SPECTRAL FEATURES

YU / HUGHES

OCTOBER 12, 2021

UMASS AMHERST

REVIEW QUESTION

- Take a look at this video clip: https://www.youtube.com/watch?v=sfT_SuDRLNE

A. Vowel quality is roughly constant across the video clip. What does this tell us about changes in the shape of the vocal tract during the video clip?

B. What is the main acoustic property changing? Is this property a property of the voice source or the vocal tract?

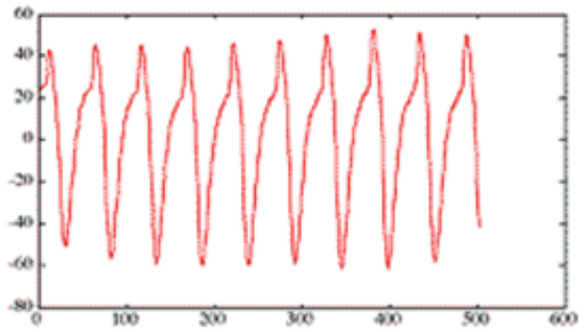
C. Around 0:31-0:33 in the clip, the singer hits some of the most famous high notes in opera. What happens to the spacing of the harmonics in the spectrum when the singer/parrot is singing at higher and higher f0?

D. Suppose that the singer is doing a warm-up exercise where she is singing different vowels at the highest note in this aria. What happens to the spacing between the harmonics? What happens to the position of peaks in the spectrum?

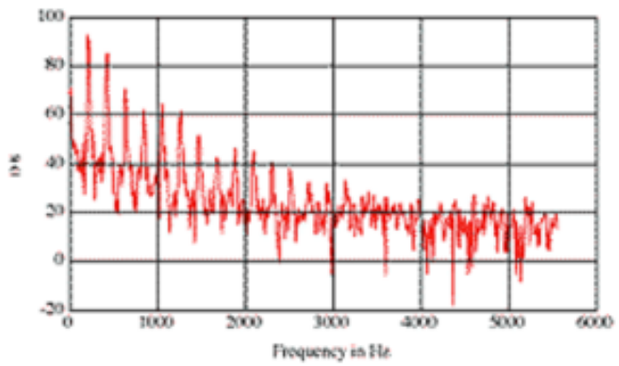
- How do these videos demonstrate independent control of f0 and vowel quality? https://youtu.be/3MH54ewvcWo, https://youtu.be/mYbFJJnJ9Q4

ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

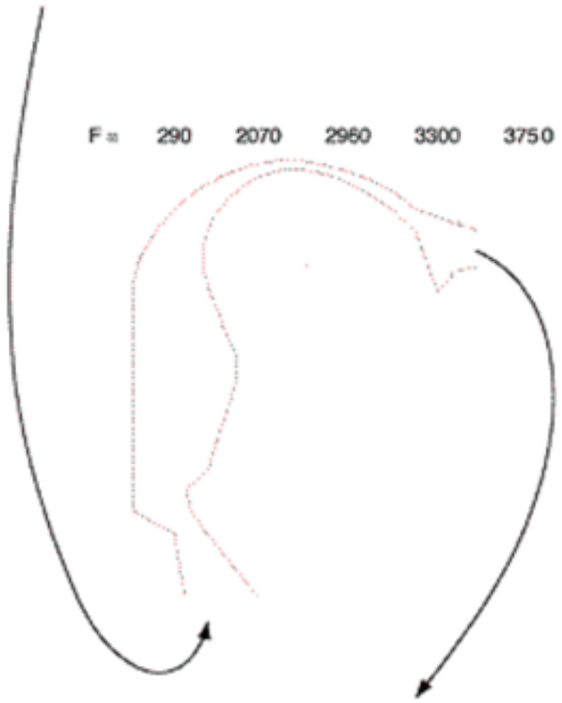
Glottal waveform



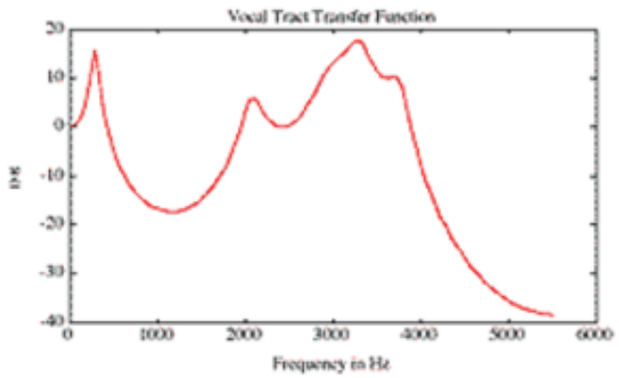
Glottal spectrum



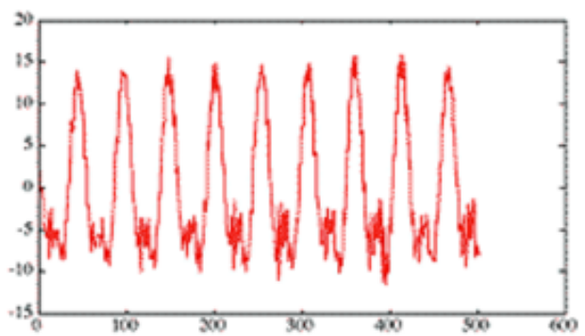
Vocal tract shape for [i]



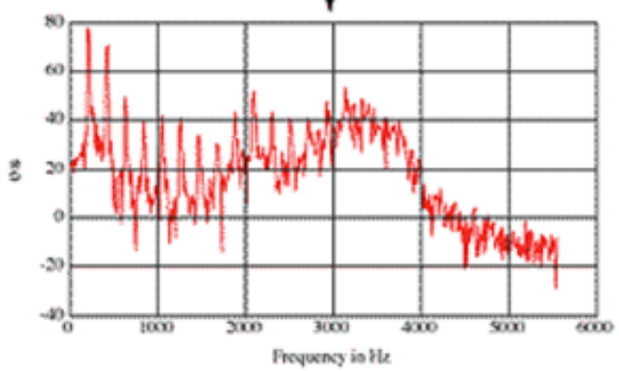
Vocal tract transfer function



Filtered waveform: [i]



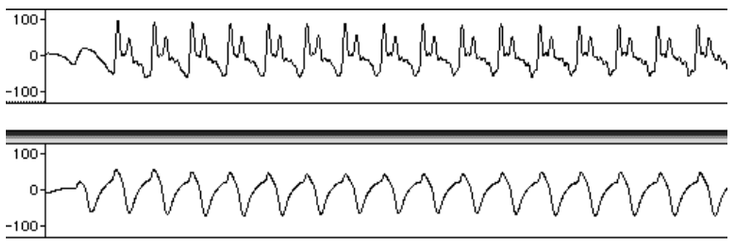
Filtered spectrum: [i]



http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter_Demo/index.html

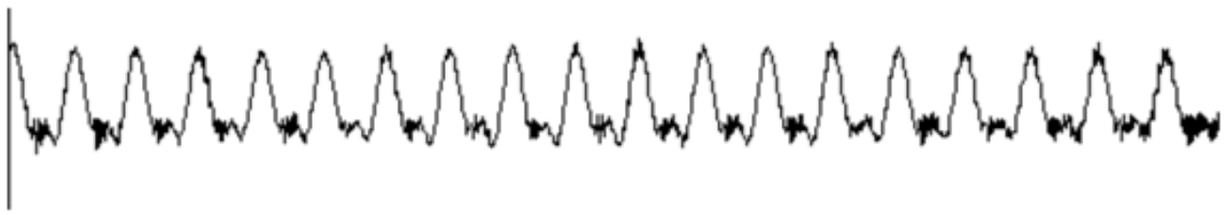
ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Mic

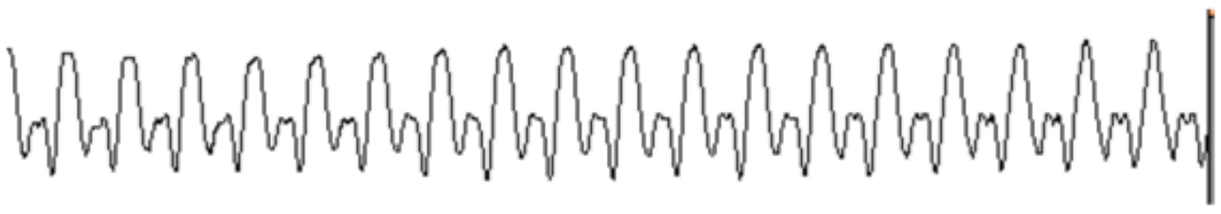


EGG source

[a]-filtered



[i]-filtered

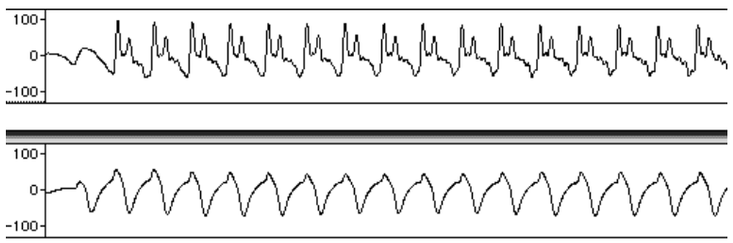


[u]-filtered

http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter_Demo/index.html

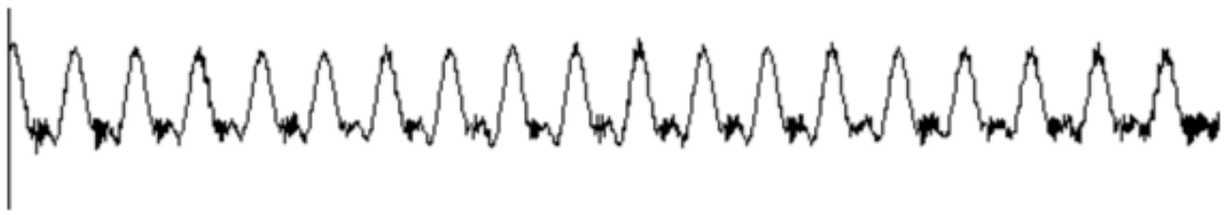
ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Mic

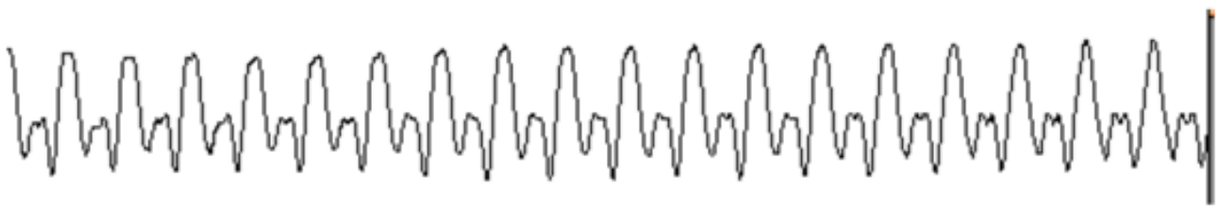


EGG source

[a]-filtered



[i]-filtered

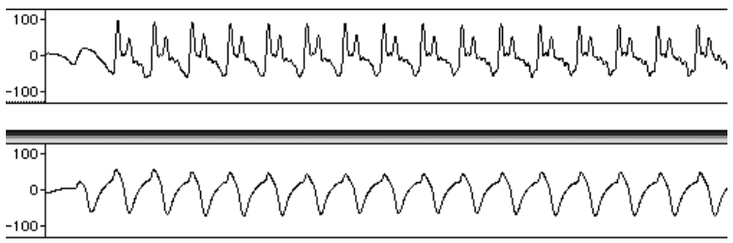


[u]-filtered

http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter_Demo/index.html

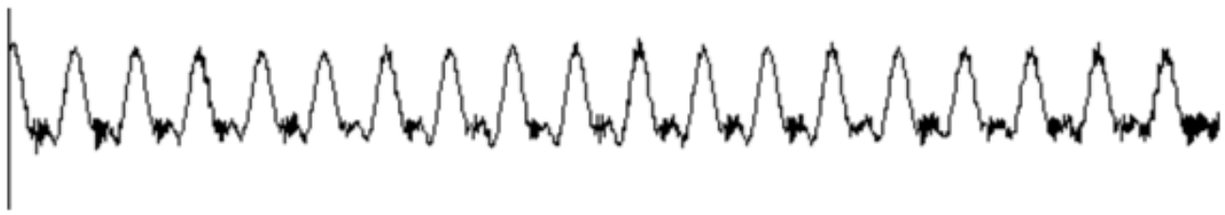
ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Mic

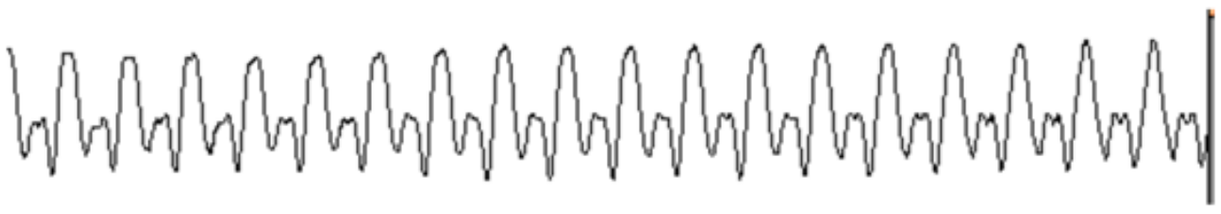


EGG source

[a]-filtered



[i]-filtered

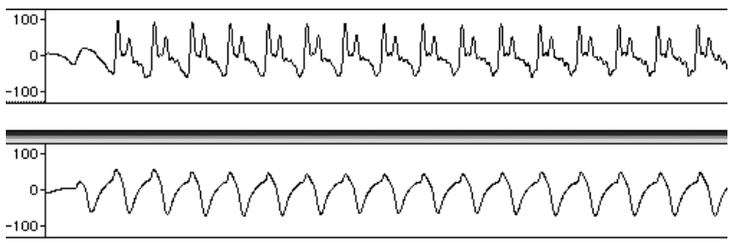


[u]-filtered

http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter_Demo/index.html

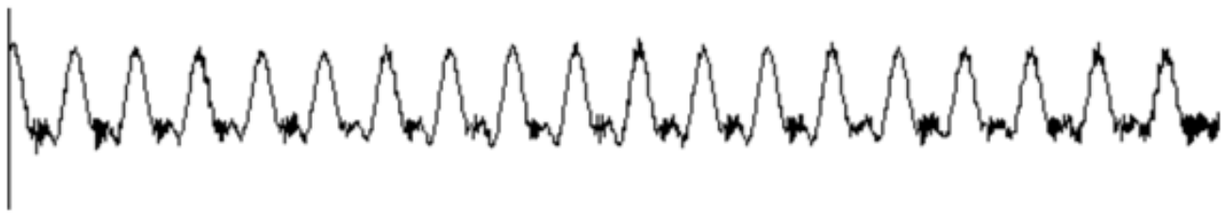
ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Mic

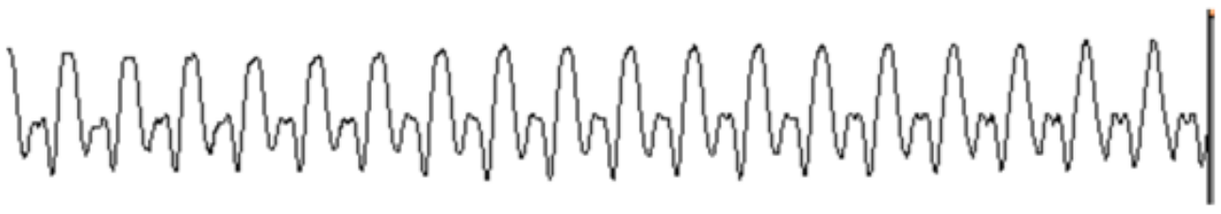


EGG source

[a]-filtered



[i]-filtered

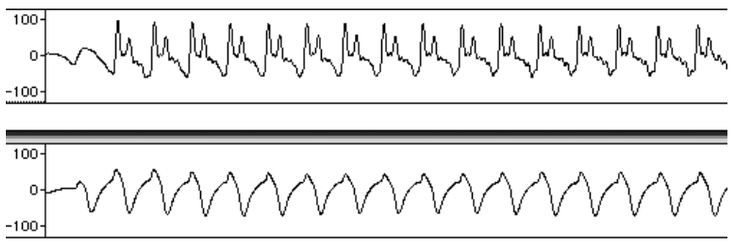


[u]-filtered

http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter_Demo/index.html

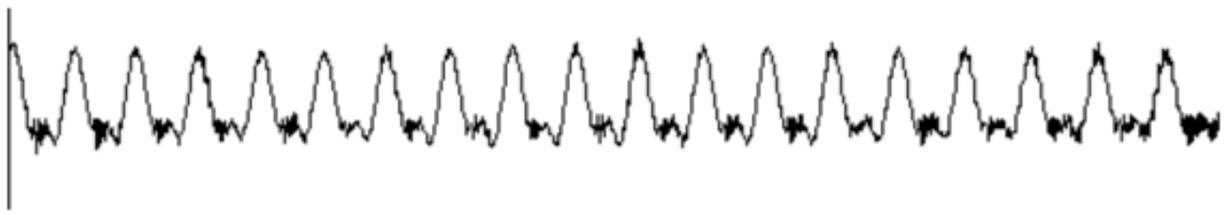
ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Mic

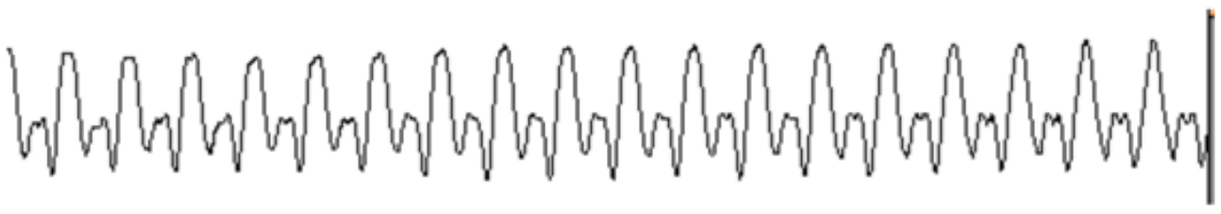


EGG source

[a]-filtered



[i]-filtered

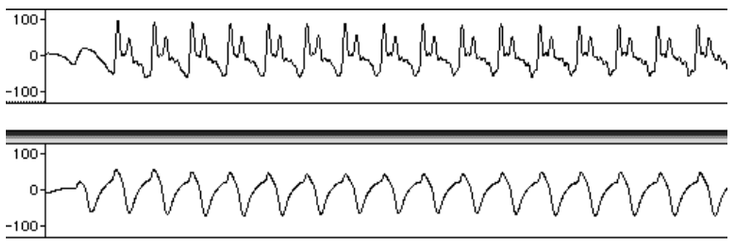


[u]-filtered

http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter_Demo/index.html

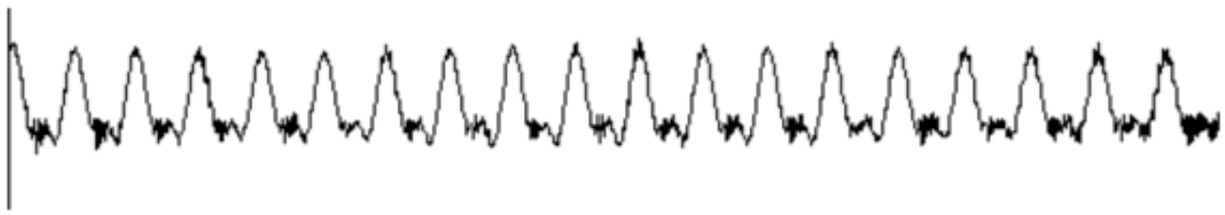
ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Mic

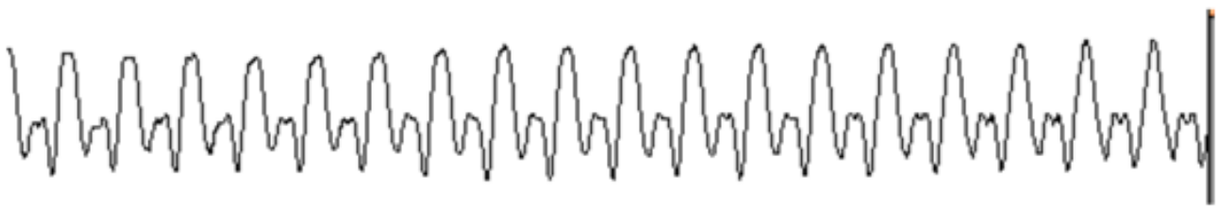


EGG source

[a]-filtered



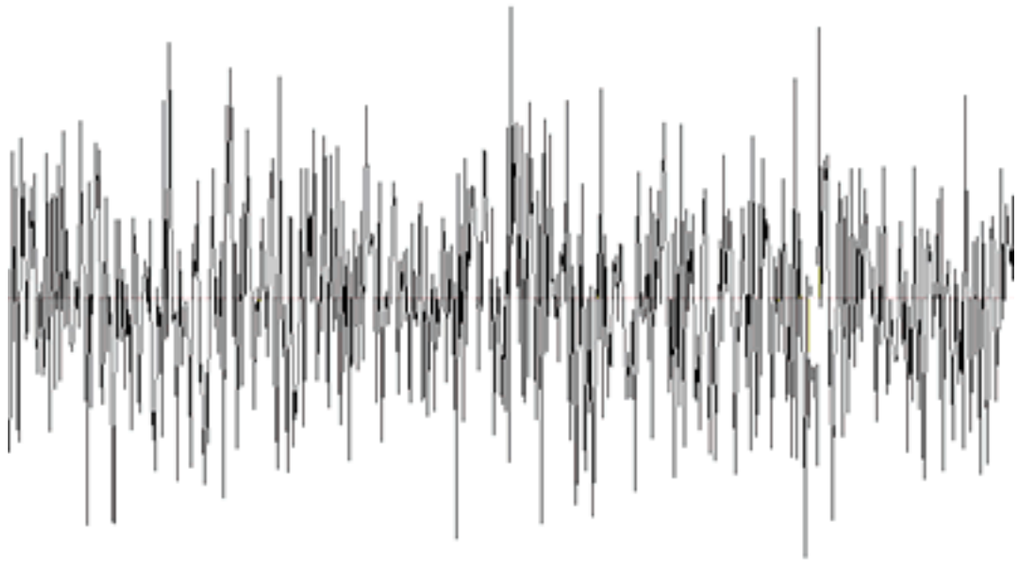
[i]-filtered



[u]-filtered

http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter_Demo/index.html

FINAL PROJECT CHECK-IN



THE CEPSTRUM

RECALL WAYS TO ESTIMATE FOQUESTION

- Download this WAV file.

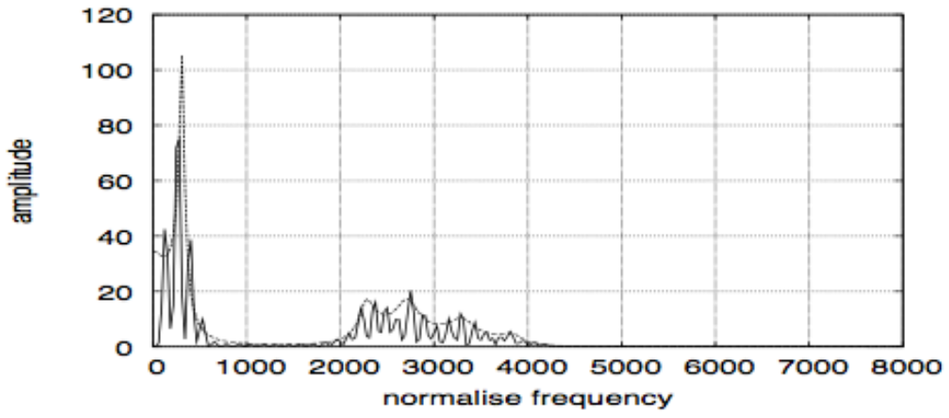
https://drive.google.com/open?id=0BxrcL0TA8FhYY2V3SWdlZjA4UVE

- Estimate the f0 in the two [ma] syllables in three ways in your team. Make TextGrids to segment out each [ma] syllable.
- Measurement methods
  - Off of the waveform: based on cycle-to-cycle distance
  - **Off of the spectrum: spacing between harmonics**
  - Off of the narrow-band spectrogram

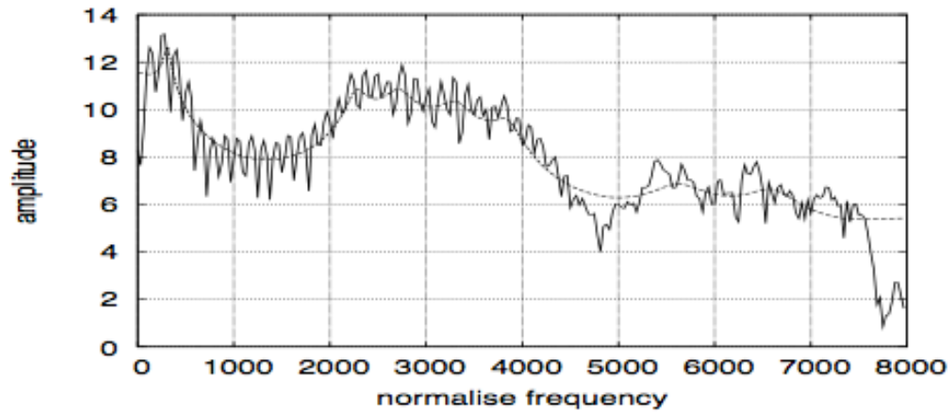
DEFINITION OF THE CEPSTRUM

The cepstrum of a signal

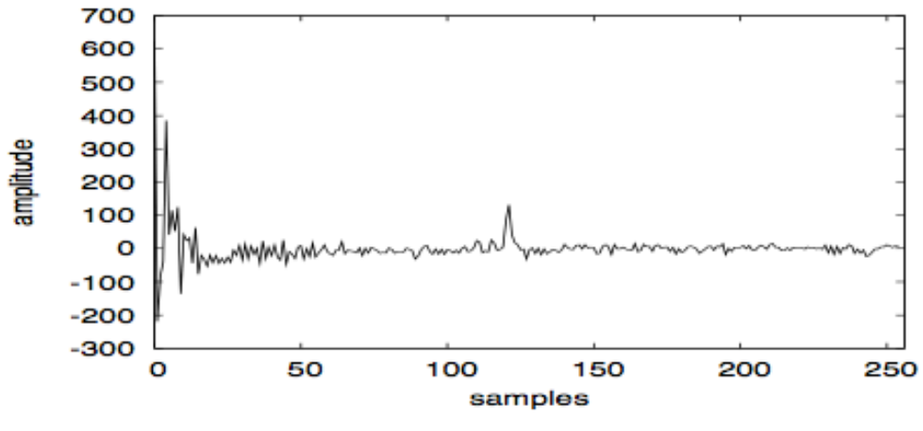
$$C(s) = Re(\mathcal{F}^{-1}(\log(|\mathcal{F}(s)|)) \quad (1)$$



(a) Magnitude spectrum and envelope



(b) log Magnitude spectrum and envelope



(c) Cepstrum

WHY THE LOG SPECTRUM?

The cepstrum of a signal

$$C(s) = Re(\mathcal{F}^{-1}(\log(|\mathcal{F}(s)|)) \quad (1)$$

- Exercise: take the spectrum of the hmong_mu.wav file in our class repository today in Python, without log transformation. Then take it with the log transformation. What's the difference?

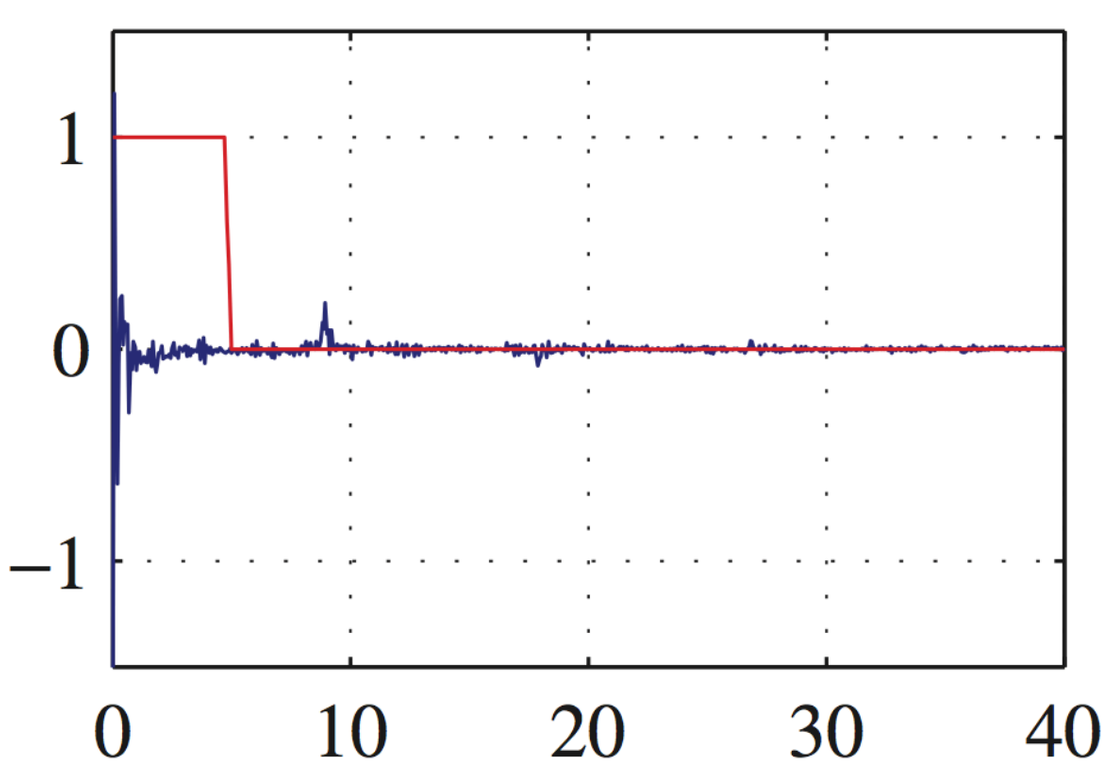
KEY IDEA FOR CEPSTRUM

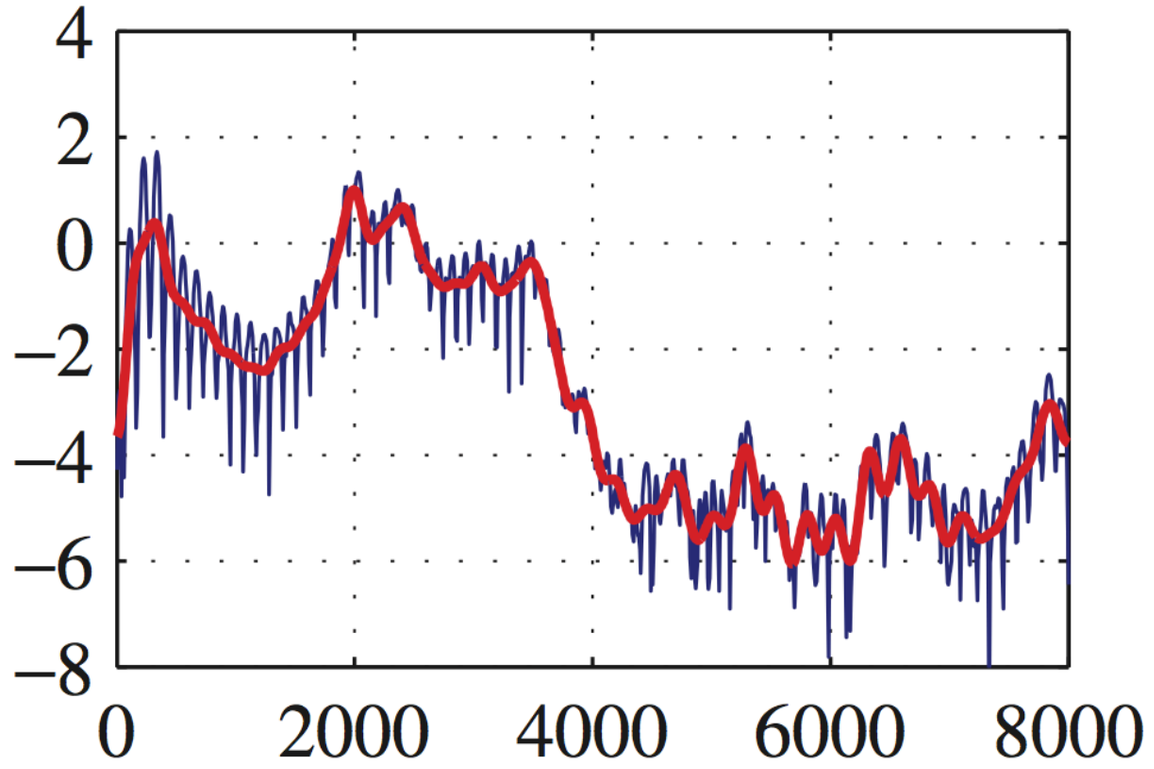
Imagine that the log spectrum were a waveform. Take the Fourier transform!

- If the log spectrum were a waveform:
  - We'd say it's (quasi)-periodic, with amplitude envelope modulation
- Use DFT/FFT to separate the two frequency components
- Expect two kinds of peaks
  - Peak for period of signal (spacing between harmonics)
  - Low-frequency peaks corresponding to amplitude modulation

LIFTERING IN THE QUEFRENCY DOMAIN

What kind of "lifter" is this? What does it do? What other kind of "liftering" might we do?





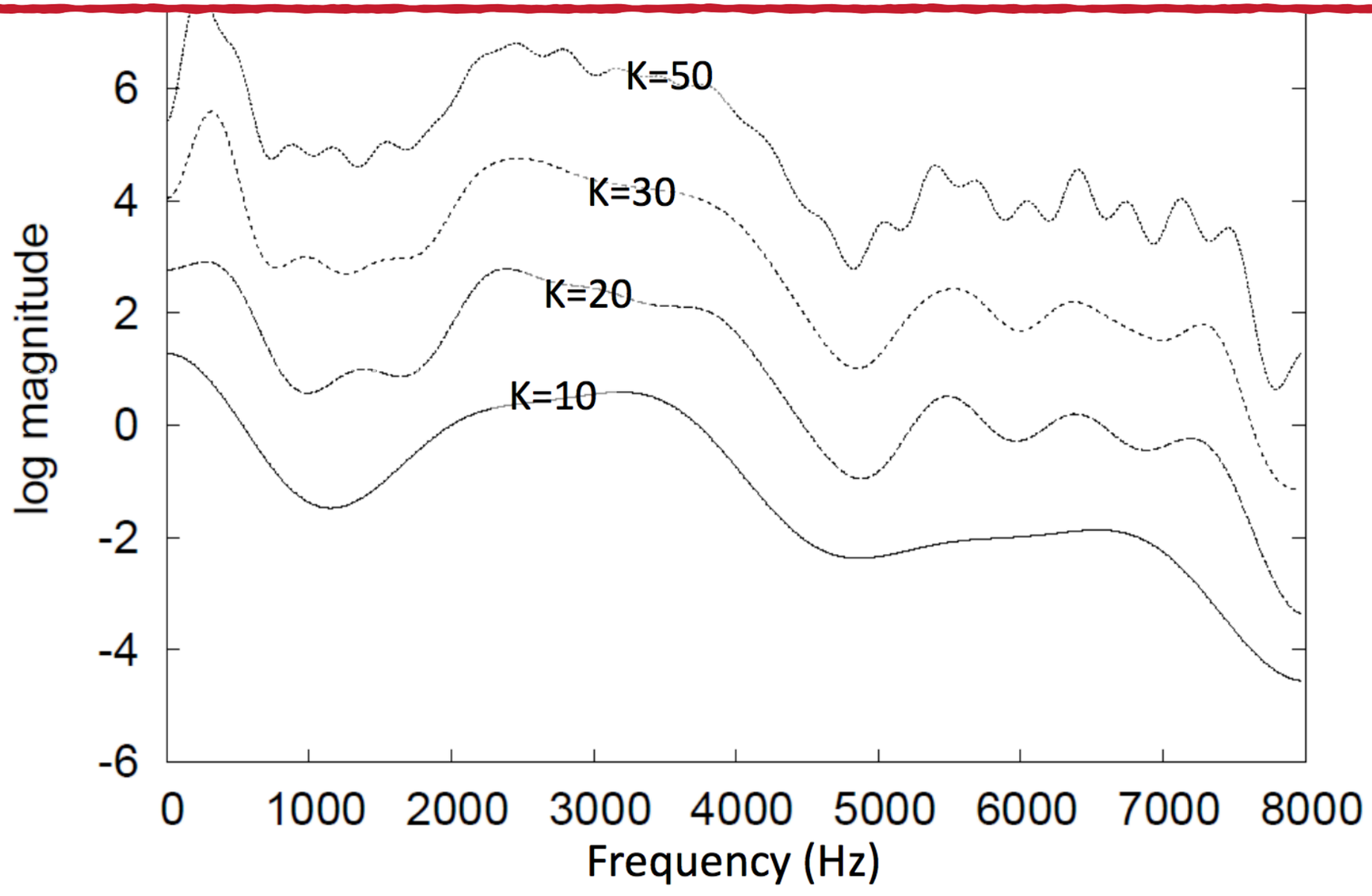
Note that the cepstrum separates out the source (fast-varying harmonic spacing) and the filter (slow-varying envelope)!

CEPSTRUM ARTICLES

- Pitch determination:
  - https://asa.scitation.org/doi/10.1121/1.1910339
- Historical overviews
  - https://ieeexplore.ieee.org/abstract/document/1328092 (speech oriented)
  - https://www.sciencedirect.com/science/article/pii/S0888327016305556 (general)

CEPSTRAL COEFFICIENTS

Cepstral coefficients compactly represent the spectral envelope (similar to principal components analysis)



CC'S FOR SPEECH RECOGNITION

- Cepstral coefficients largely uncorrelated ("orthogonality")
- Very commonly used to represent speech in speech recognition, after an auditory transform, as "MFCCs" (mel frequency cepstral coefficients)