (Baseler et al., 2011).

It has been suggested that these unexpected pRF changes may be partially due to an altered balance between the visual input and signals from feedback or lateral connections related to the phenomenon of neural ‘‘fill-in’’ (Baseler et al., 2011; Haak et al., 2012). Alternatively, previous studies have acknowledged that these pRF changes might be artifacts of the methods used in retinotopic mapping (Wandell and Smirnakis, 2009). In an effort to eliminate these

factors, our lab began a series of experiments to develop and pRF modeling analysis for obtaining unbiased measurements of cortical reorganization as a consequence of vision loss.

Accurate representation of the effective stimulus

In Binda et al. (2013), we investigated whether model estimation artifacts were responsible for the unexpected pRF changes observed in normally-sighted individuals. We began by measuring fMRI responses in cortical visual areas V1, V2, and V3 of three normally-sighted subjects (age: 26–29 years, one female) to two stimulation sequences previously used for retinotopic mapping of the visual cortex: a multifocal patch sequence (**Figure 1-1 A,C**) (Vanni et al., 2005) and a conventional drifting bar sequence (**Figure 1-1 B, D**) (Dumoulin and Wandell, 2008).

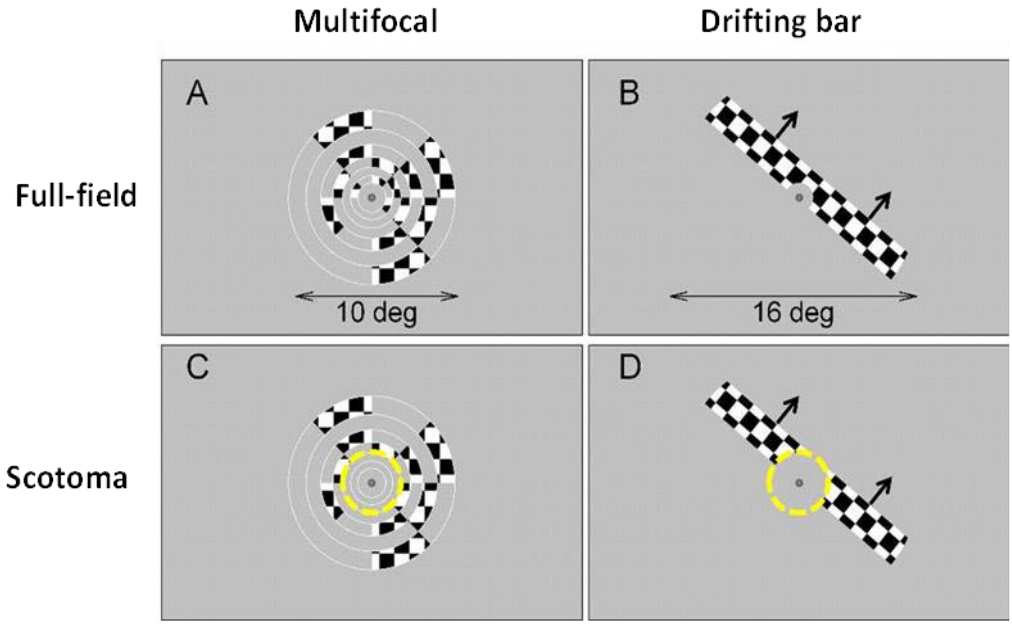


Figure 1-1. Retinotopic stimulation sequences

Taken from Binda et al 2013. Panels A and B show an individual frame of the multifocal and the drifting bars stimulus sequences in the full-field condition. Panels C and D show the same frame for the scotoma condition, with the yellow dashed circle (not part of the display) marking the extent of the mean luminance mask simulating a foveal scotoma.

Each stimulation sequence was presented under two conditions: a “full-field” condition (**Figure 1-1 A, B**) in which the area of stimulation covered the entire display region, or a “scotoma” condition (**Figure 1-1 C, D**) in which the central 2° of the same stimulus was masked by a mean luminance patch to emulate a foveal scotoma.

The fMRI responses collected with the two stimulation sequences and under the two conditions were fit as independent sets of pRF estimates. For the scotoma condition, pRFs were fit using two different methods: the previously applied (Baseler et al., 2011; Haak et al., 2012) ‘‘full-stimulus pRF method’’ in which the input stimulus to the pRF fitting algorithm represented the full-field stimulus (ignoring the presence of a scotoma); and an ‘‘effective-stimulus pRF method’’ which incorporated the scotoma into the input stimulus. Please refer to Binda et al. 2013 for full methodological details, including all stimulus descriptions, imaging parameters, and modeling techniques.

We found that modeling the fMRI responses with the “full-stimulus pRF method” resulted in pRFs that appeared shifted in position and changed in size. This suggests that the previously reported pRF changes seen with simulated scotomas were at least partially due to a methodological artifact. These artifactual changes were completely eliminated when fMRI responses collected with the multifocal stimulus sequence were modeled using the “effective-stimulus pRF method.”

However, the pattern of results was more complex for pRF estimates generated with the drifting-bar stimulus sequence: use of the “effective-stimulus pRF method” resulted in an unexpected *over-correction* of the biases. This suggested that pRF estimates could not adequately explain the pattern of fMRI responses to the drifting bars stimulus, indicating that

factors not included within the pRF model, such as spatio-temporal BOLD nonlinearities, may have been involved.

Stimuli designed with coherent movement, such as the drifting bar sequences, are known to induce a ‘‘traveling wave’’ of BOLD activity across the cortical surface (Engel et al., 1994), that can induce a spread of responses that propagates beyond the directly stimulated area (Aquino et al., 2012). Though previous studies have shown that these spatial and temporal summation of BOLD signals can be well approximated by a linear model; significant nonlinear interactions have also been found (Zenger-Landolt and Heeger, 2003; Pihlaja et al., 2008).

Alternatively, stimulus predictability and neural “fill-in” phenomena might have also affected responses to the drifting bars stimulus. Studies of attention have suggested that fMRI responses are influenced by expectations about the upcoming pattern of stimulation (Kastner et al., 1999). Furthermore, the scotoma condition stimulus might have been perceived as partially occluded object, with the drifting bar passing behind the simulated scotoma. Similar neural ‘fill-in’ effects have been shown to occur at the earliest stages of cortical processing, even when spatial attention is directed elsewhere (Meng et al., 2005).

Quantifying Neural Fill-In

‘Filling-in’ is a perceptual phenomenon by which a visual attribute such as brightness or motion is induced in a blank region of the visual field by a surrounding stimulus (Ramachandran and Gregory, 1991; Weil et al., 2007). It has been shown that filling-in can be associated with attention, amodal completion, or apparent occlusion (Meng et al., 2005; Komatsu, 2006). However, evaluating the neural correlates of ‘fill-in’ phenomena via fMRI is complicated by the spatio-temporal blurring and nonlinearities associated with the BOLD signal. To avoid these complexities, we developed a novel ‘neural image’ method to quantify the amount of neural

"fill-in" present in our measured fMRI responses to the multifocal and drifting bar stimuli. Based upon work by Kay et al. (2008), our method generates neural images through an inversion of the pRF model to estimate what stimulus best predicted the measured pattern of voxel BOLD responses at that point in time. Using this method we are able to visualize how a stimulus was represented by BOLD responses in visual cortical areas.

We began with the pRF models for areas V1-V3 that had been generated with the full-field multifocal stimulus, which as previously discussed, produced retinotopic estimates for each subject without artifactual biases. For each stimulation sequence and condition type we generated two neural image time courses comprised of either: *model* neural images or *data-driven* neural images (**Figure 1-2**).

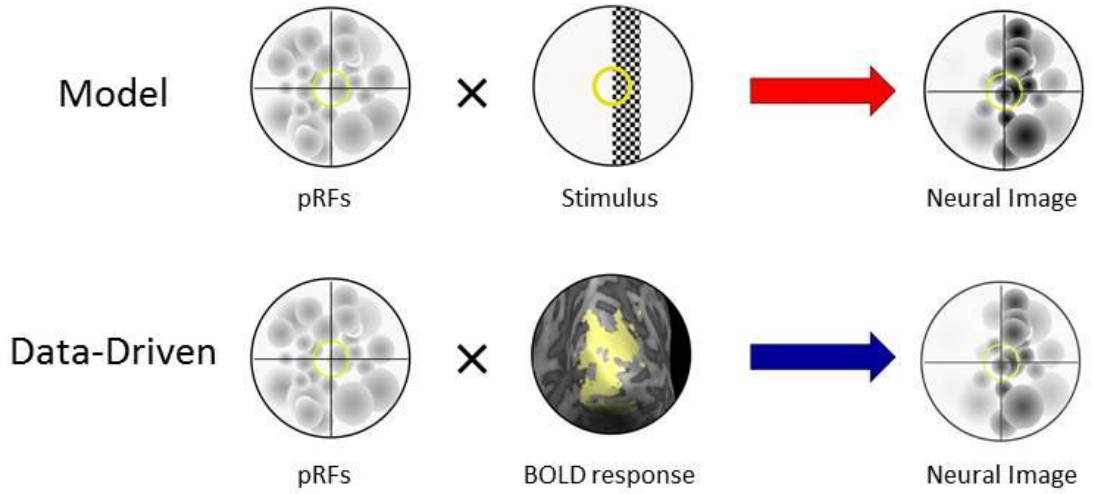


Figure 1-2. Generating Neural Images

The pRF models for each subject were generated with the full-field multifocal stimulus. *Model* neural images (red) were generated by multiplying each voxel's pRF model by each physical frame of the stimulus convolved with each subject's hemodynamic response function. *Data-driven* neural images (blue) were generated by multiplying each voxel's pRF model by the stimulus-driven BOLD signal at that point in time.

Model neural images were generated by multiplying each voxel's pRF model by each physical frame of the stimulation sequence that had been convolved with each subject's

hemodynamic response function. These *model* neural image time courses serves as a reference for a purely linear system which retains the spatio-temporal nonlinearities present in the BOLD signal, but does not contain the effects of neural fill-in. *Data-driven* neural images were generated by first separately averaging the stimulus-driven BOLD response time courses acquired for each condition (full-field or scotoma) and stimulation sequence type (multifocal and drifting bar). Then the neural images were created by multiplying each voxel's pRF model by the averaged stimulus-driven BOLD response at that point in time, with each time point corresponding to a frame in the stimulation sequence.

Next, we defined two regions-of-interest (ROIs) within the *model* and *data-driven* neural image time courses: one ROI containing the area of the scotoma (central 2°) and one equally sized ROI located outside the area of the scotoma (**Figure 1-3**).

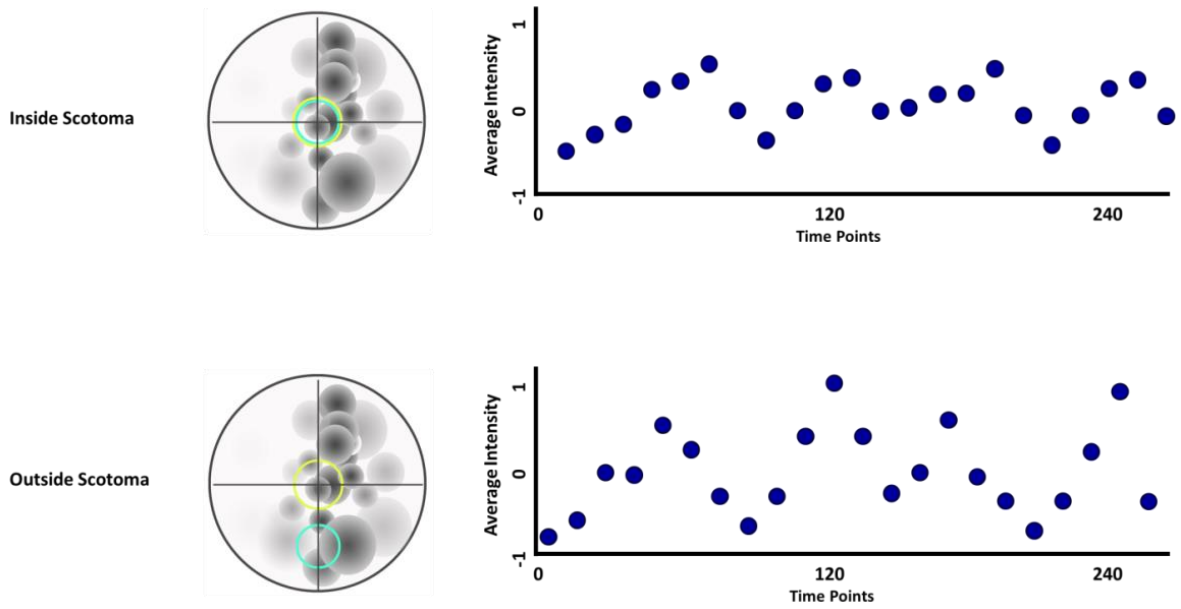


Figure 1-3. Averaged Intensity Values

We defined two regions-of-interest (ROIs) depicted here by the blue circles. One located inside the scotoma (top) and one located outside the area of the scotoma (bottom). For each time point, we averaged the normalized intensity of the neural image within each ROI (illustrated here by the blue data points).

For each time point within the neural image time courses, we averaged the intensity of the neural image (normalized between -1 to 1) within the ROIs (total of 240 average intensity values per time course). To visualize the amount of neural "fill-in" present within each ROI, we plotted the averaged intensity values from each point in time in the full-field condition against each point in time under the scotoma condition. This was acceptable since, with the exception of the scotoma mask, the frames of each stimulation sequence (240 frames in total) were identical across all scans of the same sequence type. This was crucial for isolating the effects of "fill-in," as it allowed us to directly compare each time point in the scotoma condition to each time point in the full-field condition.

Finally, a linear regression model was fit to each ROIs average intesity values, using a total least squares (TLS) method, also known as orthogonal regression. Unlike standard linear regression analyses, the TLS method takes in account observational errors on dependent and independent variables, a necessary factor given the relative noise present within the average intensity values of both the full-field and scotoma conditions.

Figure 1-4 displays the results of these comparisons for Subject 1. Plots are shown seperately for the inside scotoma and outside scotoma ROIs. For the sake of easier interpretation, only the *data-driven* values are plotted (in either light or dark blue). The best fitting lines resulting from the TLS regression, are displayed for values of the *model* neural images (red) and the *data-driven* neural images (either light or dark blue). A third line (black) (with a slope = 1) is shown as a reference for total neural fill-in, in which averaged intensity values in the scotoma and full-field conditions are perfectly correlated. This would indicate that BOLD responses did not differ across conditions, as though the scotoma was not present.

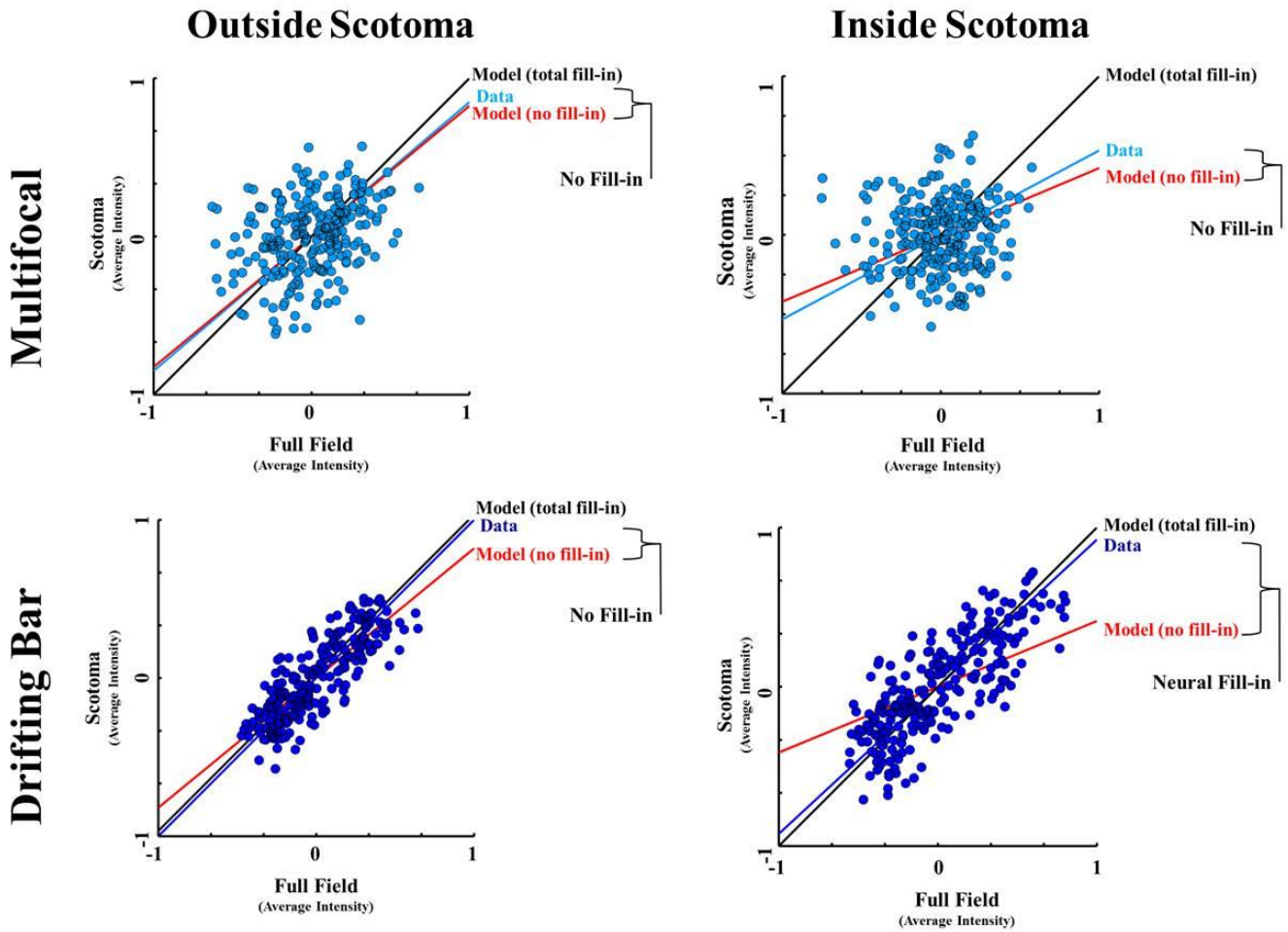


Figure 1-4. Quantifying Neural Fill-in

Plots of the averaged intensity values from each point in time in the full-field condition against each point in time under the scotoma condition. *Data-driven* values are plotted for the multifocal sequence (light blue) and the drifting bar sequence (dark blue). Best fitting lines resulting from the TLS regression are displayed for the *model* neural images (red) and the *data-driven* neural images (either light or dark blue). A third line (black) (with a slope = 1) is shown as a reference for total neural fill-in.

The amount of neural fill-in within each ROI was quantified acoording to the difference in slope between these best fitting lines. An increased slope in the *data-driven* line relative to the *model* without neural fill-in, indicates the presence of neural fill-in within that ROI. For the outside scotoma ROI, the difference in slope between the *data-driven* and *model* lines was marginal for both stimulation sequences. This is to be expected as the portions of the stimulation sequences driving responses for that area of the neural image, remained the same across the full-

field and scotoma conditions. As shown in **Figure 1-4,** results for the inside scotoma ROI differed across the stimulation sequences. Though *data-driven* values from multifocal sequences are better predicted by the *model*, some evidence for neural fill-in remained. However, the slope of the line fit to the *data-driven* values from the drifting bar sequences was considerably increased relative to the *model*, indicting a considerable amount of neural fill-in for the inside scotoma ROI. This pattern of results was present across all three subjects, as shown in **Figure 1-5**, which displays the slopes of the best fitting lines for both ROIs and both stimuluation sequences.

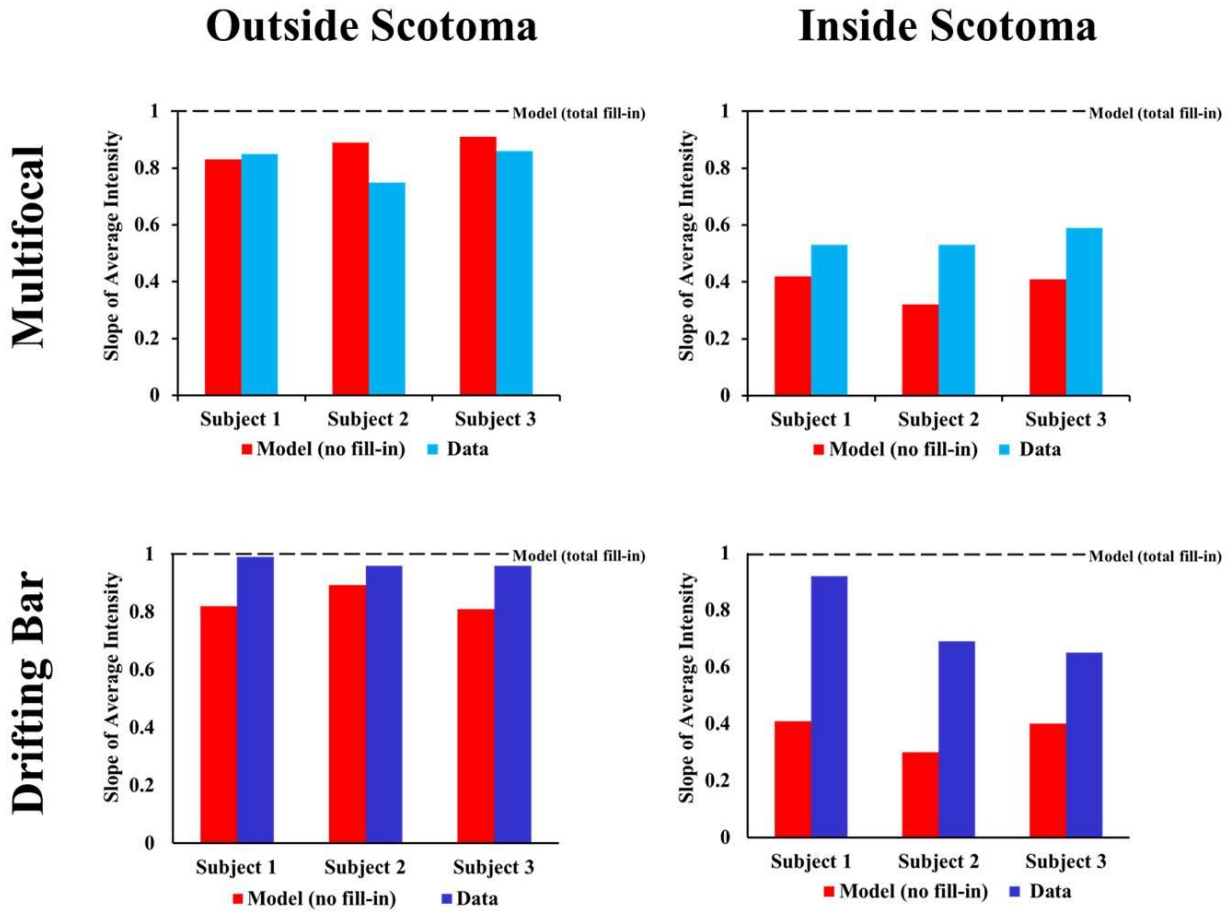


Figure 1-5. Neural Fill-In Across Subjects

Bar graphs of the slopes from the best fitting lines for the *data-driven* (either light or dark blue) and *model* (red) values are shown for each ROI and stimulation sequence. The dotted black line (slope = 1) serves as a reference for total neural fill-in.

Our results demonstrate how an inversion of the population receptive field model that generates neural images can be used to visualize and quantify the effects of neural fill-in for areas within early visual cortex. By comparing *data-driven* neural images to *model* neural images, which both contained the spatio-temporal blurring and nonlinearities associated with the BOLD signal, we were able to isolate the effects of filling-in. Using this approach we found a considerable amount of neural fill-in occurring within the area scotoma when the predictable drifting bar stimulus was used. The amount of fill-in was significantly less when the unpredictable multifocal stimulus was used. These findings suggest that we were correct in our notion that stimulus predictability and neural "fill-in" phenomena may have biased pRF estimates generated with the drifting bar stimulus. We therefore suggest that neural fill-in effects had induced neural signals within the area of the scotoma that resulted in the pRF eccentricity shifts which were over-corrected by our effective-stimulus pRF method.

This neural image method serves as compelling methodological tool for studying plasticity in the retinotopic organization of visual areas. Future directions would include comparing the effects of neural fill-in across real (e. g. macular degeneration) and simulated scotomas, as well as comparing the amount of neural fill-in present across different areas of the visual cortex.

This method is also of particular interest for those involved with sight recovery engineering technologies working to produce retinal prostheses. The ability to visualize the neural representations of visual images at the level of the cortex can serve as method for developing and validating more sophisticated forms of retinal stimulation.

Chapter 2: Population receptive field estimates of human auditory cortex

Abstract

Here we describe a method for measuring tonotopic maps and estimating bandwidth for voxels in human primary auditory cortex (PAC) using a modification of the population Receptive Field (pRF) model, developed for retinotopic mapping in visual cortex by Dumoulin and Wandell (2008). The pRF method reliably estimates tonotopic maps in the presence of acoustic scanner noise, and has two advantages over phase-encoding techniques. First, the stimulus design is flexible and need not be a frequency progression, thereby reducing biases due to habituation, expectation, and estimation artifacts, as well as reducing the effects of spatio-temporal BOLD nonlinearities. Second, the pRF method can provide estimates of bandwidth as a function of frequency. We find that bandwidth estimates are narrower for voxels within the PAC than in surrounding auditory responsive regions (non-PAC).

Introduction

Primate studies suggest that the "core" of the auditory cortex (AC), identified on the basis of the underlying cellular architecture, contains up to three subdivisions with borders delineated by tonotopic gradient reversals: A1, R, and RT (Hackett et al., 1998; Hackett, 2008). In humans, cytoarchitectonic and chemoarchitectonic studies have localized the AC core to approximately the medial two-thirds of Heschl's gyrus (HG) (Rademacher et al., 2001; Dick et al., 2012). Tonotopic organization within these areas has been measured using intracortical auditory event-related potentials (Liegeois-Chauvel et al., 1991), surgically implanted microelectrodes (Howard

et al., 1996) and non-invasive functional imaging (Formisano et al., 2003; Talavage et al., 2004; Humphries et al., 2010; Da Costa et al., 2011; Striem-Amit et al., 2011; 2013; Saenz and Langers, 2013).

A number of functional MRI studies have identified two tonotopic gradients, thought to correspond to human homologues of areas A1 (hA1) and R (hR) that form mirror-image representations reversing at a low frequency border (Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011; Langers and van Dijk, 2011; Striem-Amit et al., 2011; Moerel et al., 2012; 2013). The low frequency region is found close to Heschl's gyrus, and some uncertainty remains as to the exact orientation of these two maps with reference to the gyrus (see Saenz and Langers, 2013 for a review).

Several tonotopy studies have relied on stimuli comprised of orderly ascending or descending frequency progressions (e.g. Talavage et al., 2004; Da Costa et al., 2011; Striem-Amit et al., 2011; 2013), and a phase-encoding analysis that identifies the "best frequency" of a voxel by either finding the phase of the sinusoid or time-lagged function that best-fits the voxel time course (Engel et al., 1994). While such methods provide a robust method for identifying tonotopically organized areas, several features of this approach contribute to uncertainty in interpreting results.

Frequency progression stimuli have the potential to cause habituation and/or expectation effects. Moreover, the ''traveling wave'' of BOLD activity induced across the cortical surface (Engel et al., 1994) by frequency progressions is likely modulated by spatiotemporal nonlinearities. Previous studies have shown that while the spatial and temporal summation of BOLD signals can be well approximated by the linear model; significant nonlinear spatiotemporal interactions do occur (Zenger-Landolt and Heeger, 2003; Pihlaja et al., 2008;

Binda et al., 2013). Furthermore, estimates of best frequency based on a small number of presented frequencies tend be biased, especially near the edge of the stimulus range (Dumoulin and Wandell, 2008, Appendix B). Our data and that of Dumoulin et al. (2008, Appendix B) suggest that these effects can be somewhat reduced by fitting data using a continuous function such as a Gaussian or a sinusoid capable of assigning voxels to outside the presented frequency range rather than using a "winner-take-all" method that assigns a value based on the highest correlation to the presented frequencies only, where voxels with a best frequency outside of the stimulus range are assigned to either the highest or lowest presented frequency. Thus, frequency progression stimuli, especially when analyzed using 'winner take all' methods, have the potential to result in an overrepresentation of frequencies near the beginning or the end of the sweep. Analogous concerns have been described for visual retinotopic mapping methods (Duncan and Boynton, 2003; Dumoulin and Wandell, 2008; Haak et al., 2012; Binda et al., 2013), and recently discussed for tonotopic mapping methods (Langers et al., 2014a; 2014b).

Here, we estimate auditory frequency responses using a population Receptive Field (pRF) method developed by Dumoulin and Wandell (2008) that is less susceptible to many of these issues when utilizing a randomized stimulus, and has the added advantage of providing a means of estimating the receptive field size or bandwidth of individual voxels as a function of frequency.

Materials and Methods

Subjects

Four right-handed subjects (2 male, 2 female, ages 24-45) participated in two fMRI sessions. Subjects reported normal hearing and no history of neurological or psychiatric illness. All procedures, including recruitment, consenting, and testing, followed the guidelines of the

University of Washington Human Subjects Division and were reviewed and approved by the Institutional Review Board.

Stimulus Presentation

Auditory stimuli were generated in MATLAB using the Psychophysics Toolbox (www.psychtoolbox.org). Stimuli were delivered via MRI compatible insert earphones (S14, Sensimetrics), at a sampling rate of 44.1 kHz, with intensities adjusted to ensure flat frequency transmission from 100 Hz to 8 kHz. Subjects were instructed to keep their eyes closed during all scans.

We measured fMRI responses to two types of stimulation sequences: *ascending /descending tone progressions* and *random tone sequences*. Both stimuli were comprised of blocks of pure tone stimuli originally developed by Da Costa et al. (2011). Each frequency block lasted 2 s and contained eight pure tone bursts of the same frequency, with each burst lasting either 50 ms or 200 ms in duration (inter-stimulus interval = 50ms). Tone durations were alternated in pseudo-randomized order, switching durations at least 4 times during each 2 s block, resulting in a "Morse code" like pattern of long and short tones which served to increase the perceptual salience of the stimuli over the regular pattern of background scanner noise **(Figure 2-1 A)**.

The choice of 2 s for our frequency block duration was primarily motivated to facilitate comparison with a traditional progression paradigm used in previous studies. **Supplementary Figure 2-1** shows the results of simulations that suggest when the effects of BOLD adaptation (Soltysik et al., 2004) are included, the optimal block duration would be between 1 to 5 s.

After sound system calibration, all stimulus sound intensities were adjusted according to a standard equal-loudness curve created for insert earphones (ISO 226) to approximate equal

perceived loudness across all frequencies. Actual sound intensities (65-83 dB SPL) matched the perceived loudness of a 1000 Hz tone (reference frequency) at 70 dB SPL. Acoustic noise from the scanner was attenuated by expanding-foam eartips as well as protective ear muffs placed over the ear following earphone insertion. Subjects reported hearing all tones to be clearly and comfortably audible, and of roughly equal loudness across all frequencies.

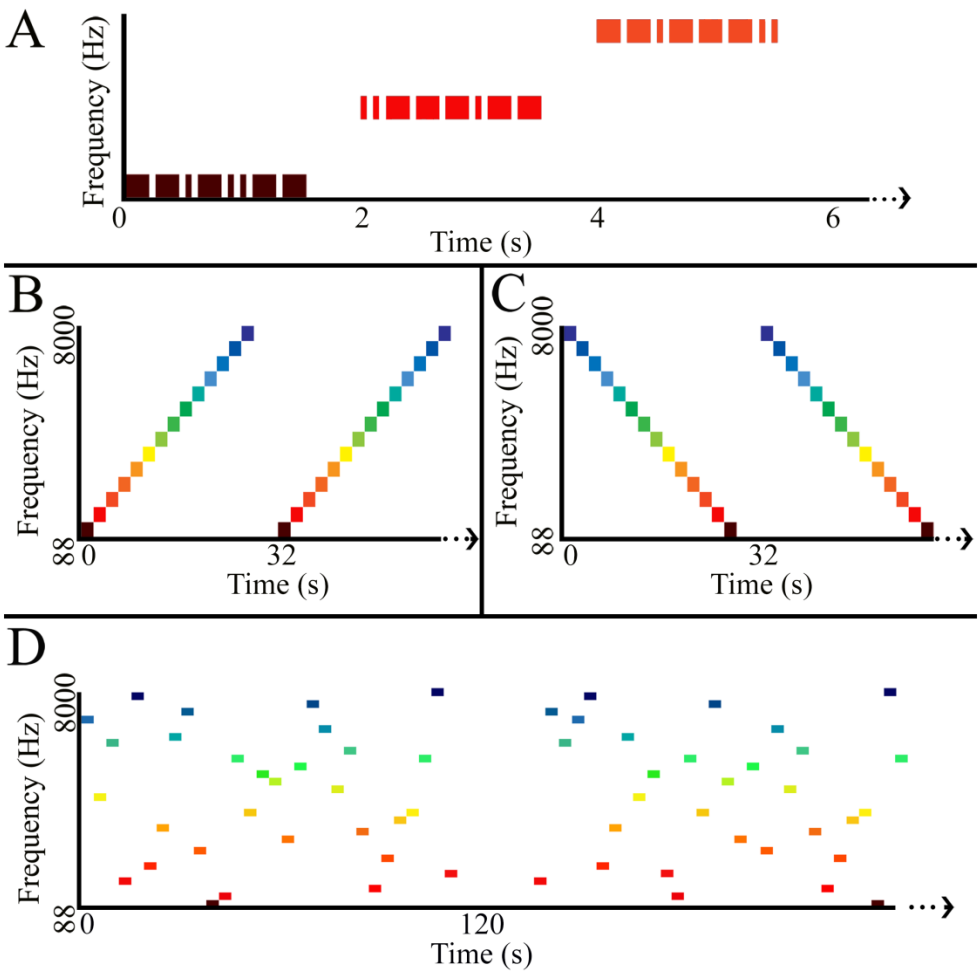


Figure 2-1. Auditory pRF mapping stimuli

(A) 3 example frequency blocks containing 8 pure tone bursts (50 or 200 ms). (B, C) 2 cycles (total 15 cycles per scan) of the ascending and descending tone progressions. (D) The first 120 frequency blocks of a random sequence scan.

Ascending /descending tone progressions

Ascending/descending progressions were identical to those of Da Costa et al. (2011). The frequencies used were: 88, 125, 177, 250, 354, 500, 707, 1000, 1414, 2000, 2828, 4000, 5657, and 8000 Hz (half-octave steps). Each frequency block was presented for 2 s before progressing to the next higher (*ascending progression*, **Figure 2-1 B)** or lower (*descending progression*, **Figure 2-1 C)** frequency until all 14 frequencies had been presented. This 28 s frequency progression was followed by a 4 s silent pause and this 32 s cycle was repeated 15 times per scan, for total scan duration of 8 min. For each subject we carried out one *ascending/descending* stimulation session, during which subjects listened to 4 *ascending progressions* and 2 *descending progressions*.

Random tone sequences

For the *random tone sequences* **(Figure 2-1 D)**, stimuli were arranged into 240 equally spaced frequency blocks (2 s) ranging from 88-8000 Hz. Each block was presented only once per scan and the frequency block order was randomly shuffled for each scan. Following every 60 frequency blocks was a 12 s silent pause which allowed the pRF algorithm to estimate the baseline fMRI response in response to scanner noise alone (Zuiderbaan et al., 2012) to improve the accuracy of bandwidth measurements. The entire scan therefore lasted a total of 8 min 48 s. Each subject participated in a single *random tone sequence* scanning session, consisting of 6 scans, each containing a different random sequence of the same tones. Our goal in presenting random tone sequences was to reduce spatial and temporal correlations between neighboring frequencies in the stimulus sequence, thereby reducing the influence of spatiotemporal nonlinearities on pRF estimates.

Magnetic resonance imaging

Functional magnetic resonance images were acquired with a 3T Phillips Achieva scanner (Philips, Eindhoven, The Netherlands) at the University of Washington Diagnostic Imaging Sciences Center (DISC) using an 8-channel head coil. Foam padding minimized head motion.

Acquisition Sequences

A common issue with tonotopic mapping protocols is the reliability of frequency measurements estimated in the presence of loud acoustic scanner noise that can interfere with or mask the hemodynamic responses to presented stimuli (Langers et al., 2005). While sparse scanning techniques limit the effects of acoustic noise, they require a marked increase in the amount of scanning time needed as compared to continuous acquisition (Hall et al., 1999; though see Petkov et al., 2006; Humphries et al., 2010).

In order to examine the reliability of the pRF estimates calculated in the presence of acoustic scanner noise, data were gathered using two acquisition sequences: a standard EPI sequence and an attenuated EPI sequence designed with Philips SofTone software (SofTone factor of 4.0) to generate quieter scanner noise. Acoustic scanner noise was recorded from inside the coil for both acquisition types (for both 27 and 35 slices, since the slice number can also influence acoustic noise) using an MR-compatible fiber-optic microphone (FOMRI-II, Optoacoustics) placed within the scanner bore and saved in WAV format using Matlab. The overall A-weighted sound pressure levels (dBA) were also collected for both acquisition types using a B&K sound level meter model 2250.

A frequency spectrum of the recorded scanner noise was obtained for each acquisition type by performing a discrete Fast Fourier Transform and calculating the resulting frequency

component magnitudes. The standard EPI sequence generated an overall louder (122 dBA) auditory scanner noise peaking at approximately 1000 Hz; while the attenuated EPI sequence generated quieter (105 dBA) scanner noise peaking at approximately 350 Hz, see **Supplementary Figure 2-2**.

Each session contained six scans. Three functional scans were collected using the standard EPI sequence (35 slices, TR/TE = 2000/25 ms, flip angle = 80°, EPI-factor = 51, no slice gap). After discarding the first 5 timeframes of each functional scan, *ascending/descending* session scans consisted of 240 acquisitions at an effective voxel size of 2.0 x 2.0 x 3.00 mm^3 (FOV= 264 x 264 x 105 mm^3, matrix size= 132 x 132 x 35). *Random tone sequence* scans consisted of 260 volumes at an effective voxel size of 2.33 × 2.33 × 2.33 mm^3 (FOV = 224 x 224 x 81.55 mm^3, matrix size= 113 x 113 x 35).

The other three scans were collected utilizing the Philips SofTone parameter, which reduces acoustic noise by controlling the shape of the gradient waveform independent of the amplitude. This change in the gradient waveform required either the number of slices to be reduced or the TR to be lengthened. We chose to maintain the 2 s TR and reduce the number of slices while increasing the voxel size to maintain roughly similar volume coverage (27 slices, TR/TE = 2000/25 ms, flip angle = 76°, EPI-factor = 39, no slice gap). Both *ascending/descending* session and *random tone sequence* scans consisted of 260 volumes at an effective voxel size of 2.75 × 2.75 × 3.00 mm^3 (FOV = 220 x 220 x 81 mm^3, matrix size= 80 x 80 x 27).

Voxel volume differed across *ascending/descending* standard EPI sequences (12 mm^3) and *random tone* standard EPI sequences (12.65 mm^3). This change was made to provide a compromise voxel size which had an intermediate in-plane resolution (2.33 x 2.33 mm^2) in

between that of the standard (2.0 x 2.0 mm^2) and attenuated (2.75 x 2.75 mm^2) EPI sequences, and was also chosen to create isotropic voxels. However, by replicating all analyses described below using only attenuated EPI acquisition sequences, we confirmed that differences in voxel size were not responsible for any observed differences between *ascending/descending* and *random tone sequences*.

For the attenuated sequence we used a significantly larger voxel volume (22.69 mm^3), which likely increased signal to noise (by a factor of ~1.6) at the cost of lower spatial resolution (Triantafyllou et al., 2005). However this decrease in non-physiological signal to noise seemed not to have a dramatic effect on pRF estimate reliability.

MR Pre-processing

For analysis, data were resampled into 1 x 1 x 1 mm^3 volumetric space. Standard pre-processing of fMRI data was carried out using BrainVoyager QX software (version 2.3.1 Brain Innovation B. V., Maastricht, The Netherlands), including 3D motion correction and high-pass filtering (cut-off: 2 cycles per scan). Functional data were aligned to the T1-weighted anatomical image acquired in the same session (MPRAGE, 1 x 1 x 1 mm^3). The anatomical images acquired in the two sessions were aligned to each other and to each subject's 3D Talairach-normalized anatomical dataset. The BrainVoyager QX automatic segmentation routine was used to reconstruct the cortical surface at the white-gray matter border (with hand-editing to minimize segmentation errors) and the resulting smooth 3D surface was partially inflated.

Voxel Selection

For each subject, large regions of interest (ROIs) were selected from the partially inflated left and right hemisphere cortical surface meshes using drawing tools within BrainVoyager QX. ROI borders were drawn generously to include all voxels within a contiguous region of auditory

cortex between the lateral border on the crown of the superior temporal gyrus, the medial border within the fundus of the lateral sulcus, the posterior border of the supramarginal gyrus, and the anterior border of the most anterior portion of the temporal lobe. Surface ROIs for each subject are shown in **Supplementary Figure 2-3**. These surface ROIs were then mapped back into the brain volume and expanded to include voxels from -1 to 3 mm around the gray-white matter boundary to assure that inaccuracies in the definition of the gray-white matter boundary did not affect the selection of voxels for data analysis. Rather than projecting this surface ROI into the functional volumes, we chose to index the voxels in anatomical space (at the cost of increased computation time). Preprocessed time-course data for each 3D anatomical voxel within the volume ROI were then exported to Matlab for further analysis.

PAC ROI

A functionally defined PAC ROI corresponding to hA1 and hR defined on the basis of the tonotopic gradients and informed by the underlying anatomy (illustrated by the solid black lines in **Figure 2-2, Figure 2-3, Figure 2-5,** and **Figure 2-7)**, was generated to compare bandwidth data across regions. A contiguous patch of cortical surface containing two primary tonotopic gradients centered on low frequencies on HG was manually selected for each hemisphere of individual subjects (Da Costa et al., 2011; 2013). Anterior and posterior borders drawn along the outer high-frequency representations were primarily based on the tonotopic maps generated using the *random tone sequences* and pRF analysis. Quantitative comparisons of the size and similarity of PAC ROI identifications when using either the phase-encoding analysis or the pRF analysis are shown in **Supplementary Figure 2-4**. ROI similarity was assessed using the Jaccard similarity coefficient (Jaccard, 1912) according to the equation:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \qquad (1)$$

where the size of the intersection between voxels included in ROI A and ROI B is divided by the size of the union of voxels included in ROI A and ROI B. Lateral and medial borders were conservatively drawn to include only the medial two-thirds of Heschl's gyrus (Rademacher et al., 2001). Auditory responsive regions outside of this ROI are referred to as non-PAC

Analyses

Phase-encoding analyses

Phase-encoding analyses on the ascending/descending tone progressions were based on Da Costa et al. (2011). Using Brain Voyager, a hemodynamic time course was predicted in response to the first 2 s sound block of each stimulus cycle, based on that individuals estimated hemodynamic response function (HDR, see below for how this was estimated). It was assumed that there was no response to the remainder of the stimulus cycle. This cyclical model function was shifted successively in time in 2 s increments (corresponding to the TR) to generate 14 time-lagged functions. Linear correlation was applied (between all 14 time-lagged functions and the measured fMRI time course) on a per-voxel basis. Each voxel was then color coded according to the lag function resulting in the highest correlation value with its time course ("winner-take-all"). Lag values were then separately averaged for ascending progression scans (4 scans) and the descending progression scans (2 scans) for each voxel within the ROI and were exported to MATLAB (Mathworks, version 7.11). As in the pRF analysis, only voxels with a correlation threshold of 0.10 were considered "tonotopic" and included for further analysis.

pRF analyses

Our pRF model computes the population receptive field for individual voxels based on a linear temporal model of the fMRI BOLD response time course. All population receptive field (pRF) analyses were carried out using custom software written in MATLAB.

The pRF analysis begins with a definition of the stimulus time course *s(f,t),* provided to the model as a matrix of binary values marking the presence or absence of auditory stimulation over frequency and time. To generate a hemodynamically blurred stimulus time course *r(f,t),* the stimulus time course was convolved with the estimated hemodynamic response function modeled as a gamma function *h(t):*

$$h(t) = \frac{\left(\frac{t-\delta}{\tau}\right)^{(n-1)} e^{-\left(\frac{t-\delta}{\tau}\right)}}{\tau(n-1)!} \quad (2)$$

with the initial parameters *n*=3, tau (τ)= 1.5, and delay (δ)=1.8, based on Boynton et al. (1996). The hemodynamic response function (HDR) parameters tau and delay were subsequently optimized for individual subjects (as described below).

$$r(f,t) = s(f,t) * h(t) \quad (3)$$

We estimated the population response using a one-dimensional Gaussian function *g(f),* defined over log frequency. The frequency center (f_0) corresponds to the best frequency, while the standard deviation (σ) was used to estimate bandwidth by transforming the values into octaves and then calculating the full width half maximum (FWHM) of the Gaussian function. To create the predicted time series, we calculated the linear sum of the overlap between the input stimulus after hemodynamic blurring with the Gaussian receptive field for each basis function, *g(f)*:

$$p(t) = \int r(f,t)g(f)\, df \quad (4)$$

Model fits for each voxel were obtained by finding values that maximized the correlation between the predicted and actual fMRI time-courses. The initial parameters for frequency centers

(f_0) spanned the range of the stimulus from 88 to 8000 Hz, and initial standard deviation (σ) values ranged from 0.5 to 4. The best fitting parameters from this set were then used as initial parameters for a nonlinear search algorithm (Matlab's *fminsearch* function) which uses unconstrained nonlinear minimization to find the pRF model parameters f_0 and σ that maximize the correlation between the pRF predicted time-series and the BOLD data.

We then estimated each individual subject's auditory HDR by holding f_0 and σ fixed and finding the best fitting parameters for τ and δ. To limit computational time this estimate was only carried out on a subset of voxels (1 out of every 6) within the anatomically defined ROI, after checking that restricting our estimation to a subset of voxels did not have an appreciable effect on the estimated HDR. Median (across all voxels with a fitted correlation value above 0.25) τ and δ parameters were used to provide the estimate of that individual's HDR. We then iteratively fit the pRF parameters (f_0 and σ) for all voxels within the ROI, using the individually fitted HDR parameters. Individual HDR parameters are reported in **Supplementary Table 2-1**.

The procedure described above had some minor modifications from the original implementation of Dumoulin and Wandell (2008). First, similar to Zuiderbaan et al. (2012) we jointly fit individual HDR and pRF parameters using a single dataset (rather than estimating the HDR using a separate dataset). Second, we convolved the hemodynamic response function before calculating overlap with the pRF rather than afterwards. This is mathematically identical but significantly reduces computation time, since once the HDR parameters have been estimated the convolution of the stimulus image with the HDR only has to be carried out once. Finally, we maximized the correlation between the predicted, $p(t)$, and the obtained time courses rather than minimizing root mean squared error. The amplitude of each Gaussian pRF was estimated after fitting, based on the linear regression between the predicted and obtained time courses.

After fitting, only voxels that met the following criteria were retained for further analyses: (1) the correlation between the observed fMRI response time-course and that predicted by the best-fitting pRF (our goodness-of-fit index) was higher than 0.10 and (2) the standard deviation (σ) of the best fitting pRF was between 0.01-2 in log frequency space, or .0332-6.64 in octaves (chosen based on the resolution and the range of the frequencies presented). These limits were chosen because at bandwidths below 0.0332 the voxel would respond to a single presented frequency, and with bandwidths above 6.64 the voxel would respond similarly to all frequencies presented.

Unlike the "winner-take-all" phase-encoded analysis, the pRF method is not limited to fitting frequency centers within the range of frequencies (88 and 8000 Hz) presented in the stimulus. However frequency center estimates outside this range are likely to be somewhat inaccurate. We included well-fit voxels with frequency center values beyond the imposed frequency cutoffs in **Figure 2-2** and **Figure 2-3** and the histograms in **Figure 2-4**, but simply categorized them as high-pass (HP) or low-pass (LP).

Comparison of different acquisition sequences

The consistency across the two acquisition sequences was calculated as a Pearson correlation coefficient (r_{xy}) across the frequency center and bandwidth estimates from three *random tone sequence* scans of both acquisition types. We additionally computed mean (across subjects) "non-attenuated" correlation coefficients ($r_{x'y'}$) for both frequency center and bandwidth values according to the equation:

$$r_{x'y'} = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}} \qquad (5)$$

where r_{xx} and r_{yy} are the correlation coefficients across single scans of the same acquisition type. Non-attenuated correlation provides a measurement of the correlation across two different scan types having factored out measurement error as estimated using scans of the same acquisition type (Spearman, 1904; Muchinsky, 1996).

Effect of frequency center distribution on bandwidth estimates

Since pRF bandwidth estimates are dependent on the distribution of frequency centers of the underlying receptive fields, we estimated the effects of systematic frequency gradients and hemodynamic spatial blurring on bandwidth values estimated from the *random tone sequence* scans. For each subject, we estimated the slope of the change in frequency center within each voxel, based on the frequency centers of neighboring voxels. For every voxel for which the pRF model was successfully fit, we found the neighboring voxels within a 3mm radius (results were not highly dependent on the choice of radius). We then assumed that the frequency sweep within the voxel of interest consisted of a linear gradient between its best fitting frequency and those of its neighbors to predict the increase in bandwidth size due these factors.

Estimation accuracy

A bootstrapping procedure (to remove variability due to selection of the subset of scans to be included in the estimate) was carried out whereby the standard error of the frequency center or bandwidth was repeatedly estimated for 2-6 of the *random tone sequence* scans, irrespective of acquisition type. Estimation variability was calculated in terms of both standard error of the mean (SEM) and the relative standard error (RSE) of the mean frequency center or mean bandwidth estimate according to the equation:

$$\mathrm{RSE} = \frac{100 \times \mathrm{STD}}{(\bar{x}\sqrt{n})} \qquad (6)$$

Where STD is the standard deviation of the frequency center or bandwidth estimate (corrected for biases due to being a small sample using Cochran's theorem), *n* is the number of scans included in the estimate and $\bar{x}$ is either the mean frequency center or mean bandwidth estimate (Cochran, 1934).

Double Gaussian model

Data were also fit with a double Gaussian center surround model (Zuiderbaan et al., 2012). The surround bandwidth was constrained (using a cost function) to be larger than the center bandwidth and to be less than 6.64 octaves. The amplitude of the surround was not constrained to be negative. The difference in fit between the single and double Gaussian model was assessed using a nested model F-test. The original single Gaussian model had 2 free parameters (frequency center and bandwidth, with the amplitude fixed at 1) while the double Gaussian model had 2 additional parameters describing the width and the amplitude of the "surround".

Results

Tonotopic maps as a function of both stimulus and analysis methods

Figure 2-2 shows tonotopic maps using the *ascending* and *descending tone progressions* on the left hemisphere cortical surface mesh for two typical subjects. Phase-encoded maps were generated using either an average of four scans of the *ascending tone progressions* (left panel) or an average of two scans of the *descending tone progressions* (middle panel). Maps using the pRF method for *ascending* and *descending tone progressions* (right panel), were generated from all six (equally weighted) functional scans. Separate maps for the right hemisphere of these subjects are shown in **Supplementary Figure 2-5.**

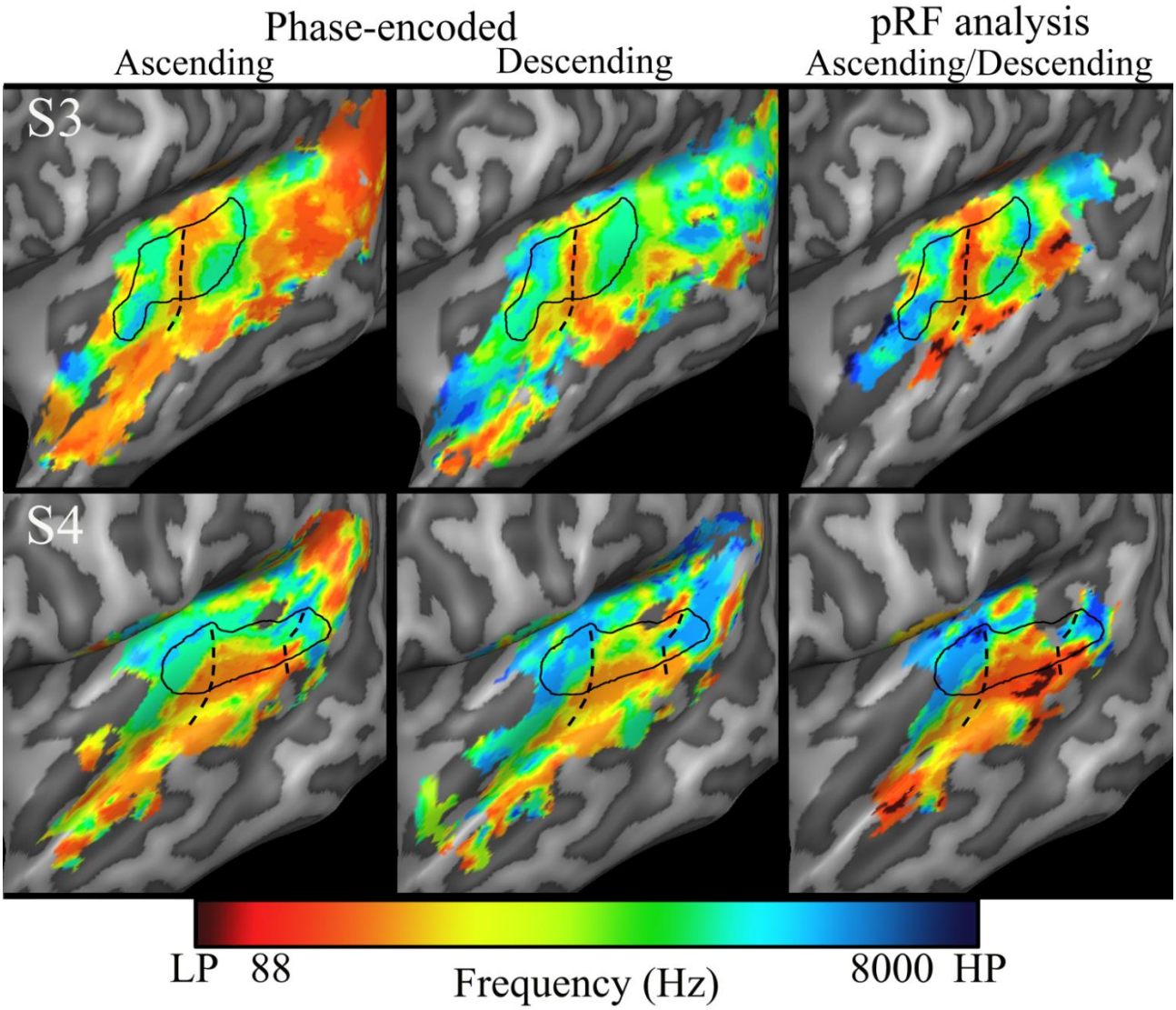


Figure 2-2. Tonotopic Maps using Ascending and Descending Progressions

Tonotopic maps for the left hemisphere of Subject 3 (first row) and Subject 4 (second row). Maps were calculated using either: *ascending tone progressions* (4 scans) analyzed using phase-encoding (left panels), *descending tone progressions* (2 scans) analyzed using phase-encoding (middle panels), or *ascending/descending tone progressions* (6 scans, therefore weighted towards the *ascending tone progressions*) analyzed using the pRF method (right panels). Frequency center (f_0) values are color-coded along a gradient, with red corresponding to the lowest frequency value (88 Hz) through blue corresponding to the highest frequency (8000 Hz). Low-pass (LP) voxels are colored dark red, while high-pass (HP) voxels are colored dark blue. The borders of the PAC ROIs are designated by the solid black lines. The crowns of gyri are indicated by the dashed black lines. Across all figures/analyses the same correlation threshold of 0.10 was used, unless otherwise stated.

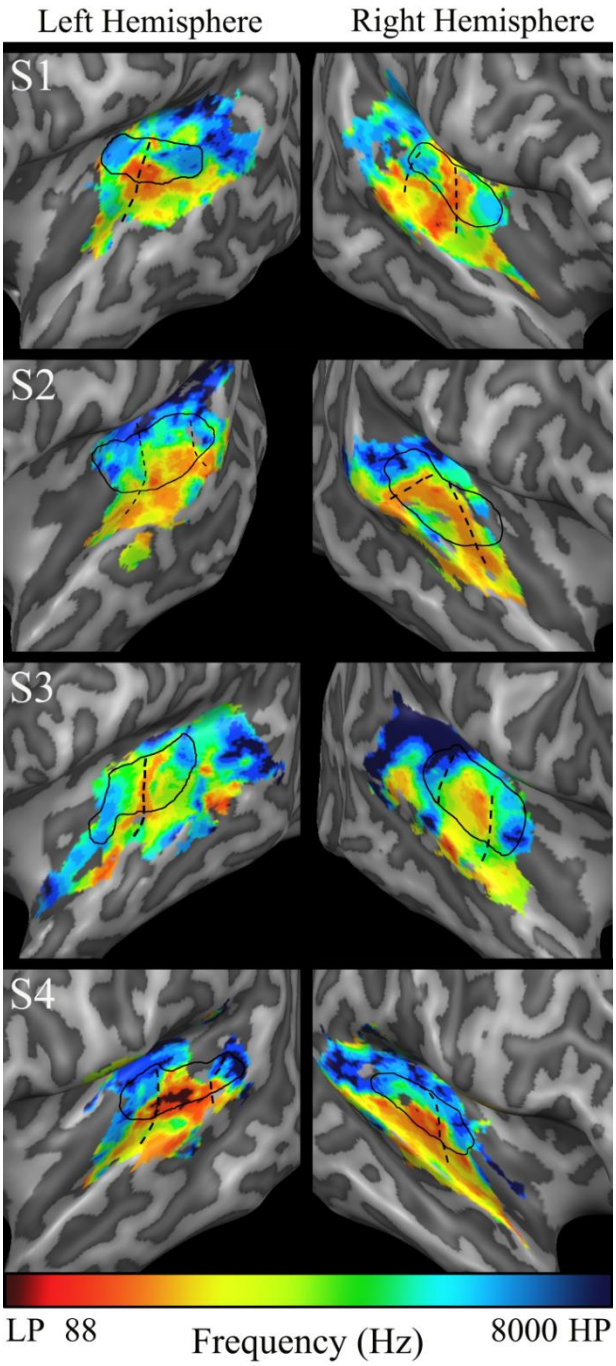


Figure 2-3. Tonotopic Maps using Random Tone progressions

Frequency center (f_0) values are color-coded along a gradient, with red corresponding to the lowest frequency value (88 Hz) through blue corresponding to the highest frequency (8000 Hz). Low-pass (LP) voxels are colored dark red, while high-pass (HP) voxels are colored dark blue. The borders of the PAC ROIs are designated by the solid black lines. The crowns of gyri are indicated by the dashed black lines.

Tonotopic maps generated using the *random tone sequences* and the pRF method from an average of six (equally weighted) functional scans are shown for both hemispheres of all four subjects in **Figure 2-3**. All maps are based on scans from both the standard and attenuated EPI acquisition protocols since, as described below, frequency center estimates were highly correlated across the two types of acquisition protocol. The frequency center values of individual voxels are color-coded along a gradient, with red corresponding to the lowest frequency value (88 Hz) through blue corresponding to the highest frequency tested (8000 Hz), as in **Figure 2-1**. Low-pass (LP) voxels are colored dark red, while high-pass (HP) voxels are colored dark blue. Across all figures/analyses the same correlation threshold of 0.10 was used. Across both stimulus types and analysis methods, clear mirror-symmetric tonotopic gradients, corresponding to the PAC subdivisions hA1 and hR, are visible. These

mirror-symmetric gradients reversed at a low frequency border, centered either on the crown of HG (Subject 1 and 3 in **Figure 2-3**) or within the sulcus intermedius (SI) between the duplications of HG (Subject 2 and 4 in **Figure 2-3**). This low frequency region was flanked both caudally (hA1) and rostrally (hR) by high frequency regions creating a "high-low-low-high" gradient reversal extending across HG. In certain cases, areas hA1 and hR appear to be oriented in a V-shape (Subject 2 and 4 in **Figure 2-3**) with the high frequency endpoints situated more medially than the low frequency center. The maps generated for *ascending* and *descending progressions* and analyzed with either phase-encoding or pRF analysis are very similar, as can be seen in **Figure 2-2** and here in **Table 2-1**.

Table 2-1. Correlation coefficients between frequency center (f_0) values

	Subject			
	S1	S2	S3	S4
PE_A vs pRF_{AD}	0.713	0.857	0.682	0.861
PE_D vs pRF_{AD}	0.805	0.861	0.758	0.830
pRF_{AD} vs pRF_R	0.534	0.892	0.765	0.785

Correlation coefficients between frequency center (f_0) values estimated using different stimulus sequences (A: *ascending progressions*, D: *descending progressions*, AD: *ascending/descending progressions*; R: *random tone sequences*) and analysis methods (PE: phase-encoding; pRF: pRF analysis). In each case correlation coefficients were calculated after having calculated frequency centers using scans collapsed across acquisition type. Separating data acquired using standard vs. attenuated acquisition EPI sequences did not change the pattern of results (data not shown).

While the general position of the gradient reversal was very consistent across *ascending/descending* and the *random tone sequence* stimuli, there are noticeable differences in the distribution of frequencies between the maps even when both are analyzed using the pRF method (**Figure 2-2** and **2-3** and **Table 2-1**).

Figure 2-4 shows stacked (across subjects) probability histograms of frequency center (f_0) values estimated for both stimulus types and analysis methods, collapsed across acquisition protocols.

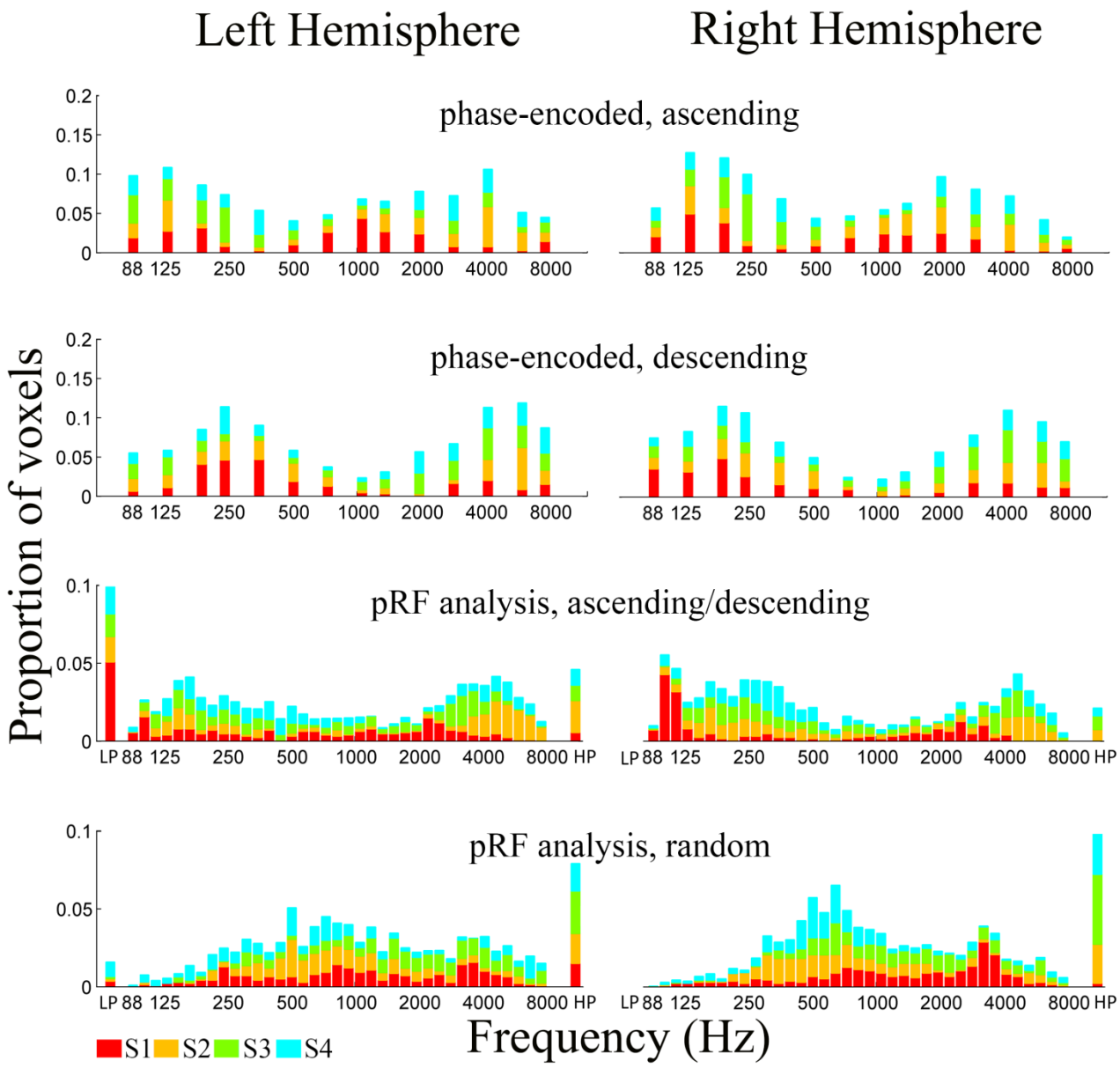


Figure 2-4. Probability histograms of frequency center (f_0) values
Probability histograms of frequency center (f_0) values for voxels in left and right hemispheres, stacked across subjects for the tonotopic maps of **Figures 2-2** and 2-**3** calculated for *ascending progressions* analyzed using phase-encoding (first panel), *descending progressions* analyzed using phase-encoding (second panel), *ascending/descending tone progressions* analyzed using the pRF method (third panel), or *random tone progressions* analyzed using the pRF method (fourth panel). The number of voxels is normalized so each subject contributes equally. The colors correspond to the individual subjects.

The histograms for the *ascending* and *descending tone progression* stimuli show a strong double peak with very strong representations of higher and lower frequencies – this double peak was visible when these data were analyzed using either "winner-take-all" phase encoding or the pRF analysis . This double peak was also robust to the choice of threshold – remaining equally salient across thresholds ranging between 0.1 (**Figure 2-4**) through 0.35. This double peak was even more noticeable in single scan data. Because the "winner-take-all" model did not allow for lag values greater than 14 or less than 1, noise within individual scans is not zero mean. That pushes f_0 away from the boundary values (towards the center) when averaging across multiple scans. In contrast, the histogram for the *random tone sequence* stimuli has a different shape, with a heavier representation of mid-range frequencies. In addition, as might be expected from the approximate range of human hearing (20 Hz-20 kHz) and primate physiology data (Cheung et al., 2001) a reasonable proportion of voxels were characterized as high pass (HP), while few voxels were characterized as low pass (LP).

pRF measurements of bandwidth

Bandwidth measurements from *random tone sequences* are shown in **Figure 2-5**. Bandwidth estimates are not reported for the *ascending/descending sequences* for reasons discussed below. Bandwidth maps were generated from an average of all six *random tone sequence* functional scans using both the standard and attenuated EPI acquisition protocols since, as described below, the two types of acquisition protocol produced similar bandwidth estimates. For bandwidths between 1-8 octaves, color coding of the bandwidth map is linear. Narrow bandwidths (< 1 octave) were color coded as red, and broad bandwidths (> 8 octaves) were coded as blue.

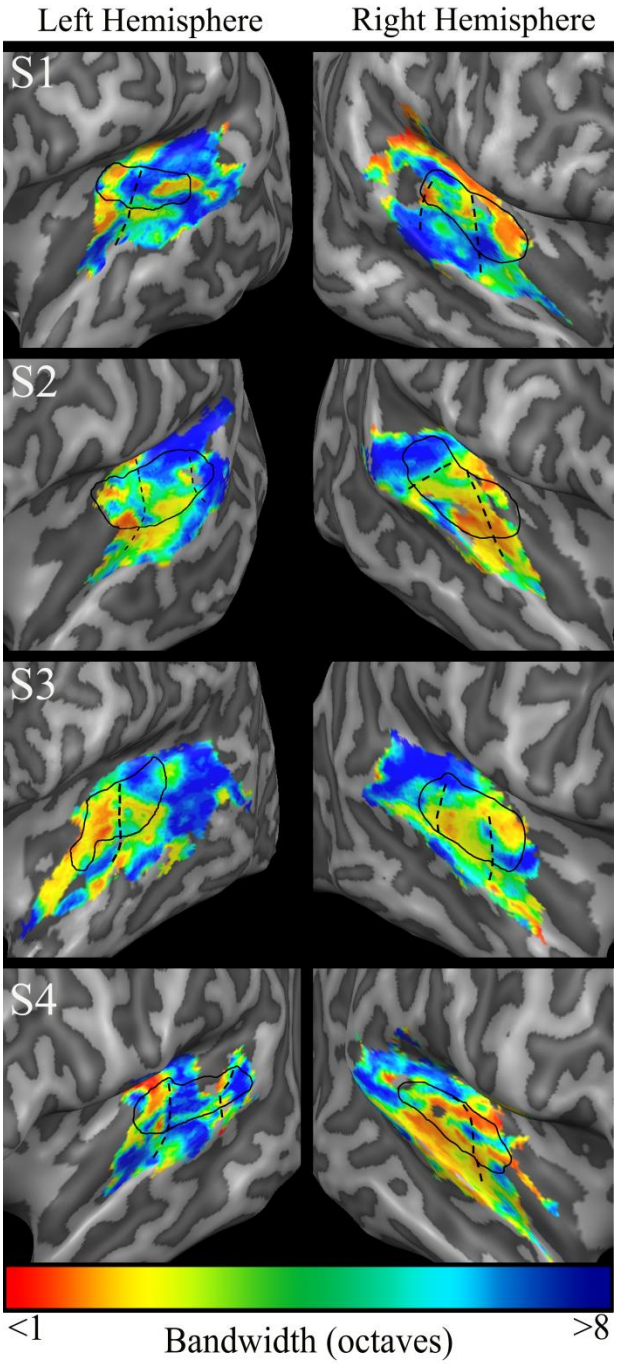


Figure 2-5. Bandwidth maps

Bandwidth maps for all four subjects. Bandwidth values are color-coded along a gradient with red corresponding to the narrow bandwidth values (<1 octave) through blue corresponding to the broader bandwidth values (>8 octaves).

We did not find a significant correlation between bandwidth and frequency center values (r = 0.046). Nor did we find a clear bandwidth gradient running orthogonal to the tonotopic gradient for any individual subjects (using a gradient analysis similar to that used by Petkov et al. (2006), data not shown). However, as shown in **Figure 2-6,** we did find that bandwidth values were smaller in the PAC than in non-PAC regions. A 2-way ANOVA testing Area (PAC vs. Non-PAC) x Hemisphere (Left vs. Right) found a significant main effect of Area [F (1, 12) = 16.36 p = 0.002] on mean bandwidth value. There were no significant main effect of Hemisphere [F (1, 12) = 0.54, p < 0.476] and no significant interaction of Area vs Hemisphere [F (1, 12) = 0.07, p < 0.8017]. While bandwidths were overall narrower in PAC than in surrounding non-PAC areas, our findings do not support the use of bandwidths as the sole criteria for defining PAC borders since there was significant

heterogeneity of bandwidths within our tonotopically defined PAC ROI. We estimated the effects of systematic frequency gradients and hemodynamic spatial blurring on bandwidth size. These estimates predicted that shifts in frequency preference as a function of distance across the cortical surface is likely to have increased our bandwidth estimates by a median value of ~10%.

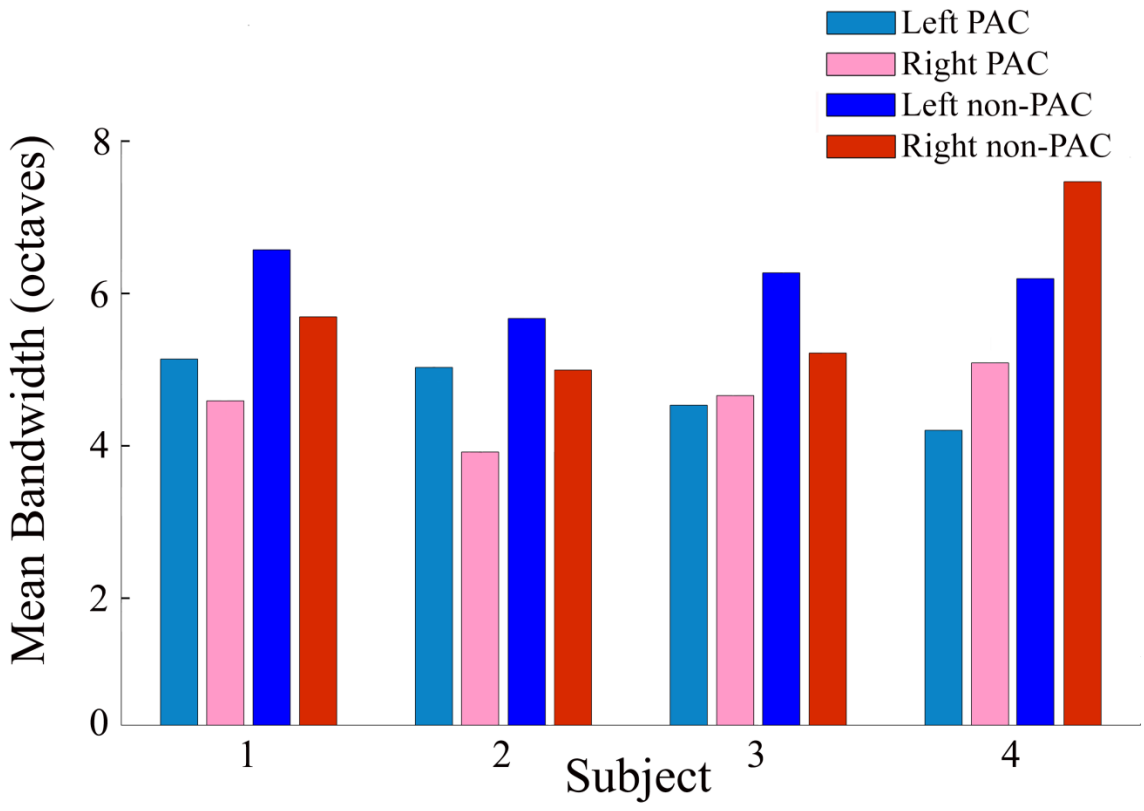


Figure 2-6. Bandwidth values across cortical areas
Bandwidths are narrower within the PAC than in outlying non-PAC areas. As described in the main text, a 2-way ANOVA found a significant main effect of cortical area (PAC vs. non PAC) on mean bandwidth value.

Finally, we found a positive correlation between activation levels (GLM t-value, shown in **Supplementary Figures 2-6** and **2-7**) and bandwidth estimates (*random tone sequences* only), as shown in **Supplementary Figure 2-8**. Broader tuning results in a higher activation level across the duration of the sound condition, leading to stronger responses in the sound vs. silence contrast.

Comparison of different acquisition sequences

Tonotopic maps and bandwidth maps looked very similar across standard and attenuated EPI sequences. **Figure 2-7** shows an example from Subject 3, generated from the *random tone sequences* and fitted using the pRF analysis technique.

The scatter plots of **Figure 2-8** directly compare frequency center and bandwidth values between the two acquisition sequences across all voxels that met our selection criteria for all four subjects. Neither the maps, nor the scatter plots show any obvious mis-estimation of frequency values near the peak of the scanner noise.

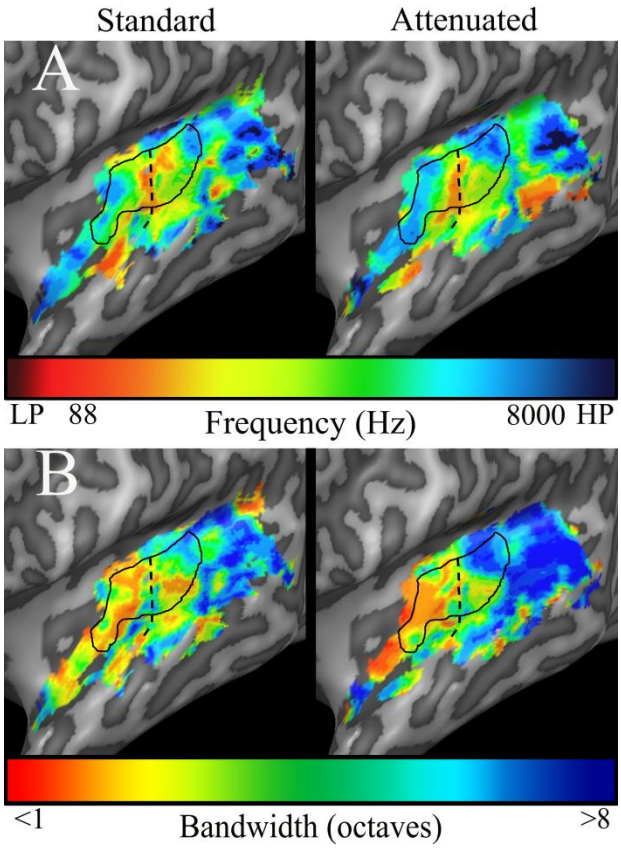


Figure 2-7. Maps across Acquisition Sequences

(A) Tonotopic maps for example Subject 3 calculated using random tone progressions averaged across separate data sets of three scans using a standard EPI sequence and an attenuated EPI sequence, analyzed using the pRF method. Color coding is the same as **Figure 2- 3**. **(B)** Bandwidth maps for the same data and analysis. Color coding as in **Figure 2-5**.

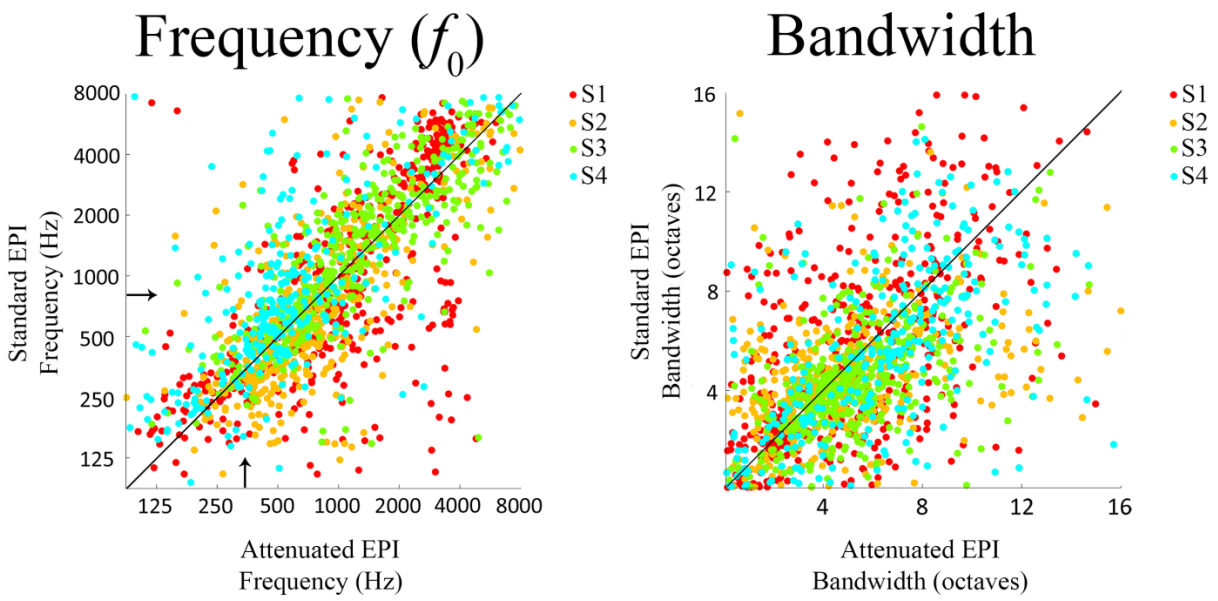


Figure 2-8. Across Acquisition Types

Scatter plots showing the correlation of frequency center (f_0) (left panel) or bandwidth values (right panel) between the two acquisition sequences. Each data point represents a single voxel whose correlation (between the predicted and obtained time-course, calculated across the three scans of standard the attenuated acquisition protocols) was larger than our threshold of 0.1 for both acquisition sequences. The colors correspond to the individual subjects.

Individual subject correlation coefficients between three scans of the attenuated EPI sequence and three scans of the standard EPI sequences are shown in **Table 2-2**. The mean (across subjects) non-attenuated correlation coefficient (which represents the correlation after parsing out measurement error) between standard and attenuated sequences was 0.813 (varying between 0.742-0.965 across subjects) for frequency centers and 0.726 (varying between 0.566-0.813 across subjects) for bandwidth estimates.

Table 2-2. Correlation Coefficients

Subject	frequency (f_0)	bandwidth
S1	0.716	0.590
S2	0.745	0.401
S3	0.797	0.484
S4	0.711	0.560

Correlation coefficients for frequency center (f_0) values and bandwidth estimates from *random tone sequence* data collected using standard vs. attenuated acquisition protocols for all subjects. In each case correlation coefficients were calculated after having calculated pRFs using all three scans of a given acquisition type.

Mean correlation coefficients between single scans of the same or different type of EPI sequence are shown in **Supplementary Table 2-2**. While frequency center and bandwidth estimates were similar across the two acquisition sequences, further testing, including presenting stimuli using a sparse sequence, will be necessary to determine the extent to which acoustic scanner noise affects pRF estimation.

Estimation accuracy

Figure 2-9 shows pRF estimation variability as a function of the number of scans included in the estimate, for both frequency center (f_0) and bandwidth. PRF estimation variability is presented in terms of standard error (SEM) of the mean frequency center (f_0) or mean bandwidth estimate (leftward y-axis) and the percentage standard error (% STE) of the mean frequency center (f_0) or mean bandwidth estimate (rightward y-axis).

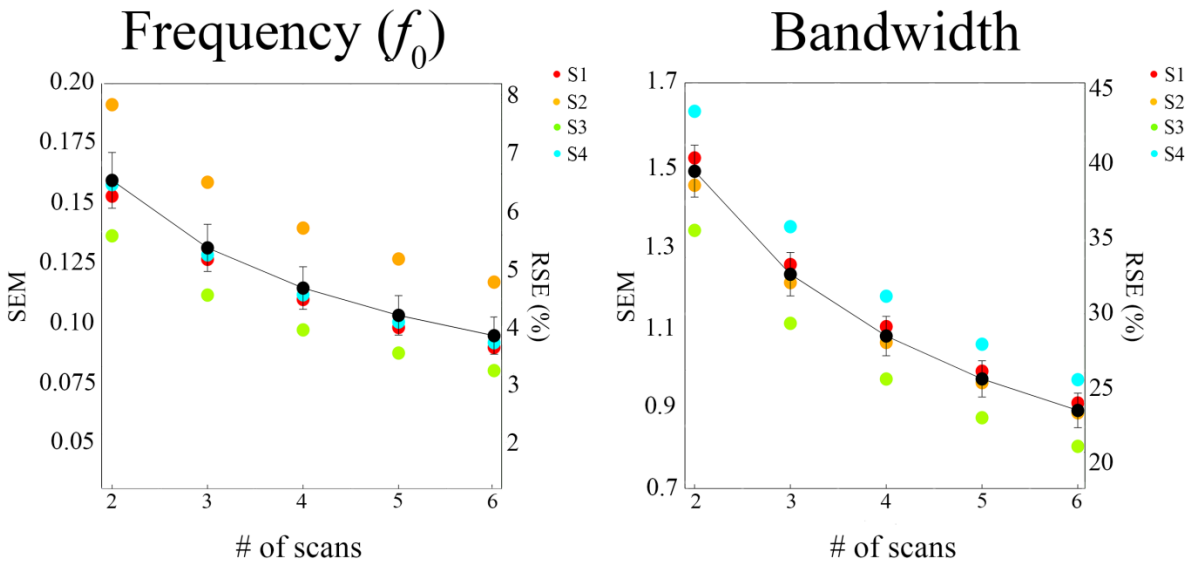


Figure 2-9. Estimated Standard Error

Estimated standard error of frequency (f_0) (left panel) and bandwidth (right panel) estimates as a function of the number of scans. The x-axis represents the number of scans included in the estimate, the leftward y-axis represents the standard error of the mean, and the rightward y-axis represents the relative standard error, as described in the main text. The colored markers correspond to the individual subjects. The mean (across subjects) is shown in black and error bars represent standard error across subjects.

Frequency estimates were remarkably reliable – with only 2 scans (less than 9 minutes each) the estimated standard error was less than 5% of the estimated frequency center (f_0), suggesting that a single scan is adequate to obtain a basic tonotopic map. Unsurprisingly, we find that bandwidth estimates were less accurate than frequency estimates: for our dataset a full session was required to obtain bandwidth estimates with standard error variability of less than 25%. Although we corrected for overlap in the sampled distributions using the Cochran correction, given the small number of scans on which these simulations were carried out, these estimates should be considered a rough guideline and may underestimate variance.

Double Gaussian model

The difference between the single and double Gaussian model was assessed using a nested model F test that examined the percentage of voxels that were fit significantly better using a double Gaussian model. In 2 of the four subjects less than 5% of voxels were better fit by the double Gaussian model – suggesting that the model did not provide any additional explanatory power. For Subjects 1 and 3, 14% and 17% of voxels respectively were significantly better fit with the double Gaussian model. For voxels that were significantly better fit by the double Gaussian model, Subject 1 had a median improvement of 33% (goodness-of- fit value of 0.24 for the single Gaussian model and 0.32 for the double Gaussian model). Subject 3 had a median improvement of 36% (goodness-of- fit value of 0.22 for the single Gaussian model and 0.30 for the double Gaussian model). For these better fit voxels, the median amplitude of the inhibitory surround compared to the center amplitude (fixed at 1) was -0.51 for Subject 1 and -0.50 for Subject 3 (note that the surround was not constrained to be negative).

Discussion

We show here that a modified version of Dumoulin and Wandell's (2008) pRF technique can be used to generate tonotopic and bandwidth maps of human auditory cortex. When using traditional *ascending/descending tone progressions* this method results in tonotopic maps that are very similar to those obtained using traditional phase-encoding analysis methods. However, unlike phase-encoding methods, the pRF method does not require an orderly sequence; instead, functional maps of auditory cortex may be derived from responses to a wide array of stimuli, including random or pseudo-random sequences. While maps generated using *random tone progressions* identified roughly similar tonotopic areas as those obtained using *ascending/descending tone progressions*, the exact boundaries of these areas, and the representation of frequencies within these areas was somewhat different.

One advantage of the *ascending/descending tone progression* stimuli is that the maps generated from these stimuli are very robust. This makes progression stimuli well suited for determining whether an area is tonotopic and/or providing rough estimates of area boundaries based on frequency reversals when scan time is limited. However, as previously described, *ascending/descending tone progressions* are susceptible to habituation and/or expectation effects (Kastner et al., 1999) as well as other types of spatiotemporal nonlinearities (Zenger-Landolt and Heeger, 2003; Pihlaja et al., 2008) due to the "traveling wave" of BOLD activity on the cortical surface. This "traveling wave" improves signal to noise: goodness-of-fit index values were higher for *ascending/descending tone progressions* than for *random tone sequences*. However these effects also result in an over-emphasis of frequencies near the beginning or the end of the sweep (Duncan and Boynton, 2003; Dumoulin and Wandell, 2008; Haak et al., 2012; Binda et al., 2013; Langers et al., 2014a; 2014b). We therefore do not report bandwidth estimates for the

ascending/descending sequences as these effects can result in an overestimation of receptive field size, probably due to spatiotemporal blurring. Additionally the *ascending/descending sequences* had a half-octave separation between frequency blocks. This under-sampling made it impossible to accurately estimate narrower bandwidths (estimated bandwidths became infinitely small). In contrast, *random tone sequences* do not suffer from these biases and likely provide a more accurate method for measuring the finer-scale tonotopic and bandwidth organization.

The organization of tonotopic gradients and tuning bandwidth

Like previous studies, we identified two tonotopic gradients, thought to correspond to hA1 and hR (Formisano et al., 2003; Da Costa et al., 2011; Langers and van Dijk, 2011; Moerel et al., 2012; Da Costa et al. (2013)). These "high-low-low-high" mirror-symmetric gradients followed the morphology of the cortex, with the more posterior gradient corresponding to hA1 and the more anterior gradient corresponding to hR (Humphries et al., 2010; Da Costa et al., 2011; Langers and van Dijk, 2011; 2013).

Ours is the first fMRI study to examine bandwidth using the traditional psychoacoustic approach of randomly presented pure tones. A variety of plausible organizations for bandwidth have been suggested in the literature, for example: (1) bandwidth varies systematically with best frequency (Recanzone et al., 1999; Cheung et al., 2001), (2) bandwidth gradients run orthogonal to tonotopic gradient (Petkov et al., 2006; Kayser et al., 2007b), and (3) tuning widths are narrower within the PAC than surrounding non-PAC regions (Petkov et al., 2006; Seifritz et al., 2006; Moerel et al., 2012).

While we did not find a significant correlation between bandwidth and frequency center values nor a clear bandwidth gradient, we did find that bandwidth values were significantly

smaller within PAC than in surrounding non-PAC regions. This is consistent with evidence from animal electrophysiology (Tian and Rauschecker, 1994; Rauschecker et al., 1995) and neuroimaging (Petkov et al., 2006; Kayser et al., 2007b). Additionally, a recent human study by Moerel et al. (2012) found some indication that bandwidths were narrower along HG. Here we find that the area of narrow bandwidths extends beyond the gyrus itself. However, there was some heterogeneity of bandwidths within our tonotopically defined PAC, with regions of broad bandwidth consistently contained within the PAC borders. Therefore, unlike Moerel et al. (2012) we do not limit the PAC borders to the areas of narrowest tuning bandwidth.

One caveat is that bandwidth estimates are more variable than frequency estimates, even when based on a full session of data. We believe this is not a limitation of our particular methods, since the reliability of our frequency estimates compare favorably with previous studies using similar techniques (Petkov et al., 2006; Humphries et al., 2010; Da Costa et al., 2011; Striem-Amit et al., 2011; Moerel et al., 2012). Thus, it is likely that additional methodological advances might reveal further topographical structure in bandwidth organization.

What do auditory population receptive fields represent?

It is not straightforward to extrapolate single neuron tuning-width data from responses within human fMRI voxels, since voxel-level BOLD signals include the responses of hundreds of thousands of neurons collectively across cortical layers and neuronal types (10^4-10^5 neurons per cubic mm in cortex). As described in Dumoulin and Wandell (2008), pRF estimates depend on the average receptive field of the population of neurons that drive the BOLD response. This population may be a biased sample of the underlying population. For example, the neurons driving the response may depend on the properties of the stimulus being used – it is possible that

different pRF estimates will be obtained using band pass noise or naturalistic stimuli (Moerel et al., 2012) than with our pure tone frequency stimulus.

Dumoulin and Wandell (2008) found that pRF estimates of visual receptive field size were systematically larger than single unit recording estimates, but were relatively closely matched to local field potential (LFP) estimates of receptive field size. Similarly, our bandwidth estimates were generally larger than those from single-unit recordings, which have been shown to range between one tenth of an octave to wider than five octaves in primates (Recanzone et al., 1999; Cheung et al., 2001), but were consistent with those measured using auditory field potentials (Kayser et al., 2007a).

The use of forward models to examine auditory processing

Similar to other "forward models", the pRF method models the response properties of individual voxels, an approach that is becoming increasingly popular, see Naselaris et al. (2011) for a review. One advantage of such forward models is that they allow for flexibility in the choice of stimulus sequence. As previously mentioned, Moerel et al. (2012) used a forward model to examine tonotopic organization using naturalistic stimuli and found that the regions that showed preferential responses to human voice and speech were tuned to the lower range of the frequency spectrum. This tuning was apparent using both natural sound and simple tone stimuli. In our case we took advantage of this stimulus flexibility to examine tonotopic mapping using a more traditional psychoacoustic approach, using stimuli that consisted of randomly ordered pure tones.

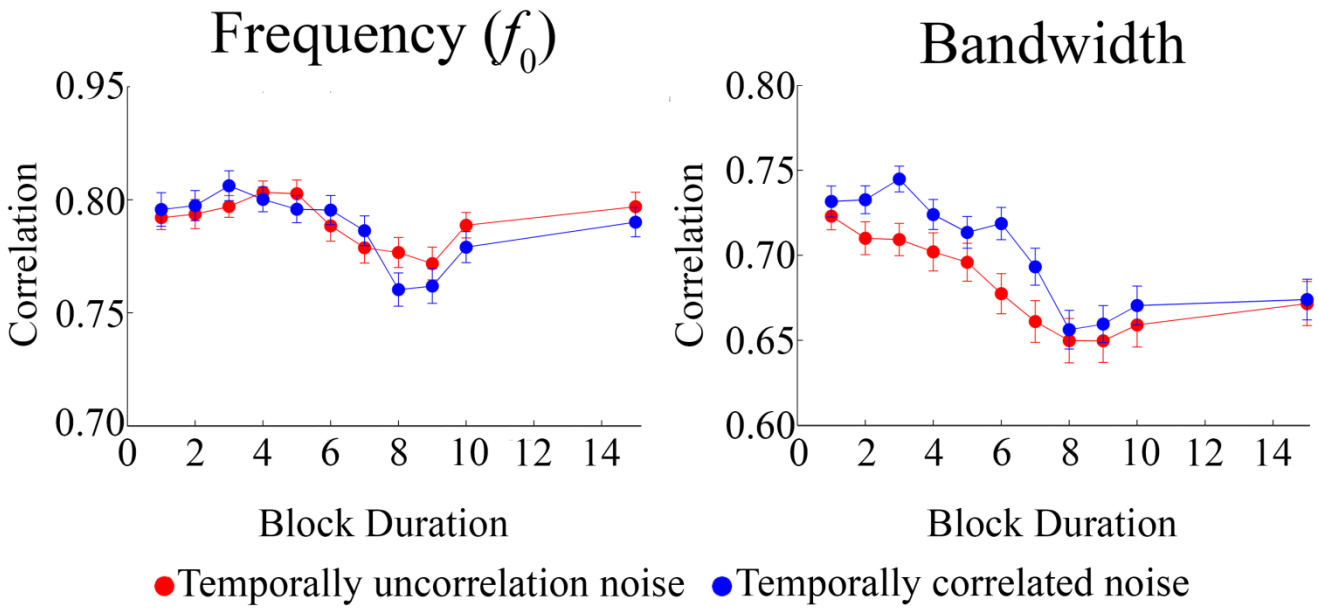
A second advantage of forward models is that they allow for the modeling of tuning properties of individual voxels. In our case, because we were specifically interested in frequency tuning and bandwidth, we used a very simple model, assuming a Gaussian in log space. In

contrast, Moerel et al. (2012) used a two-stage model that allowed them to directly estimate the spectral tuning of the voxel. However, to quantify frequency and bandwidth tuning a Gaussian was fit to this spectral response profile– thereby reducing their model to a close equivalent of ours. More recently, this same group did away with fitting a Gaussian to characterize multi-peaked spectral tuning profiles, as opposed to single bandwidth value (Moerel et al., 2013).

It is also possible to create more complex pRF models than a simple Gaussian – an approach intermediate between the spectral tuning model and the simple Gaussian pRF model we focus on in this paper. The improved fits for the double Gaussian model found within a subset of voxels in 2 of our 4 subjects suggest that some additional power may be obtained by permitting more complex frequency response profiles than a simple Gaussian (Zuiderbaan et al., 2012). Other interesting future directions will include presenting multiple tones at once, thereby allowing measurement of response saturation (Kay et al., 2013) and the inclusion of alternative temporal response profiles including onset and offset responses (Harms and Melcher, 2003). One of the advantages of the pRF approach is that it provides an elegant framework within which to assess whether more complex models can provide additional insight into auditory BOLD responses.

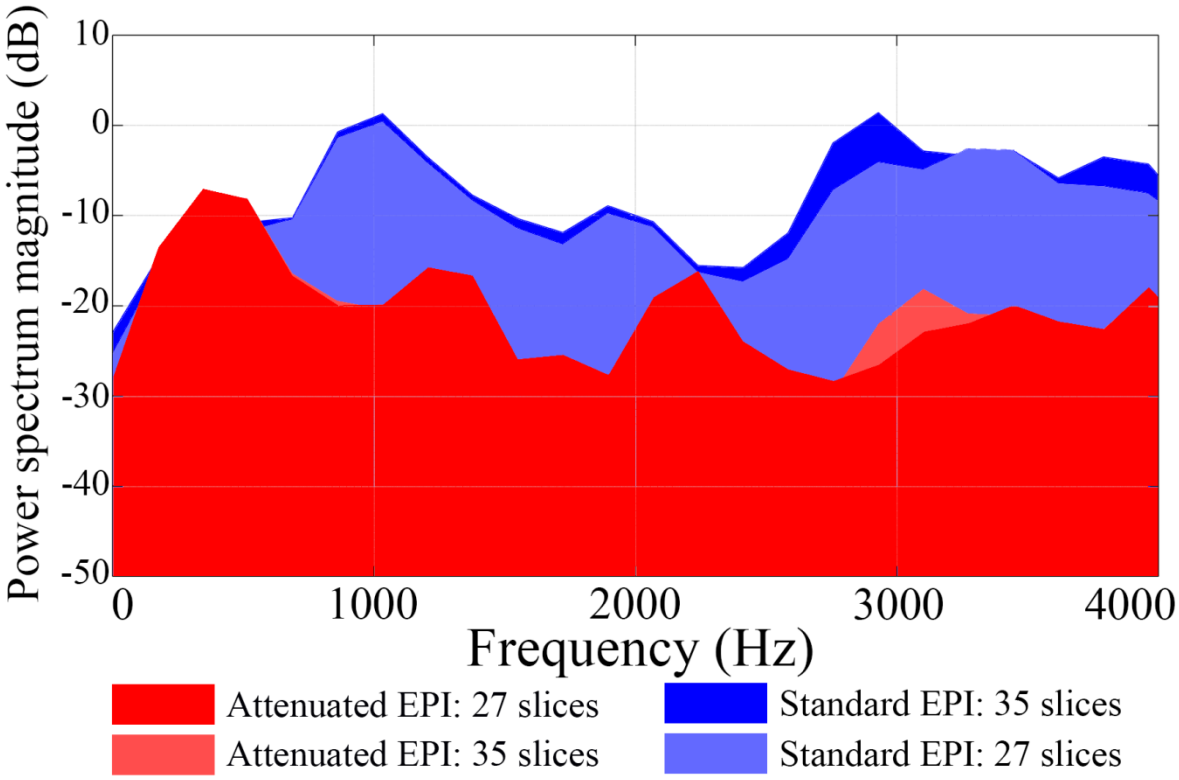
The pRF method described here also provides a natural methodology to examine differences in tuning as a function of various factors known to affect auditory processing such as attention (Da Costa et al., 2013), auditory training (Menning et al., 2000; Jäncke et al., 2001), musical experience (Pantev and Herholz, 2011), or loss of vision early in life (Elbert et al., 2002).

Supplementary Material



Supplementary Figure 2-1. Simulation Results

Results of a simulation examining the effects of frequency block duration on signal to noise. We assumed a fixed TR of 2 s and a fixed scan length of 240 s. The auditory stimulus was modeled as a series of frequencies in random order, presented in tone bursts (blocks) of 1, 2 …10 or 15 s duration. Each voxel's population receptive field was modeled as a Gaussian with a frequency center and bandwidth selected from the range found in our data. We assumed adaptation of the BOLD signal as a function of block duration based on data from Soltysik et al. (2004) who measured adaptation using a stimulus was roughly similar to ours: 440 Hz tones, each lasting 125 ms followed by a rest period of 125 ms, producing four tones per second. Two model predictions of the effects of tone duration are shown. For the *temporally uncorrelated noise* model (red curve) the stimulus was convolved with the auditory HDR (tau =1.5, delay = 1.8) and the simulated time series was calculated as the linear sum of the overlap between the input stimulus after hemodynamic blurring with the Gaussian receptive field of that voxel. Zero mean Gaussian noise was then added to the time series. For the *temporally correlated noise* model (blue curve) zero mean Gaussian noise was added to the stimulus. This noisy stimulus was convolved with the auditory HDR (producing temporally correlated noise) and the simulated time series was calculated as the linear sum of the overlap between the noisy input stimulus after hemodynamic blurring with the Gaussian receptive field of that voxel. For both models, the amplitude of the noise was chosen to produce signal to noise similar to that found in our dataset for the 2 s block duration. As in the main paper, model fits for each voxel were obtained by finding the frequency center and bandwidth values that maximized the correlation between predicted and simulated fMRI time-courses responses in each voxel over time. For each frequency block simulations for *temporally uncorrelated* and *temporally correlated noise* were repeated 60 times: each repetition consisted of a different set of 100 voxels and a different random tone sequence. The x-axis represents frequency block duration and the y-axis represents the mean cross-correlation (over the 100 voxels) between simulated and estimated pRF frequency center and bandwidth values across the 60 repetitions. Error bars represent standard error of the mean across repetitions. The effect of block duration was similar for *temporally correlated* and *temporally uncorrelated noise* (at least for the degree of temporal correlation that we simulated): according to both models the improvement in estimation accuracy as a function of block duration is highest between 1 to 5 s.

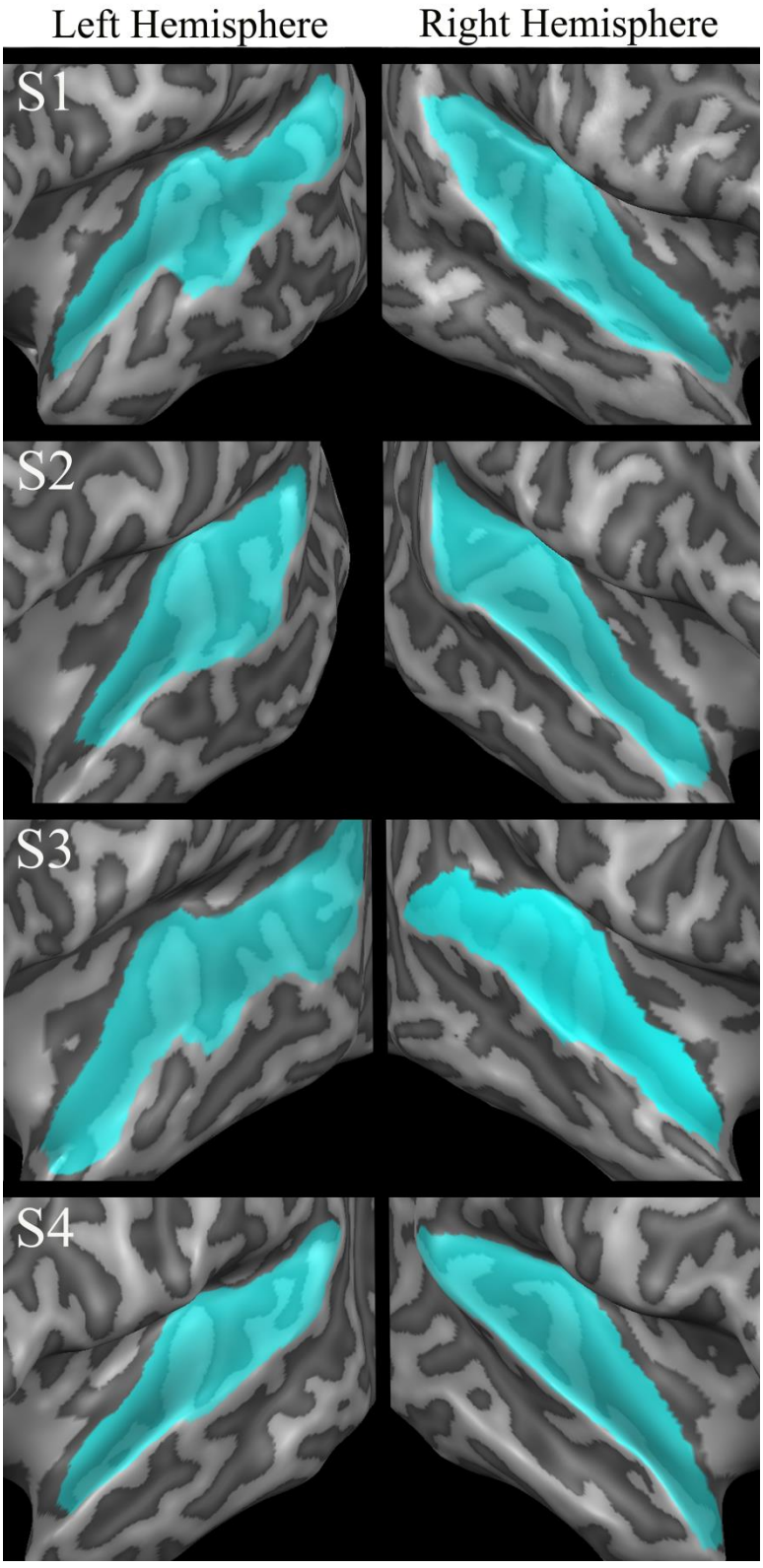


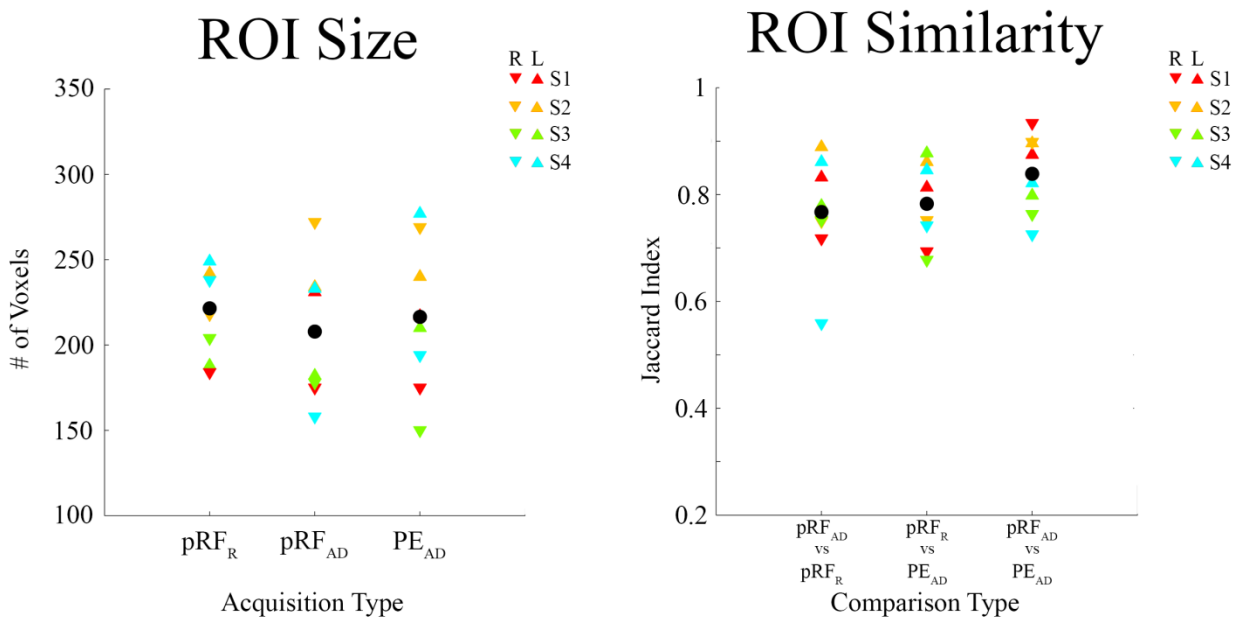
Supplementary Figure 2-2. Power Spectrums of Scanner Noise

A power spectrum of the recorded scanner noise for both the attenuated EPI (dark red) and standard EPI (dark blue) sequences. Noise recordings were made from inside the coil as described in the main text. A frequency spectrum of the recorded scanner noise was obtained for each acquisition type by performing a discrete Fast Fourier Transform and calculating the resulting normalized frequency component magnitudes. For demonstration purposes only, measurements of scanner noise were generated for both acquisition types using both 35 slices and 27 slices. Data in the main paper were collected using 35 slices for the standard EPI sequence which produced an overall louder (122 dBA) auditory scanner noise peaking at approximately 1000 Hz, with a normalized frequency component magnitude around 0 dB; and 27 slices for the attenuated EPI sequence which produced an overall quieter (105 dBA) scanner noise peaking at approximately 350 Hz with a normalized frequency component magnitude around -8 dB.

Supplementary Figure 2-3. Anatomical ROIs

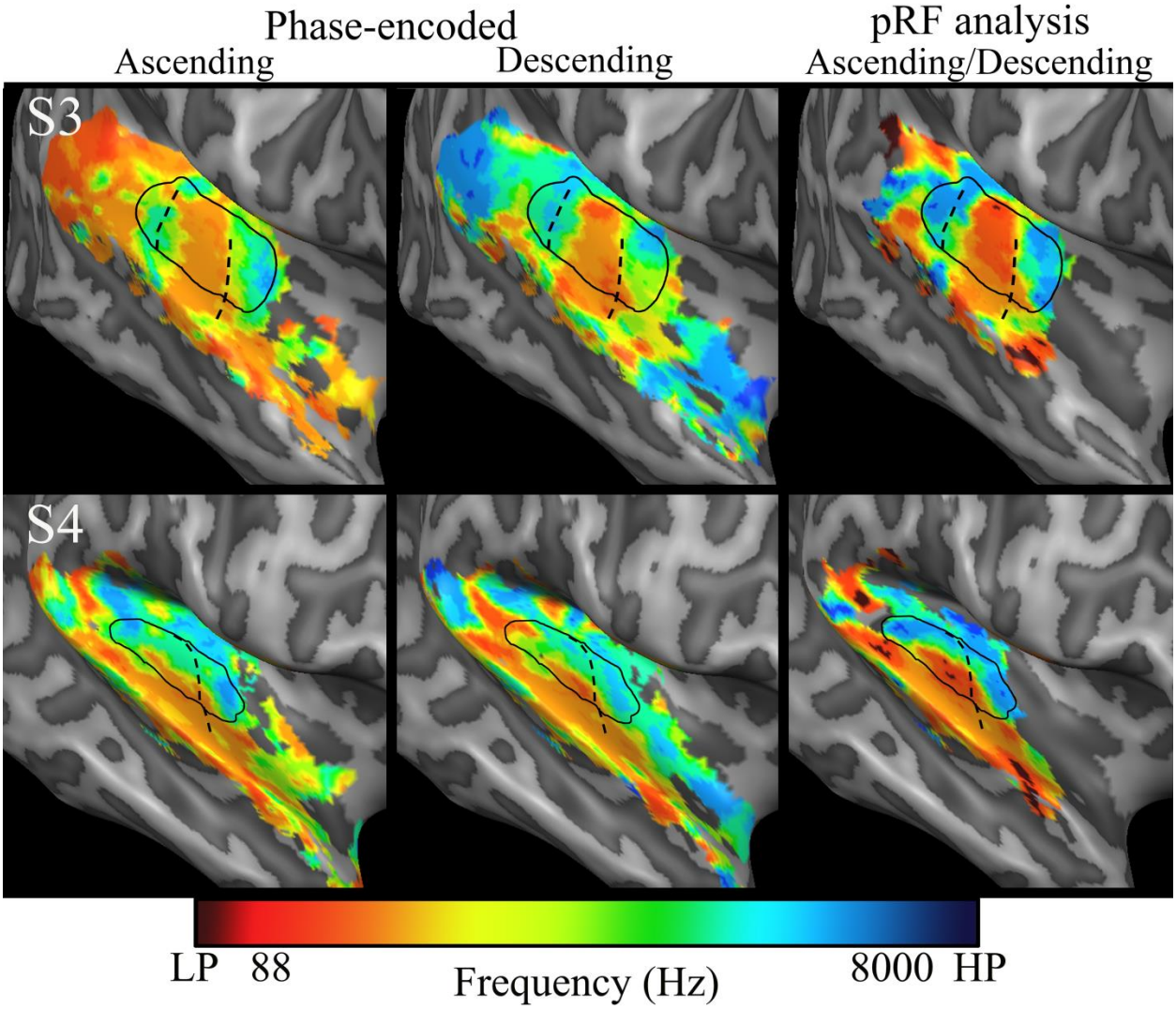
ROI borders were drawn generously to include all voxels within a contiguous region of auditory cortex between the lateral border on the crown of the superior temporal gyrus, the medial border within the fundus of the lateral sulcus, the posterior border of the supramarginal gyrus, and the anterior border of the most anterior portion of the temporal lobe.





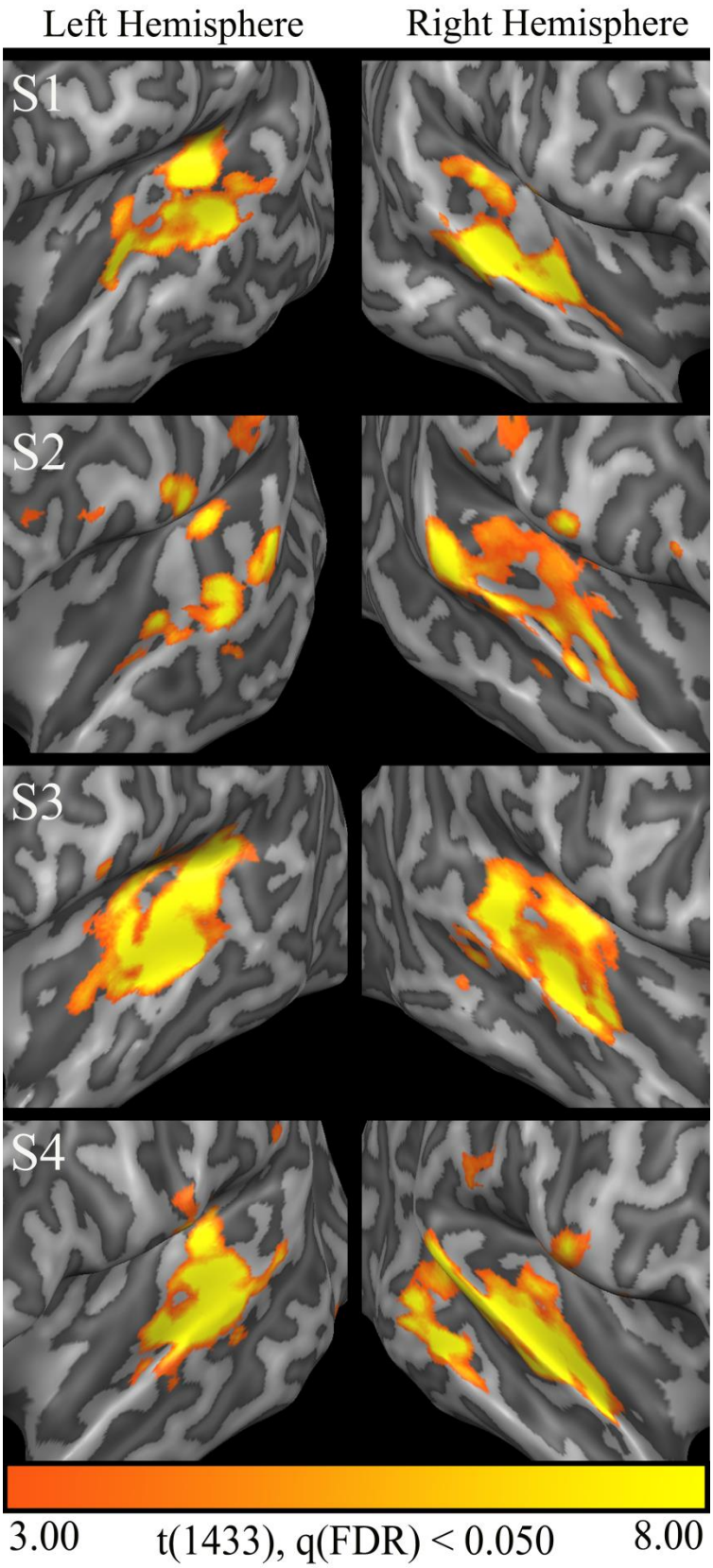
Supplementary Figure 2-4. Comparisons of PAC ROI Size and Similarity

A naïve lab member experienced in drawing cortical boundaries, delineated PAC ROIs for both hemispheres of each subject based on: (1) six *random tone sequence* scans using the pRF analysis (pRF_R), (2) six *ascending/descending tone progression scans* using the pRF analysis (pRF_{AD}), and (3) six *ascending/descending tone progression scans* using the phase-encoded analysis (PE_{AD}). Mean values across subjects and hemispheres are shown in black. **Panel A** shows PAC ROI size as measured by number of anatomical voxels for both hemispheres of all four subjects. The right (R) hemispheres of individual subjects are shown with colored downward pointing arrows, left (L) hemisphere with colored upward pointing arrows. The defined size of PAC remained relatively consistent across both auditory stimulus and analysis types. **Panel B** shows PAC ROI overlap across different types of stimuli and analyses. Overlap was measured using the Jaccard similarity coefficient for all three possible comparisons.



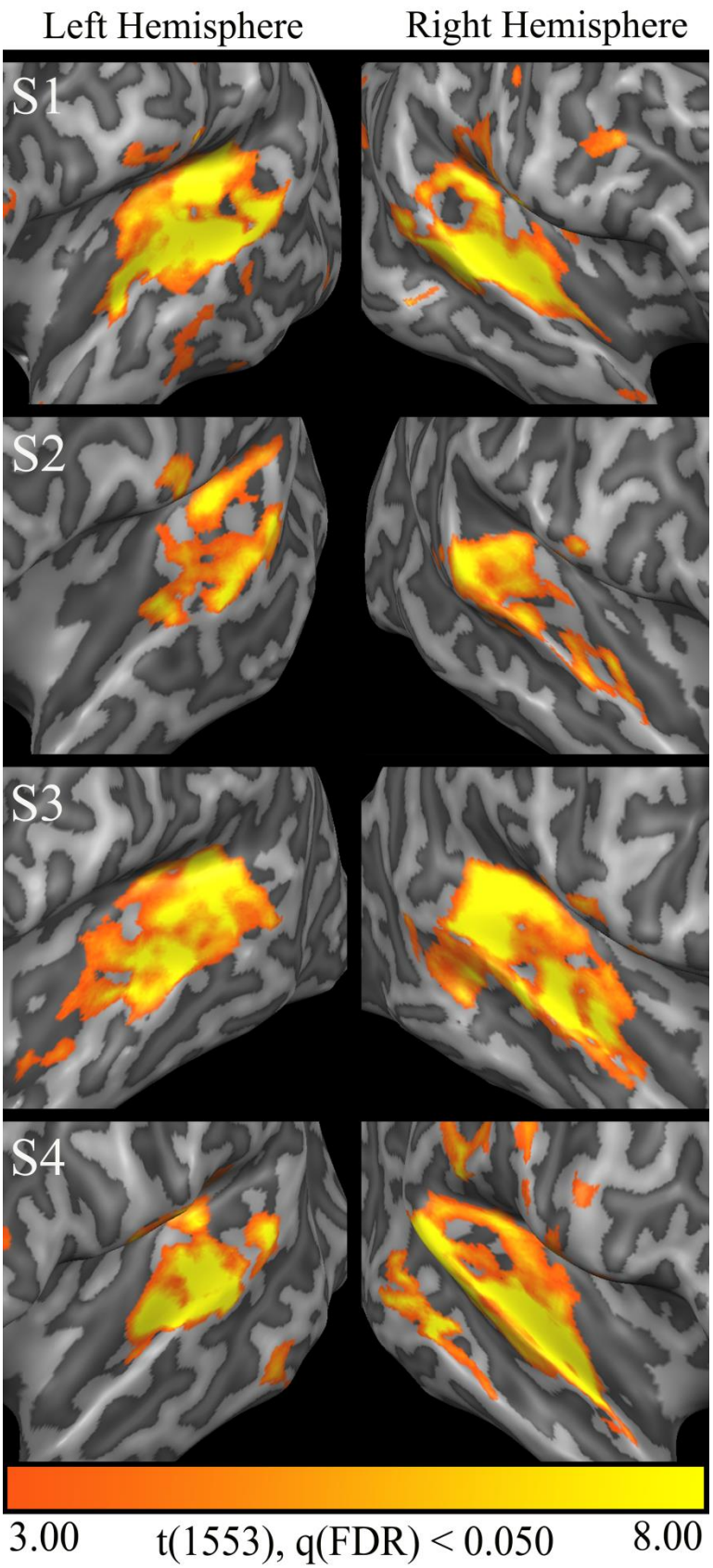
Supplementary Figure 2-5. Right Hemisphere Tonotopic Maps

Tonotopic maps for the right hemisphere of Subject 3 (first row) and 4 (second row). Maps were calculated using either: *ascending tone progressions* (4 scans) analyzed using phase-encoding (left panels), *descending tone progressions* (2 scans) analyzed using phase-encoding (middle panels), or *ascending/descending tone progressions* (6 scans) analyzed using the pRF method (right panels). Frequency center (f_0) values are color-coded along a gradient, with red corresponding to the lowest frequency value (88 Hz) through blue corresponding to the highest frequency (8000 Hz). Low-pass (LP) voxels are colored dark red, while high-pass (HP) voxels are colored dark blue. The borders of the PAC ROIs are designated by the solid black lines. The crowns of gyri are indicated by the dashed black lines. Across all figures/analyses the same correlation threshold of 0.10 was used, unless otherwise stated.



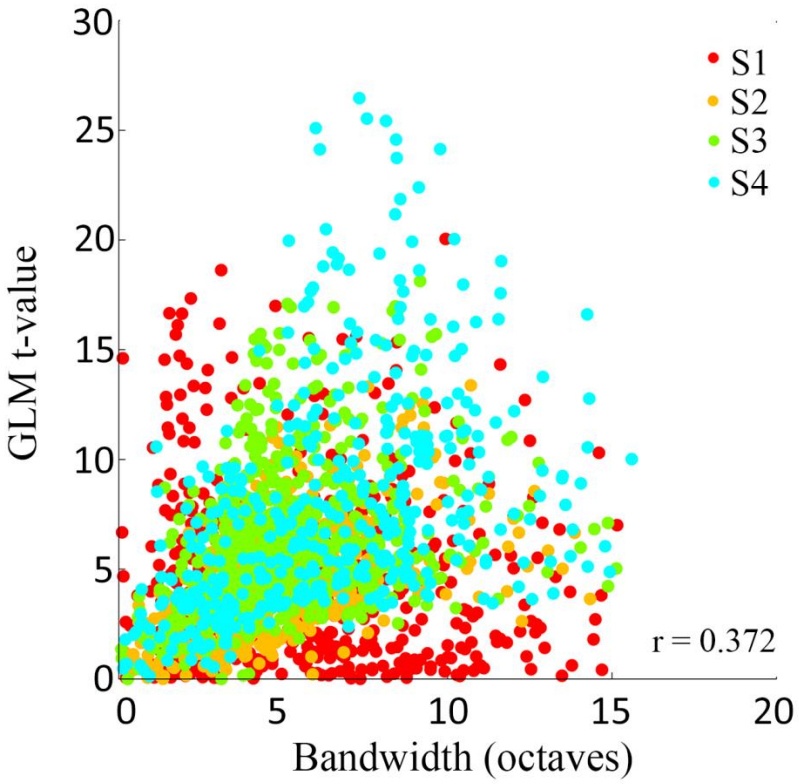
Supplementary Figure 2-6. GLM Results for Ascending/Descending Tone Progressions

A general linear regression model (GLM) using the single contrast of sound vs. silence was calculated from all six *ascending/descending tone progression* scans for each subject. Using a false discovery rate (FDR) criterion of $q < 0.05$, the corrected thresholds for significance equaled $p < 0.001$.



Supplementary Figure 2-7. GLM Results for Random Tone Sequences

A general linear regression model (GLM) using the single contrast of sound vs. silence was calculated from all six *random tone sequence* scans for each subject. Using a false discovery rate (FDR) criterion of $q < 0.05$, the corrected thresholds for significance equaled $p < 0.006$.



Supplementary Figure 2-8. Scatter Plot of Correlation Values

Scatter plot showing the correlation between the single contrast sound vs. silence GLM activation levels (t-value) and the bandwidth values for single voxels calculated from all six *random tone sequence* scans for each subject. The colors correspond to the individual subjects. There is a positive correlation between activation levels (t-value) and bandwidth estimate. Broader tuning results in a higher activation level across the duration of the sound condition leading to stronger responses in the sound vs. silence contrast.

Supplementary Table 2-1. Individually fitted HDR parameters

Subject	tau (τ)	delay (δ)
S1	1.025	3.276
S2	0.201	3.983
S3	1.073	3.231
S4	0.928	3.138

Supplementary Table 2-2. Average correlation coefficients

Subject	frequency (f_0)	bandwidth
S1	0.716	0.590
S2	0.745	0.401
S3	0.797	0.484
S4	0.711	0.560

Average correlation coefficients for frequency center (f_0) values and bandwidth estimates from *random tone sequence* data calculated between either two single scans of the same acquisition type, or two single scans of different acquisition types.

Chapter 3: Reconstructing auditory sequences from fMRI BOLD responses within human primary auditory cortex.

Abstract

Using fMRI BOLD responses in human primary auditory cortex (PAC) we show that it is possible to not only decode a song-like sequence of auditory frequencies a person had been listening to, but also estimate the individual frequencies of the 'song' over time. First, we characterized the tonotopic organization of each subject's auditory cortex by measuring auditory responses to randomized set of pure tone stimuli and modeling the frequency tuning of each fMRI voxel as a Gaussian in log frequency space. Then, we tested our model by examining its ability to work in reverse. Auditory responses were re-collected in the same subjects, except this time while they listened to sequences of tones taken from simple songs (e.g. 'Somewhere Over the Rainbow'). By comparing the model's prediction of BOLD responses to actual BOLD responses, we were able to estimate the frequency of each note in the song and reconstruct the stimulus over time. This allowed us to determine which song subjects had been listening to. The ability to accurately reconstruct the auditory experience of an individual over time using non-invasive measurements of brain activity should provide a powerful method for studying the neurological basis of both enhancements and deficits in auditory perception.

Introduction

A variety of BOLD imaging studies have identified a pair of mirror-symmetric tonotopic gradients on the cortical surface thought to be the human homologues of primary areas A1 and R

(Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011; Langers and van Dijk, 2011; Moerel et al., 2012; 2013). The general organization of these tonotopic gradients has been shown to be highly replicable across diverse imaging paradigms (Langers et al., 2014b; Saenz and Langers, 2014) and a range of stimulus types including orderly frequency progressions (Talavage et al., 2004; Da Costa et al., 2011; Striem-Amit et al., 2011; 2013; Langers et al., 2014a) as well as more complex natural stimuli (Moerel et al., 2012). While it is apparent this overall gradient pattern is highly replicable, the accuracy in which these maps have modeled the explicit tonotopic organization is unclear.

For example, it is possible to obtain robust tonotopic maps by evaluating BOLD responses to only a few discrete frequencies using a general linear model (GLM) (Formisano et al., 2003; Woods and Alain, 2009; Humphries et al., 2010). However, these models likely fail to capture the explicit representation of frequency selectivity in the auditory cortex, which is thought to represent a rather wide range of auditory frequencies. Stimulus-specific biases can also alter the frequency preference assigned to a given fMRI voxel. Frequency "sweep" stimuli have been shown to induce a "traveling wave" of BOLD activity across the cortex (Engel et al., 1994) that is susceptible to biases induced by habituation and/or expectation effects as well as spatio-temporal BOLD nonlinearities (Binda et al., 2013; Thomas et al., 2015). Moreover, the complex morphology and small size of auditory cortical areas makes them highly susceptible to these biases (Saenz and Langers, 2014). As such, tonotopic maps produced with these ordered stimuli can vary dramatically depending on the on the direction of the frequency sweep (Da Costa et al., 2011).

Stimulus-specific biases are particularly concerning for studies relating cortical responses to acoustic behavioral performance, which clearly would require the tonotopic

organization of the cortex to be accurately modeled. Previously, we described a 'population receptive field' or *pRF* analysis that makes it possible to use an unpredictable randomized stimulus to estimate the frequency tuning curves of individual voxels (Thomas et al. 2015, also see Dumoulin and Wandell, 2008), and suggested that this unpredictable stimulus might serve to reduce stimulus-specific biases in estimates of tonotopic organization (Binda et al., 2013).

Here, we present a method for investigating the accuracy of our voxel-wise frequency tuning curves by examining how well our model can generalize to predict responses to novel stimuli. Specifically, we examined whether tonotopic maps generated using randomized tones could be used to decode and reconstruct the sequence of notes in a 'song-like' stimulus on the basis of an individual subjects' BOLD responses over time. Next we measured cortical responses in the same subjects to novel song-like stimuli, based on the melodies "When You Wish Upon A Star" (Harline et al., 1940) and "Over The Rainbow" (Arlen and Harburg, 1939). Then, using a parametric decoding method, we reconstructed these songs by determining what frequency would best maximize the correlation between predicted (based on our pRF models) and obtained BOLD activity patterns for each point in the stimulus time course.

The quality of the stimulus reconstruction was quantified in two ways: Identification performance and reconstruction accuracy. Identification performance was assessed as the ability to correctly identify the actual song over other similar song-like sequences. Using an algorithm based on first-order Markov chains, we simulated 1000 song-like sequences analogous to the tested sequences. Identification performance was then determined as the number of times in which the actual sequence was correctly identified over any of the simulated song-like sequences. Reconstruction accuracy was assessed as the residual difference in cents (1200 cents per octave) between each note in the reconstructed and actual sequences.

Using a combined auditory pRF encoding/decoding approach, we found that we could accurately identify and reconstruct the 'song-like' stimuli over time on the basis of BOLD responses, thereby demonstrating the predictive accuracy of our tonotopic model of PAC.

Materials and Methods

Three right-handed subjects (2 male, 1 female, ages 27-46) participated in two fMRI sessions. Subjects reported normal hearing and no history of neurological or psychiatric illness. All procedures, including recruitment, consenting, and testing, followed the guidelines of the University of Washington Human Subjects Division and were reviewed and approved by the Institutional Review Board.

MRI data acquisition and analysis

BOLD imaging was performed using a 3 Tesla Phillips Achieva scanner (Philips, Eindhoven, The Netherlands) at the University of Washington Diagnostic Imaging Sciences Center (DISC). Subjects were instructed to keep their eyes closed throughout all scans and foam padding was used to minimize head motion. fMRI data were acquired using a 32-channel head coil and a continuous EPI pulse sequence (2.8 x 2.8 x 2.8 mm^3, TR/TE = 2000/25 ms, flip angle = 60°, EPI-factor = 35, no slice gap) designed with Philips SofTone software (SofTone factor of 4.0) to generate less acoustic scanner noise (Thomas et al., 2015).

Standard pre-processing of fMRI data was carried out using BrainVoyager QX software (version 2.3.1 Brain Innovation B. V., Maastricht, The Netherlands), including slice scan time correction, temporal high-pass filtering, and 3D motion correction. Functional data were aligned to a T1-weighted anatomical image acquired in the same session (MPRAGE, 1 x 1 x 1 mm^3).

The anatomical images acquired in the two sessions were aligned to each other and to each subject's 3D Talairach-normalized anatomical dataset. The BrainVoyager QX automatic segmentation routine was used to reconstruct the cortical surface and the resulting smooth 3D surface was partially inflated. For each subject, large anatomical regions of interest (ROIs) were selected from both hemispheres of the auditory cortical surface using drawing tools within BrainVoyager QX. Preprocessed time-course data for each 3D anatomical voxel within the volume ROI were then exported to Matlab for further analysis.

Auditory stimulus presentation

Sound stimuli were generated in MATLAB using the Psychophysics Toolbox (www.psychtoolbox.org). Stimuli were delivered via MRI compatible insert earphones (S14, Sensimetrics), at a sampling rate of 44.1 kHz, with intensities calibrated to ensure flat frequency transmission from 100 Hz to 8 kHz. After sound system calibration, all stimulus sound intensities were adjusted according to a standard equal-loudness curve created for insert earphones (ISO 226) to approximate equal perceived loudness across all frequencies. Acoustic noise from the scanner was attenuated by expanding-foam eartips as well as protective ear muffs placed over the ear following earphone insertion. Subjects reported hearing all stimuli at a clear and comfortably audible level, with roughly equal loudness across all tones.

pRF estimation

To reduce the influence of spatiotemporal nonlinearities on pRF estimates, we measured fMRI responses to a high-resolution randomized pure tone sequences consisting of 240 frequency blocks. As shown in **Figure 3-1 A**, each block lasted 2 s and consisted 8 pure tone

bursts of a single frequency. Each burst lasted either 50 ms or 200 ms in duration (inter-stimulus interval = 50ms) and was presented in a pseudo-randomized order, resulting in a "Morse code" like pattern of tones. This served to increase the perceptual salience of the tone bursts over the background scanner noise. The frequencies presented in the blocks ranged from 88-8000 Hz, each frequency block was presented only once per scan and block order was randomly shuffled for each scan. Following every 60 blocks was a 12 s silent pause. This silent period allows the pRF algorithm to better estimate the baseline fMRI response to scanner noise. Each subject participated in a single pRF estimation scanning session, consisting of 6 scans, each containing a different randomized sequence of the same 240 frequency blocks.

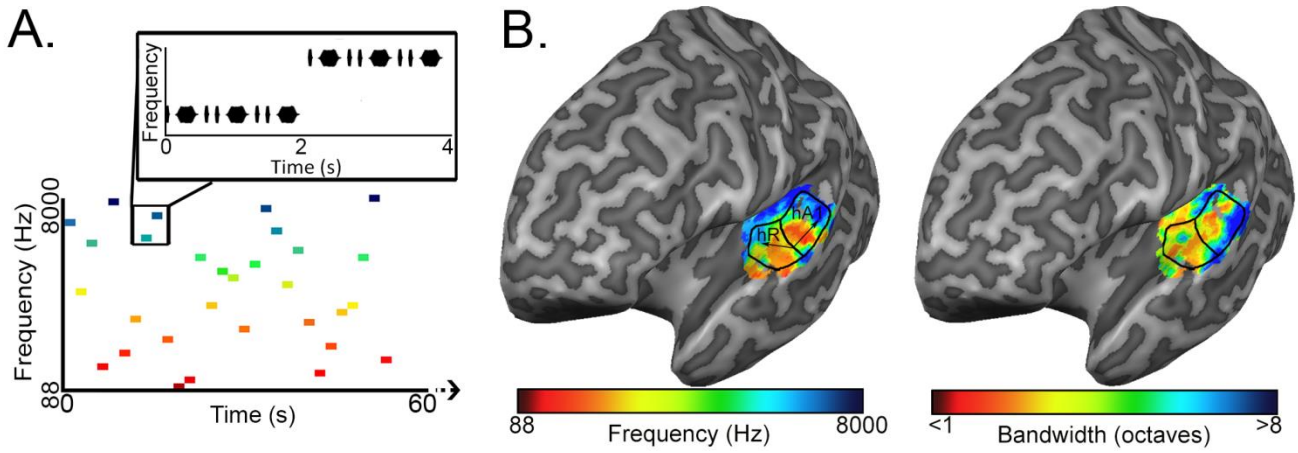


Figure 3-1. pRF estimation

A. The first 60 s of a *random sequence* stimulus used during pRF estimation. Each block lasted 2 s and consisted 8 pure tone bursts of a single frequency. Bursts lasted either 50 ms or 200 ms in duration (inter-stimulus interval = 50ms) and were presented in a pseudo-randomized order. **B.** Tonotopic and bandwidth maps for the left hemisphere of example Subject 1. As indicated by the black arrows, pRF frequency center (Hz) values formed two mirror-symmetric tonotopic gradients corresponding to the primary auditory fields A1 and R, outlined here by solid black lines. No clear organization was observed for pRF bandwidth (octaves) values.

Following previously described methods, we used customized MATLAB software to estimate the frequency tuning curves for individual voxels based on a linear temporal model of

the fMRI BOLD response time course (Thomas et al., 2015). Briefly, analysis began by defining a stimulus time course, which indicates the presence or absence of a particular frequency over time. This stimulus time course was convolved with each subject's estimated hemodynamic response function (HDR) modeled as a gamma probability density function (Boynton et al., 1996). Each voxel's response was modeled using the 1-dimensional Gaussian function g*(f),* defined over frequency (in log space). The center (f_0) of each Gaussian corresponds to the best frequency of the voxel, while the standard deviation (σ) was transformed into bandwidth values by calculating the full width half maximum (FWHM) in terms of octaves. A predicted time course was then generated for each voxel by calculating the linear sum of the overlap between the hemodynamically blurred stimulus time course and the pRF model. Finally, model fits for each voxel were obtained using a nonlinear search algorithm that iterated towards model parameters that maximize the correlation value (goodness-of-fit) between the voxel's pRF predicted time course and the acquired fMRI BOLD response time course.

The procedure described above included a few modifications from our original pRF analysis. First, we applied static power-law nonlinearity to the Gaussian model by including a free exponent parameter (*n*) to account for non-linear summation of the BOLD response according to the compressive spatial summation (CSS) model developed by Kay et al. (2013). The incorporation of this static non-linearity, which is applied after the initial fitting of the linear model, has been shown to more accurately explain BOLD activity and improve overall receptive field fits. Second, we constrained the Gaussian standard deviation (σ) to values greater than 0.015 (due to the resolution of frequencies presented) and exponent parameter (n) to values between 0 and 1.

After fitting, only voxels with a pRF correlation value (goodness-of-fit) above 0.15 were retained for song decoding (526, 529, 244 voxels for subjects S1-S3, respectively). As demonstrated in **Figure 3-1 B**, pRF center (f_0) values formed two mirror-symmetric tonotopic gradients corresponding to the primary auditory fields A1 and R in both hemispheres of all subjects. No clear organization was observed for either pRF bandwidth values (average bandwidth in octaves ± SD, S1= 3.385 ± 2.807, S2 = 3.732 ± 1.634, S3 = 2.219 ± 1.201), or exponent parameters (average value of n ± SD; S1=0.587 ± 0.310, S2 = 0.611 ± 0.228, and S3 = 0.726 ± 0.318).

Frequency Decoding

During a separate scanning session, we collected fMRI responses to two pure tone song-like sequences based on two familiar melodies: "When You Wish Upon A Star" (Wish) and "Somewhere Over The Rainbow" (Rainbow). Each song-like sequence was generated using 2 s frequency blocks with frequencies ranging from 880-2349 Hz (corresponding to the notes A5-D7 on the western music scale). Each 2 s block contained 13 tone bursts of the same frequency, each lasting 75 ms in duration (inter-stimulus interval = 75 ms). This created a vibrato-like effect which served to increase the perceptual salience of each block, without interrupting the melodic feel of the song-like sequence. A single presentation of each song-like sequence contained either 25 (Wish) or 23 (Rainbow) frequency blocks followed by 8 s of silence, and the entire presentation was repeated 8 times per scan. Averaged fMRI BOLD time courses were then generated for each song-like sequence by averaging data responses across the eight presentations within each scan, and across two scans of the same sequence type.

We decoded both song-like sequences by reconstructing each sequence one block at a time. To do this, we utilized the pRF models previously estimated with the randomized sequence to generate the predicted voxel activity patterns elicited by a set of 14 frequencies which sampled 88-8000 Hz in half-octave steps. The best fitting frequency from this set is then used as the initial parameter for a nonlinear optimization fitting procedure (Matlab's 'fminconv' function) that determined what frequency produced the predicted voxel activity pattern best correlated with the measured voxel activity pattern for each 2 s block. This process was then repeated for each block in the sequence, until all frequency blocks had been reconstructed. Finally, to account for the delayed hemodynamic blurring of BOLD signal, we applied a fixed temporal lag of 6 s was applied to the reconstructed sequence (Kay et al., 2008).

It is important to note that our method only depends on the frequency selectivity of individual voxels, not their physical locations within auditory cortex. This method is therefore not dependent upon any particular model for the overall structure of the tonotopic maps. This is significant because there is currently a lack of consensus in the literature about the orientation of tonotopic fields in human cortex with respect to Heschl's gyrus (Moerel et al., 2014; Saenz and Langers, 2014).

The quality of the reconstructed sequences was quantified in two ways: *Identification performance* and *reconstruction accuracy*. Identification performance was assessed as the ability to correctly identify the actual song over other similar song-like sequences. For each reconstructed sequence, we applied an algorithm based on first-order Markov chains to randomly generate 1000 simulated song-like sequences that reflected the frequency content and note-to-note probabilities of the Rainbow and Wish sequences. Then we determined how well correlated (Pearson's *r*) the reconstructed sequence was with the actual presented sequence, as well as each

of the simulated sequences. Identification performance was then determined as the number of times in which the actual sequence was correctly selected for, on the basis of a higher correlation with the reconstructed sequence, over any of the simulated song-like sequences.

For the purpose of this study, we selected the Markov chain approach as we were primarily concerned with generating simulated sequences with the basic statistical properties of the actual songs used in our experiments. However, other more advanced methods for generating simulated sequences, including probabilistic models of melodic intervals, would likely be better for generating more continuous and melodic sequences (Temperley, 2008, 2014).

Reconstruction accuracy was assessed as the ability to recreate each note in the actual sequence. This was calculated as the residual difference in cents (1200 cents per octave) between each note in the reconstructed and actual sequences. To determine if any systematic over or underestimation was present in the reconstructed sequences, we performed a two-tailed t-test on the means of the residual errors. Any mean that was significantly different from zero reflected an overall bias in reconstruction accuracy.

Results

Figure 3-2 depicts identification performance. We began by simply determining the correlation between reconstructed and actual frequencies for each subject for both the Rainbow (**Figure 3-2 A**) and Wish (**Figure 3-2 B**) sequences. For all subjects, reconstructed sequences were well correlated with the actual song-like sequences. Histograms containing the correlation values between the reconstructed Rainbow (**Figure 3-2 C)** and Wish **(Figure 3-2 D)** sequences and 1000 simulated sequences were generated with a first-order Markov chain algorithm (new sequences were generated for each subject). The correlation value of the actual sequence is represented by a black line in each histogram, indicating the threshold for which the actual

sequence is correctly identified. Identification performance of both reconstructed sequences was at near perfect levels for all three subjects, demonstrating how the identity of a song-like sequence can be readily be decoded by the similar pattern of frequencies in the reconstructed sequence.

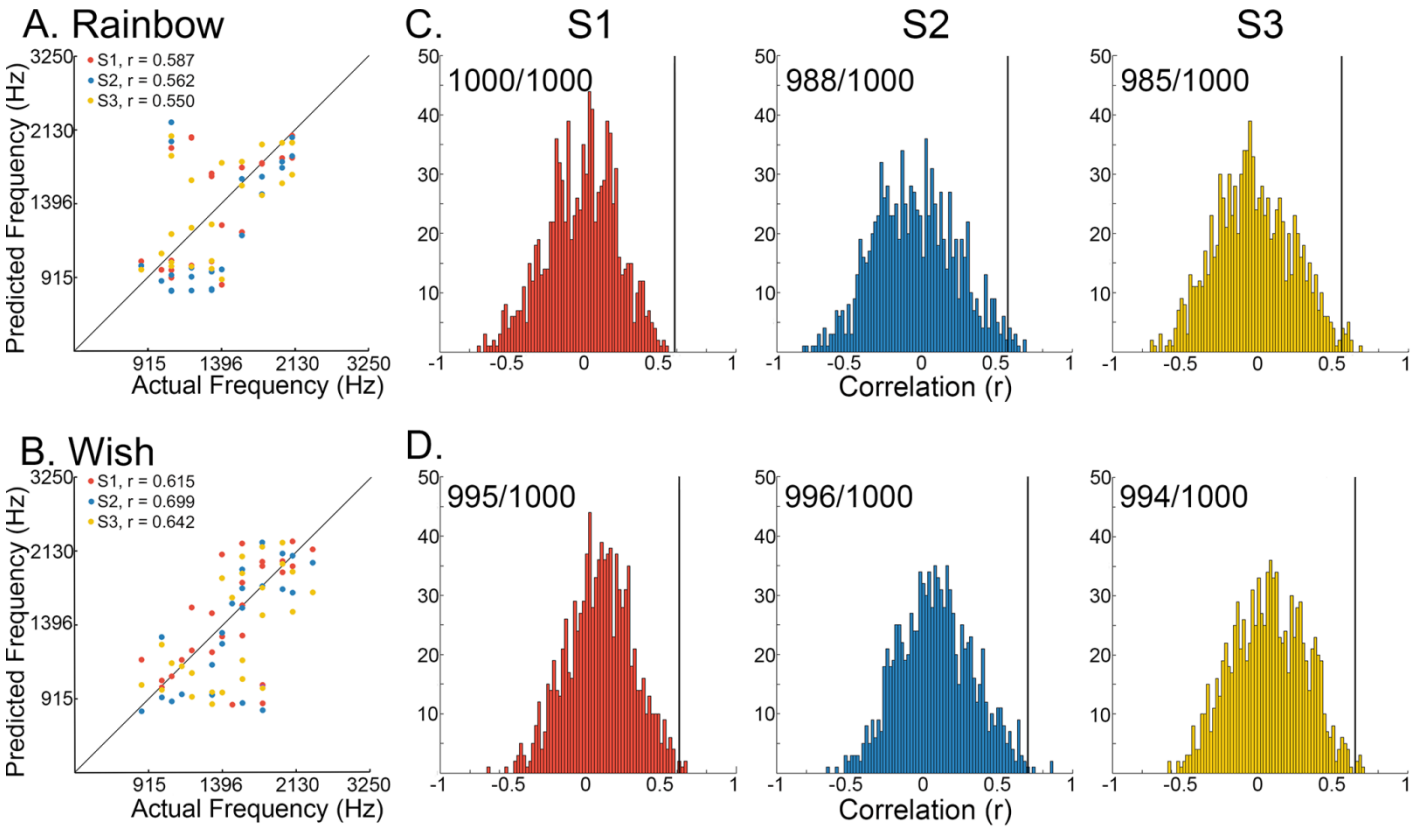


Figure 3-2. Identification Performance

A&B, Scatter plots showing the correlation between reconstructed and actual frequencies for each subject for both the Rainbow (A) and Wish (B) sequences. **C&D,** Using a method based on a first order Markov chain algorithm, we simulated 1000 song-like sequences reflecting the frequency content and note-to-note probabilities of the Rainbow (C) and Wish (D) sequences. Histograms of the correlation values (Pearson's r) between each of the simulated sequences and either reconstructed sequence. The line in black designates the correlation value between the actual song-like sequences and the reconstructed sequences, indicating the degree to which the correct sequence had been successfully identified. The number of correct identifications (out of 1000) is reported for each reconstructed sequence. Colors correspond to individual subjects.

The precision of our pRF decoding method was determined by how accurately each song-like sequence had been reconstructed in terms of musical intervals or cents. **Figure 3-3** displays the notes of the actual and reconstructed sequences of each subject according to modern musical notation. Purely for illustration purposes, the reconstructed frequencies in **Figure 3-3** were rounded to the nearest semitone (12 semitones per octave), or "note". We also lowered all notes (actual and reconstructed) one octave for better representation on the treble clef.

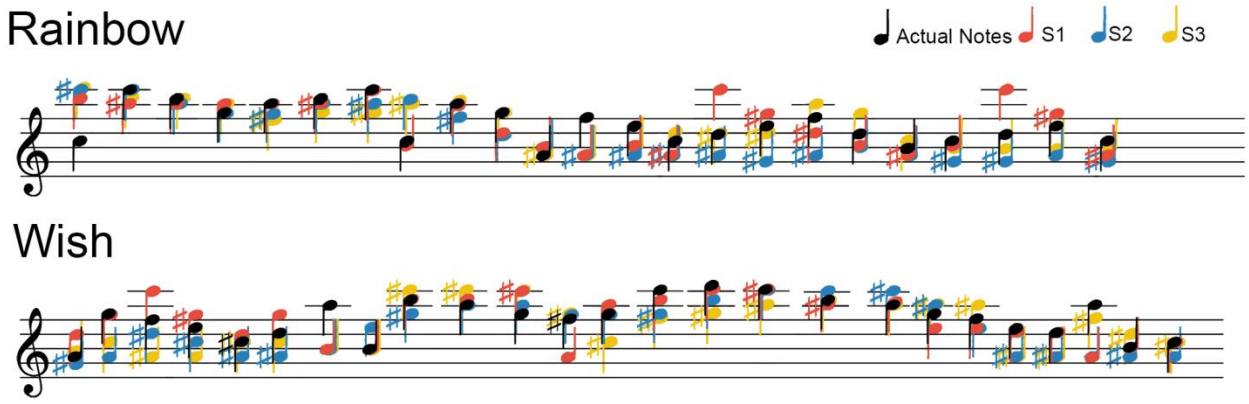


Figure 3-3. Sequence Reconstruction

For easier visualization on a treble clef, all frequencies (Hz) were rounded the nearest semitone and lowered one octave. Actual notes from each song-like sequence are in black, while the color of notes in the reconstructed sequences corresponds to individual subjects.

We report two measures of reconstruction accuracy based on the residual errors between the reconstructed frequency and the actual frequency. The first is the mean of the residual errors (**Table 3-1**). A mean that is different from zero reflects an overall bias in our reconstruction accuracy. Of the six means, only one reached statistical significance with a two-tailed t-test (Subject 2, Wish, t(24) = -215.54 cents, p=.0173). Thus, there does not appear to be a systematic

over or underestimation of the reconstructed frequencies (at least as far as the power of our experimental design can provide). The second measure of reconstruction accuracy is the standard deviation of the residual errors, also reported in **Table 3-1**. Standard deviations ranged between 434 and 512 cents across subjects and songs (around a third of an octave).

Table 3-1. Residual Errors

Subject	Song	Mean	Standard Deviation
S1	Rainbow	25.98	465.44
	Wish	-30.71	448.05
S2	Rainbow	-210.12	512.17
	Wish	-215.54	421.35
S3	Rainbow	14.6	456.24
	Wish	-155.78	434.02

Means and standard deviations of residual errors in cents between the reconstructed and actual frequencies.

Discussion

Here, we demonstrate how a pRF model of tonotopic organization in the human primary auditory cortex estimated from random tone sequences can accurately generate predicted BOLD responses to successfully decode song-like sequences. Not only was it possible to identify what song-like sequence a person had been listening to, but it was also possible to estimate the frequencies played over time.

Auditory decoding models have previously been used to classify speech content and speaker identity (Formisano et al., 2008) as well as the emotional content of speech (Ethofer et al., 2009). However, these studies employed linear classifier algorithms trained to discriminate

between stimulus categories according to the patterns of activity across fMRI voxels. While these decoding methods can readily identify or classify acoustic stimuli from brain activity, they are limited to candidate stimulus sets and cannot be generalized to substantially novel stimuli (Naselaris et al., 2011). Moreover, these decoding models do not provide insight into the feature space over which these complex stimuli are functionally organized within auditory cortex.

It is well known that neurons in auditory cortex respond selectively to stimulus dimensions other than frequency, including spectral and temporal modulation rate (Schönwiesner and Zatorre, 2009; Santoro et al., 2014). The inclusion of additional dimensions in our analysis would undoubtedly improve our decoding accuracy. However, our goal was to determine the extent to which stimuli could be reconstructed our predictions of fMRI responses to frequency alone. Our results demonstrate that our population-based model of voxel-wise frequency tuning is quite accurate, considering that the model assumes a simple unimodal Gaussian-shaped tuning curve based on responses to a random stimulus sequence.

Our analysis is best described as a combination *encoding/decoding* model (Naselaris et al., 2011). We began with our *encoding* pRF model to describe the frequency selectivity of individual voxels in each subject's primary auditory cortex (Thomas et al., 2015). Then we applied a parametric *decoding* method to our pRF models to identify and reconstruct novel 'song-like' sequences, similar to methods previously used in the visual domain to identify (Kay et al., 2008) and reconstruct novel image stimuli (Thirion et al., 2006; Miyawaki et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011). While the identification performance of our technique was very near perfect, (98.5 -100 % correct identification), it was our ability to *reconstruct* novel auditory stimuli that allowed us to examine the validity of our tonotopic encoding model. We were able to reconstruct the perceived song-like stimuli of all three subjects

within a half of an octave or less, with little evidence of systematic biases in frequency estimation.

As far as future directions are concerned, our results were obtained using relatively standard acquisition protocols on a 3 Telsa scanner. Increased fMRI spatial resolution (especially given the relatively small size of the PAC) and reductions in measurement noise would likely further improve reconstruction accuracy (Kay and Gallant, 2009). Further improvements might also be made by developing more sophisticated decoding models to include (for example) spatiotemporal non-linearities and/or expectation and habituation effects. Such models will likely require the use of more diverse auditory sequences specifically designed to elicit these phenomenon.

Finally, we only considered voxel tuning along the single dimension of frequency. Other auditory fMRI studies have developed more complex encoding models that include the spectro-temporal modulations common to natural sound stimuli (Schönwiesner and Zatorre, 2009; Moerel et al., 2012; 2013; Santoro et al., 2014). A natural extension to our research would be to expand our pRF encoding algorithm to allow us to model more complex and behaviorally relevant stimuli and examine the response properties in non-primary auditory areas (Moerel et al., 2014; Saenz and Langers, 2014). Future investigations may also be aimed at linking cortical responses to perceptual experience for ambiguous auditory stimuli, and examining the effects of attention (Da Costa et al., 2013).

Final Remarks

Here we have demonstrated how voxel-based encoding and decoding models of functional MRI responses can be used to quantify the representation of sensory information within the human visual and auditory cortex. Computational models such as these provide a useful framework for not only investigating the processing of stimulus information across different cortical areas, but also for examining individual differences in sensory perceptions.

As the field of computational neuroscience moves forward, more sophisticated encoding and decoding models are sure to evolve. To increase the accuracy and predictive power of these techniques, computational models will likely become more complex as they take into account an increasing number of stimulus features. In doing so, these models will generalize across a broader range of stimulus types, and may eventually provide researchers with the necessary tools for testing the more abstract concepts of human cognition and perception.

References

Aquino KM, Schira MM, Robinson PA, Drysdale PM, Breakspear M (2012) Hemodynamic Traveling Waves in Human Visual Cortex. PLoS Comput Biol 8:e1002435.

Arlen H, Harburg EY (1939) Over the rainbow : from the M-G-M picture, the Wizard of Oz. In. New York: L. Feist.

Baker CI, Peli E, Knouf N, Kanwisher NG (2005) Reorganization of visual processing in macular degeneration. J Neurosci 25:614-618.

Baker CI, Dilks DD, Peli E, Kanwisher N (2008) Reorganization of visual processing in macular degeneration: replication and clues about the role of foveal loss. Vision Res 48:1910-1919.

Baseler HA, Morland AB, Wandell BA (1999) Topographic organization of human visual areas in the absence of input from primary cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 19:2619-2627.

Baseler HA, Brewer AA, Sharpe LT, Morland AB, Jagle H, Wandell BA (2002) Reorganization of human cortical maps caused by inherited photoreceptor abnormalities. Nat Neurosci 5:364-370.

Baseler HA, Gouws A, Haak KV, Racey C, Crossland MD, Tufail A, Rubin GS, Cornelissen FW, Morland AB (2011) Large-scale remapping of visual cortex is absent in adult humans with macular degeneration. Nat Neurosci 14:649-655.

Binda P, Thomas JM, Boynton GM, Fine I (2013) Minimizing biases in estimating the reorganization of human visual areas with BOLD retinotopic mapping. Journal of Vision 13.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. The Journal of Neuroscience 16:4207-4221.

Boynton GM, Demb JB, Glover GH, Heeger DJ (1999) Neuronal basis of contrast discrimination. Vision research 39:257-269.

Buracas GT, Fine I, Boynton GM (2005) The relationship between task performance and functional magnetic resonance imaging response. The Journal of neuroscience : the official journal of the Society for Neuroscience 25:3023-3031.

Cheung SW, Bedenbaugh PH, Nagarajan SS, Schreiner CE (2001) Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. Journal of neurophysiology 85:1732-1749.

Cochran WG (1934) The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. Mathematical Proceedings of the Cambridge Philosophical Society 30:178-191.

Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M (2013) Tuning In to Sound: Frequency-Selective Attentional Filter in Human Primary Auditory Cortex. The Journal of Neuroscience 33:1858-1863.

Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. The Journal of Neuroscience 31:14067-14075.

Dick F, Tierney AT, Lutti A, Josephs O, Sereno MI, Weiskopf N (2012) In vivo functional and myeloarchitectonic mapping of human primary auditory areas. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:16095-16105.

Dilks DD, Baker CI, Peli E, Kanwisher N (2009) Reorganization of visual processing in macular degeneration is not specific to the "preferred retinal locus". J Neurosci 29:2768-2773.
Dilks DD, Serences JT, Rosenau BJ, Yantis S, McCloskey M (2007) Human adult cortical reorganization and consequent visual distortion. The Journal of neuroscience : the official journal of the Society for Neuroscience 27:9585-9594.
Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647-660.
Duncan RO, Boynton GM (2003) Cortical magnification within human primary visual cortex correlates with acuity thresholds. Neuron 38:659-671.
Elbert T, Sterr A, Rockstroh B, Pantev C, Muller MM, Taub E (2002) Expansion of the tonotopic area in the auditory cortex of the blind. J Neurosci 22:9941-9944.
Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. Nature 369:525.
Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009) Decoding of emotional information in voice-sensitive cortices. Current biology : CB 19:1028-1033.
Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. Science (New York, NY) 322:970-973.
Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40:859-869.
Haak KV, Cornelissen FW, Morland AB (2012) Population receptive field dynamics in human visual cortex. PLoS One 7:e37686.
Hackett TA (2008) Anatomical organization of the auditory cortex. Journal of the American Academy of Audiology 19:774-779.
Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. The Journal of comparative neurology 394:475-495.
Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7:213-223.
Harline L, Washington N, Disney W, Edwards C, Young V, Victor Young O, Ken Darby S (1940) When you wish upon a star (from Walt Disney's "Pinocchio"). In. United States: Decca.
Harms MP, Melcher JR (2003) Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. Hum Brain Mapp 20:168-183.
Hoffmann MB, Kaule FR, Levin N, Masuda Y, Kumar A, Gottlob I, Horiguchi H, Dougherty RF, Stadler J, Wolynski B, Speck O, Kanowski M, Liao YJ, Wandell BA, Dumoulin SO (2012) Plasticity and stability of the visual system in human achiasma. Neuron 75:393-401.
Howard MA, 3rd, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA (1996) A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. Brain Res 724:260-264.
Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. Neuroimage 50:1202-1211.
Jaccard P (1912) The Distribution of the Flora in the Alpine Zone. New Phytologist 11:37-50.
Jäncke L, Gaab N, Wüstenberg T, Scheich H, Heinze HJ (2001) Short-term functional plasticity in the human auditory cortex: an fMRI study. Cognitive Brain Research 12:479-485.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nature neuroscience 8:679-685.
Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751-761.
Kay KN, Gallant JL (2009) I can see what you see. Nature neuroscience 12:245-245.
Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355.
Kay KN, Winawer J, Mezer A, Wandell BA (2013) Compressive spatial summation in human visual cortex. Journal of neurophysiology 110:481-494.
Kayser C, Petkov CI, Logothetis NK (2007a) Tuning to Sound Frequency in Auditory Field Potentials. Journal of neurophysiology 98:1806-1809.
Kayser C, Petkov CI, Augath M, Logothetis NK (2007b) Functional imaging reveals visual modulation of specific fields in auditory cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 27:1824-1835.
Komatsu H (2006) The neural mechanisms of perceptual filling-in. Nature reviews Neuroscience 7:220-231.
Langers DR, Van Dijk P, Backes WH (2005) Interactions between hemodynamic responses to scanner acoustic noise and auditory stimuli in functional magnetic resonance imaging. Magn Reson Med 53:49-60.
Langers DRM, van Dijk P (2011) Mapping the Tonotopic Organization in Human Auditory Cortex with Minimally Salient Acoustic Stimulation. Cerebral Cortex.
Langers DRM, Krumbholz K, Bowtell RW, Hall DA (2014a) Neuroimaging paradigms for tonotopic mapping (I): The influence of sound stimulus type. NeuroImage 100:650-662.
Langers DRM, Sanchez-Panchuelo RM, Francis ST, Krumbholz K, Hall DA (2014b) Neuroimaging paradigms for tonotopic mapping (II): The influence of acquisition protocol. NeuroImage 100:663-675.
Liegeois-Chauvel C, Musolino A, Chauvel P (1991) Localization of the primary auditory area in man. Brain : a journal of neurology 114 (Pt 1A):139-151.
Masuda Y, Dumoulin SO, Nakadomari S, Wandell BA (2008) V1 projection zone signals in human macular degeneration depend on task, not stimulus. Cereb Cortex 18:2483-2493.
Masuda Y, Horiguchi H, Dumoulin SO, Furuta A, Miyauchi S, Nakadomari S, Wandell BA (2010) Task-dependent V1 responses in human retinitis pigmentosa. Invest Ophthalmol Vis Sci 51:5356-5364.
Meng M, Remus DA, Tong F (2005) Filling-in of visual phantoms in the human brain. Nature neuroscience 8:1248-1254.
Menning H, Roberts LE, Pantev C (2000) Plastic changes in the auditory cortex induced by intensive frequency discrimination training. NeuroReport 11:817-822.
Miyawaki Y, Uchida H, Yamashita O, Sato MA, Morito Y, Tanabe HC, Sadato N, Kamitani Y (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60:915-929.
Moerel M, De Martino F, Formisano E (2012) Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. The Journal of Neuroscience 32:14205-14216.
Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. Frontiers in Neuroscience 8.

Moerel M, De Martino F, Santoro R, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:11888-11898.

Morland AB, Baseler HA, Hoffmann MB, Sharpe LT, Wandell BA (2001) Abnormal retinotopic representations in human visual cortex revealed by fMRI. Acta psychologica 107:229-247.

Muchinsky PM (1996) The Correction for Attenuation. Educational and Psychological Measurement 56:63-75.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. NeuroImage 56:400-410.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63:902-915.

Nishimoto S, Vu An T, Naselaris T, Benjamini Y, Yu B, Gallant Jack L (2011) Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Current Biology 21:1641-1646.

Pantev C, Herholz SC (2011) Plasticity of the human auditory cortex related to musical training. Neurosci Biobehav Rev 35:2140-2154.

Petkov CI, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. PLoS biology 4:e215.

Pihlaja M, Henriksson L, James AC, Vanni S (2008) Quantitative multifocal fMRI shows active suppression in human V1. Hum Brain Mapp 29:1001-1014.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13:669-683.

Ramachandran VS, Gregory RL (1991) Perceptual filling in of artificially induced scotomas in human vision. Nature 350:699-702.

Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science (New York, NY) 268:111-114.

Recanzone GH, Schreiner CE, Sutter ML, Beitel RE, Merzenich MM (1999) Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. The Journal of comparative neurology 415:460-481.

Saenz M, Langers DR (2013) Tonotopic mapping of human auditory cortex. Hear Res.

Saenz M, Langers DR (2014) Tonotopic mapping of human auditory cortex. Hear Res 307:42-52.

Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E (2014) Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. PLoS Comput Biol 10:e1003412.

Schönwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proceedings of the National Academy of Sciences 106:14611-14616.

Seifritz E, Di Salle F, Esposito F, Herdener M, Neuhoff JG, Scheffler K (2006) Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. Neuroimage 29:1013-1022.

Soltysik DA, Peck KK, White KD, Crosson B, Briggs RW (2004) Comparison of hemodynamic response nonlinearity across primary cortical areas. Neuroimage 22:1117-1127.

Spearman C (1904) The proof and measurement of association between two things. The American journal of psychology 15:72-101.

Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding FMRI. PLoS One 6:e17832.

Sunness JS, Liu T, Yantis S (2004) Retinotopic mapping of the visual cortex using functional magnetic resonance imaging in a patient with central scotomas from atrophic macular degeneration. Ophthalmology 111:1595-1598.

Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic Organization in Human Auditory Cortex Revealed by Progressions of Frequency Sensitivity. Journal of neurophysiology 91:1282-1296.

Temperley D (2008) A Probabilistic Model of Melody Perception. Cognitive Science 32:418-444.

Temperley D (2014) Probabilistic Models of Melodic Interval. Music Perception: An Interdisciplinary Journal 32:85-99.

Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, Lebihan D, Dehaene S (2006) Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage 33:1104-1116.

Thomas JM, Huber E, Stecker GC, Boynton GM, Saenz M, Fine I (2015) Population receptive field estimates of human auditory cortex. NeuroImage 105:428-439.

Tian B, Rauschecker JP (1994) Processing of frequency-modulated sounds in the cat's anterior auditory field. J Neurophysiol 71:1959-1975.

Triantafyllou C, Hoge RD, Krueger G, Wiggins CJ, Potthast A, Wiggins GC, Wald LL (2005) Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. Neuroimage 26:243-250.

Vanni S, Henriksson L, James AC (2005) Multifocal fMRI mapping of visual cortical areas. Neuroimage 27:95-105.

Wandell BA, Smirnakis SM (2009) Plasticity and stability of visual field maps in adult primary visual cortex. Nature reviews Neuroscience 10:873-884.

Weil RS, Kilner JM, Haynes JD, Rees G (2007) Neural correlates of perceptual filling-in of an artificial scotoma in humans. Proceedings of the National Academy of Sciences 104:5211-5216.

Woods DL, Alain C (2009) Functional imaging of human auditory cortex. Current opinion in otolaryngology & head and neck surgery 17:407-411.

Zenger-Landolt B, Heeger DJ (2003) Response suppression in v1 agrees with psychophysics of surround masking. The Journal of neuroscience : the official journal of the Society for Neuroscience 23:6884-6893.

Zuiderbaan W, Harvey BM, Dumoulin SO (2012) Modeling center-surround configurations in population receptive fields using fMRI. J Vis 12:10.