Ari Brown

Ralph Adolphs Lab

Co-Mentor: Julien Dubois

Caltech SURF Program

September 2016

Investigation of Neural Responses to Physical and Theoretical Musical Features with functional Magnetic Resonance Imaging

ABSTRACT: The visual system of humans has been studied extensively using functional magnetic resonance imaging (fMRI); scientists are now able to coarsely reconstruct visual movie stimuli from the brain activity of human subjects watching them in the scanner. In comparison, the perception of complex auditory stimuli is as yet less well understood. This SURF project lays the foundations for the ambitious goal of reconstructing complex music stimuli that a human subject listens to in the scanner. The first step towards this goal is to parameterize music, decomposing it into a collection of time-varying features that may be represented in the brain. Some of these features capture physical properties of music, such as frequency content or loudness, which can be extracted using signal processing techniques (music information retrieval). Music can also be analyzed at a theoretical level using keys, scales, harmony, and melody. We describe the collection of features that we extracted from 1.5h of music excerpts (cello suites by Bach). We also describe two optimized experiments which we designed to test the response of the brain to specific aspects of music theory, namely the perception of octaves (using an event-related fMRI adaptation design) and the perception of key context (using an event-related design). We scanned one pilot subjects while they listened to 1.5h of music excerpts, and performed our optimized, music theory-based designs for 30min. We report preliminary analyses of the data and discuss future steps.

INTRODUCTION:

The overarching goal of this project is to understand how music is processed by the brain. An optimal outcome would be to demonstrate the reconstruction of audio stimuli from fMRI data, which establish that we have captured some key aspects of the representation of music in the brain. The main approach that we adopt directly follows recent studies in Jack Gallant's laboratory at UC Berkeley who successfully reconstructed visual stimuli from fMRI data in the early visual cortex (Kay et al. 2008; Nishimoto et al. 2011), and coined their approach "voxelwise modeling". What they did was simply record brain responses while subjects looked at images or watched movies, passively. Then, the analysis of the data relied on a detailed decomposition of the visual stimuli into low-level features (using a bank of "Gabor" filters), motion measures, and, more recently, semantic information such as the presence of a face or other object (Huth et al. 2012). The response of the brain to the hundreds of extracted features can be learned on a subset of the data using regularized regression models. Then, using the learned models, the response of the brain to other visual stimuli (the remainder of the data) can be predicted – and compared to the measured response. A good prediction is indicative of the appropriateness of the feature set in capturing something about the response of the brain. We describe in the report how we started adapting this approach to the study of music, using both physical features of sound as well as theoretical aspects of music.

A more classical approach to fMRI research is the design of optimized hypothesis-based experiments, in which the effect of a well-defined manipulation of a dependent variable is measured. As a learning exercise and a complementary way to study the representation of some aspects of music theory, we designed two specific experiments which we also describe in this report.

APPLYING THE VOXELWISE MODELING APPROACH TO MUSICAL STIMULI:
The experimental acquisition could not be simpler: a person lies in the scanner and listens attentively to several pieces of music. To keep them focused, we ask them to count the number of movements in each excerpt that we play. We chose to use the cello suites by Bach (Fig. 2) because...

To extract physical, low-level features from the music, we used the MIRtoolbox in Matlab (mirToolbox). Features extracted by the toolbox include spectral centroid, zero crossing rate, fundamental frequency, and amplitude.
Music can also be analyzed at a theoretical level based upon how musicians classify notes. Western music theory, our system of classification, is based on centuries of previous practice in identifying "what sounds good".

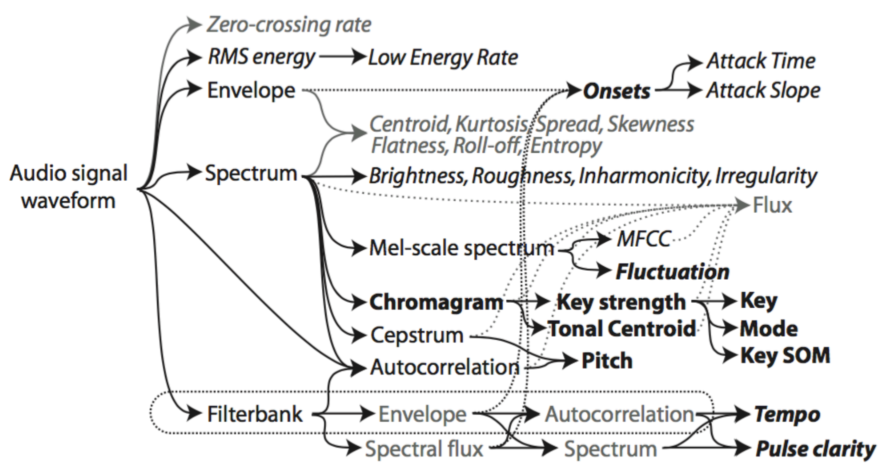

Figure 1. Features Extraction of Audio Signal Using MIRtoolbox The flowchart above shows the many features that MIRtoolbox can extract from an audio signal. Then these features can be correlated with fMRI BOLD data.

The mapping between the features of the music and neural activity across the brain is learned with a regularized regression model from all data but excerpt 7, which is reserved for testing the model.

Excerpt	Suite No.	Excerpt	Length	Length with pauses	
1	5	Prelude, Allemenade	12:33	12:53	
2	5	Sarabande to Gigue	11:55	12:15	
3	2	Prelude, Allemande	6:52	7:12	
4	2	Sarabande to Gigue	12:07	12:37	
7	6	Sarabande 0:00 - 1:45	9:02	9:22	Test

Figure 2. Naturalistic Stimuli for fMRI Scanning The above excerpts were listened to by subjects in the fMRI scanner. The test piece is the segment we are interested in reconstructing, based upon the training set (excerpts 1-4 above).

A successful prediction of the brain responses in the test dataset means that the features used in our model capture the response of the brain to the music in a reliable way. Inverting the model, one could predict the features of music that a subject listens to in the scanner from brain activity alone. Reconstruction of the actual music based on the features requires a re-synthesis of audio components. Additive synthesis of sine waves may be a useful way of reconstructing sound (as most audio information is retrieved based on harmonic content). In addition, we may try to using an optimizing algorithm by editing a randomized sound incrementally and assessing its fitness to each of the features we want to synthesize into one audio sample. We hope that these methods will help create a sound image that might have similar contour or texture features to the test piece's audio signal.

TARGETED, HYPOTHESIS-BASED EXPERIMENTS

A) NEURAL REPRESENTATION OF OCTAVES:

The first targeted experiment investigates octave varying pitches in the brain. We focus on the question of what makes a "note" sound the same at any octave, regardless of register. We aim to see if there is physical and perceptual basis for our response to the same note at different octaves (given that an octave's frequency ratio is 2:1). We hypothesize and search for a region of the brain that responds to the octave interval in the same way that the region responds to a repeated note at the same register.

We use a fMRI adaptation paradigm to design our experiment. When a condition is repeated, the BOLD response will be decreased in the area of the brain that responds to the condition. Three different types of stimuli were used: the same note twice, two notes separated by an octave interval, and two notes that are not separated by an octave interval. Regions of the brain where neurons respond to a specific note in a given register with adapt to the repetition of a note, but not to a note preceded by the same note in a different register. We expect to find such a response pattern in primary auditory cortex, which is believed to code absolute frequency. We hypothesize that some regions will show a decreased response also for the same note played in a different register: these are the regions that are responsible for our subjective perception of the same note. Figure 3 gives a breakdown of the design we used to present the stimuli.

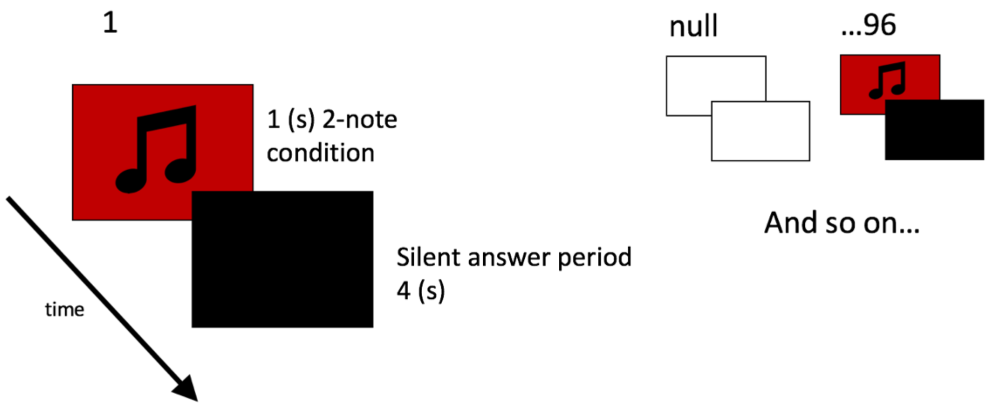

Figure 3. Octave Experiment Design The two notes in each trial fit in a one second duration, and the BOLD response was recorded within a four second silence after. 25% of the trials were null (blank trials). In this experimental design, all 12 notes for each of the three conditions was tested twice, resulting in 96 trials.

We ran a standard General Linear Model (GLM) analysis with FSL FEAT on the octave BOLD response data, and our next steps are to continue to analyze the contrast between response to the same note and a different note, and the contrast between the octave interval and two different notes. We ran the octave stimuli in the fMRI scanner twice for one subject, and so another step would be to obtain more data in order to boost our statistical power in correlating a region of the brain with octave recognition.

KEY CONTEXT AND SCALE DEGREE RECOGNITION:

In the second theoretical experiment, we are interested in identifying the role of a key in note perception. A usual key consists of 7 notes (1-7 scale degrees), as shown in Figure 4. The key can be established by playing sequences of these 7 notes, because the relationship between these 7 notes is always distinct (no matter which of the 12 notes the scale starts on).

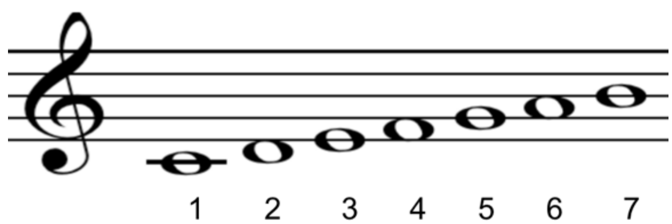

Figure 4. Seven-Note Series Creates a Key The key of C-Major is shown above. The distinct relationship between the 7 notes (labeled as the seven "scale degrees"), gives rise to the Western tuning system and Western music theory.

We would like to find regions of the brain that respond to key context, and to do so we came up with a design that presents the same note in may different keys. The note is physically the same (same frequency) but the context of key changes (notes played temporally around the note of interest), which makes the note of interest sound different (creating a different response). Figure 4 shows the conditions that we used in our experiment. We used four distinct end notes of interest (C, B, B-flat, and A), and the notes of interest were preceded by a phrase of notes in each of the seven keys listed in the note's row (as shown in Figure 5). The seven keys listed next to each note of interest make the note of interest a particular scale degree 1-7 within each key context. "0" represents a control set of pitches, which give rise to no particular key perceptually.

End Note	Scale Degrees 0	1	2	3	4	5	6	7
C	[0	C	Bb	Ab	G	F	Eb	Db]
B	[0	B	A	G	Gb	E	D	C]
Bb	[0	Bb	Ab	Gb	F	Eb	Db	B]
A	[0	A	G	F	E	D	C	Bb]

Figure 5. Key Experimental Design in Relation to End Notes The elements of the above matrix show letters referring to the key context that a melody will present. The four end notes chosen will be ended on by 7 different key contexts (where the end note is scale degree 1-7). Trials marked by 0 represent a control keyless phrase, that will end with the note without being a scale degree.

Figure 6 below gives a picture of the note sequences in each trial, for the C note of interest (row 1 in Figure 5).

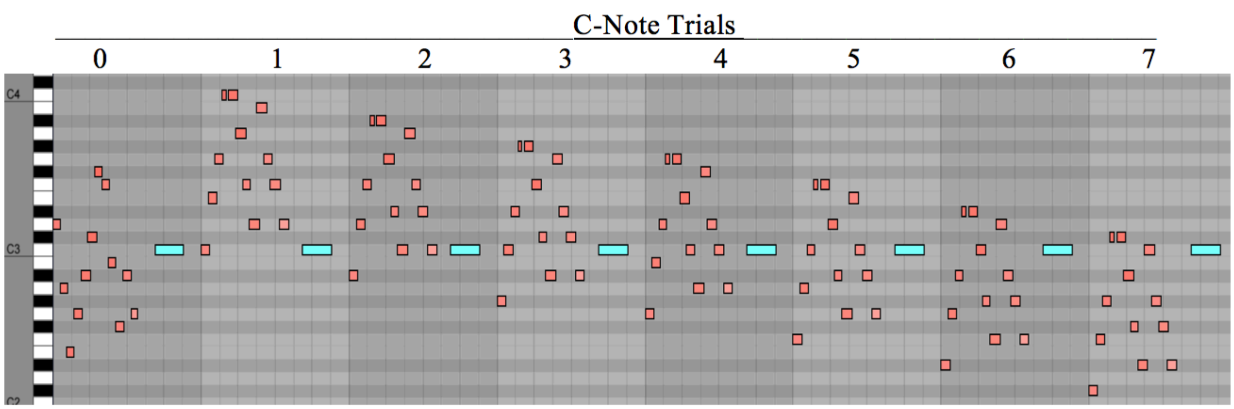

Figure 6. MIDI Representation of C-Note Trials The sequence of notes for the eight C-Note trials are displayed above. The end note for each trial is colored in blue, and is always the note C. However, each time the C-Note is played, it is preceded by a different key context, and so the note is reflected as a different scale degree (1-7). "0" marks the atonal trial (random notes without a key relationship). The specific keys played in each trial are given in row 1 of Figure 5.

All above trials (Figure 5) were repeated 2 times each during the scanning session (two runs doubled that amount to 4 times each). Each trial took 4 seconds, and the trials were separated by a jitter silence between 4 and 8 seconds. The order of presentation of the trials was randomized using Bob Spunt's easy-optimize-x MATLAB tool. This tool finds the most efficient order by taking desired contrasts into account. The stimuli design can be seen in Figure 7.

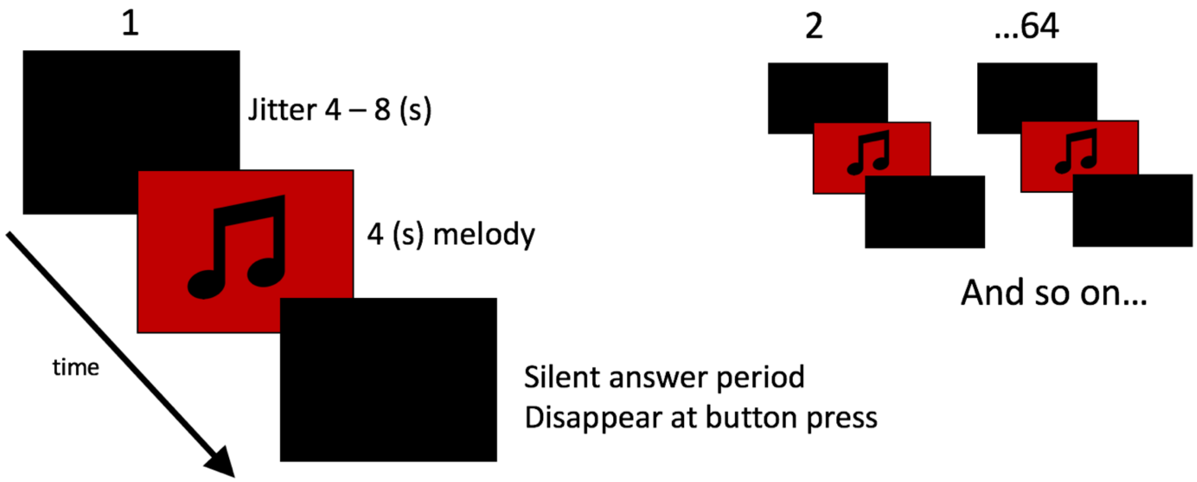

Figure 7. Key Context Experiment Design The sequence of notes in each trial fit in a 4 second duration. Pauses between trials jittered between 4 and 8 seconds. Each trial in Figure 5 was repeated twice, which resulted in 64 trials total.

For the key context experiment, there are three main conceptual variables that can be analyzed in FSL FEAT in order to detect brain regions that correlate with our conditions:

1. *Scale degrees 1 – 7 (columns in Figure 5)* The times that the same scale degrees were used (regardless of end note) can be compiled into a list and used as a repressor in FSL. We contrast these 7 variables (scale degree 1-7) with the trials in which the note had no scale degree context.
2. *Arrival at the same note (rows in Figure 5)* A list of timings can be made based on when a note of interest was the same note (8 times for each note). After FEAT analysis using these features, we may be able to report regions of the brain that respond to just the physical note regardless of context (or perhaps the response always changes, which would further prove that the brain characterizes notes based on scale degrees).
3. *The same key (all of the same letter within the matrix in Figure 5)* These variables can be modeled based on when the exact same sequence of pitches (same key) were repeated across trials. This may isolate response to just the melodic portion of the stimuli (and not the last note). Contrast of response to sequences in a key to the atonal sequences also may reveal brain regions implicated in key recognition. Preliminary results of this contrast can be seen in Figure 8, in which the response to each of the 12 keys was contrasted against the atonal melody response.

We hypothesize that these three types of variables will provide us with interesting information on how the brain represents structures in music theory. We continue to analyze this data in FSL FEAT and a possible next step would be to collect more data in more than one subject.

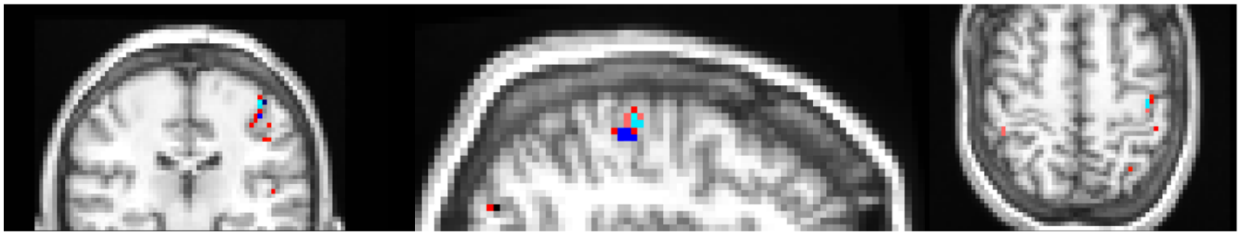

Figure 8. Contrast Between Key and Atonal Response The fMRI 3-dimensional image above shows a coronal view (left), sagittal view (middle), and transverse view (right) of the brain. Each color displays the contrast between response to a key (1-12) and the atonal melody.

CONCLUSION:

In this SURF project, we designed and executed two hypothesis-based music theory experiments and collected naturalistic data in order to analyze the brain's representation of various musical features. These features can be based on the physical audio signal, in which we used Music Information Retrieval techniques to extract the features. They can also be based on perceptual music theory as defined by our Western tuning system. We continue to investigate the preliminary data we obtained, and may collect more data based on what kind of brain responses we see.

REFERENCES:

McDermott, J.H., Lehr, A. J., Oxenham, Andrew J. (2008) Is Relative Pitch Specific to Pitch? Association for Psychological Science Vol. 19 No. 12. http://mcdermottlab.mit.edu/papers/McDermott_Lehr_Oxenham_2008_relative_pitch.pdf

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J. L. (2011) Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Current Biology Vol. 21, 1641-1646. http://gallantlab.org/_downloads/2011a.Nishimoto.etal.pdf

Norman-Haignere, S., Kanwisher, N., McDermott, J.H. (2013) Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex. The Journal of Neuroscience http://mcdermottlab.mit.edu/papers/Norman-Haignere_Kanwisher_McDermott_2013_pitch_fMRI.pdf

Norman-Haignere, S., Kanwisher, N., McDermott, J.H. (2015) Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. Neuron Vol.88, 1281-1296. http://mcdermottlab.mit.edu/papers/NormanHaignere_etal_2015_voxel_decomposition.pdf