Exploring the Role of Temporal Fine Structure and Envelope in Timbral Coding

Presented By Andrew Sivaprakasam

12/07/2020

Music Coding by the Auditory Periphery

To appreciate an art, we must first be able to sense it. But what happens when our senses are impaired? What kind of signals do we receive from our sensory inputs, and in what fidelity do we receive them? What musical features are most important for us to "lock on" to, and what happens when we can't lock on to them?

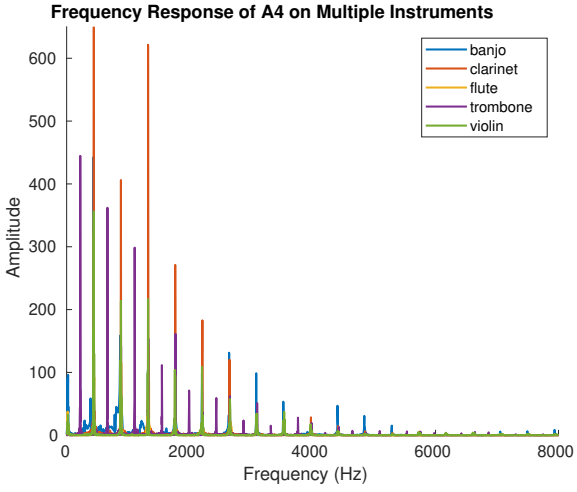
- Musical stimuli are not simple, they are **often non-periodic and spectrally complicated**
- If we look *purely* at how we encode components of music, and how important the fidelity we encode those components is, **we can drive innovative processing algorithms used in hearing aids and cochlear implants to better represent musical sounds**

  A start would be looking at how **timbre** is encoded by the auditory system, and what the signals "look like" before being sent to the brain

What is *timbre*?

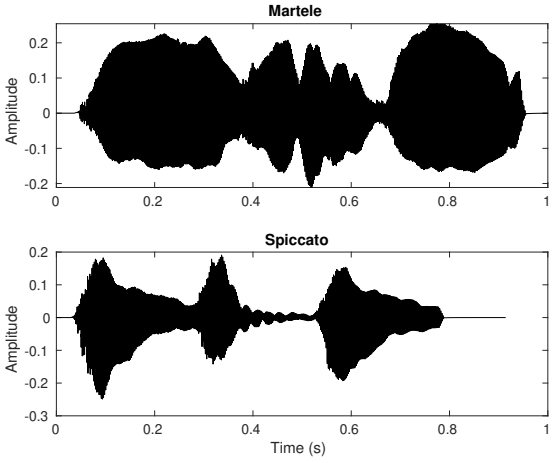
Timbre is the *color* or *quality* of a tone perceived by a listener

- It is a complex psychoacoustic phenomenon that still is not completely understood
- Timbre is what helps listeners differentiate when the same tone is played by a different instrument
- Most consider instrumental timbre to be due to the difference in the magnitude of the harmonics of a given tone



What is *timbre*?

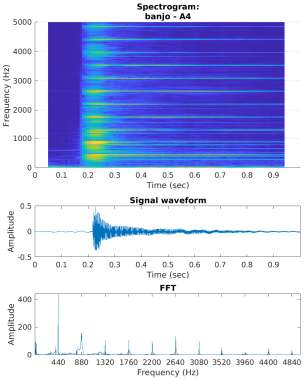
- However, envelope is also important
- In fact, a 2011 study by Heng *et al.* demonstrated that cochlear implant users can differentiate between instruments when 0% TFS is used in a music-noise chimaera
- Normal hearing users, however, rely primarily on TFS cues to differentiate instruments

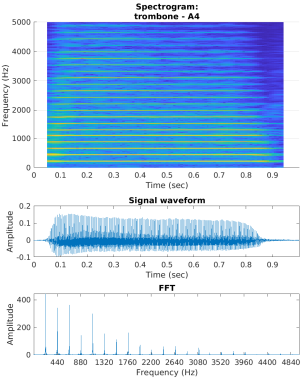


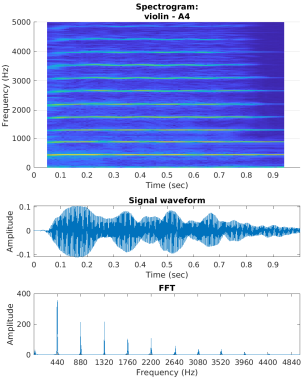
Envelope cues could help us differentiate between different articulations in music, such as *martele* and *spiccato* played by the violin

Spectrotemporal Analysis

Spectrograms are an useful tool to visualize differences in instruments. Here are a few of the stimuli I used. The sound samples were collected by the Philharmonia Orchestra in London.







The Idea

Goal: Use a spectrally-specific framework designed to inspect the relevance of TFS and ENV in auditory nerve responses to investigate timbral coding in normal hearing and hearing impaired conditions.

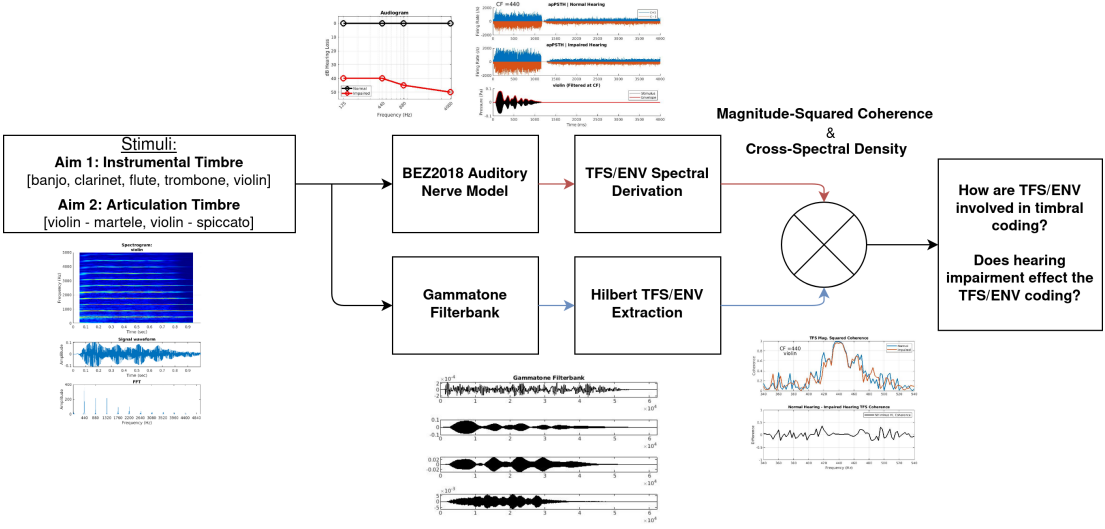
Aim 1: **Instrumental** Timbral Coding:

- Compare timbral coding differences between **banjo, clarinet, flute, trombone, violin** for a common pitch, **A4** (440 Hz)

Aim 2: **Articulation** Timbral Coding:

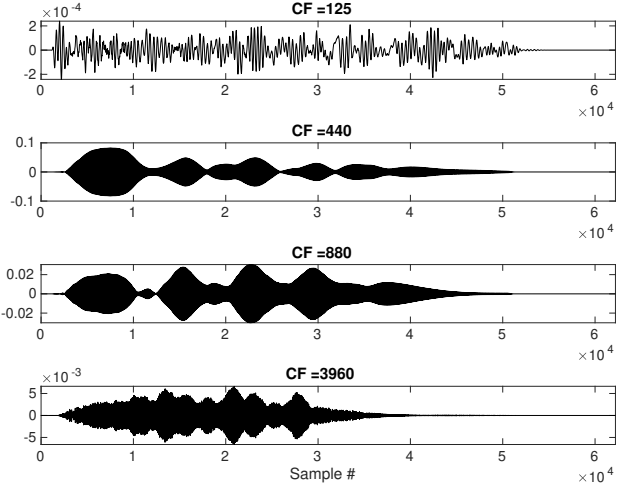
- Compare timbral coding differences between two articulations on the violin, **martelé** and **spiccato**

General Methods Overview



Gammatone Filterbank and Envelope Extraction

- Used a computationally efficient gammatone filterbank developed by Ning Ma
- Filtered each stimulus using same CFs used for model
- Extracted TFS and ENV of each channel using Hilbert analytical signal



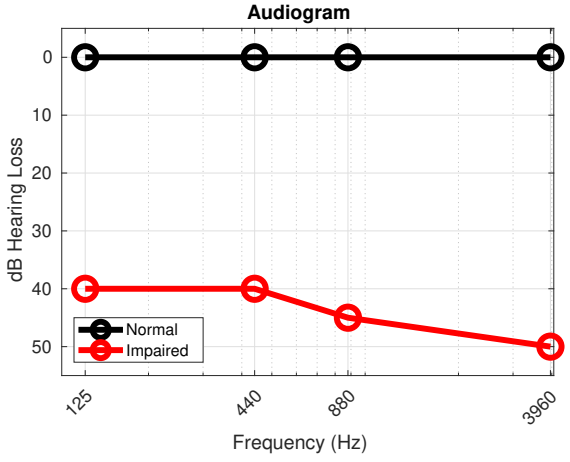
Modeling Approach

- Using a model of the auditory nerve developed by Zilany et al. and updated by Bruce et al. in 2018, I was able to quickly collect data.
- This model can be run in Matlab and also has a variety of customizable parameters.
- I simulated auditory nerve responses at 4 different Characteristic Frequencies (CF):

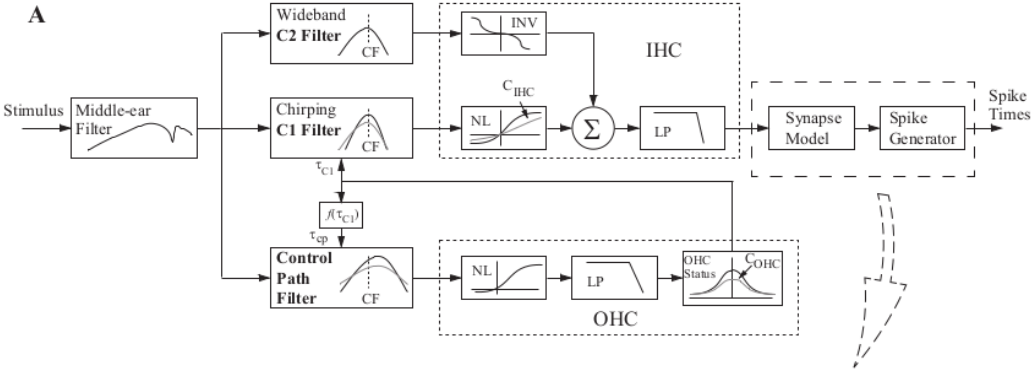
  125 Hz, F_0 (440 Hz), F_1 (880 Hz), and F_8 (3960 Hz)
- The Organ of Corti in the cochlea can be thought of as a filterbank with filters centered around a given CF (the further down the cochlea you go, the lower the CF)

Modeling Approach

This model also allowed adjustment of Outer Hair Cell and Inner Hair Cell parameters based on an audiogram. I chose hearing loss threshold shifts roughly based on data reported by Tufts et al. in a 2005 study of musical interval consonance:



Modeling Approach



A schematic of the original Zilany Model. Sound stimulus goes in, spike train comes out.

Model Parameters and Implementation

Using BEZ2018 Model:

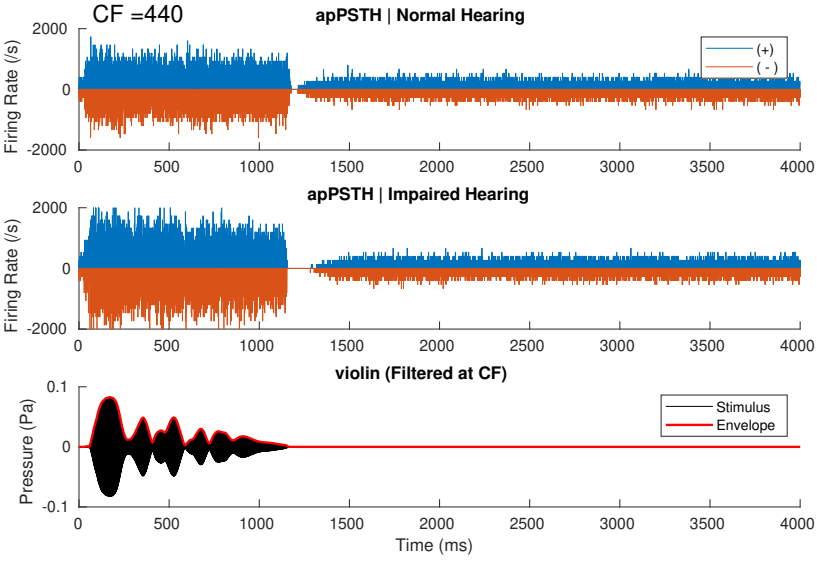
Parameters:

- CF (Hz) = 125, 440, 880, 3960
- Repetitions/CF = 75
- Spont. Rate = 70
- Abs and Baseline mean relative refrac. period = 0.6 ms
- IHC/OHC health = set by `fitaudiogram2`
- Species = Human - Shera tuning
- Fractional Gauss. noise type = 0 (fixed)
- Power-law functions = 0 (approximate)
- **Spike time resolution = 10** μs

Other Details:

- Compensated for threshold shift in Hearing Loss condition by increasing stimulus gain by mean threshold shift (43.75 dB)

Model Output



apPSTH Analysis

Used the rationale developed by Parida et al. to compute $s(t)$ and $\phi(t)$ and their spectrograms corresponding to ENV and TFS characteristics:

ENV-related response, the *sum* PSTH (polarity-tolerant):

$$s(t) = \frac{p(t)+n(t)}{2}$$

TFS-related response to PSTH (polarity-sensitive):

$$\phi(t) = \sqrt{2} \times rms[d(t)] \times \cos[\angle d(t)]$$

Where:

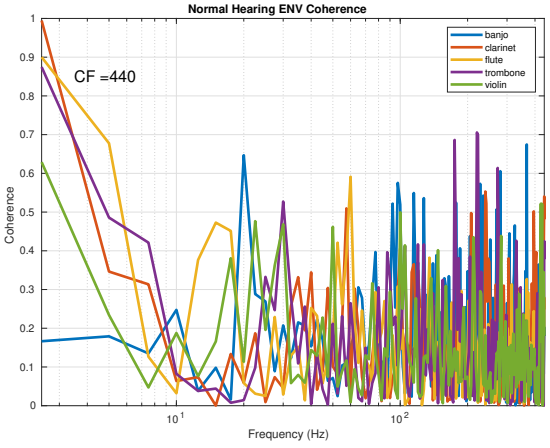
$$d(t) = \frac{p(t)-n(t)}{2}$$

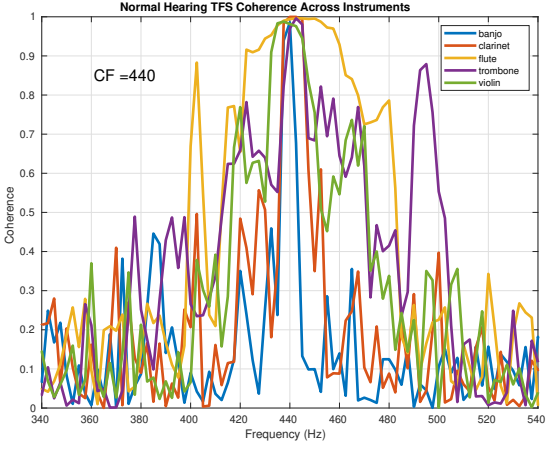
Measuring the "strength" of TFS/ENV Coding

The coding "strength" of the ENV and TFS nerve responses can then be assessed using either cross-spectral density or coherence with the Hilbert ENV or TFS. Magnitude-squared Coherence, $C_{xy}(f)$, is a nice way to compare these, since it is essentially a normalized cross-spectral density:

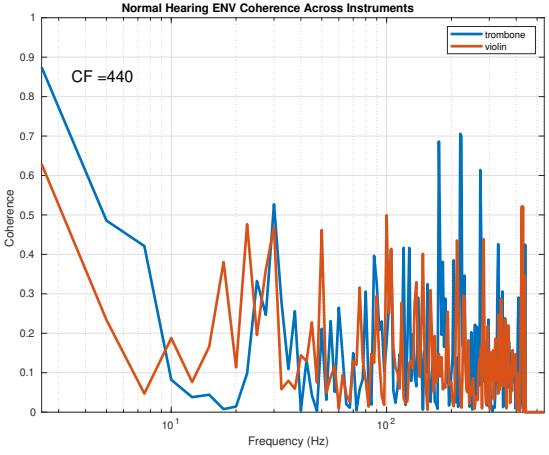
$$C_{xy}(f) = \frac{|P_{xy}(f)^2|}{P_{xx}(f)P_{yy}(f)}$$

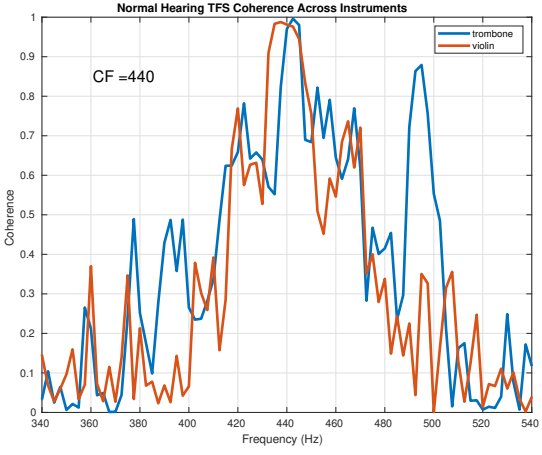
ENV and TFS Coherence Across Instruments



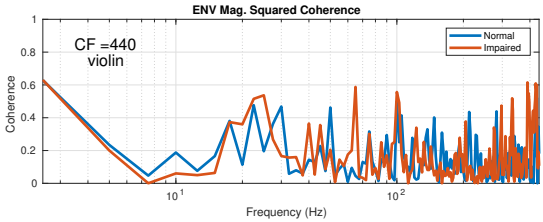


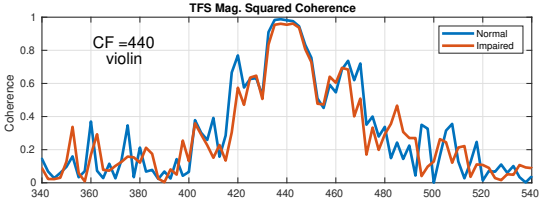
ENV and TFS Coherence Across Instruments

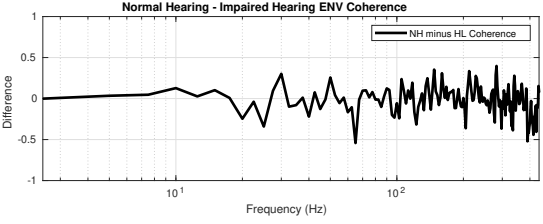


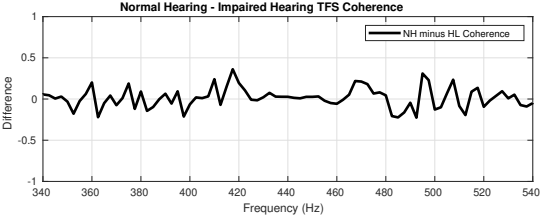


Effects of Hearing Loss on ENV and TFS Coherence

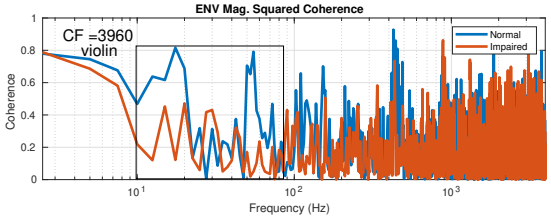


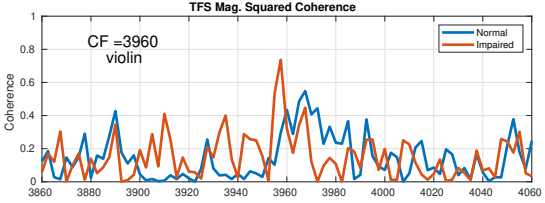


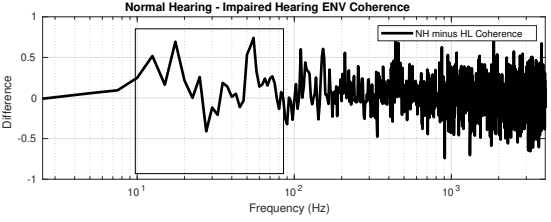


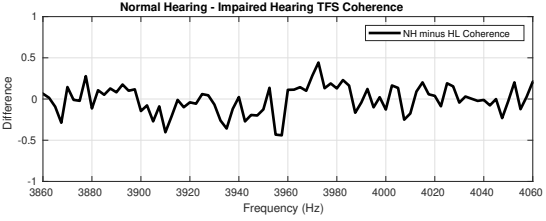


Effects of Hearing Loss on ENV and TFS Coherence

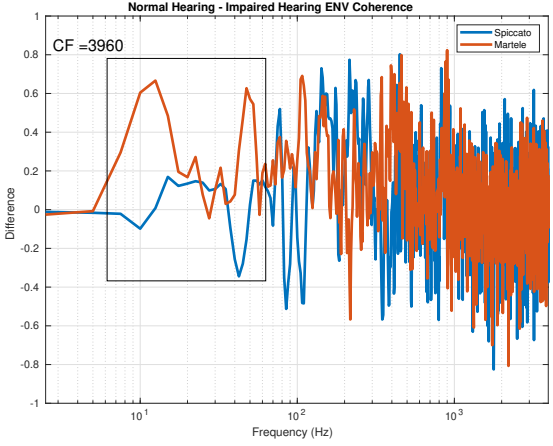


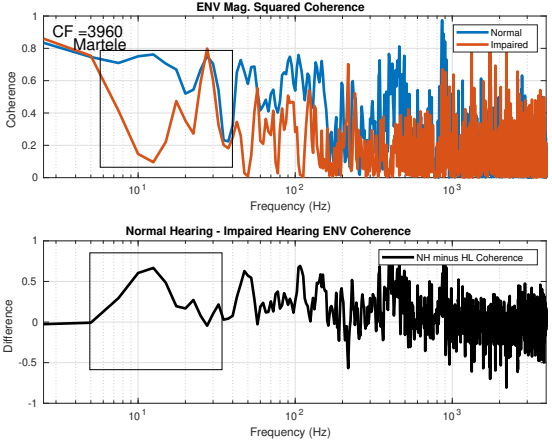




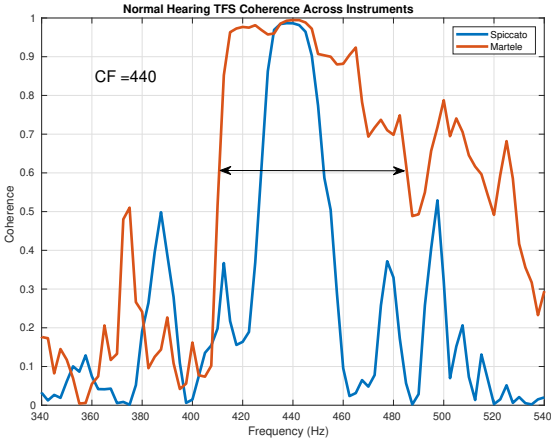


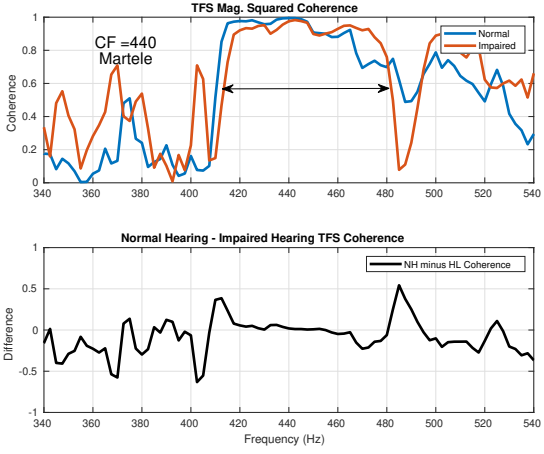
Hearing Loss May Effect Timbral Coding of ENV in Articulation





Hearing Loss May Not Affect Timbral Coding of TFS in Articulation





Conclusion

- A framework was designed to investigate TFS/ENV coding of timbre
- The spectral composition of instrumental tones and articulation is non-trivial
- Understanding the way we perceive sound beyond the auditory periphery is likely very important in solidifying the relevance of TFS/ENV in timbral perception
- Even at the peripheral level, it appears coding of ENV and TFS may be affected by hearing loss
- The framework used to analyze ENV/TFS coding can be extended to *in vivo* auditory nerve measurements or non-invasive measures collected using EEG, so could be fun to try this in the lab

Future Considerations/Work

- There are a lot of coherence plots I can generate with my code...but how to sort through them
- These timbral coding effects vary from instrument to instrument so finding a unifying, non-noisy, metric to measure across multiple conditions would be helpful
- Other parameters to mess with? Any shortcomings of model to consider?
- How can I reduce variance in the coherence? Multi-tapered Coherence?
- Worth studying further? Other ideas of things to look at?

Questions/Other points of Discussion??