Spectro-Temporal Information Distinguishes Between Speech and Music

NYU POEPPEL [lab]
yl7138@nyu.edu

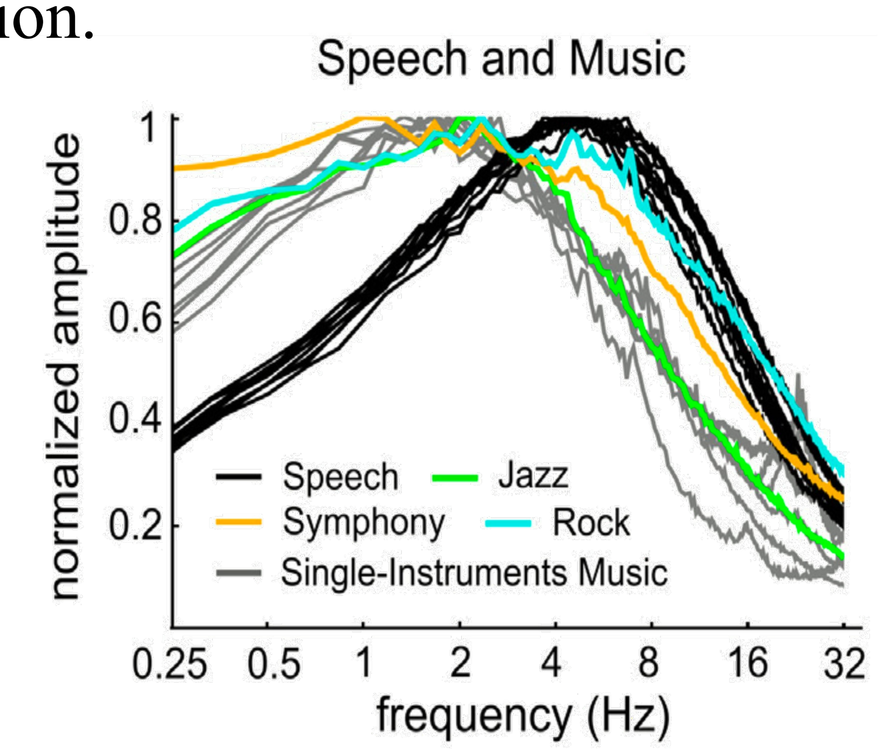
Yike Li[1], Andrew Chang[1], & David Poeppel[1,2,3]

[1] Department of Psychology, New York University, [2] NYU Center for Language, Music, and Emotion (CLaME), [3] Ernst Strüngmann Institute for Neuroscience

Introduction

Speech and music are frequent and typical signals for human audition.

The human brain can distinguish them based on low- to mid-level acoustic properties (e.g., amplitude, frequency).

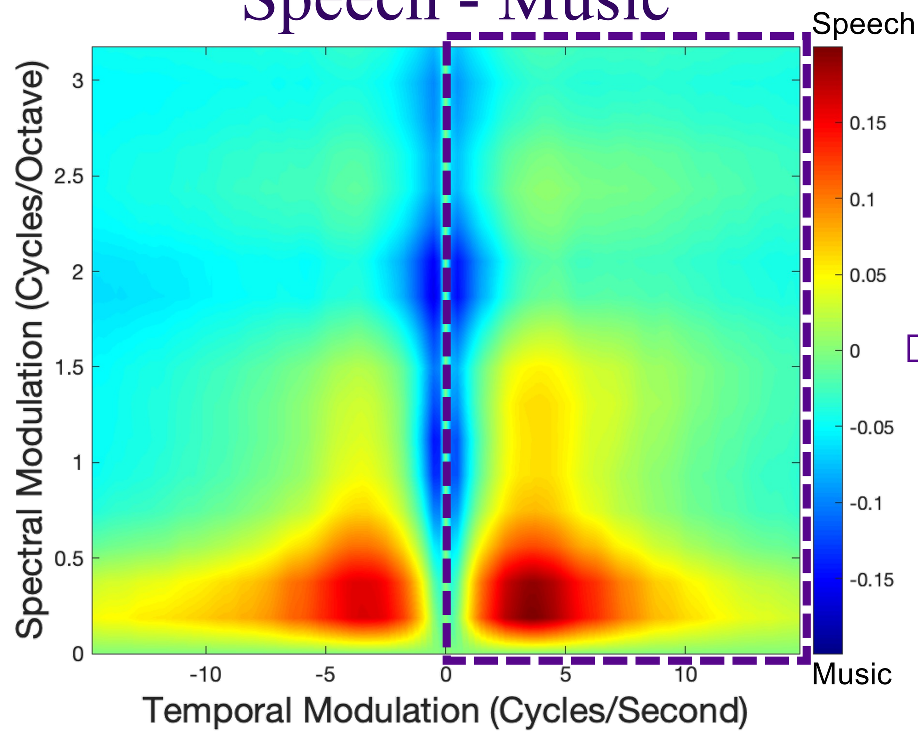


(Ding et al., 2017)

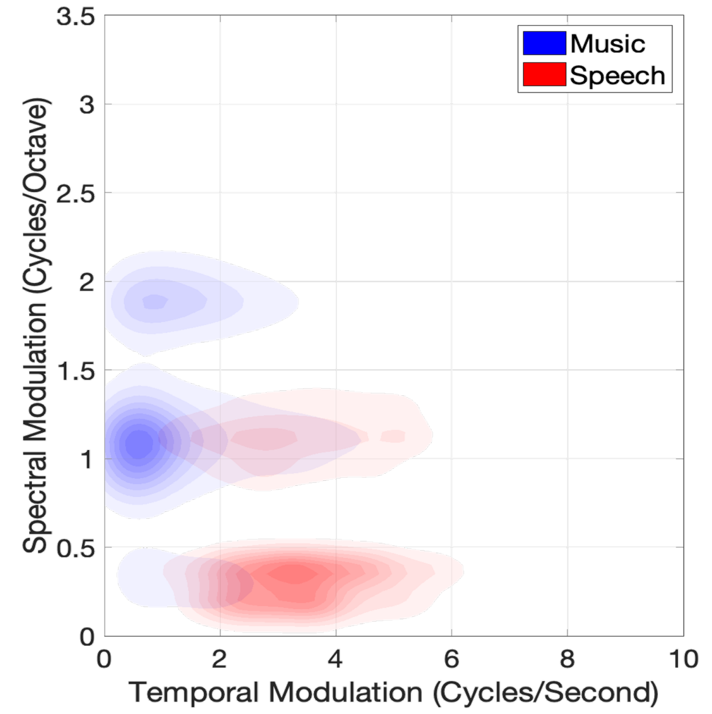
The acoustic differences between speech and music have not been fully quantified in a neuro-physiologically supported way (e.g., spectro-temporal modulation representations).

Results

Spectro-Temporal Modulation: Speech - Music



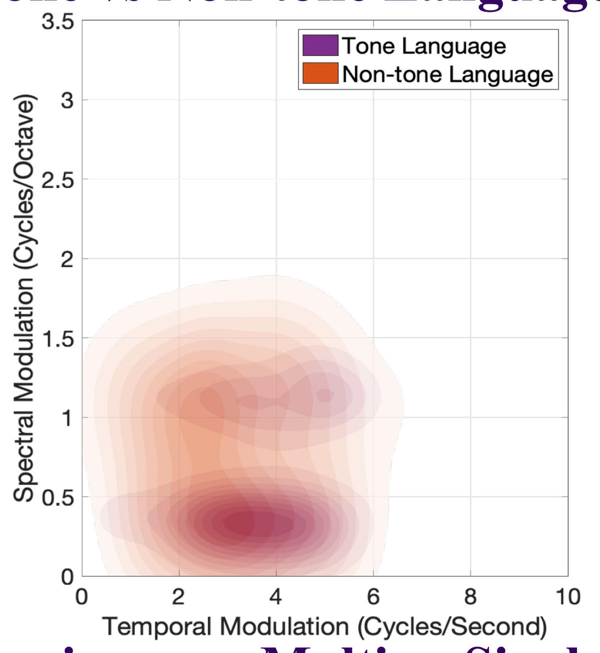
Peak Distribution of Spectro-Temporal Power

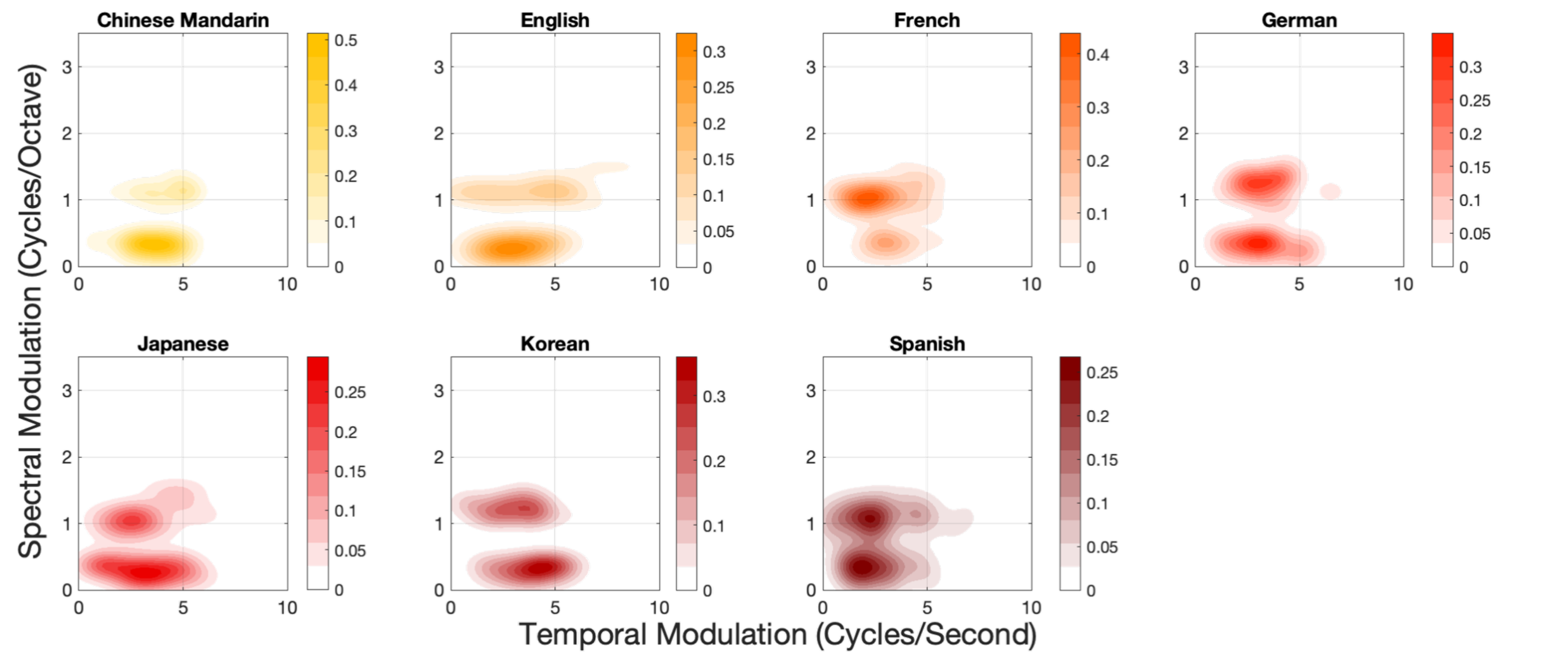


Speech

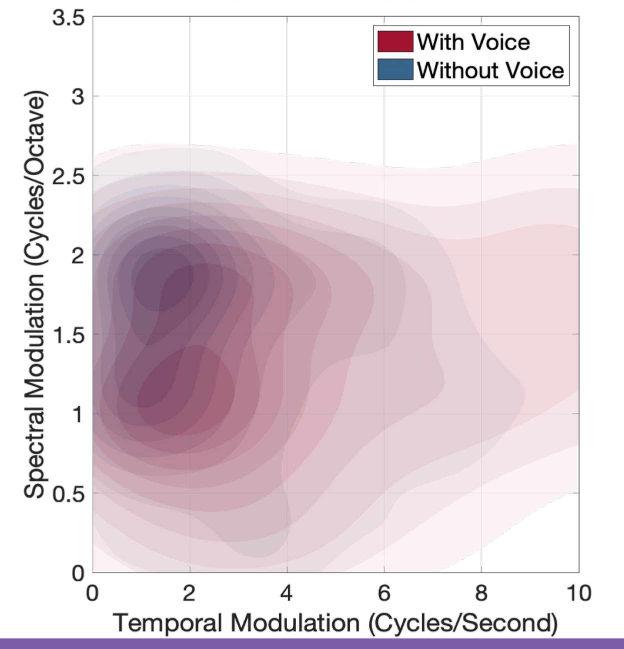
Music

Tone vs Non-tone Languages

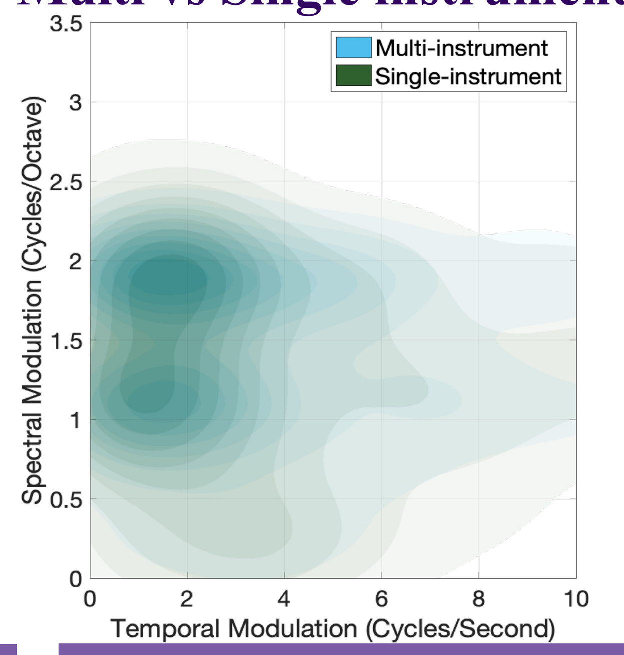




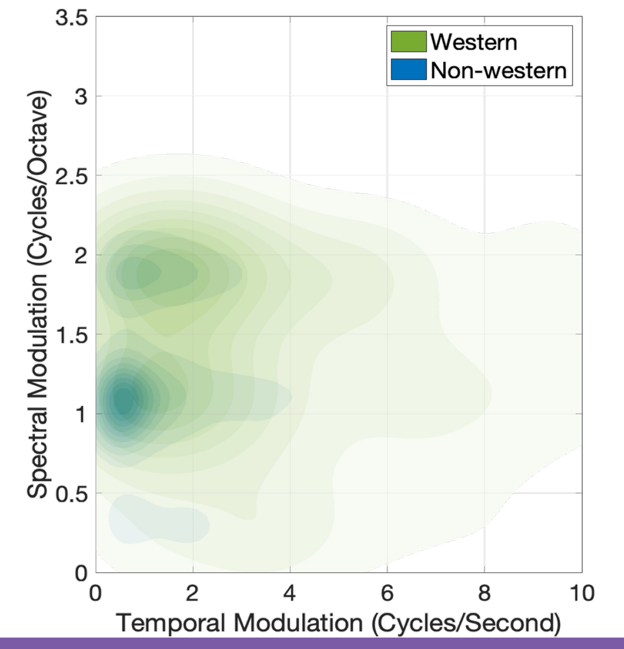
With vs Without voice

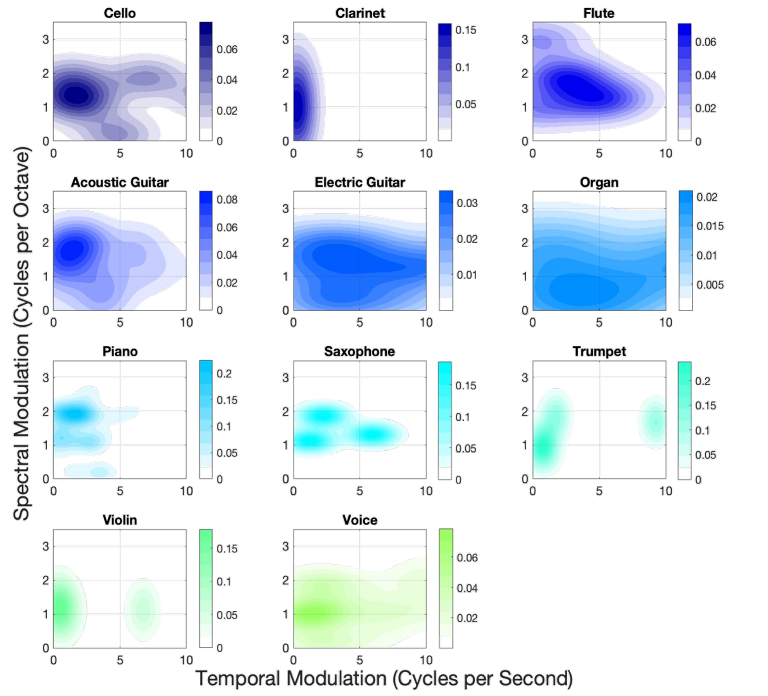


Multi vs Single instrument



Western vs Non-western





Sampling criteria:
4s audio excerpts (without silent gaps > 1s)

Sample size:

- LibriVox: 20 min/language (speaker gender balanced)
- TIMIT: 636 recordings
- Clarity Speech: 555 recordings
- IRMAS: 533 recordings
- Garland: 304 recordings

Methods

Materials

3 speech corpora (LibriVox, TIMIT, The Clarity Speech Corpus)

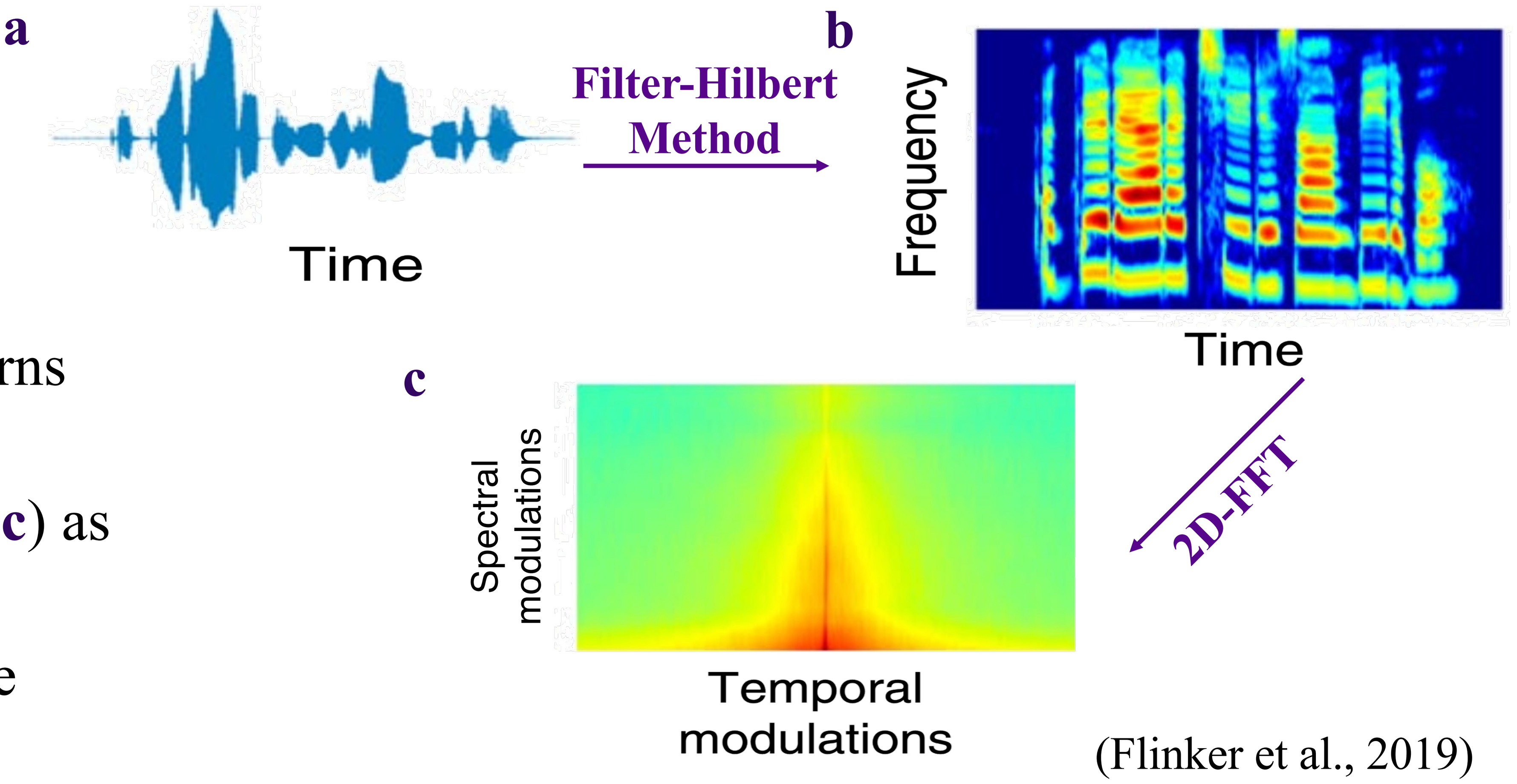
2 music corpora (IRMAS, Garland Encyclopedia of World Music)

Spectro-Temporal Modulation

Sound (waveform, **a**) can be plotted as a spectrogram (**b**) to show how spectral patterns (frequency) change over time.

A spectrogram (**b**) can then be decomposed and depicted in the modulation domain (**c**) as temporal (cycles per second) and spectral (cycles per octave) modulations.

Modulation power peaks are extracted for each recording, and the distributions of the peaks of different groups are then estimated.



(Flinker et al., 2019)

Discussion

- Speech and music show distinct spectro-temporal modulation peaks.
- Within speech or music, the spectro-temporal modulation peaks show high consistency across languages and cultures.
- Consistent with the asymmetrical neurophysiological features of the auditory cortex (Albouy et al., 2020):
  - Speech: finer temporal details (left auditory cortex)
  - Music: finer spectral details (right auditory cortex)

References & Acknowledgement

Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. Science, 367(6481), 1043–1047.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81*, 181–187.

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour*, *3*(4), 393–405.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, *88*(6), 1281–1296.