Lecture 3: Audio Applications

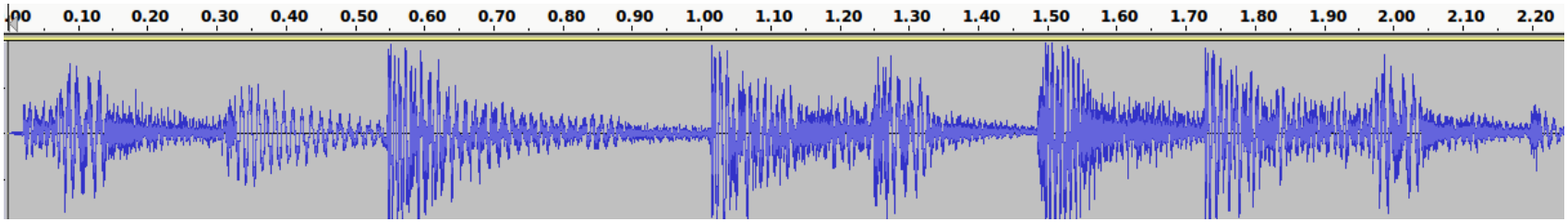
Topological Time Series Analysis - Theory And Practice

Jose Perea, Michigan State University. Chris Tralie, Duke University

7/20/2016

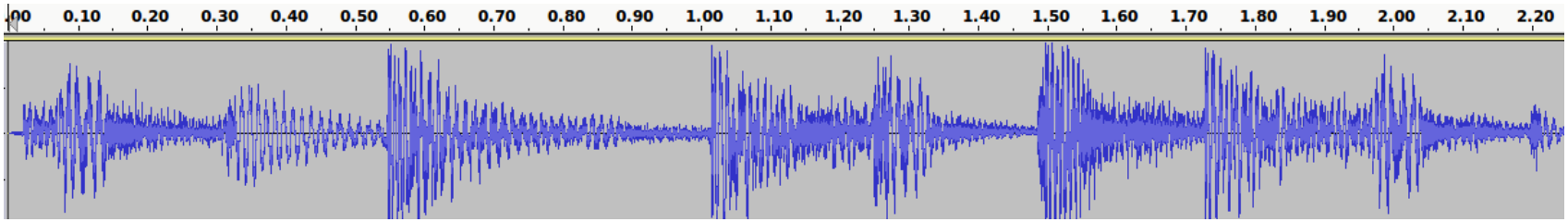
Table of Contents

Digital Audio Basics: Representation/Sampling



- 1D time series $x[n]$, sampled at 44100hz
  - Shannon Nyquist: Need to sample at at least twice the highest frequency of a bandlimited signal to avoid aliasing

Digital Audio Basics: Representation/Sampling

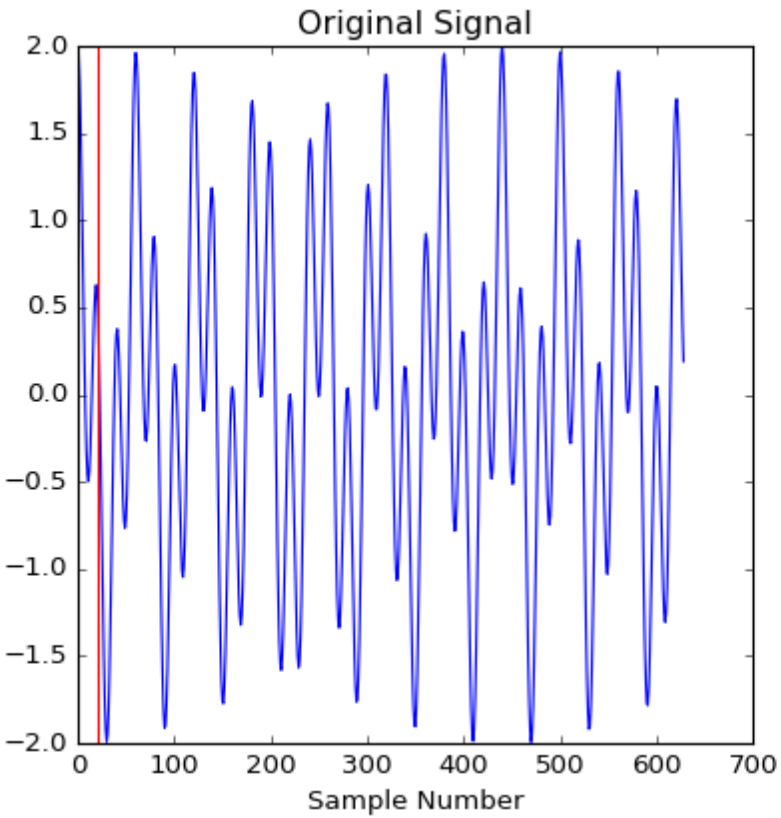


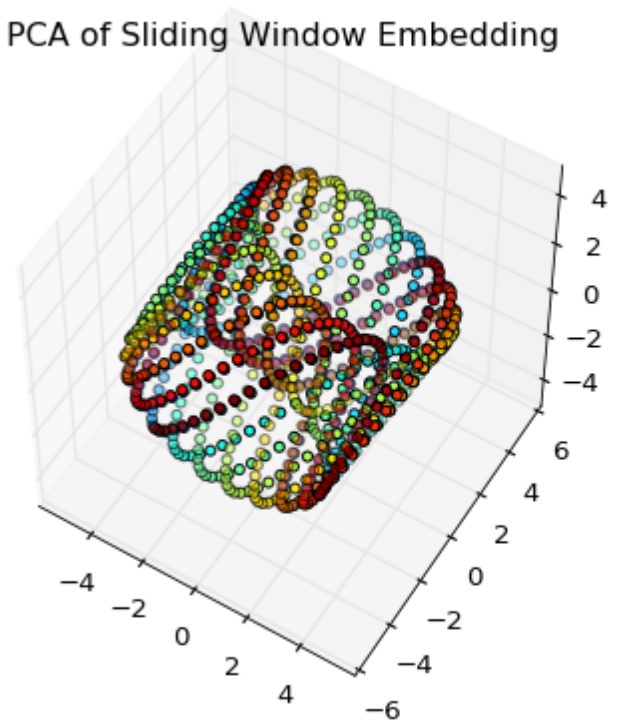
- 1D time series $x[n]$, sampled at 44100hz
  - Shannon Nyquist: Need to sample at at least twice the highest frequency of a bandlimited signal to avoid aliasing
- Very high sampling rate!
  - 1 second chunk lives in $\mathbb{R}^{44100}$
  - 3 second chunk lives in $\mathbb{R}^{132300}$!

Biphonation

- 2 noncommensurate frequencies present at the same time in biological phenomena

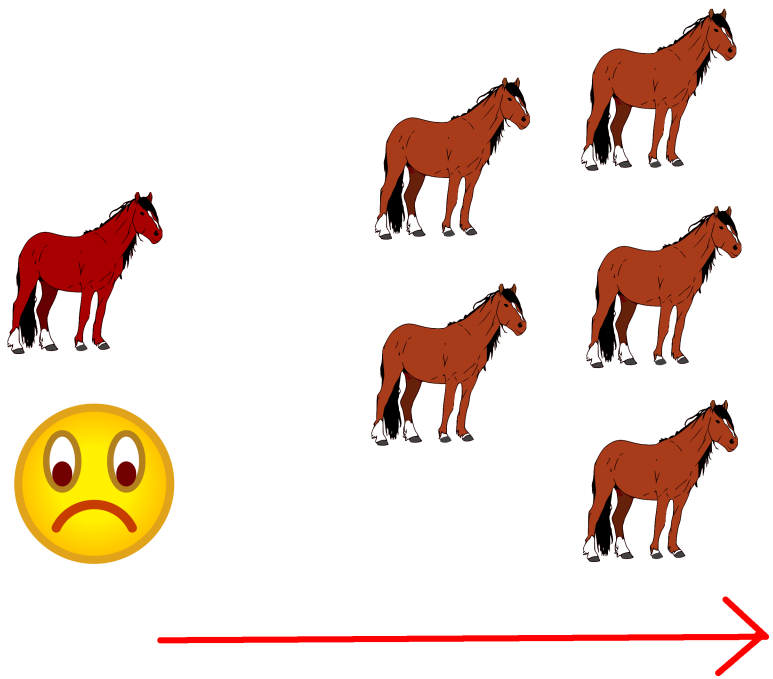
e.g. $\cos(t) + \cos(\pi t)$





Horse Whinnies

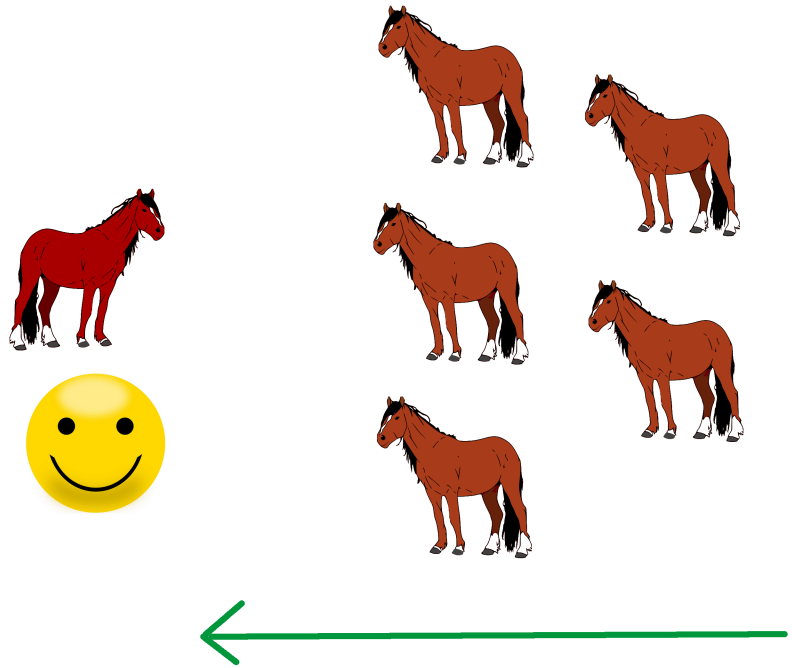
High Valence Negative



Briefer, Elodie F., et al. ”Segregation of information about emotional arousal and valence in horse whinnies.” Scientific reports 4 (2015).

Horse Whinnies

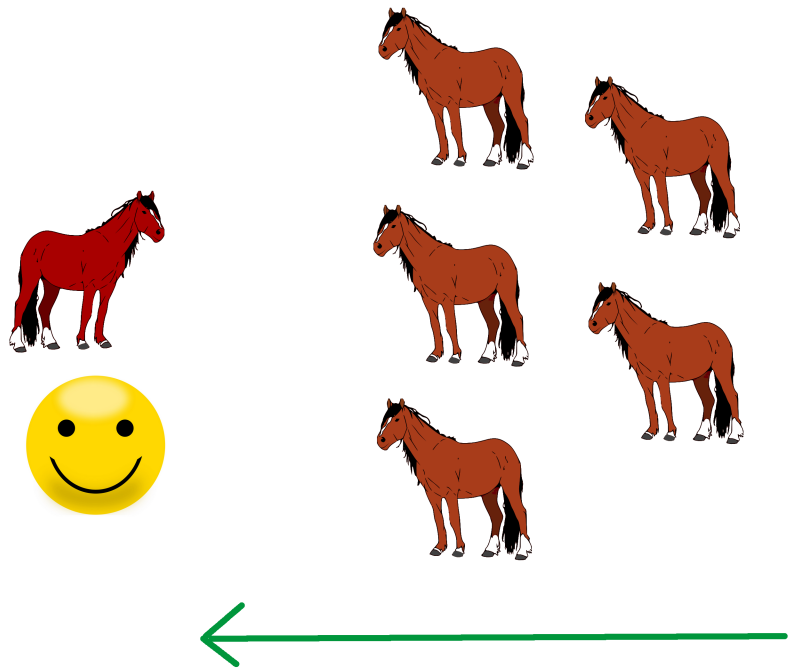
High Valence Positive



Briefer, Elodie F., et al. "Segregation of information about emotional arousal and valence in horse whinnies." Scientific reports 4 (2015).

Horse Whinnies

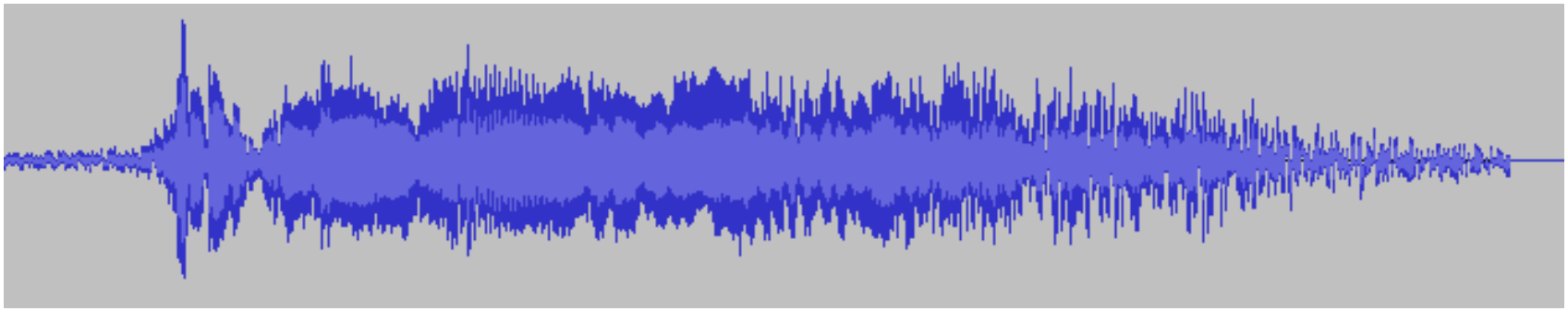
High Valence Positive



- We'll be focusing on the positive clip today...

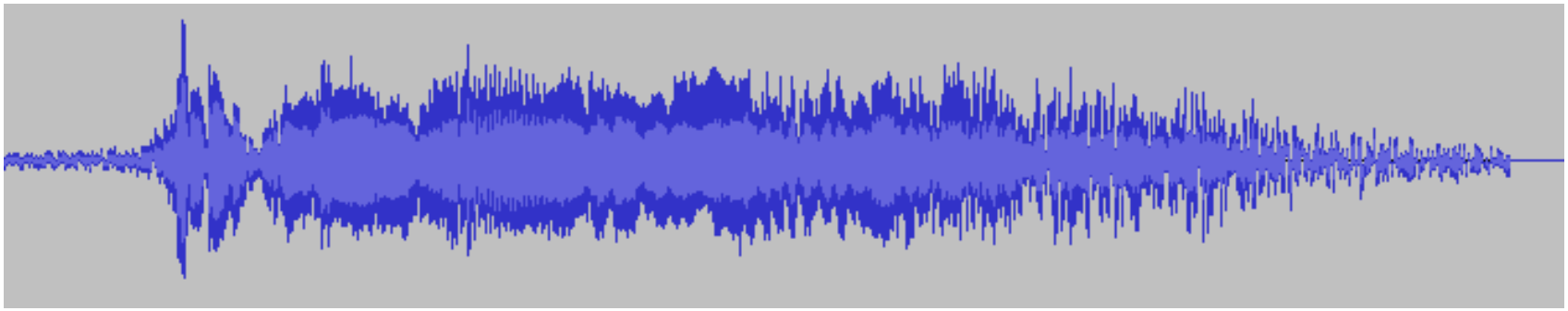
Briefer, Elodie F., et al. "Segregation of information about emotional arousal and valence in horse whinnies." Scientific reports 4 (2015).

Horse Whinnie Audio



Interactively Show Audio File

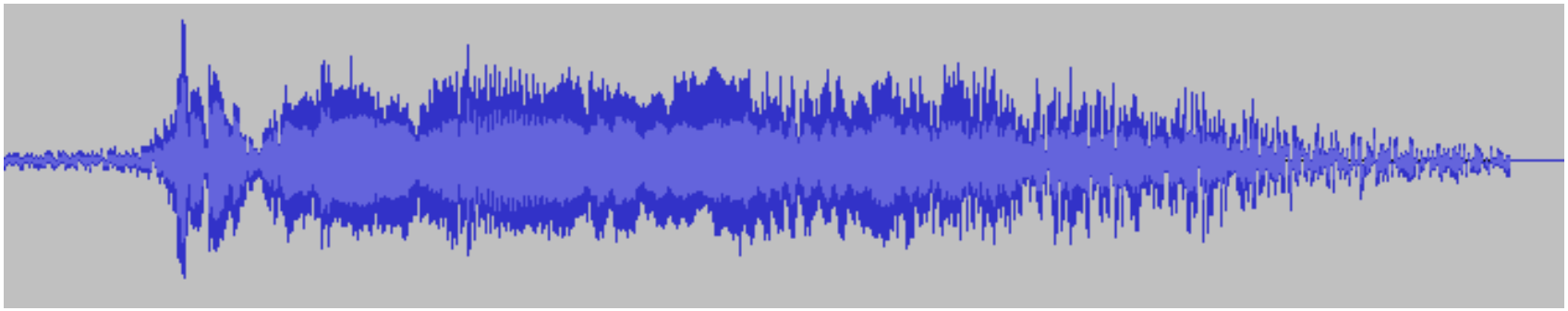
Horse Whinnie Audio



Interactively Show Audio File

- Base frequencies on the order of 1000hz (Window size?)

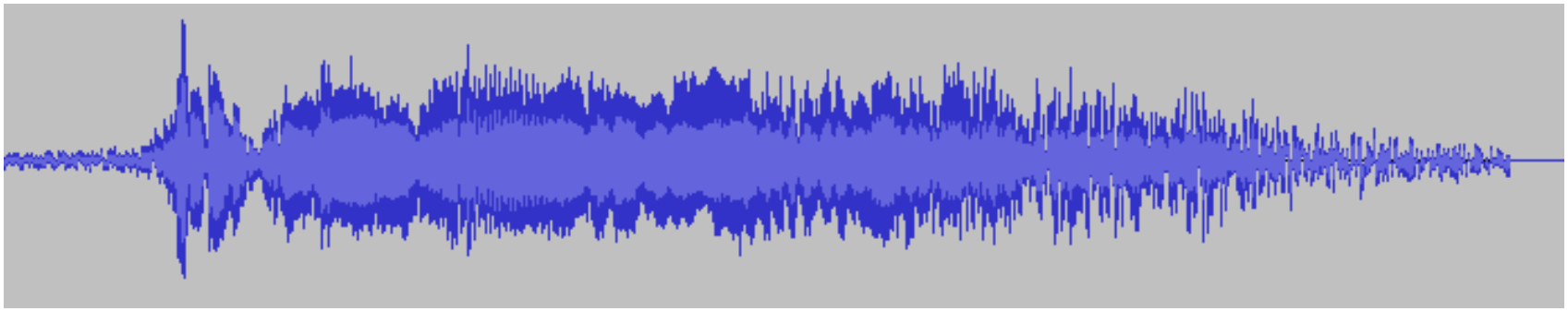
Horse Whinnie Audio



Interactively Show Audio File

- Base frequencies on the order of 1000hz (Window size?)
- By default, only using 512 samples after the starting time (≈23 milliseconds of audio)

Horse Whinnie Audio



Interactively Show Audio File

- Base frequencies on the order of 1000hz (Window size?)
- By default, only using 512 samples after the starting time (≈23 milliseconds of audio)

Have Students Find Steady State Region

Biphonation Finding Competition

- Pan through audio file to find best region of biphonation, as measured by persistence of second most persistent class
- May be corrupted due to noise
- Will keep a running tab of best score on the board!

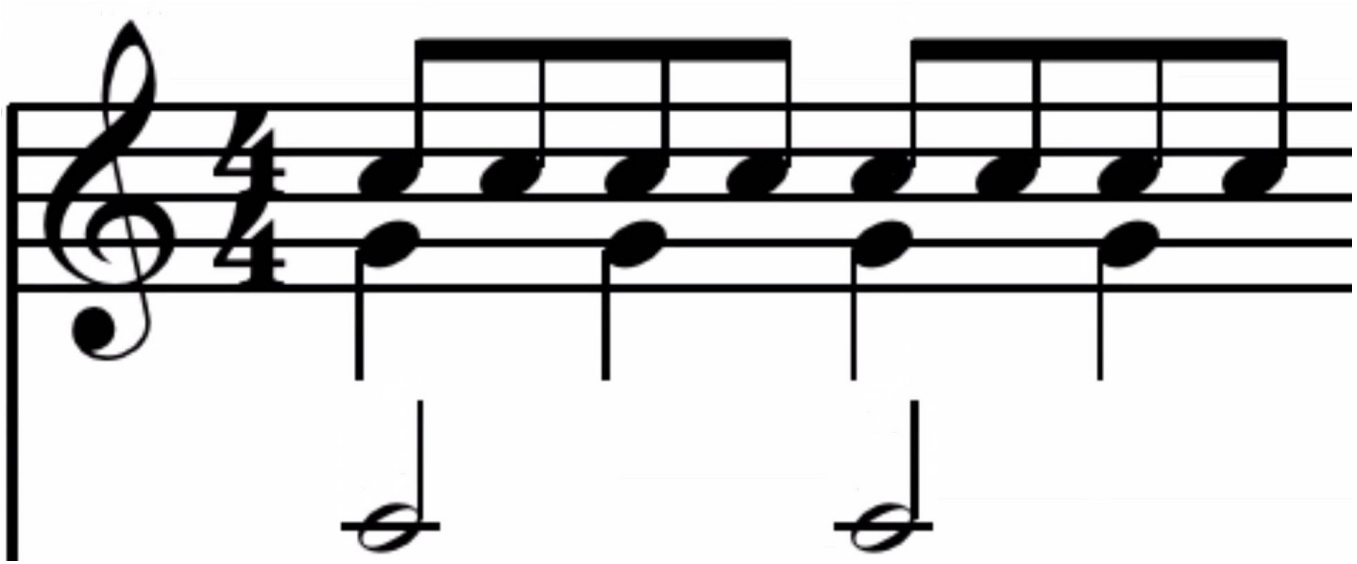
Table of Contents

Tempo / Repetition

- Music is full of repetition

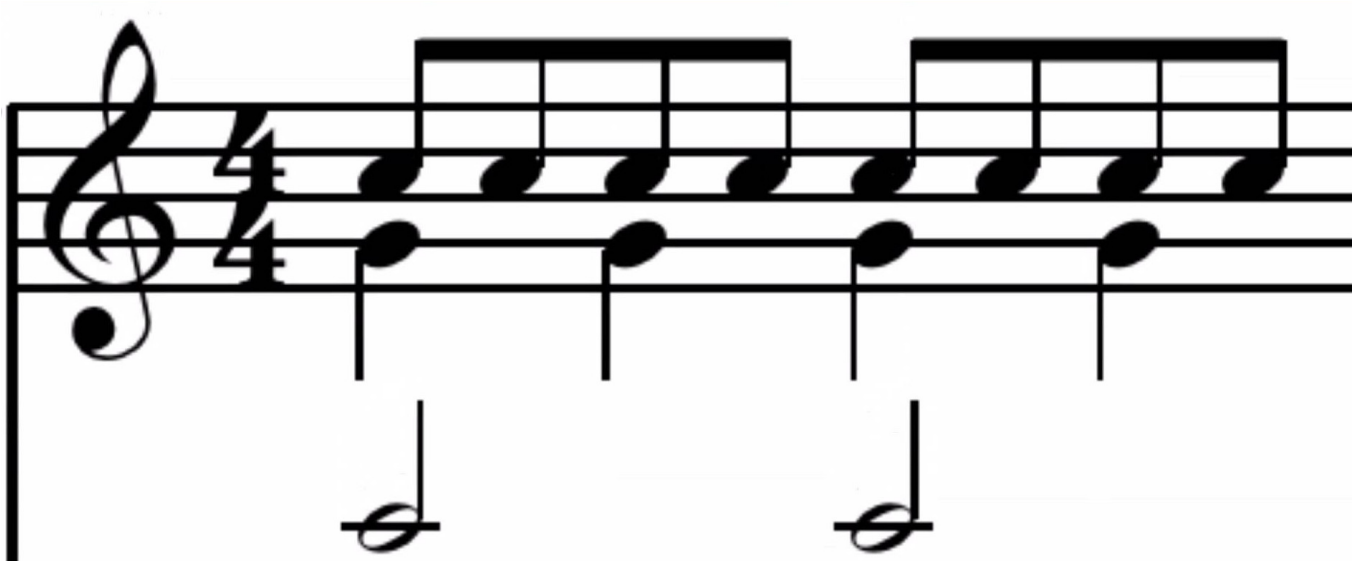
Tempo / Repetition

- Music is full of repetition
- Tempo is determined by a train of music “pulses”/“beats” in a periodic pattern



Tempo / Repetition

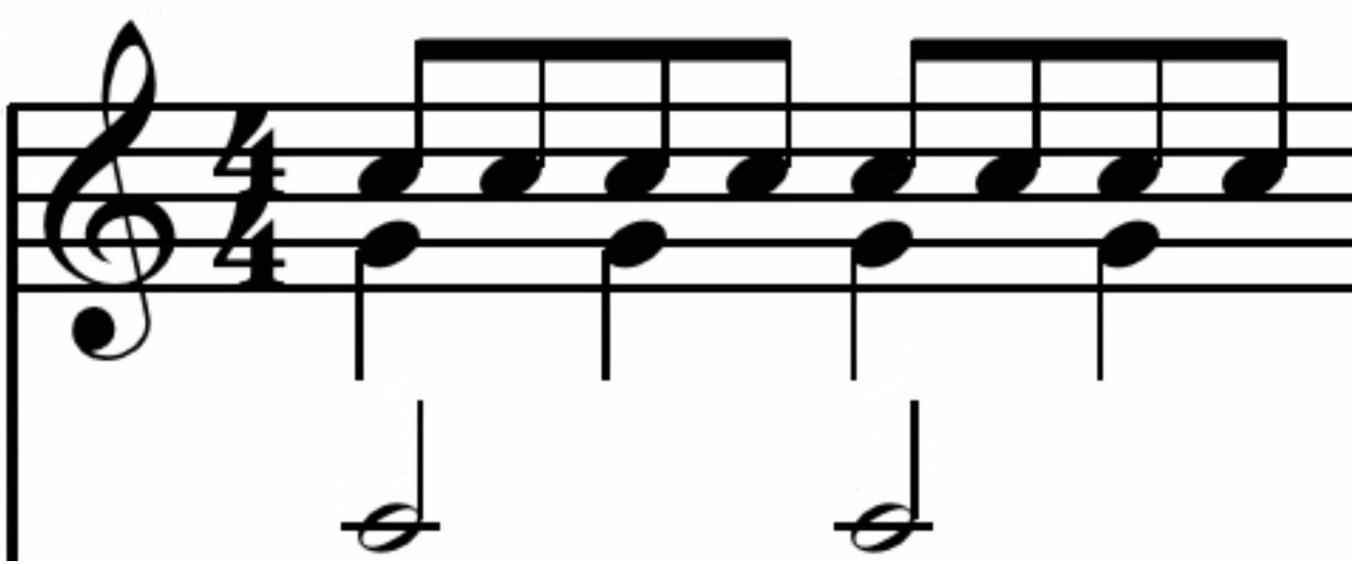
- Music is full of repetition
- Tempo is determined by a train of music “pulses”/“beats” in a periodic pattern



- Foot tapping

Tempo / Repetition

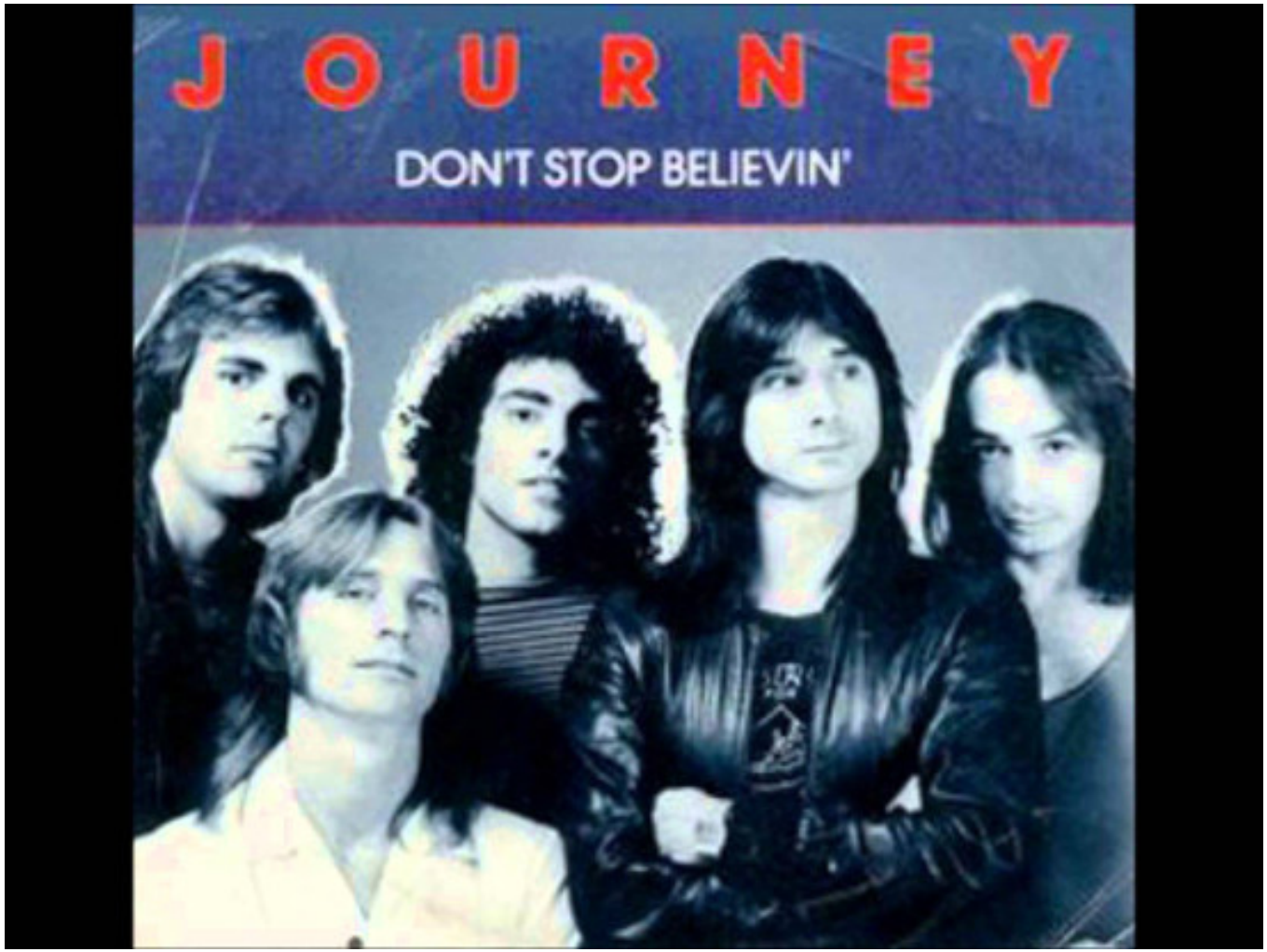
- Music is full of repetition
- Tempo is determined by a train of music “pulses”/“beats” in a periodic pattern



- Foot tapping
- Tempo usually 50 - 200 beats per minute

Tempo / Repetition

Don’t Stop Believin (120 beats per minute)



Raw Audio Delay Embedding

- $\tau * \text{dim} = 22050$ (why?)

Raw Audio Delay Embedding

- $\tau * \text{dim} = 22050$ (why?)
- $dT = 441$

Raw Audio Delay Embedding

- $\tau * \text{dim} = 22050$ (why?)
- $dT = 441$
- Taking first 3 seconds of audio

Raw Audio Delay Embedding

- $\tau * \text{dim} = 22050$ (why?)
- $dT = 441$
- Taking first 3 seconds of audio

Run it! What happens?

Audio Spectrograms: Definition

Aka the Squared Magnitude Short-Time Fourier Transform. Given

- A discrete signal x
- A *window size* W (implicitly $\tau = 1$)
- A *hop size* H (like dT)

Audio Spectrograms: Definition

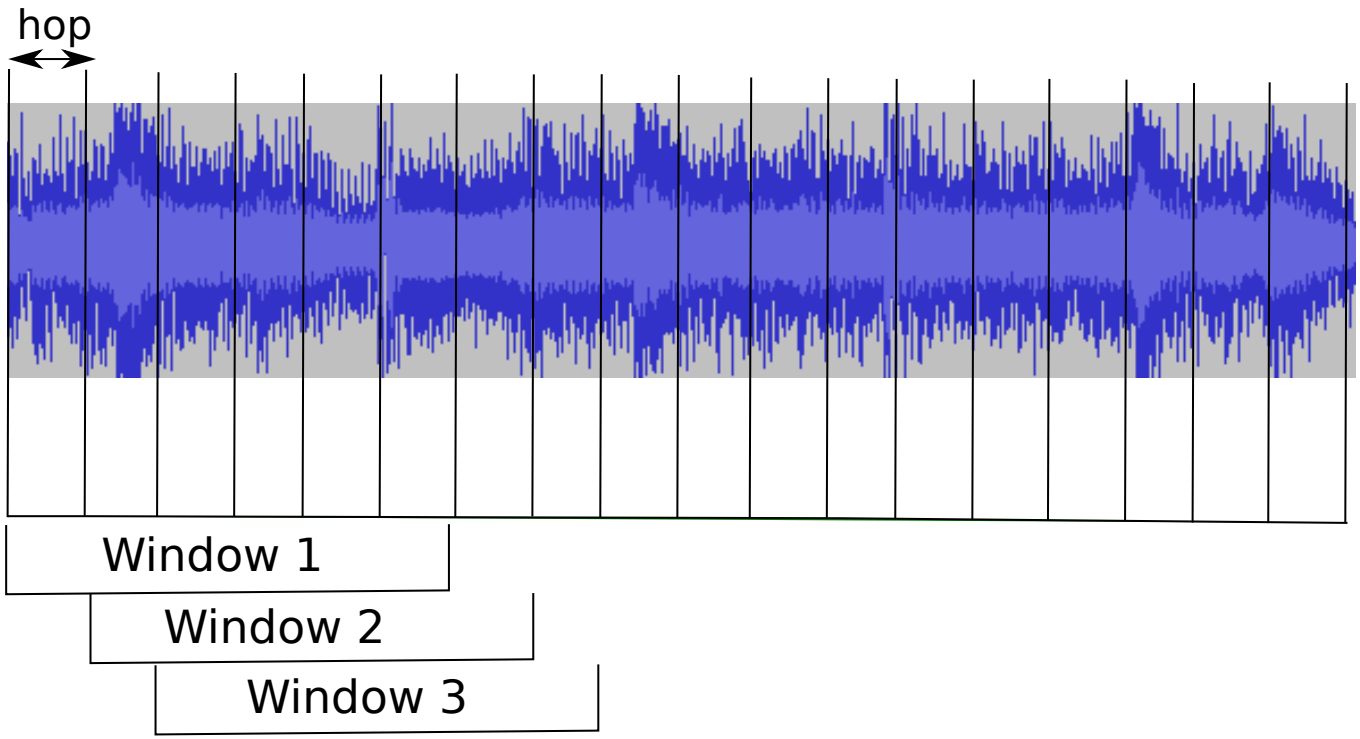
Aka the Squared Magnitude Short-Time Fourier Transform. Given

- A discrete signal x
- A *window size* W (implicitly $\tau = 1$)
- A *hop size* H (like dT)

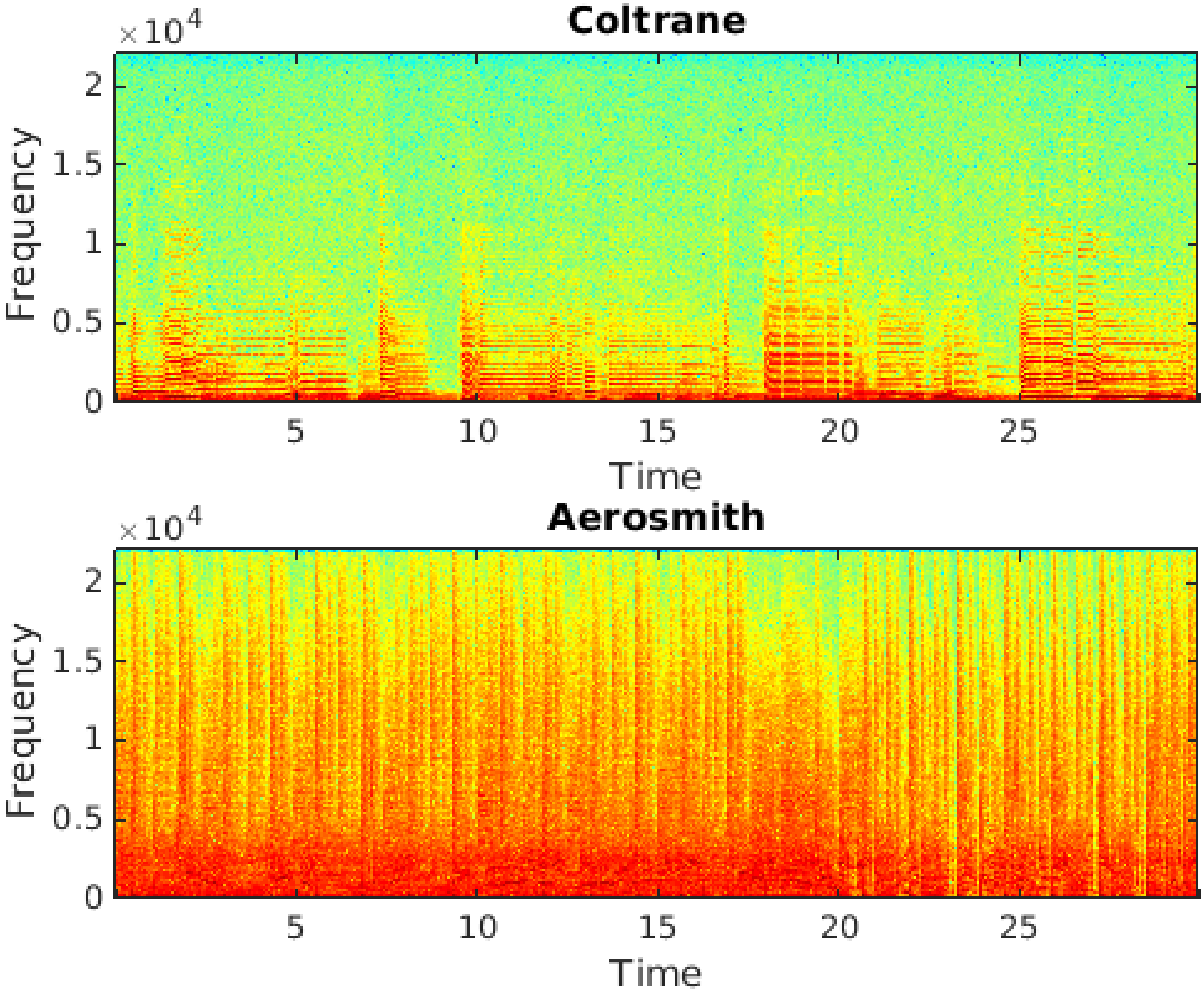
$$S[k, n] = \left| \text{FFT}\left(x\begin{bmatrix} kH \\ kH+1 \\ \vdots \\ kH+W-1 \end{bmatrix} \right)[k] \right|^2$$

Audio Spectrograms: Definition

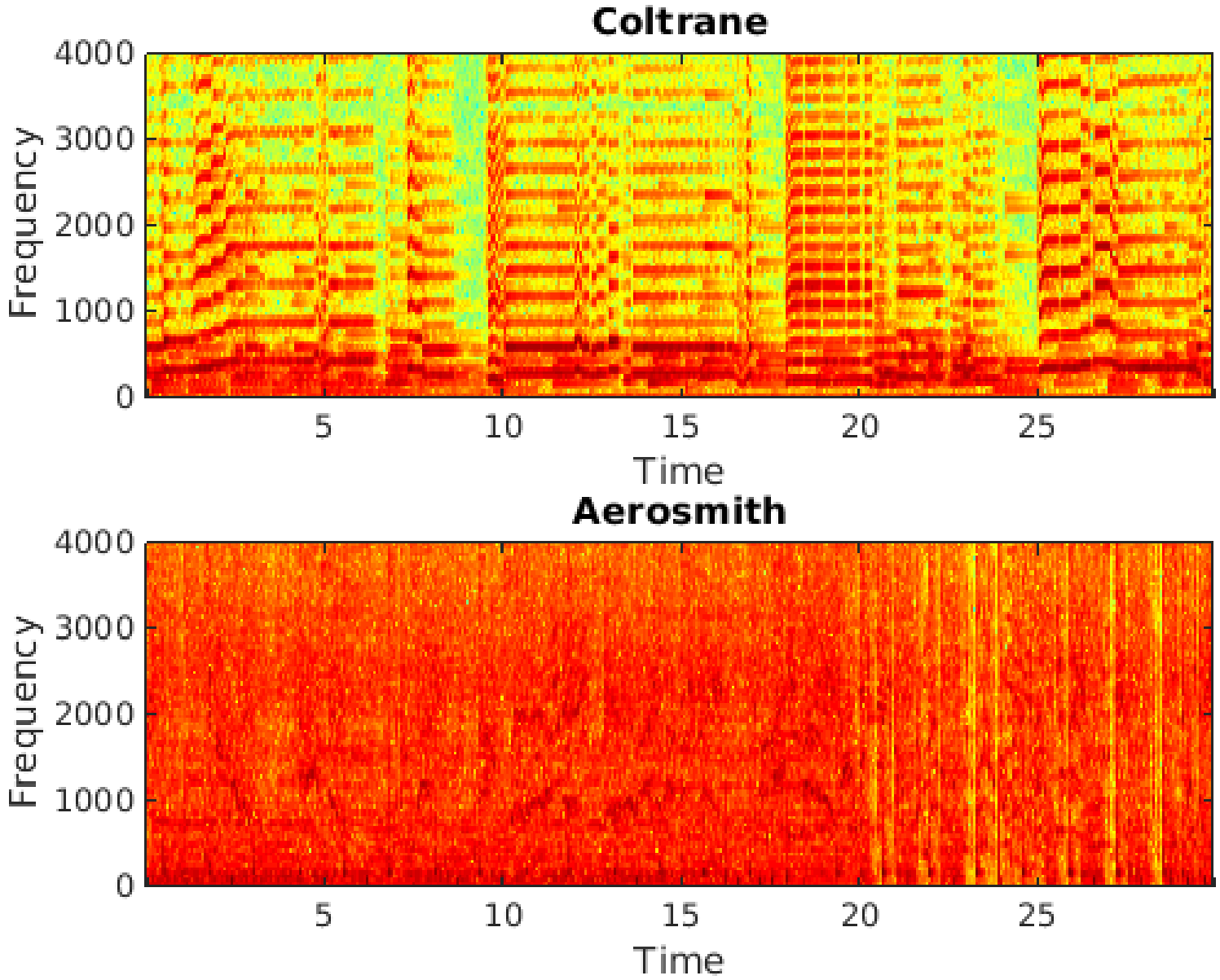
$$S[k,n] = \left| \text{FFT}\left(x \begin{bmatrix} kH \\ kH+1 \\ \vdots \\ kH+W-1 \end{bmatrix} \right)[k] \right|^2$$



Audio Spectrograms



Audio Spectrograms



Audio Spectrograms

- Look at Journey example, show percussion

Audio Novelty Functions

$$f[n] = \sum_{k=0}^{W-1} s(\log(S[k+1,n]) - \log(S[k,n]))$$

where

$$s(x) = \left\{ \begin{array}{cc} x & x > 0 \\ 0 & \text{otherwise} \end{array} \right\}$$

Indicator function for “audio onsets”

Audio Novelty Functions

Show module, show Journey example

Audio Novelty Functions

Show module, show Journey example

- By what factor have we reduced the sampling rate?

Audio Novelty Functions

Show module, show Journey example

- By what factor have we reduced the sampling rate?
- Show synchronized audio

Audio Novelty Functions

Lots of variants

1 Ellis, Daniel PW. ”Beat tracking by dynamic programming.” Journal of New Music Research 36.1 (2007): 51-60.

2 Gouyon, Fabien, Simon Dixon, and Gerhard Widmer. ”Evaluating low-level features for beat classification and tracking.” 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP’07. Vol. 4. IEEE, 2007.

3 Boeck, Sebastian, and Gerhard Widmer. ”Maximum filter vibrato suppression for onset detection.” Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland. 2013.

Audio Novelty Functions

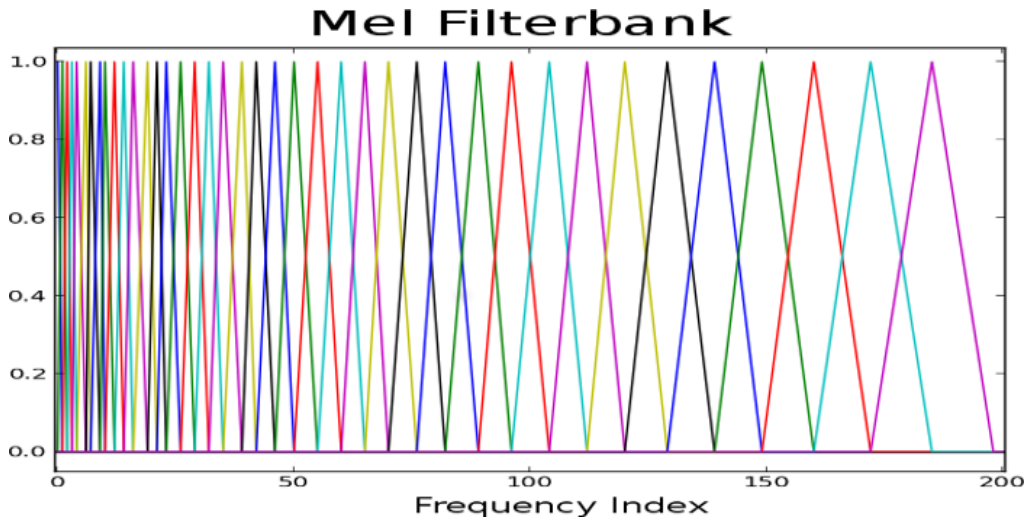
Lots of variants

1 Ellis, Daniel PW. ”Beat tracking by dynamic programming.” Journal of New Music Research 36.1 (2007): 51-60.

2 Gouyon, Fabien, Simon Dixon, and Gerhard Widmer. ”Evaluating low-level features for beat classification and tracking.” 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP’07. Vol. 4. IEEE, 2007.

3 Boeck, Sebastian, and Gerhard Widmer. ”Maximum filter vibrato suppression for onset detection.” Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland. 2013.

e.g. in [1]



Music Vs Speech

Show module

Music Vs Speech

Show module

- A sliding window of sliding windows!

Conclusions

- Quasiperiodicity (biphonation) is present in nature

Conclusions

- Quasiperiodicity (biphonation) is present in nature
- Due to noise/artifacts, sometimes necessary to search around

Conclusions

- Quasiperiodicity (biphonation) is present in nature
- Due to noise/artifacts, sometimes necessary to search around
- Summary features often better than raw data

Conclusions

- Quasiperiodicity (biphonation) is present in nature
- Due to noise/artifacts, sometimes necessary to search around
- Summary features often better than raw data
- After proper preprocessing, TDA on sliding window embeddings can pick up on rhythmic periodicities in music