# Reconstructing tone sequences from fMRI BOLD responses within human primary auditory cortex.

Jessica M. Thomas, Ione Fine, Geoffrey M. Boynton

Department of Psychology, University of Washington, Seattle WA , USA 98195-1525

Corresponding author:

Jessica M. Thomas

University of Washington, Department of Psychology, Guthrie Hall, Seattle WA 98195-1525

Box 351525; phone no. (206)-543-4095, Email: [jmthomas@uw.edu](mailto:jmthomas@uw.edu)

Number of pages: 17

Number of figures: 3

Number of tables: 1

Total number of words: 4722 ( NEEDS TO BE 4500)

Conflict of Interest: The authors declare no competing financial interests

.Acknowledgements: This work was supported by the NIH 2T32DC005361-11 grant awarded to J.M.T and the NEI EY-014645 grant awarded to I.F. ( NEED GEOFF’S GRANT)

# Abstract

Using fMRI BOLD responses in human primary auditory cortex (PAC) we show that it is possible to not only decode the 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 frequencies 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. 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.

Keywords: auditory, decoding, population receptive field, tonotopy,

# Introduction

A variety of BOLD imaging studies have identified a pair of mirror-symmetric tonotopic gradients centered on Heschl’s gyrus on the cortical surface thought to be the human homologues of primary areas A1 and R ([Formisano et al., 2003](#_ENREF_10); [Humphries et al., 2010](#_ENREF_12); [Da Costa et al., 2011](#_ENREF_5); [Langers and van Dijk, 2011](#_ENREF_16); [Moerel et al., 2012](#_ENREF_20); [2013](#_ENREF_4)). The general organization of these maps has been shown to be highly replicable across diverse imaging paradigms ([Langers et al., 2014b](#_ENREF_18); [Saenz and Langers, 2014](#_ENREF_26)) and a range of stimulus types including orderly frequency progressions ([Talavage et al., 2004](#_ENREF_30); [Da Costa et al., 2011](#_ENREF_5); [Striem-Amit et al., 2011](#_ENREF_29); [2013](#_ENREF_4); [Langers et al., 2014a](#_ENREF_17)) as well as more complex natural stimuli ([Moerel et al., 2012](#_ENREF_20)).

However, while the overall pattern of frequency gradients is highly replicable, the accuracy with which these maps have modeled the actual frequency preferences of individual voxels 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](#_ENREF_10); [Woods and Alain, 2009](#_ENREF_36); [Humphries et al., 2010](#_ENREF_12)). 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](#_ENREF_7)) that is susceptible to biases induced by habituation and/or expectation effects as well as spatio-temporal BOLD nonlinearities ([Binda et al., 2013](#_ENREF_2); [Thomas et al., 2015](#_ENREF_35)). Moreover, the complex morphology and small size of auditory cortical areas makes them highly susceptible to these biases ([Saenz and Langers, 2014](#_ENREF_26)). 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](#_ENREF_5)).

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](#_ENREF_2)).

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 tones on the basis of an individual subjects’ BOLD responses over time. Next we measured cortical responses in the same subjects to novel stimuli containing a sequence of tones based on the melodies “When You Wish Upon A Star” ([Harline et al., 1940](#_ENREF_11)) and “Over The Rainbow” ([Arlen and Harburg, 1939](#_ENREF_1)). 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 tone-sequence over other similar tone-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 tone 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 tone sequences 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 mm3, 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. 2014).

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 mm3). 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 1A**, 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.

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., 2014](#_ENREF_34)). 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](#_ENREF_3)). Each voxel’s response was modeled using the 1-dimensional Gaussian function g*(f),* defined over frequency (in log space). The center (f0) 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)](#_ENREF_15). 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 1B**, pRF center (f0) 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 ‘fmincon’ 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](#_ENREF_14)).

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](#_ENREF_21); [Saenz and Langers, 2014](#_ENREF_26)).

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](#_ENREF_31), [2014](#_ENREF_32)).

*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 2** depicts identification performance. We began by simply determining the correlation between reconstructed and actual frequencies for each subject for both the Rainbow (**Figure 2A**) and Wish (**Figure 2B**) sequences. For all subjects, reconstructed sequences were well correlated with the actual song-like sequences. Histograms containingthe correlation values between the reconstructed Rainbow (**Figure 2C)** and Wish **(Figure 2D)** 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.

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** 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** 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.

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 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 1**. Standard deviations ranged between 434 and 512 cents across subjects and songs (around a third of an octave).

# 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](#_ENREF_9)) as well as the emotional content of speech ([Ethofer et al., 2009](#_ENREF_8)). 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](#_ENREF_23)). 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](#_ENREF_28); [Santoro et al., 2014](#_ENREF_27)). 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](#_ENREF_23)). 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](#_ENREF_35)). 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](#_ENREF_14)) and reconstruct novel image stimuli ([Thirion et al., 2006](#_ENREF_33); [Miyawaki et al., 2008](#_ENREF_19); [Naselaris et al., 2009](#_ENREF_24); [Nishimoto et al., 2011](#_ENREF_25)). 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](#_ENREF_13)). 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](#_ENREF_28); [Moerel et al., 2012](#_ENREF_20); [2013](#_ENREF_22); [Santoro et al., 2014](#_ENREF_27)). 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](#_ENREF_21); [Saenz and Langers, 2014](#_ENREF_26)). 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](#_ENREF_4)).

# References

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

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.

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.

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647-660.

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.

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.

Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. Neuroimage 50:1202-1211.

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.

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.

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.

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.

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.

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.

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 (2014) Population receptive field estimates of human auditory cortex. NeuroImage.

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.

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

# Legends

**Figure 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

**Figure 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.

**Figure 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.

**Table 1. Residual Errors.** Means and standard deviations of residual errors in cents between the reconstructed and actual frequencies.