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

Presented By Andrew Sivaprakasam— BME 511 Fall 2020

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?

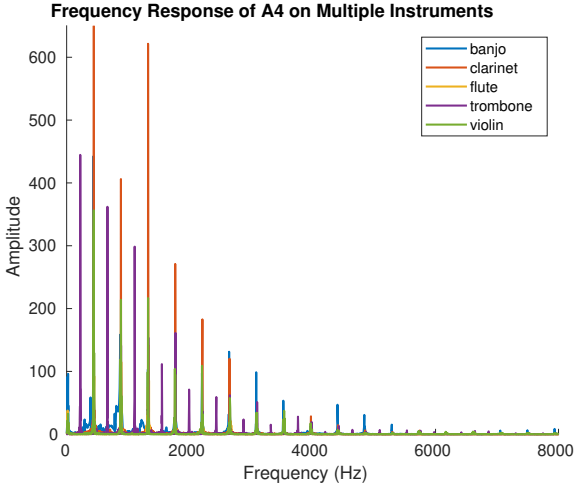
- Musical stimuli are not simple, they are **often non-periodic and spectrally complicated**
- If we look *purely* at how we encode components of music, and how important the fidelity we encode those components is, **we can drive innovative processing algorithms used in hearing aids and cochlear implants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

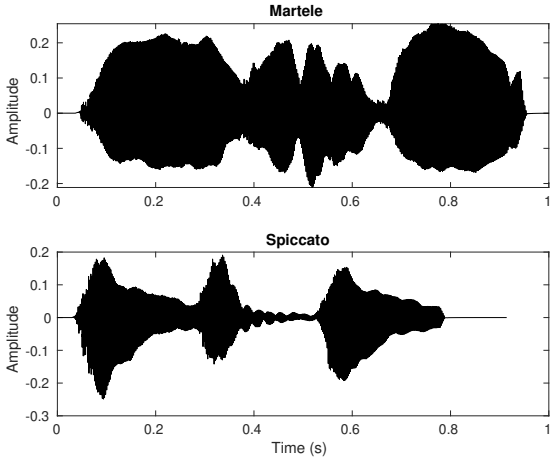
Timbre is the *color* or *quality* of a tone perceived by a listener

- 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*?

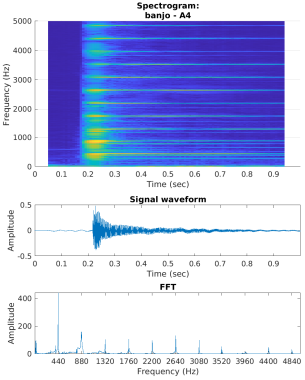
- However, envelope is also important
- In fact, a 2011 study by Heng *et al.* demonstrated that cochlear implant users can differentiate between instruments when 0% TFS is used in a music-noise chimaera
- Normal hearing users, however, rely primarily on TFS cues to differentiate instruments

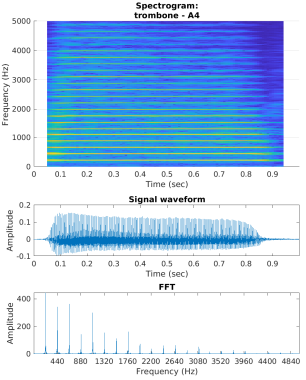


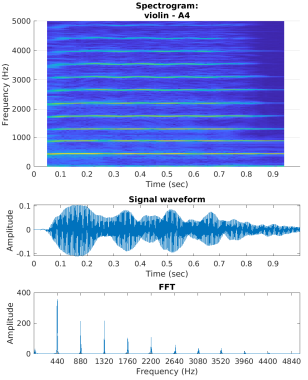
Envelope cues could help us differentiate between different articulations in music, such as *martele* and *spiccato* played by the violin

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.

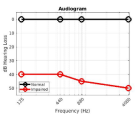
Aim 1: **Instrumental** Timbral Coding:

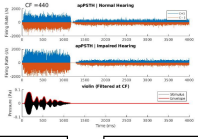
- Compare timbral coding differences between **banjo, clarinet, flute, trombone, violin** for a common pitch, **A4** (440 Hz)

Aim 2: **Articulation** Timbral Coding:

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

General Methods Overview





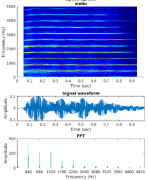
Stimuli:

Aim 1: Instrumental Timbre

[banjo, clarinet, flute, trombone, violin]

Aim 2: Articulation Timbre

[violin - martele, violin - spiccato]

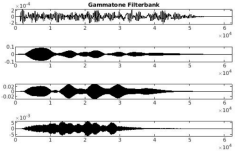


BEZ2018 Auditory Nerve Model

TFS/ENV Spectral Derivation

Gammatone Filterbank

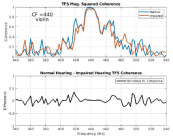
Hilbert TFS/ENV Extraction



Magnitude-Squared Coherence

&

Cross-Spectral Density



How are TFS/ENV involved in timbral coding?

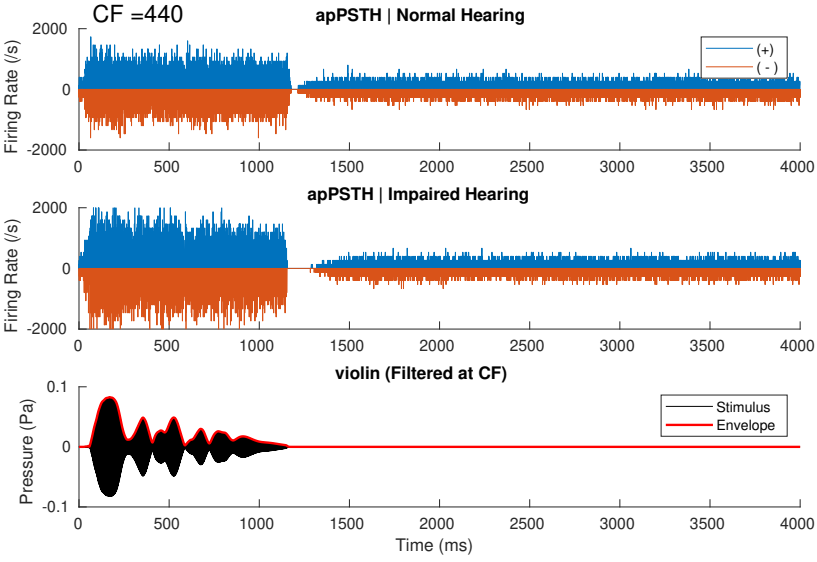
Does hearing impairment effect the TFS/ENV coding?

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)

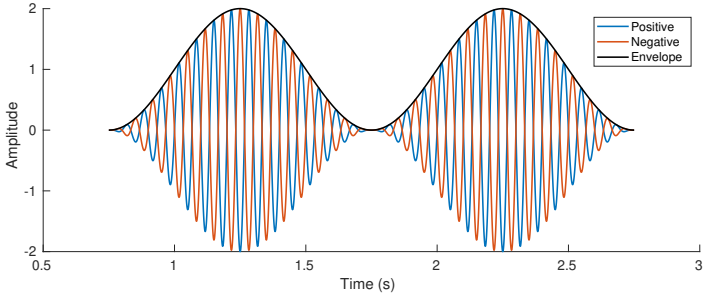
Model Output



Why Use Different Polarities?

Using alternate polarities allows us to isolate the effects of the envelope of a stimulus (the polarity-tolerant component) and the temporal fine structure (the polarity-sensitive component).

By summating the responses to both polarities (and dividing by two) the envelope response can be isolated, while taking the difference removes the envelope and isolates the fine structure.

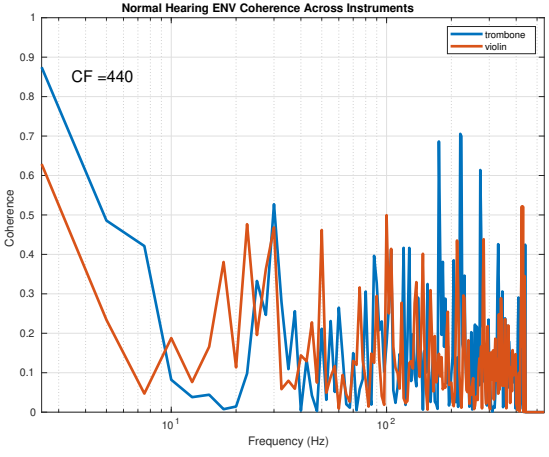


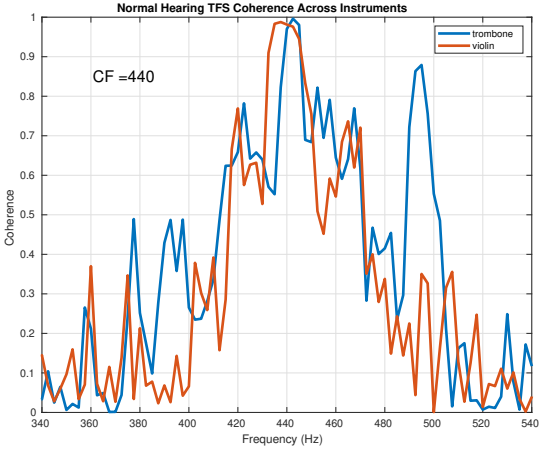
Measuring the "Strength" of TFS/ENV Coding

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:

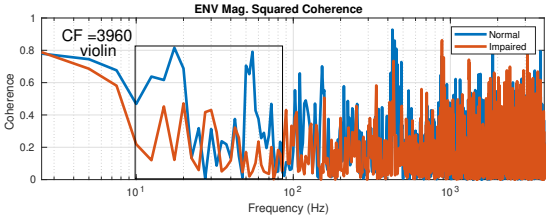
$$C_{xy}(f) = \frac{|P_{xy}(f)^2|}{P_{xx}(f)P_{yy}(f)}$$

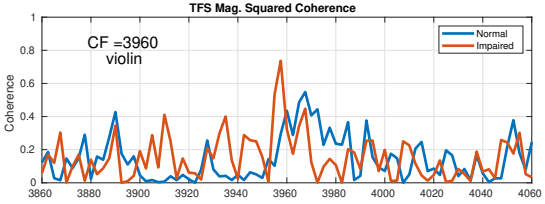
ENV and TFS Coherence Across Two Instruments

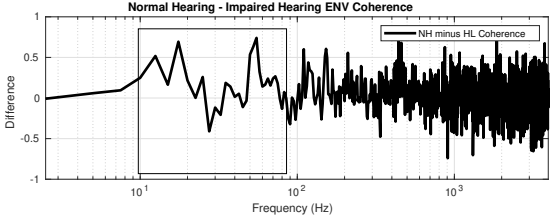


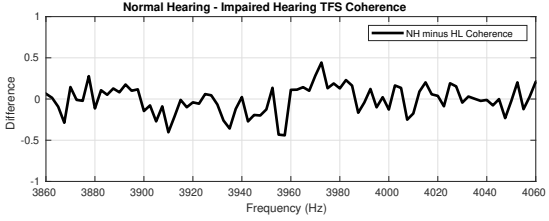


Effects of Hearing Loss on ENV and TFS Coherence

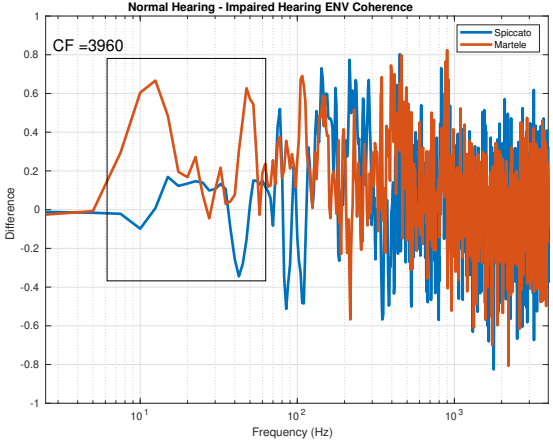


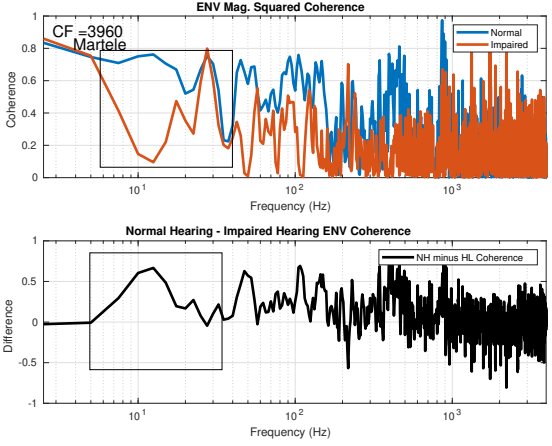






Hearing Loss May Effect Timbral Coding of ENV in Articulation





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 is affected by hearing loss
- There are a lot of coherence plots in my code, so finding a unifying metric to analyze across conditions would be useful
- The framework used here can be extended to look at EEG measures *in vivo* or compared to auditory nerve experiments