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

Andrew Sivaprakasam

Weldon School of Biomedical Engineering

Purdue University

West Lafayette, USA

asivapr@purdue.edu

Abstract—While the neural response to simple, stationary, and periodic auditory signals can be fairly well investigated, responses to auditory stimuli that are more complex, such as speech and music, are less well-characterized. Particularly, music is psychoacoustically complex. It is not well-understood how humans perceive the nuances of music, and how hearing impairment may affect the perception of such nuances. Before we can fully understand perception, we must first investigate how musical attributes like timbre are coded by the auditory periphery. By using a simulated auditory nerve model and comparing neural responses to stimulus envelope (ENV) and temporal fine structure (TFS), it is possible to see how timbral coding might be affected by hearing impairment. In this project, both instrumental timbre and articulation timbre were considered, and variations in coherence spectra of neural responses and Hilbert TFS/ENV were observed across instruments, articulations, and hearing impairment conditions.

Index Terms—auditory, neuroscience, music, modeling, envelope, temporal fine structure, timbre

I. INTRODUCTION

The field of auditory neuroscience has made leaps and bounds in understanding how we percieve and code sounds that reach the cochlea. However, despite much study of perception of speech intelligibility and discrimination, the perception and coding of *music* still is quite under-investigated and remains an enigma. Though many auditory neuroscientists are musically inclined, music is quite complicated to study. Music is non-periodic and spectrally complicated. In the real-world setting, music is complex. Whether it be a symphony, rock concert, or music festival the soundstage varies, as does the articulation and instrumentation of the music itself, which makes it especially difficult to come up with a general set of rules or ideas that govern the way we hear and appreciate music. Additionally, psychoacoustics and musical training play a strong role in hearing [1]. Therefore, it is important to consider the differences in neural coding and perception that exist between varying instrumentation and articulation.

One way of assessing these difference is investigating the *timbre* and how we respond to sounds of varying timbre. Timbre is the complex psychoacoustic phenomenon that allows us to discriminate between different "colors" or "qualities" of music. It is also the phenomenon that allows us to know when the same pitch is played by a different instrument. Though the perception of timbre is likely dependent on perception and
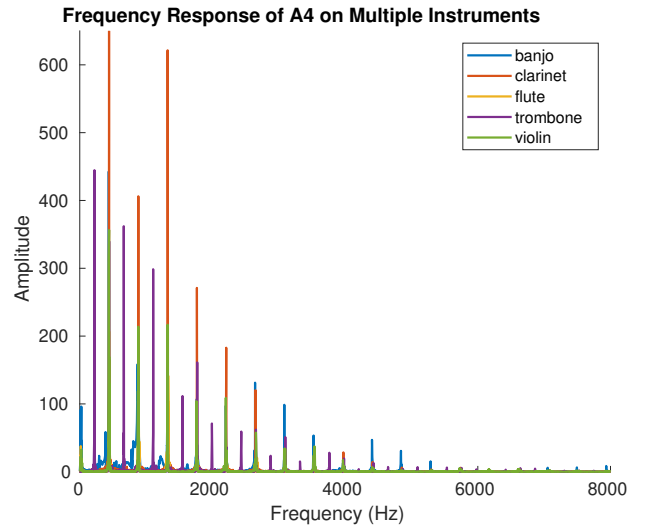

Fig. 1. Spectra of five different instruments playing an A4 (440 Hz) tone. Variations in harmonic magnitude between instruments play a role in their characteristic sounds.

musical training, there is no doubt that physical properties of timbre and the coding of these features are important, as well. In fact, recent work shows that normal hearing subjects rely on temporal fine structure to differentiate between instruments, while cochlear implant users rely almost exclusively on the envelope [2]. But the question remains, *what features of timbral coding are most relevant, and are these features impacted by hearing loss?*

Answering this question could lead to innovations in hearing-assistive technology to improve the representation of music in devices such as cochlear implants and hearing aids.

A. The Physics of Timbre

Variations in spectrotemporal content between instruments and their articulations is a good first step in assessing the differences in timbre from instrument to instrument. Isolating the physical characteristics of a sound played by one instrument that makes it unique from another is a key step in being able to study the coding of those characteristics.

Figure 1 shows the variations in the magnitude of the harmonics present in the same A4 (440 Hz) tone played on
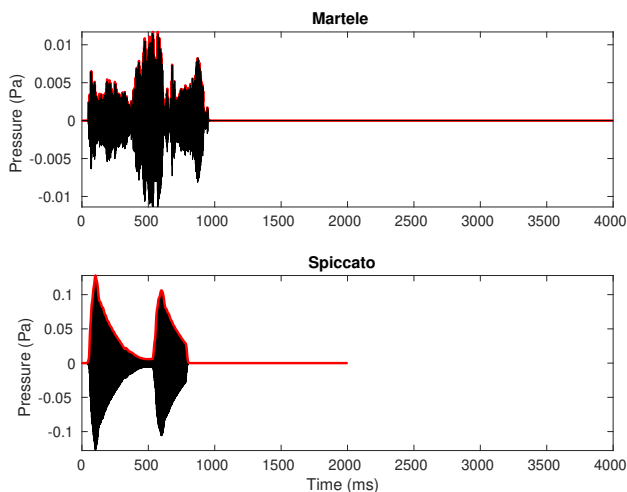
Fig. 2. Variations in envelope (red) for two different articulations played by the violin. Waveforms were gammatone-filtered with a central frequency of 440 Hz.

five different instruments. These same harmonics are also very pronounced in the spectrograms illustrated in Figure 3. Though it is somewhat obvious that these harmonic differences alter the physics and fine structure of a musical note, it less known how important this fine structure is in driving the coding and perception of that note by the auditory system.

The envelope of a tone also varies from instrument to instrument. Changes in envelope, even within a single instrument, can be observed by playing different articulations of that instrument. Figure 2 demonstrates variations in the envelope between the two violin articulations *martelé* and *spiccato*. Martelé is a more "hammered" style of bowing, while spiccato is more "bouncy, light, and separate."

B. Methods of Analysis for Investigating TFS and ENV Neural Responses

Much progress has been made on developing methods and analyses by which we can study how features of a particular sound stimulus reach the auditory nerve, and how they may be processed by higher order systems within the brain and brainstem. Particularly, the envelope (ENV) and temporal fine structure (TFS) of a given stimulus have been established as having relevance in its perception [3]. However, *perception* is different from *coding*, and further signal processing methods were designed to characterize how TFS and ENV features are encoded by the auditory nerve [4]. More recently, Parida et al. have developed a spectrally-specific framework for investigating ENV and TFS that can be applied to spike trains *and* non-invasive far-field measures like frequency following responses (FFRs) [5]. This framework can also be applied to non-stationary signals.

C. Modeling the Auditory Nerve

To record neural spike data at the level of the auditory nerve is an acute and invasive process. Each experiment may span

several hours, or even days. Therefore, being sure that the stimulus presentation and signal acquistion match the question being investigated is important. One way to reduce the time needed to get data and animals needed for experimentation is using an auditory nerve model.

Zilany et al. created the auditory nerve model that was last updated in 2018, BEZ2018 [6]. This model is highly regarded by the field of auditory neuroscience. Though there is no replacement for true electrophysiological data, it is helpful to have an *in silico* model with customizable parameters that may guide *in vivo* experimentation in the future. At the very least, it allows researchers to test the validity of spike train analysis so that when *in vivo* experiments may be conducted, the analysis is refined enough to answer the question at hand.

D. A Generalizable Framework for Studying Timbre

While studies have investigated timbral *perception* and have even proposed means of quantifying this perceptual ability, we are not aware of any studies that have investigated timbral *coding* [7]. By combining the modeling work from Zilany et al. and the spectrally specific framework designed by Parida et al., a new framework for analysis can be created to study ENV and TFS coding in various instrumental sounds and how they may or may not be impacted by hearing loss. By using this framework to identify relevant timbral coding features *in silico*, recommendations for future *in vivo* timbral coding experiments can be suggested. Additionally, the spectral analysis methods outlined in this new framework can be applied to *both* invasive auditory nerve experiments and non-invasive FFR experiments due to the generalizabilty of the methods developed by Parida et al.

II. METHODS

A framework was developed to investigate timbral coding using a combination of auditory nerve modeling and application of a spectrally-specific framework to study the strength of ENV and TFS coding across various instruments and articulations. Figure 5 represents a general flow diagram of the methods used.

A. Sound Stimuli

All sound stimuli were acquired from the Philharmonia Orchestra's (UK) sound sample database [8]. This sound sample library has thousands of neatly categorized and concise recordings that span the instrumental range of the standard orchestra, over several pitches, articulations, and expressions. The sound samples were resampled in MATLAB (Natick, MA) to 100 kHz in order to be processed by the auditory nerve model. To assess timbral coding variations between instruments, banjo, clarinet, flute, trombone, and violin A4 (440 Hz) tones were used. To assess differences in timbral coding between articulation, a martelé and spiccato sound at A4 on the violin was used.
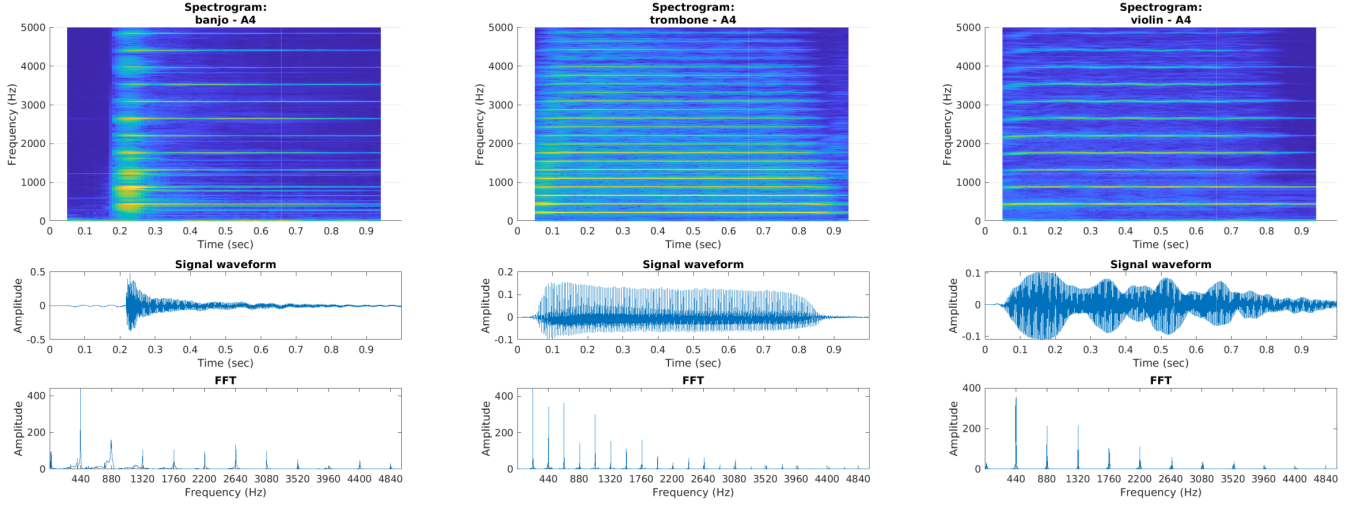
Fig. 3. Spectrograms of three different instruments playing A4. Differences in onset and decay characterstics, modulations, and harmonic content may be visualized. Interestingly, harmonics of A3 (220 Hz) were found in the trombone recording for A4, demonstrating that timbral features may not necessarily be limited to the harmonics of the fundamental frequency of the note being played.

B. Auditory Nerve Model

The BEZ2018 model was used to generate spike histograms in response to both positive and negative polarities of the stimuli. Auditory neurons with characteristic frequencies (CFs) at 125, 440, 880, 3960 Hz were included in this simulation. In the normal hearing model, sound was played at 65 dB. The hearing-impaired condition had an average threshold shift of 43.75 dB, and this gain was added to the stimulus for a total stimulus gain of 108.75 dB. The outputs of this model were normal hearing and impaired hearing alternating-polarity peristumulus time histograms (apPSTHs) for a given stimulus, shown in Figure 4. model parameters in table?

C. Stimulus Filtering and Hilbert Extraction

Stimuli were passed through a computationally efficient gammatone filterbank developed by Ning Ma [9]. The center frequencies for each filter band were chosen to correspond with the auditory neuron characteristic frequentions mentioned above (125, 440, 880, 3960 Hz). The resultant channels were then processed using the Hilbert analytic signal to extract the stimulus ENV and TFS for a given CF.

D. Spectral Computation and Analysis

Using the same rationale developed by Parida et al., the ENV and TFS response of the apPSTHs generated by the auditory nerve model were computed as $s(t)$ and $\phi(t)$, respectively:

$$s(t) = \frac{p(t) + n(t)}{2} \qquad (1)$$

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

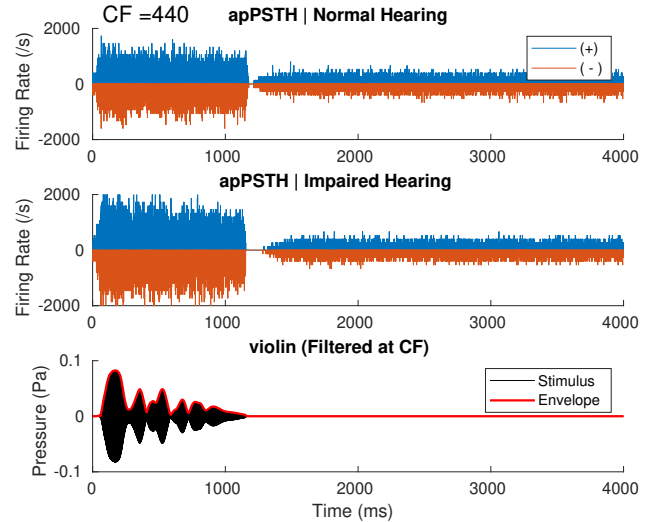$$\phi(t) = \sqrt{2} \times rms[d(t)] \times \cos[\angle d(t)] \qquad (3)$$

Fig. 4. Normal and impaired hearing apPSTHs in response to a A4 on the violin filtered at CF = 440 Hz. Stimulus envelope highlighted in red in bottom figure.

Where $p(t)$ corresponds to the positive-polarity PSTH and $n(t)$ corresponds to the negative-polarity response.

The normalized cross spectral density, magnitude-squared coherence, C_{xy}, was used to assess the strength of ENV and TFS coding:

$$C_{xy}(f) = \frac{|P_{xy}(f)^2|}{P_{xx}(f)P_{yy}(f)} \qquad (4)$$

Where x corresponds to either $s(t)$ or $\phi(t)$ and y corresponds to the Hilbert ENV or TFS of the filtered stimulus, respectively.
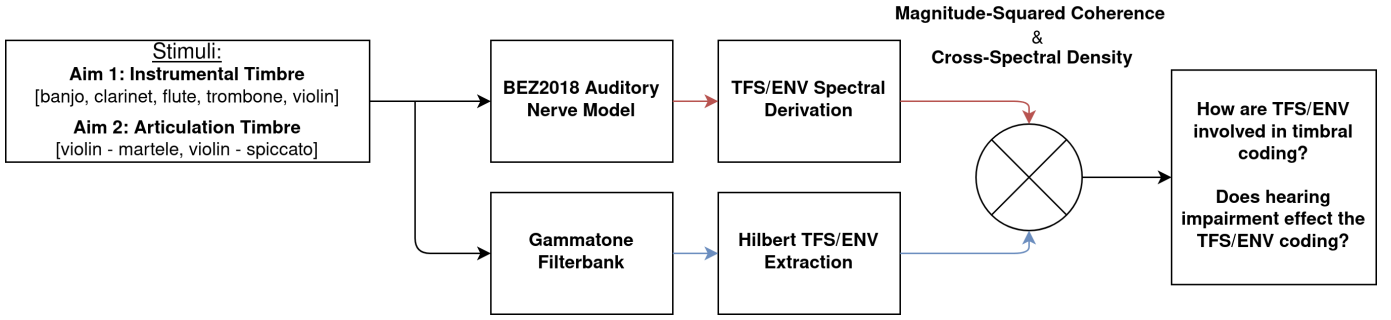
Fig. 5. Methods flow diagram. Different instrument and articulation stimuli were processed through the BEZ2018 auditory nerve model and through a gammatone filterbank. Alternating-polarity peristimulus time histograms (apPSTHs) (red) from the auditory nerve model were processed to extract TFS and ENV-related responses, and these were compared to the Hilbert TFS and ENV (blue) extracted from the gammatone-filtered stimuli by means of Magnitude-Squared Coherence.

III. RESULTS

Magnitude-squared coherence of ENV and TFS characteristics were visually inspected in MATLAB for the different instruments and two articulations. Differences between normal and impaired hearing ENV and TFS coherence spectra were also computed. The data analysis is ongoing, but a few notable findings to date are presented here.

A. Timbral Coding of Instrumental Features

Figure 6 demonstrates that TFS coding of A4 in a simulated healthy auditory neuron with CF = 440 Hz played on trombone compared violin showed an additional coherence peak at approximately 490 Hz. High coherence at 440 Hz was expected in both instruments, given that there should be strong coding and phase locking of a 440 Hz tone by a neuron with CF = 440 Hz. However, a peak at 490 Hz is not a linear combination of the fundamental A4, and could be coherence related to a timbre-specific attribute present in the trombone neural response.
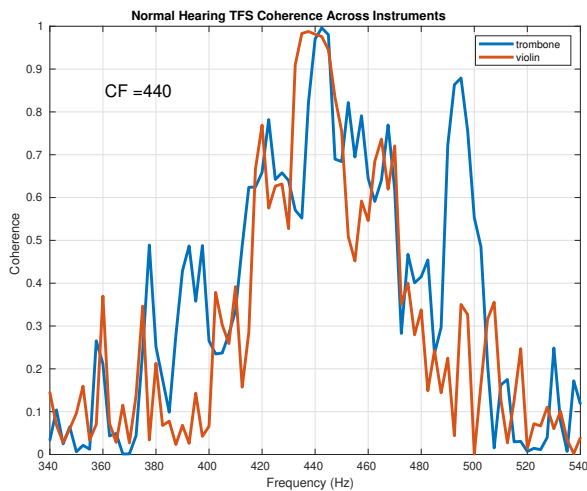

Fig. 6. TFS coherence differences between violin and trombone in a CF = 440 Hz auditory neuron. Of particular note is a 490 Hz peak in the trombone that is absent in the violin TFS coherence.

Additionally, differences were observed in ENV coherence of violin-specific attributes in normal and impaired hearing conditions, especially in the higher-frequency neuron (CF = 3960 Hz). Figure 7 shows that the decreased ENV coherence in hearing impaired neurons is most pronounced in the 10-100 Hz region.
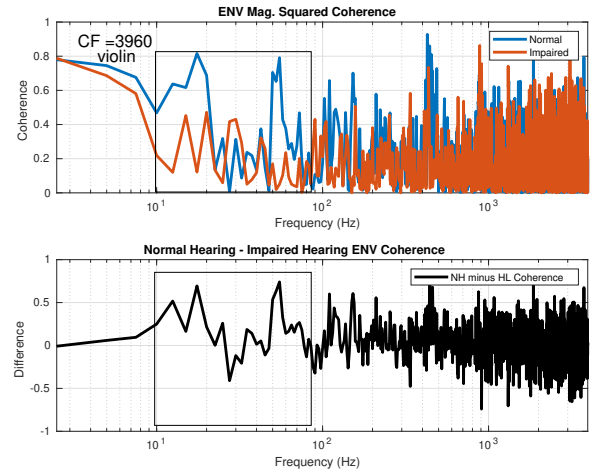

Fig. 7. ENV coherence differences in violin A4 at CF = 3960 Hz. ENV coherence is decreased in the 10-100 Hz region (black box) in impaired neurons compared to normal hearing neurons.

B. Timbral Coding of Articulation

Martelé and spiccato also showed differences in ENV coherence at CF = 3960. Figure 9 demonstrates hearing impairment particularly decreasing ENV coherence of martelé in the 6-30 Hz region. This difference in ENV coherence was not apparent in ENV coherence of spiccato at the same CF, shown in Figure . These finding may indicate that ENV coding is particularly important in conveying timbral information related to martelé, but that this ENV coding is less important or relevant in the coding of spiccato. This may also indicate that hearing impaired listeners should more easily be able to discern "bouncier" articulations like spiccato, as opposed to heavier articulations like martelé.
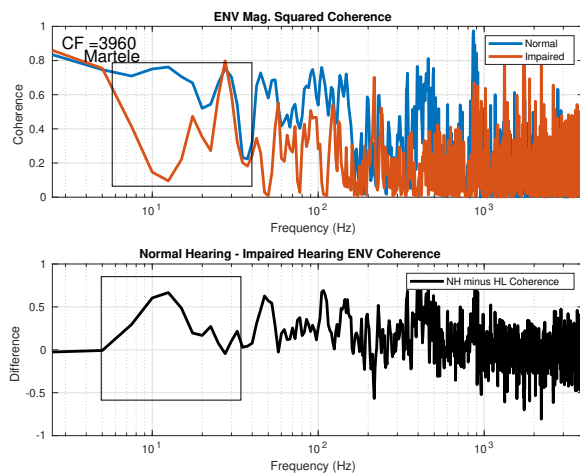
Fig. 8. Hearing impairment markedly decreases ENV coherence of martelé at CF = 3960 Hz.
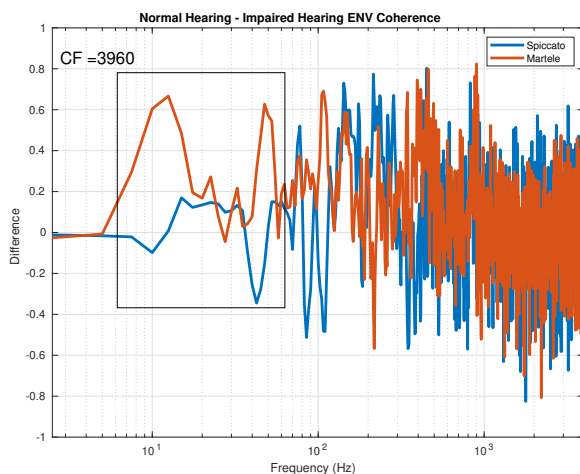

Fig. 9. Hearing impairment differentially impacts martelé and spiccato ENV coherence of frequencies in the 6-30 Hz region.

IV. DISCUSSION

ACKNOWLEDGMENT

REFERENCES

REFERENCES

[1] G. M. Bidelman, A. Krishnan, and J. T. Gandour, "Enhanced brain-stem encoding predicts musicians' perceptual advantages with pitch," *European Journal of Neuroscience*, vol. 33, no. 3, pp. 530–538, 2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-9568.2010.07527.x.

[2] J. Heng, G. Cantarero, M. Elhilali, and C. J. Limb, "Impaired perception of temporal fine structure and musical timbre in cochlear implant users," *Hearing Research*, vol. 280, pp. 192–200, Oct. 2011.

[3] Z. M. Smith, B. Delgutte, and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, pp. 87–90, Mar. 2002.

[4] M. G. Heinz and J. Swaminathan, "Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech," *J Assoc Res Otolaryngol*, vol. 10, pp. 407–423, Sept. 2009.

[5] S. Parida, H. Bharadwaj, and M. G. Heinz, "Spectrally specific temporal analyses of spike-train responses to complex sounds: A unifying framework," *bioRxiv*, p. 2020.07.17.208330, July 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[6] I. C. Bruce, Y. Erfani, and M. S. A. Zilany, "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites," *Hearing Research*, vol. 360, pp. 40–54, Mar. 2018.

[7] H. Lee and D. Müllensiefen, "The Timbre Perception Test (TPT): A new interactive musical assessment tool to measure timbre perception ability," *Atten Percept Psychophys*, vol. 82, no. 7, pp. 3658–3675, 2020.

[8] "Sound samples."

[9] N. Ma, "An Efficient Implementation of Gammatone Filters."