LINGUISTIC TIMBRE: FORMANTS

YU / LAMONT

FEBRUARY 27 & MARCH 1, 2018

UMASS AMHERST

PARSING WORDS FROM CONTINUOUS SPEECH

Perception and music
by Reid McKinney - Monday, February 19, 2018, 11:58 PM

Since folks with musical training are more adept at discerning timbre, would they be better at hearing separate words in another language? I tend to have problems finding where one word ends and the other begins. Could my inexperience with music be a factor in my language perception?

Permalink | Edit | Delete | Reply | Export to portfolio

Re: Perception and music
by Chin Wah Yip - Tuesday, February 20, 2018, 1:13 AM

I remember reading some papers that suggest people with musical training perform better in speech segmentation. Rhythm, stress, intonation are some of the most important things in a speech that can be used as a way to separate words. As music training increases people's exposure of various tones and beats, it seems that music can really help language perception!

VOWEL MERGERS IN SPECTROGRAMS

Timbre in different dialects
by Melissa Karp - Sunday, February 18, 2018, 1:11 AM

I would really like to look at spectrograms of English speakers from the Midwestern United States. A typical midwestern accent includes fewer vowel differences, so that the words "pin," and "pen" sound nearly identical. I think it would be really interesting to examine a spectrogram of someone with this dialect saying these two words to see if their formants actually differ like they would in standard American English, or if the difference is too small to be perceived.

Permalink | Edit | Delete | Reply | Export to portfolio

Re: Timbre in different dialects
by Reid McKinney - Tuesday, February 20, 2018, 12:11 AM

Finding a minute difference in a vowel merger would be odd. Imagine if there was a difference between the vowels that people make without noticing. I'm pretty sure I'm not doing anything differently when I say "cot" and "caught," but now I have to check a spectrogram to be sure.

Permalink | Show parent | Edit | Split | Delete | Reply | Export to portfolio

RATE OF PHONEMES PER SECOND

Phonemes per Second

by Ayden Moczydlowski - Tuesday, February 20, 2018, 9:17 AM

It is really interesting that one can speak ten phonemes per second. With how short a second is, each phoneme would have to be spoken extremely quickly, faster than conscious thought for each individual sound. In music, comparatively, a typical tempo is 120 bpm, which would equate to 2 beats per second. Playing eighth notes at this temp is not difficult necessarily (meaning 4 notes in a second at this tempo), but once it becomes sixteenth notes (8 notes in a second at this tempo), it becomes difficult to consciously think of each note. In speech, if 10 phonemes a second is common, then the mind must think quickly to produce each phoneme, faster than is typically done in music.

SPEECH INFORMATION RATE ARTICLE

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.433.3226&rep=rep1&type=pdf

A CROSS-LANGUAGE PERSPECTIVE ON SPEECH INFORMATION RATE

FRANÇOIS PELLEGRINO

Laboratoire Dynamique du Langage, Université de Lyon and *CNRS*

CHRISTOPHE COUPÉ

Laboratoire Dynamique du Langage, Université de Lyon and *CNRS*

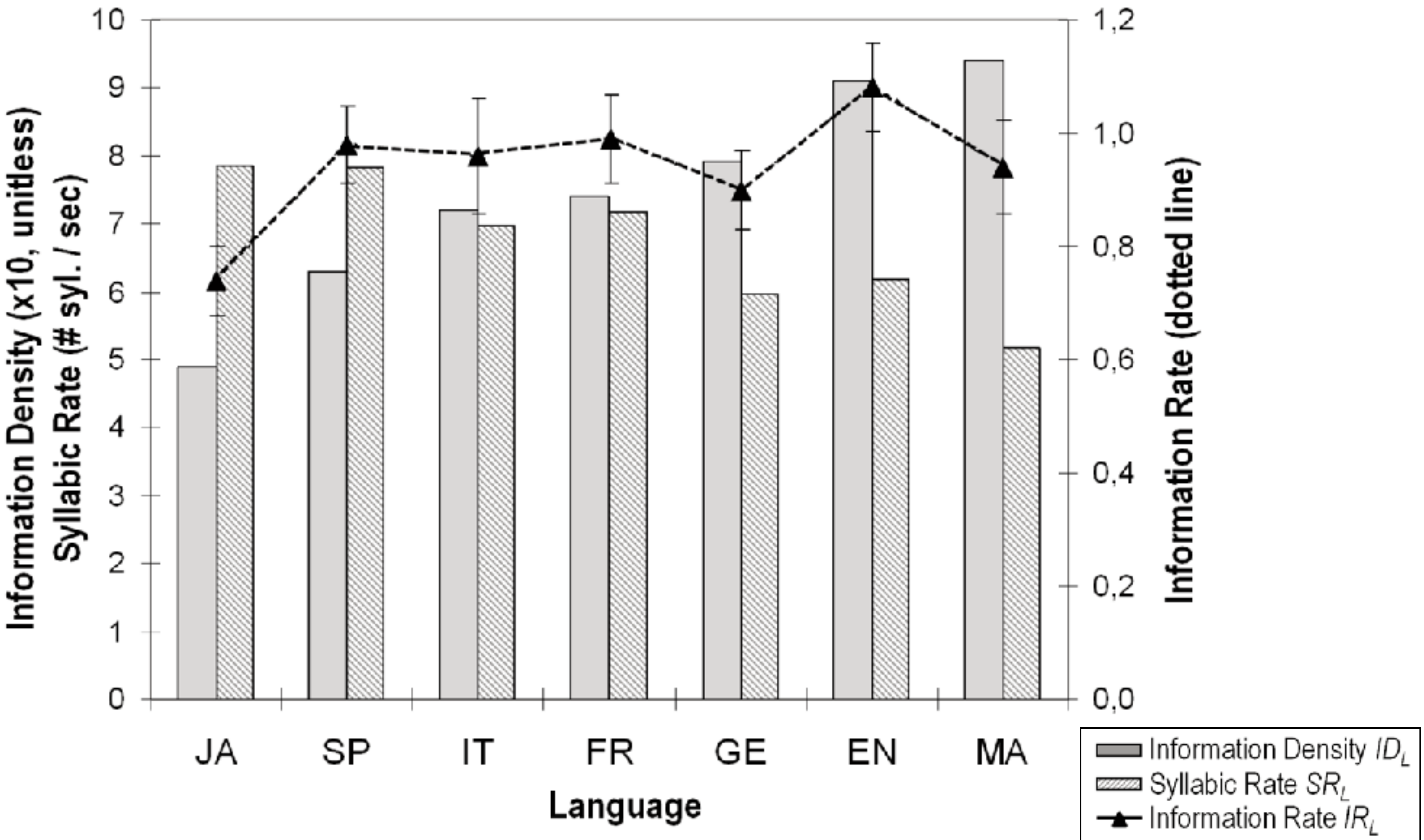
EGIDIO MARSICO

Laboratoire Dynamique du Langage, Université de Lyon and *CNRS*

This article is a crosslinguistic investigation of the hypothesis that the average information rate conveyed during speech communication results from a trade-off between average information density and speech rate. The study, based on seven languages, shows a negative correlation between density and rate, indicating the existence of several encoding strategies. However, these strategies do not necessarily lead to a constant information rate. These results are further investigated in relation to the notion of syllabic complexity.*

BRIEF ARTICLE SUMMARY: BLOG POST

http://rosettaproject.org/blog/02012/mar/1/language-speed-vs-density/



HOW MANY PHONEMES A LANGUAGE USES

the human voice and timbre
by Megan Schelb - Monday, February 19, 2018, 10:06 PM

I think that the fact that "the human voice is capable of producing timbres corresponding to ~ 800 distinct phonemes" is so incredible. I never knew that the number was that high. Since this number is so high, it makes me wonder why languages are so limiting when using phonemes. For example, why would a language only use 30 of the 800 possible phonemes we could produce? My guess would be due to language complexity, but I would be interested to see what other explanations there could be.

Phoible: http://phoible.org/

UPSID: http://web.phonetik.uni-frankfurt.de/upsid_info.html

LAPSyD: http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/index.php

THE VOCAL TRACT AND TIMBRE: VOWELS

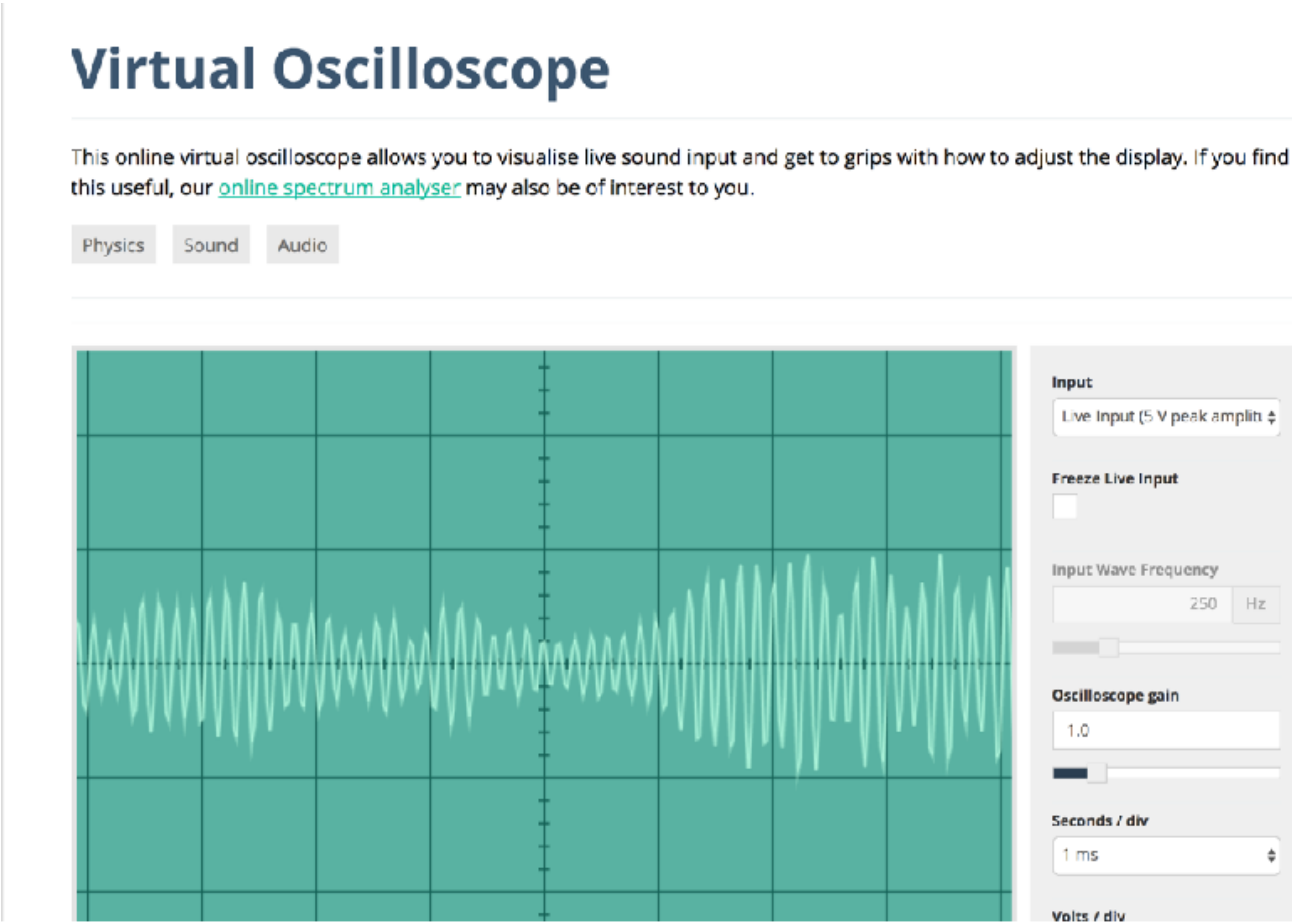
MRI VIDEOS OF VOWEL PRODUCTIONS!

front close unrounded vowel

back open unrounded vowel

http://sail.usc.edu/span/rtmri_ipa/pk_2015.html

REAL-TIME OSCILLOSCOPE

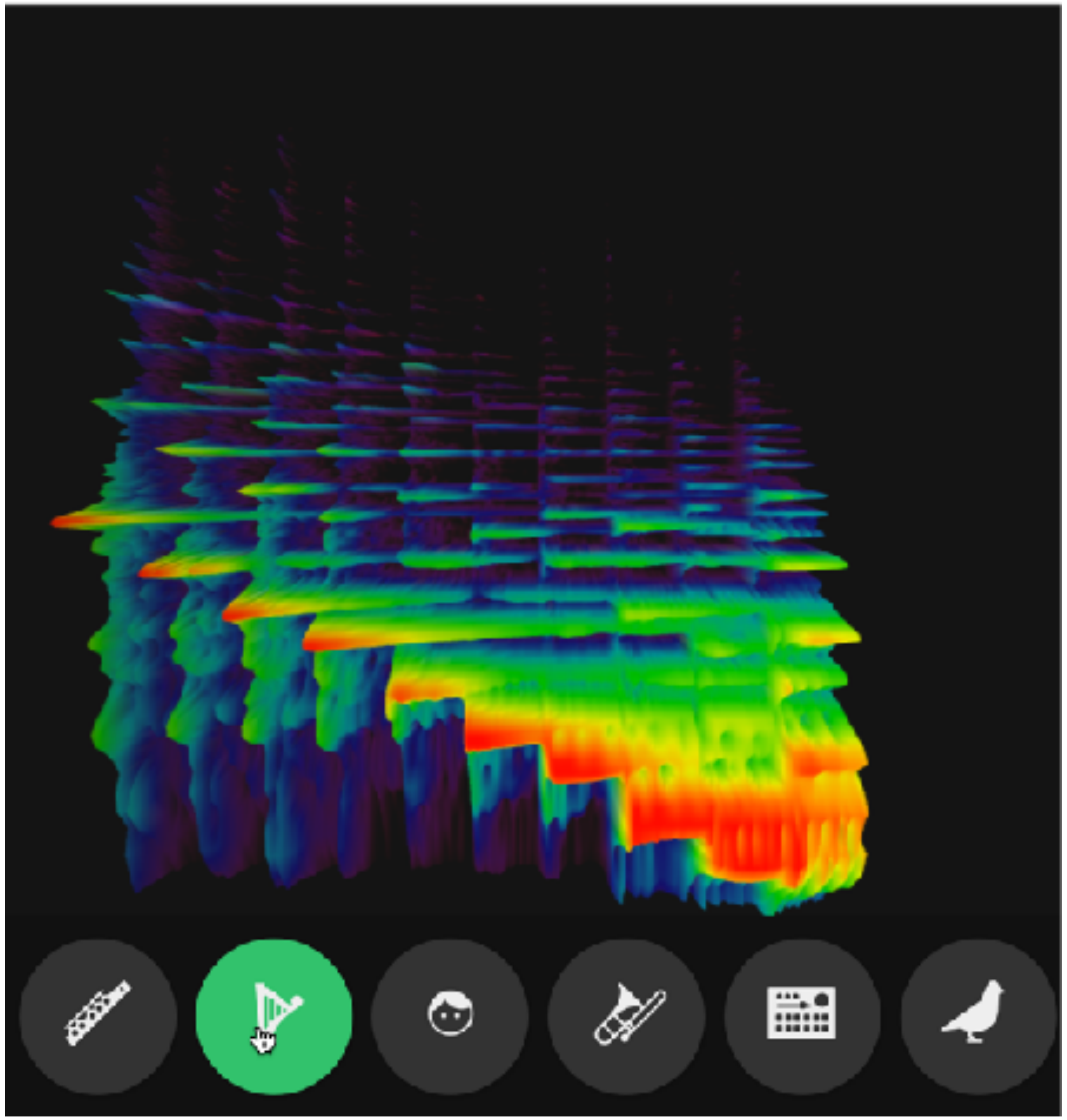


http://academo.org/demos/virtual-oscilloscope/

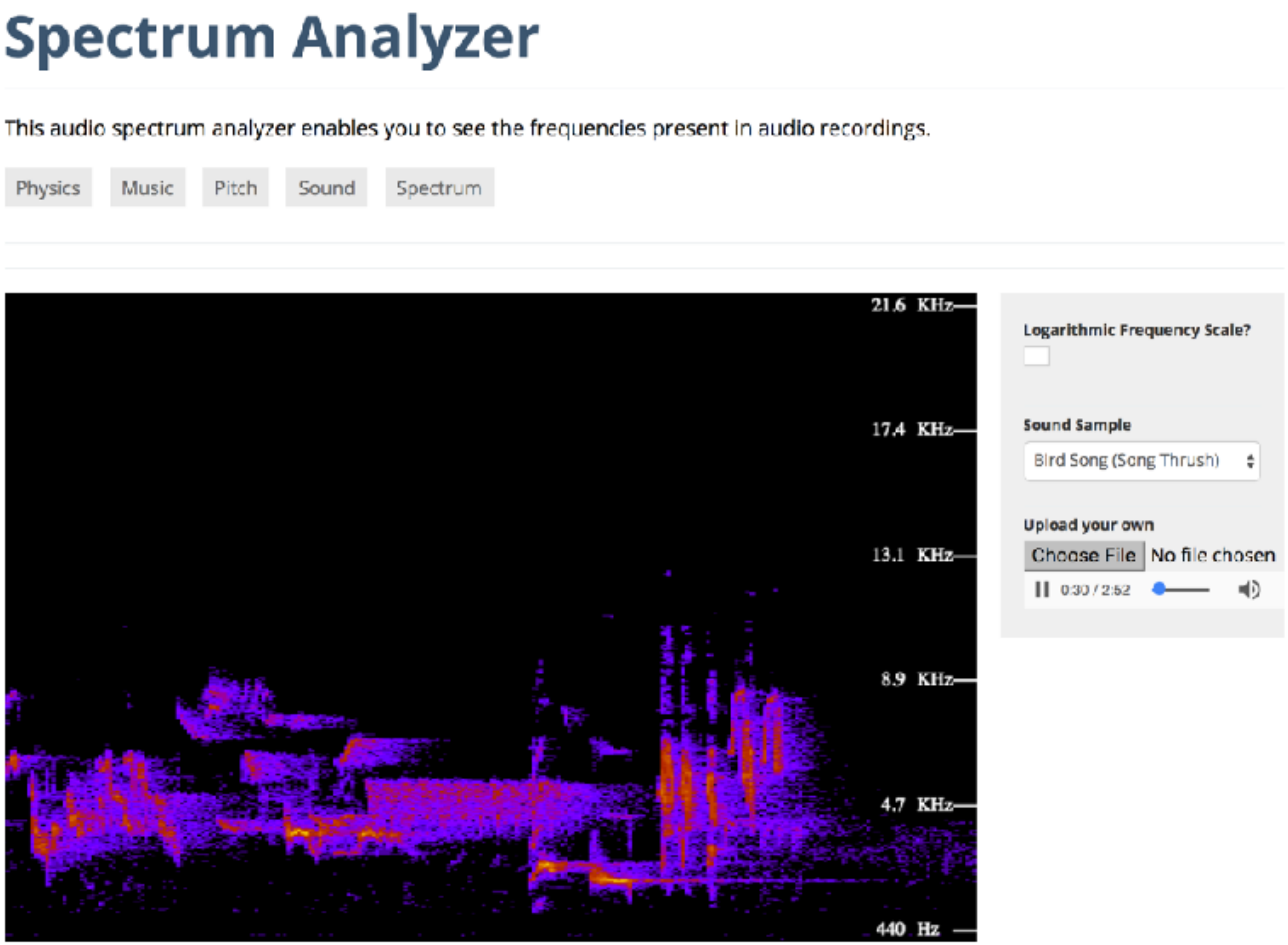
REAL-TIME SPECTRUM ANALYZER II

http://musiclab.chromeexperiments.com/Spectrogram

SPECTROGRAM



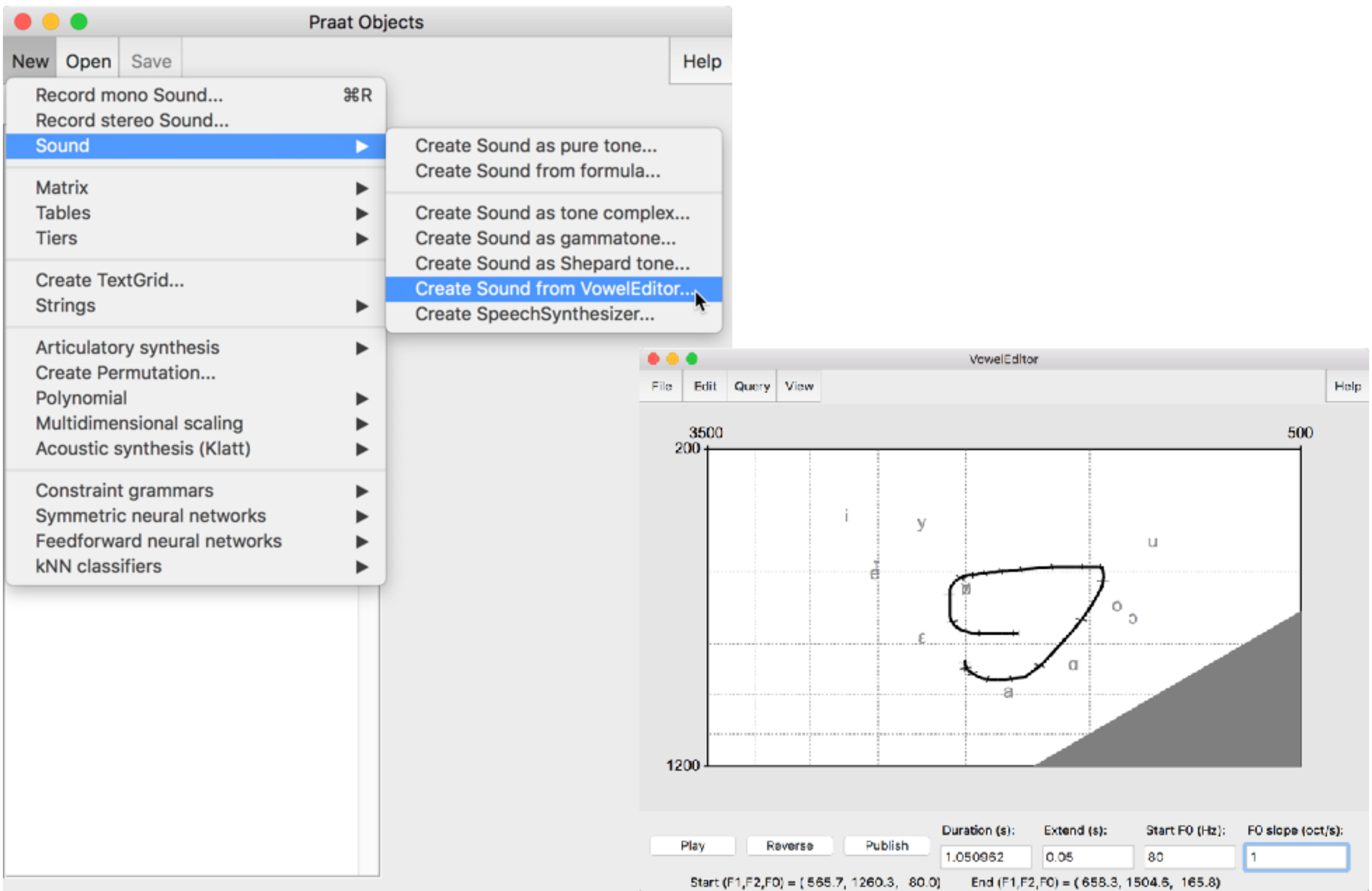
REAL-TIME SPECTRUM ANALYZER



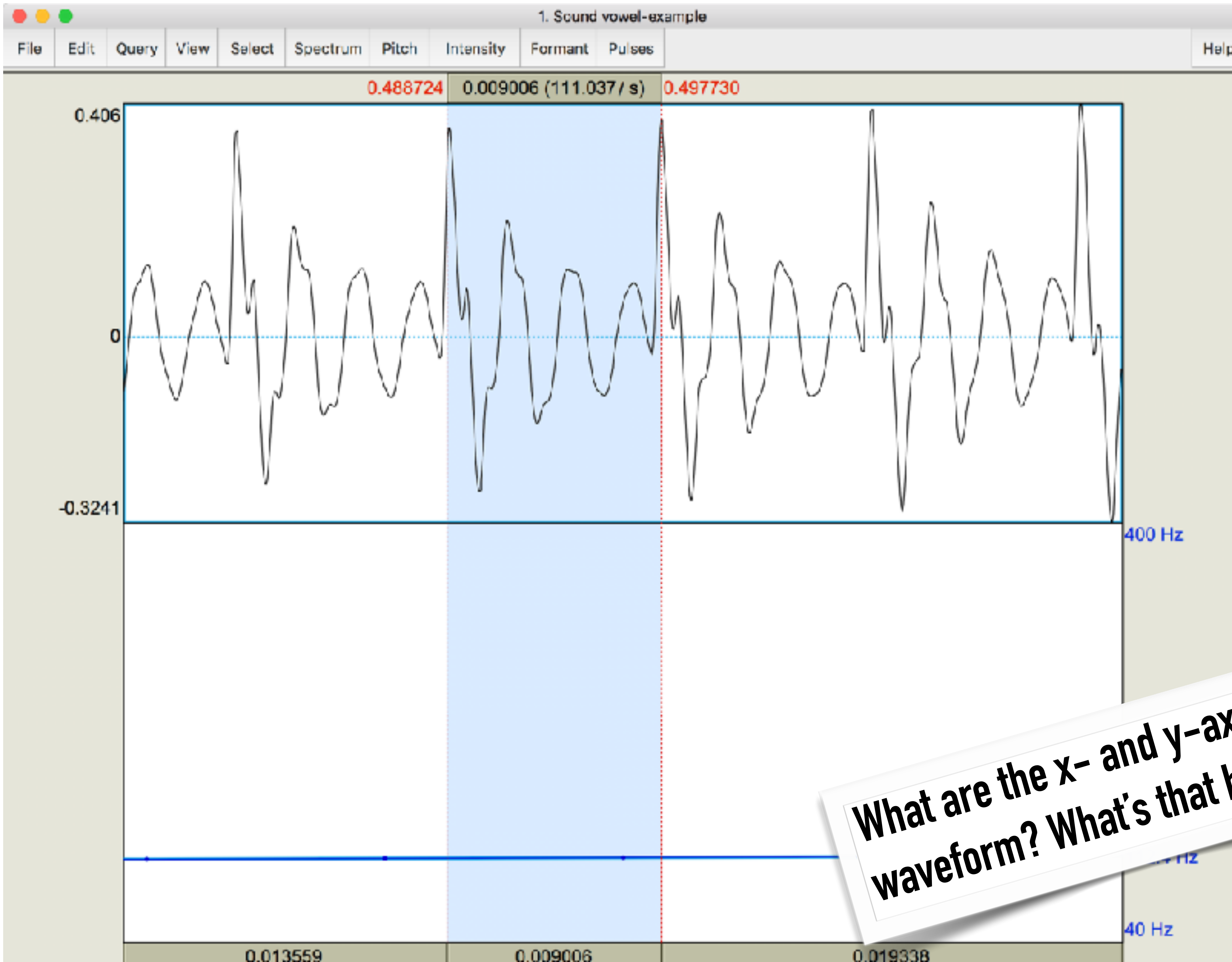
http://academo.org/demos/spectrum-analyzer/

WAVEFORMS, SPECTRA, AND SPECTROGRAMS

SYNTHESIZE A VOCALIC SOUND…

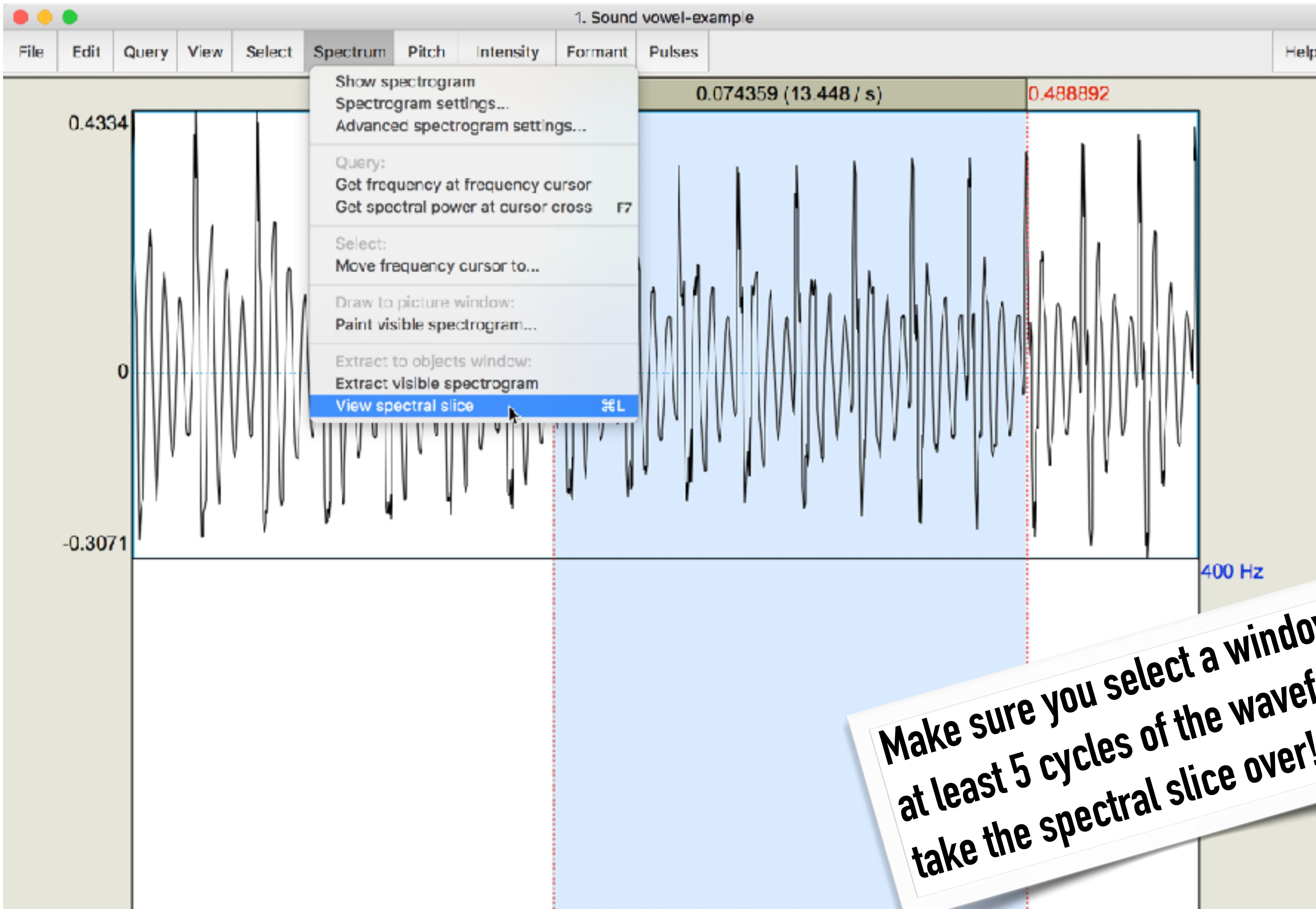


REVIEW: FIND PERIOD FROM THE WAVEFORM



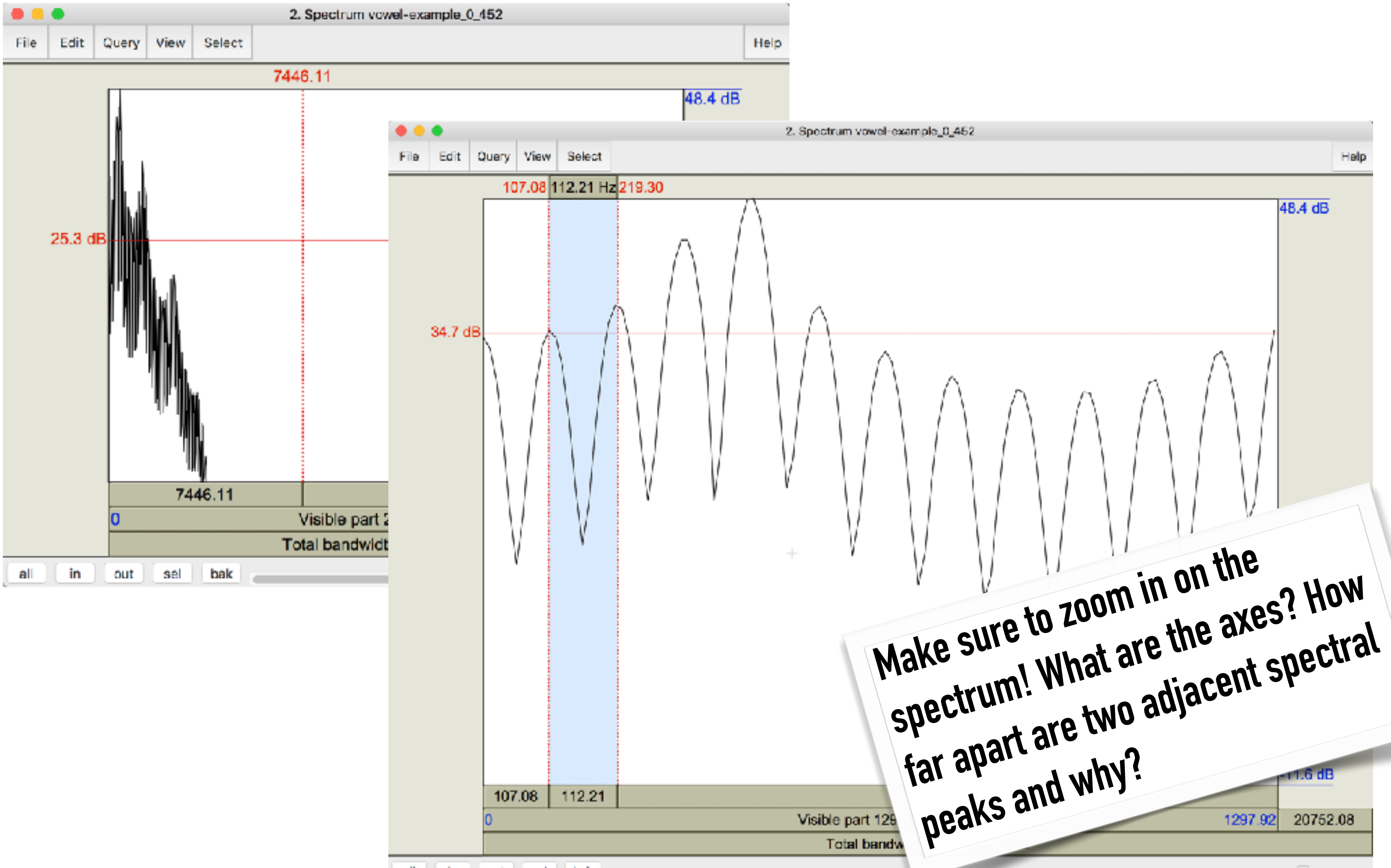
What are the x- and y-axes in the waveform? What's that blue line?

REVIEW: TAKE SPECTRAL SLICE (SPECTRUM)

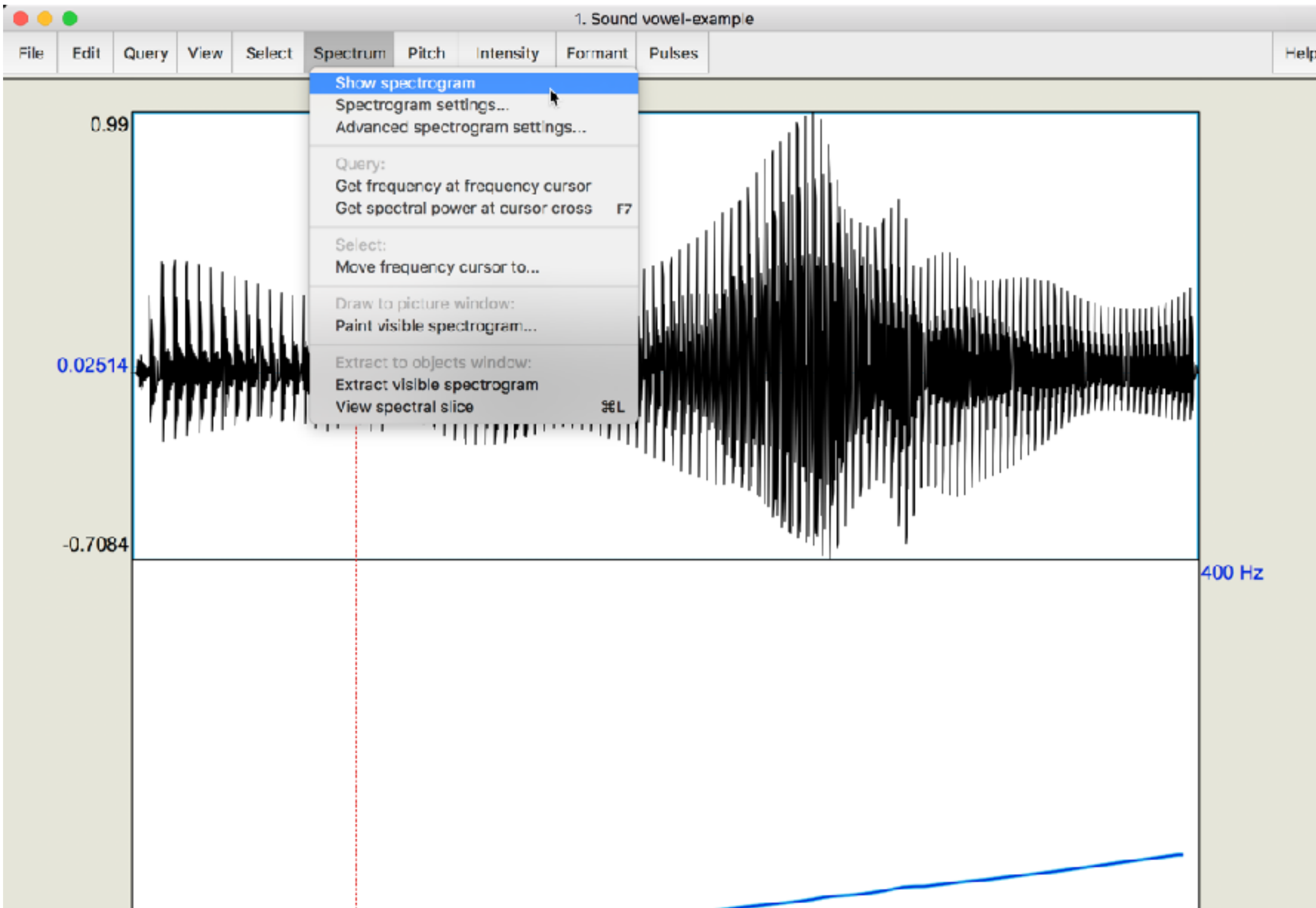


Make sure you select a window of at least 5 cycles of the waveform to take the spectral slice over!

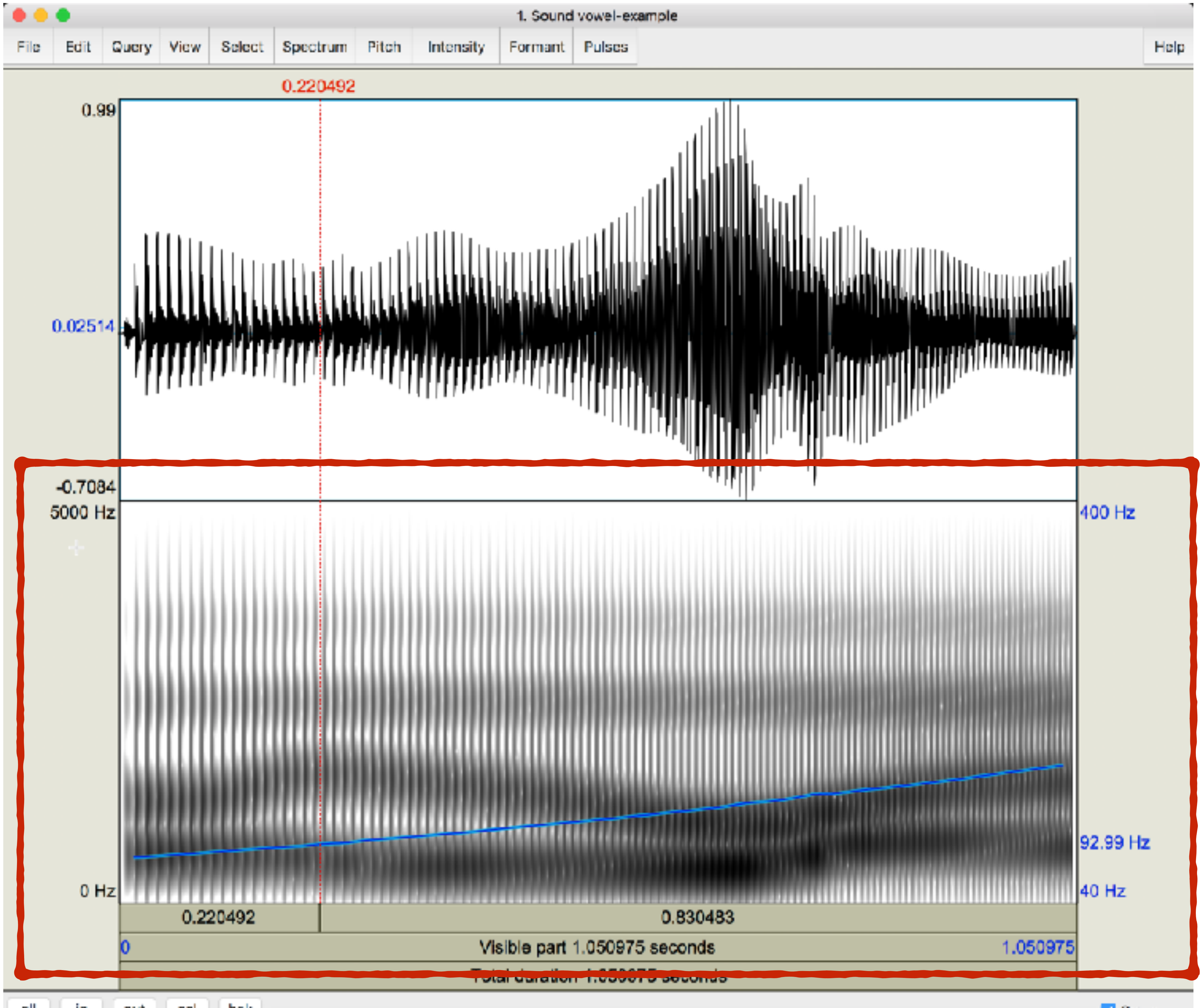
REVIEW: UNDERSTANDING THE SPECTRUM



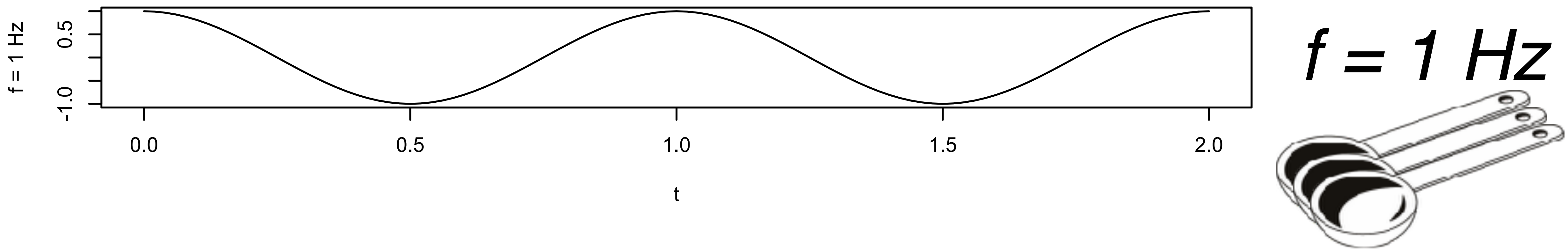
NEW: THE SPECTROGRAM!

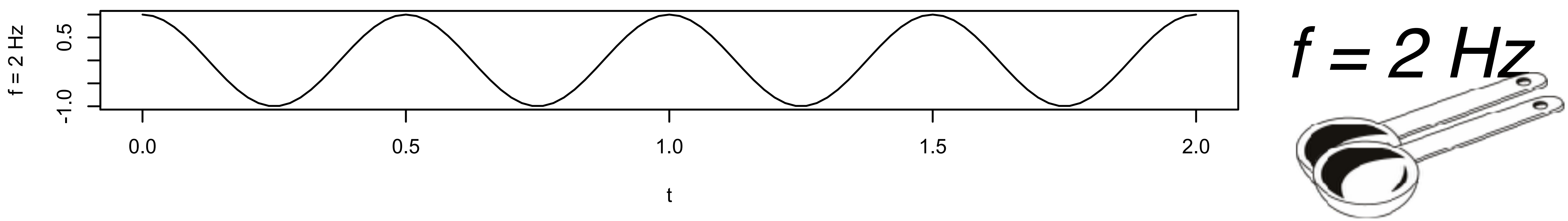


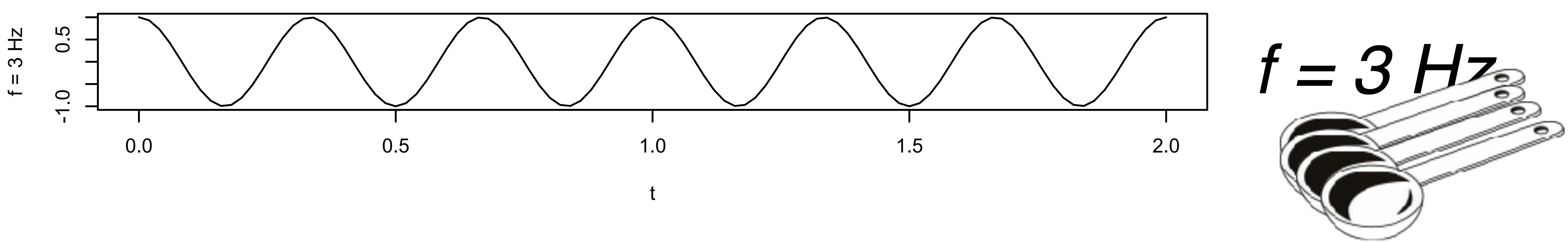
NEW: THE SPECTROGRAM!

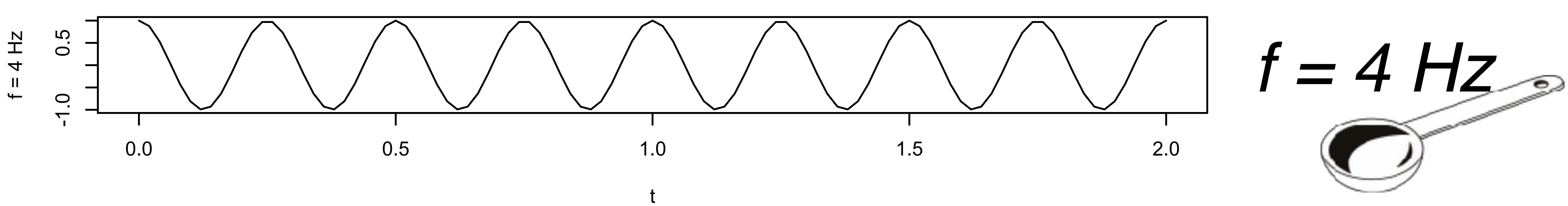


FROM THE SPECTRUM TO THE SPECTROGRAM

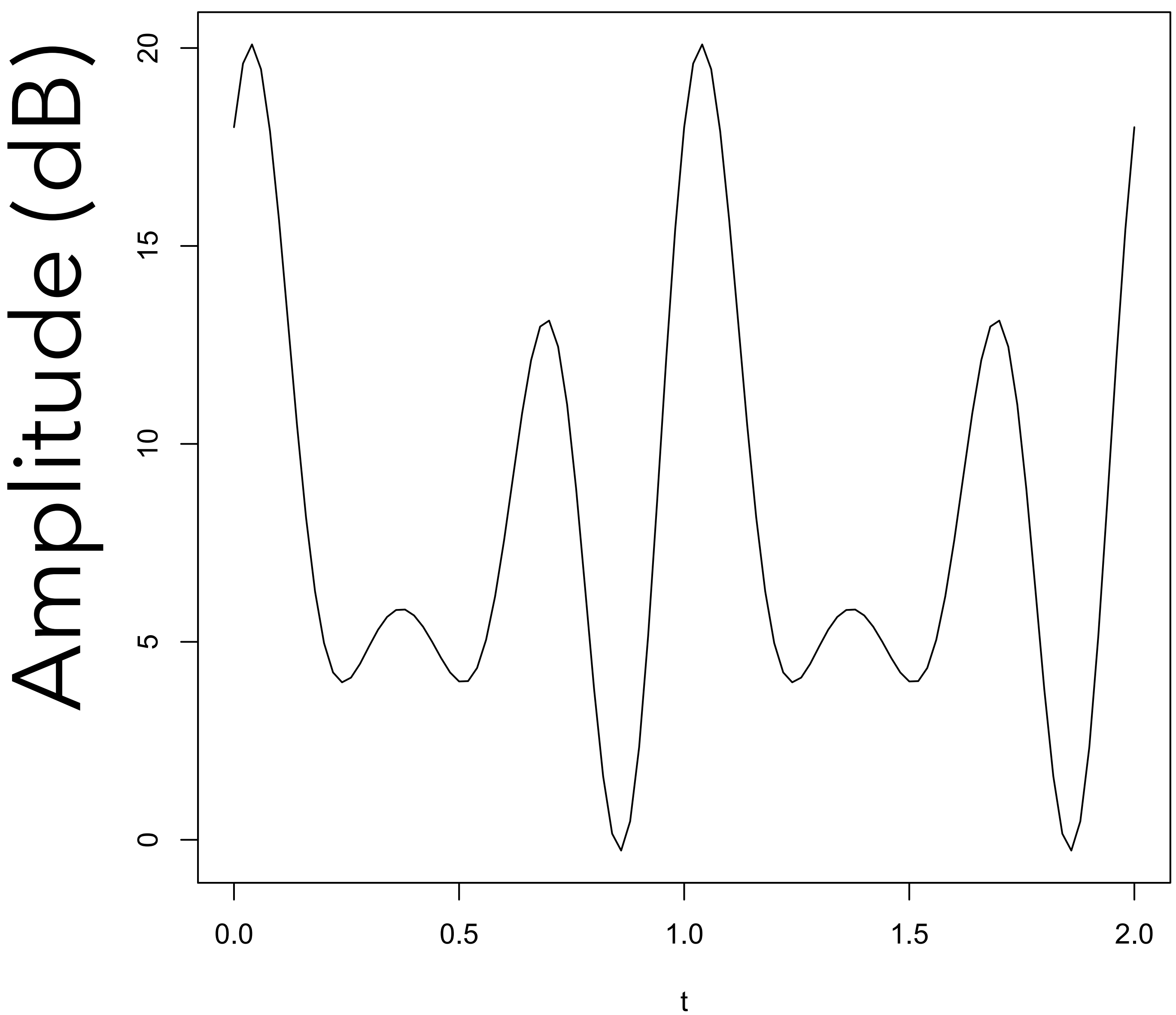

f = 1 Hz
0.5
-1.0
0.0
0.5
1.0
1.5
2.0
t
f = 1 Hz


f = 2 Hz
0.5
-1.0
0.0
0.5
1.0
1.5
2.0
t
f = 2 Hz

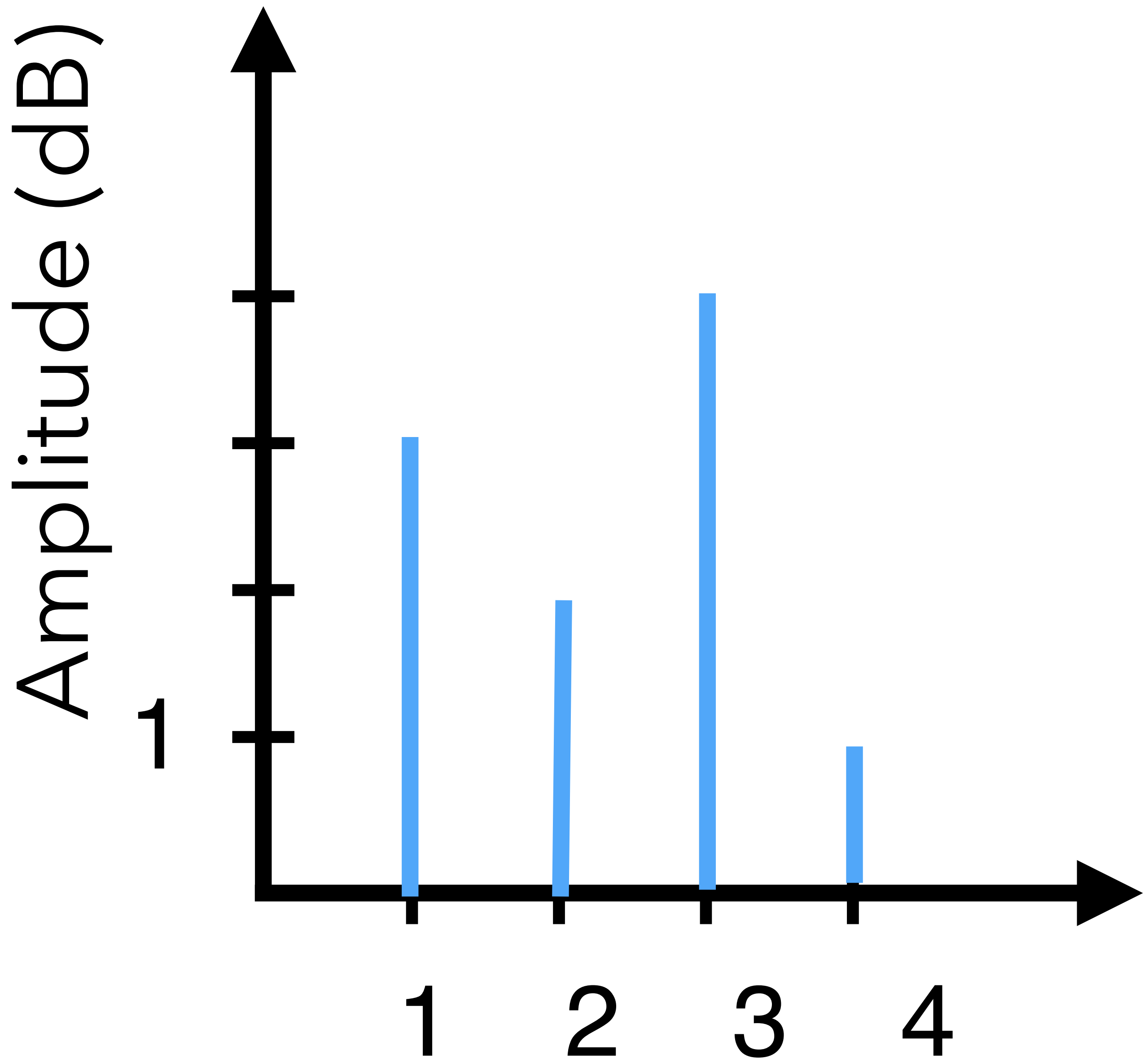

f = 3 Hz
0.5
-1.0
0.0
0.5
1.0
1.5
2.0
t
f = 3 Hz


f = 4 Hz
0.5
-1.0
0.0
0.5
1.0
1.5
2.0
t
f = 4 Hz

SPECTRUM OF COMPLEX WAVEFORM

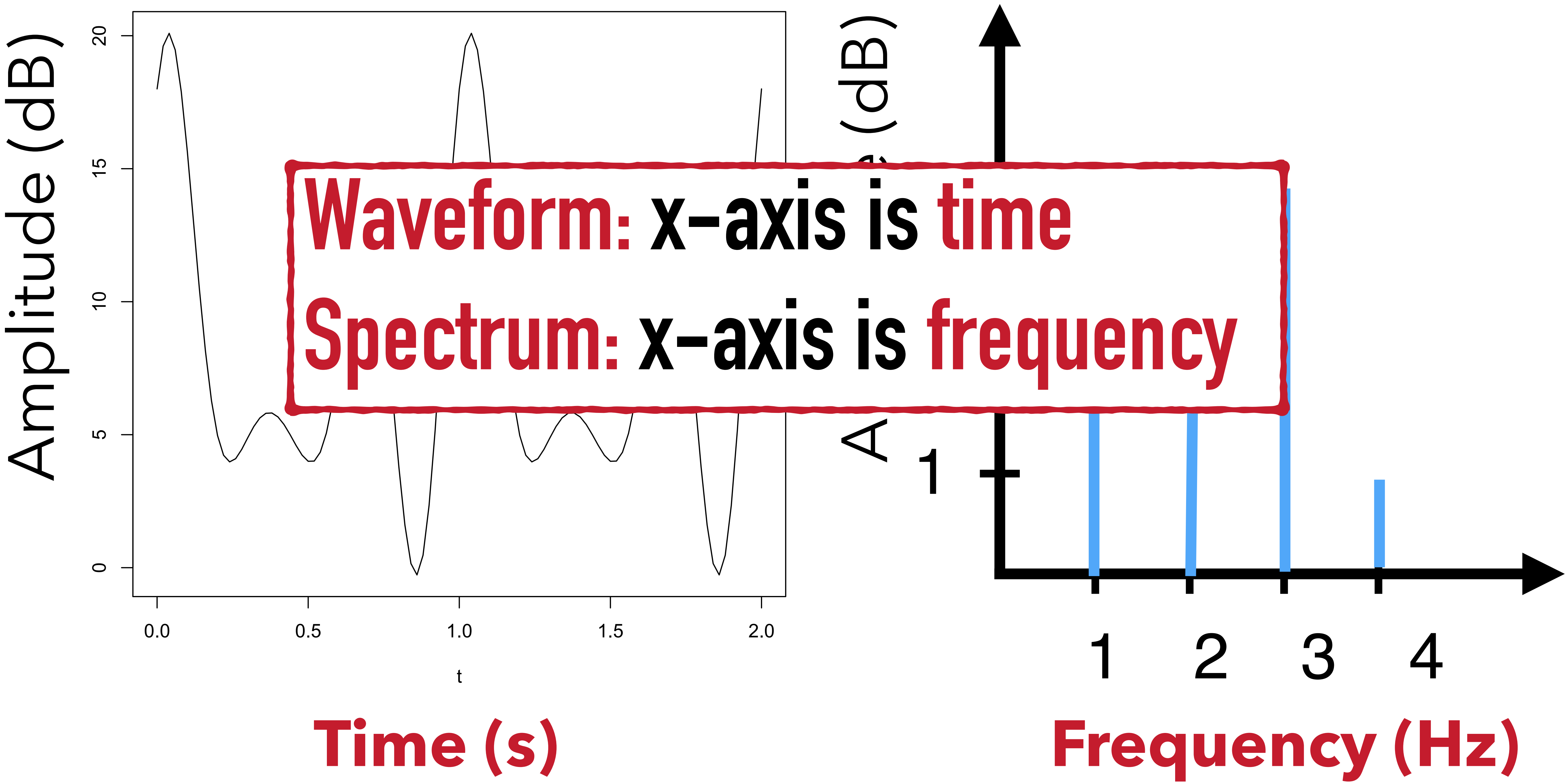


Time (s)



Frequency (Hz)

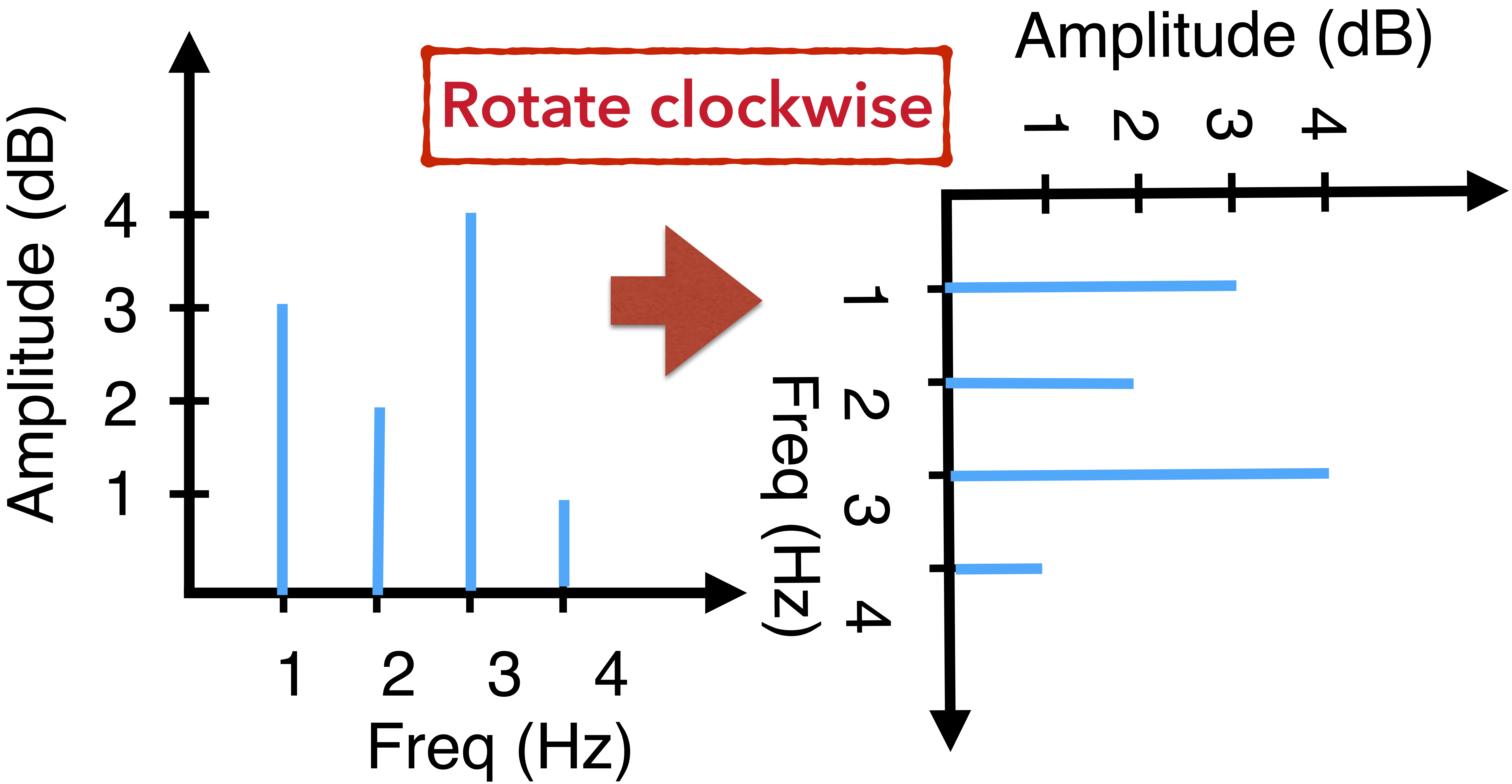
SPECTRUM OF COMPLEX WAVEFORM



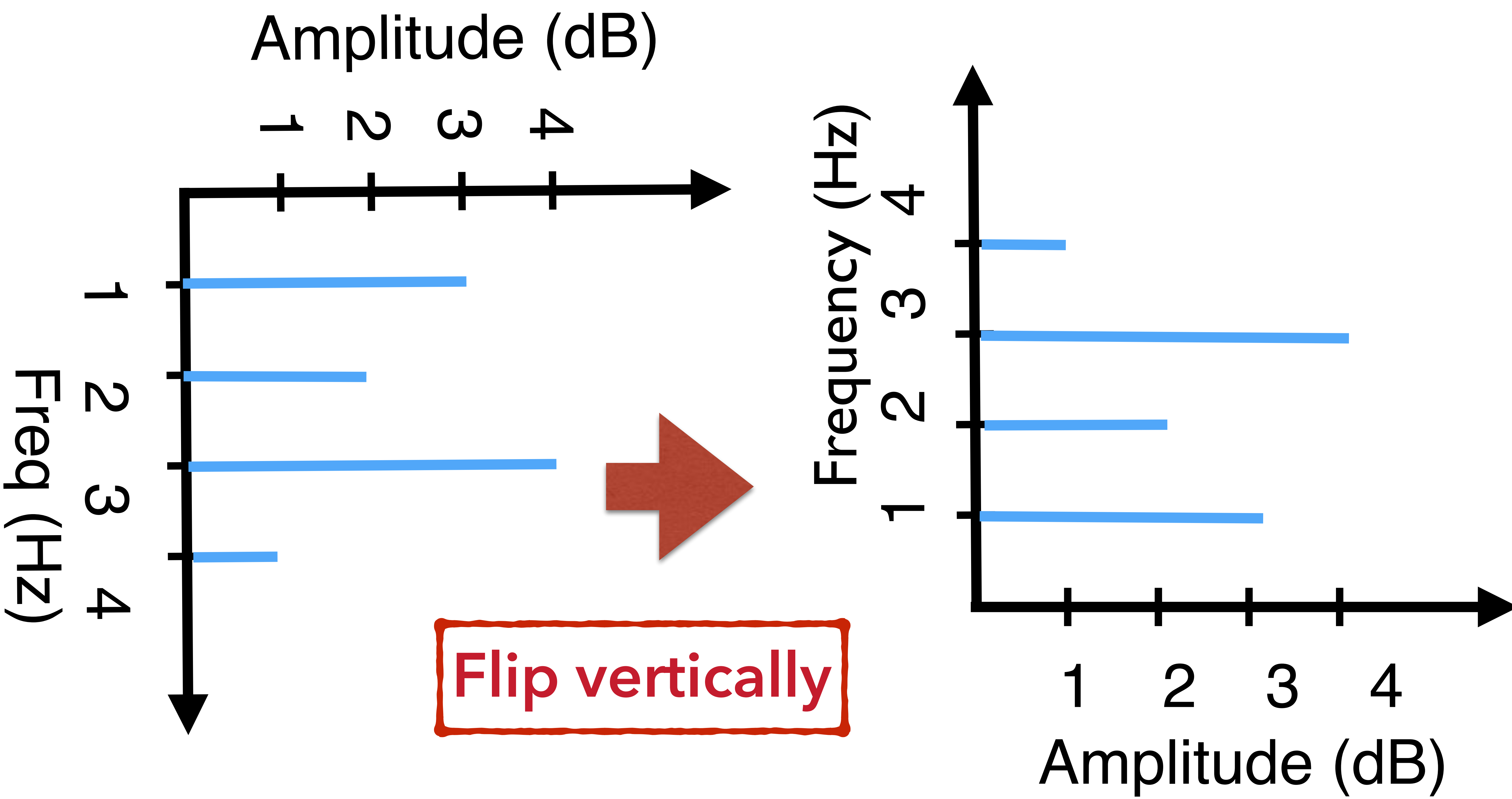
Waveform: x-axis is time

Spectrum: x-axis is frequency

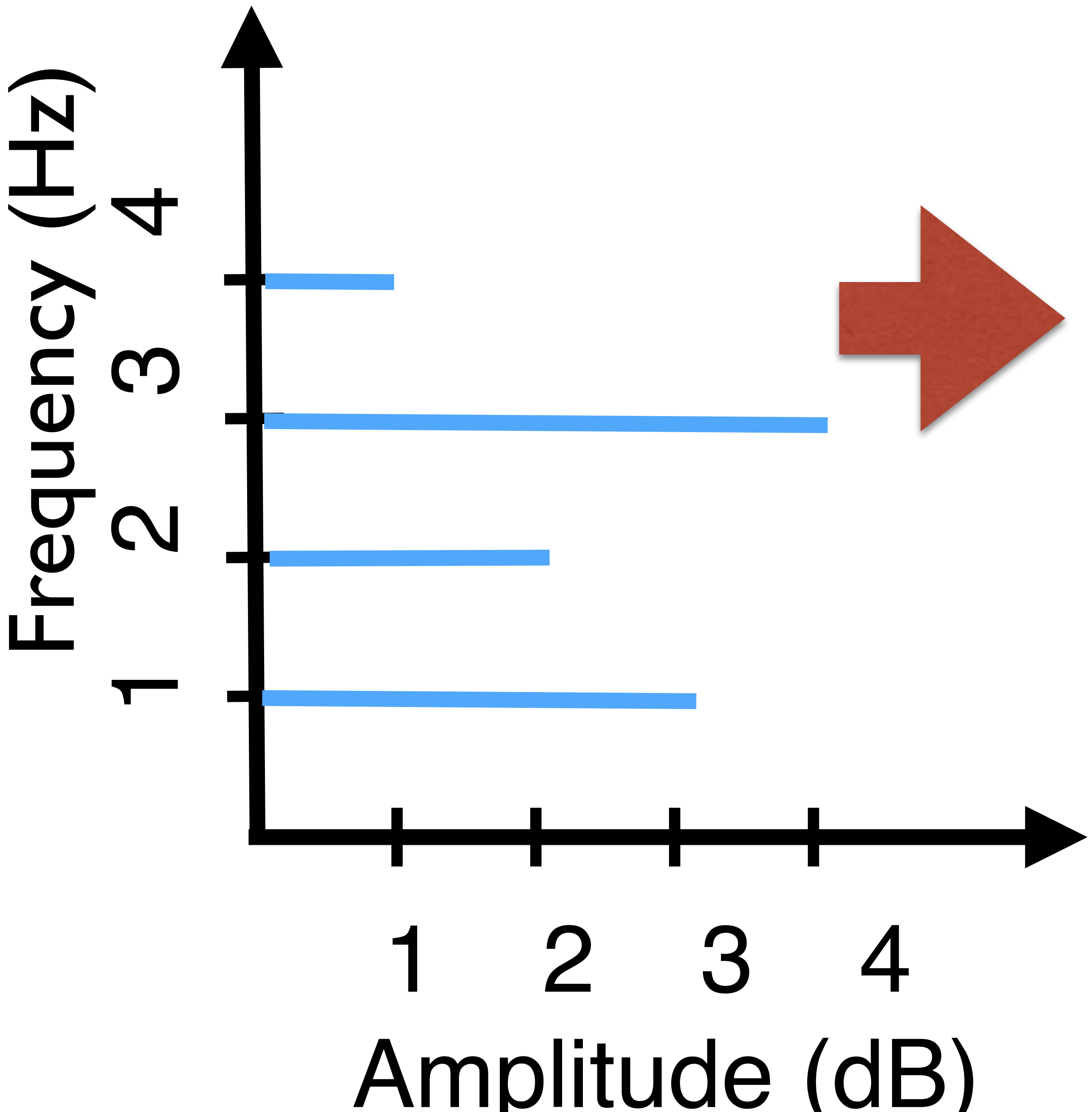
FROM SPECTRUM TO SPECTROGRAM: STEP 1



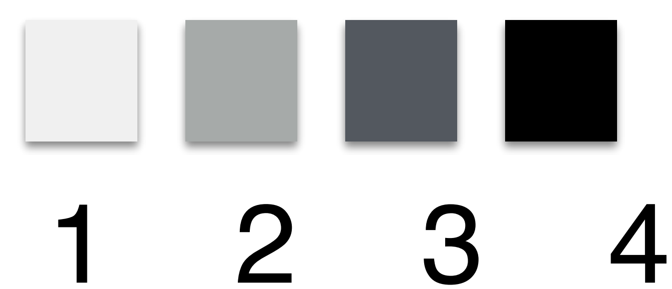
FROM SPECTRUM TO SPECTROGRAM: STEP 2



FROM SPECTRUM TO SPECTROGRAM: STEP 3



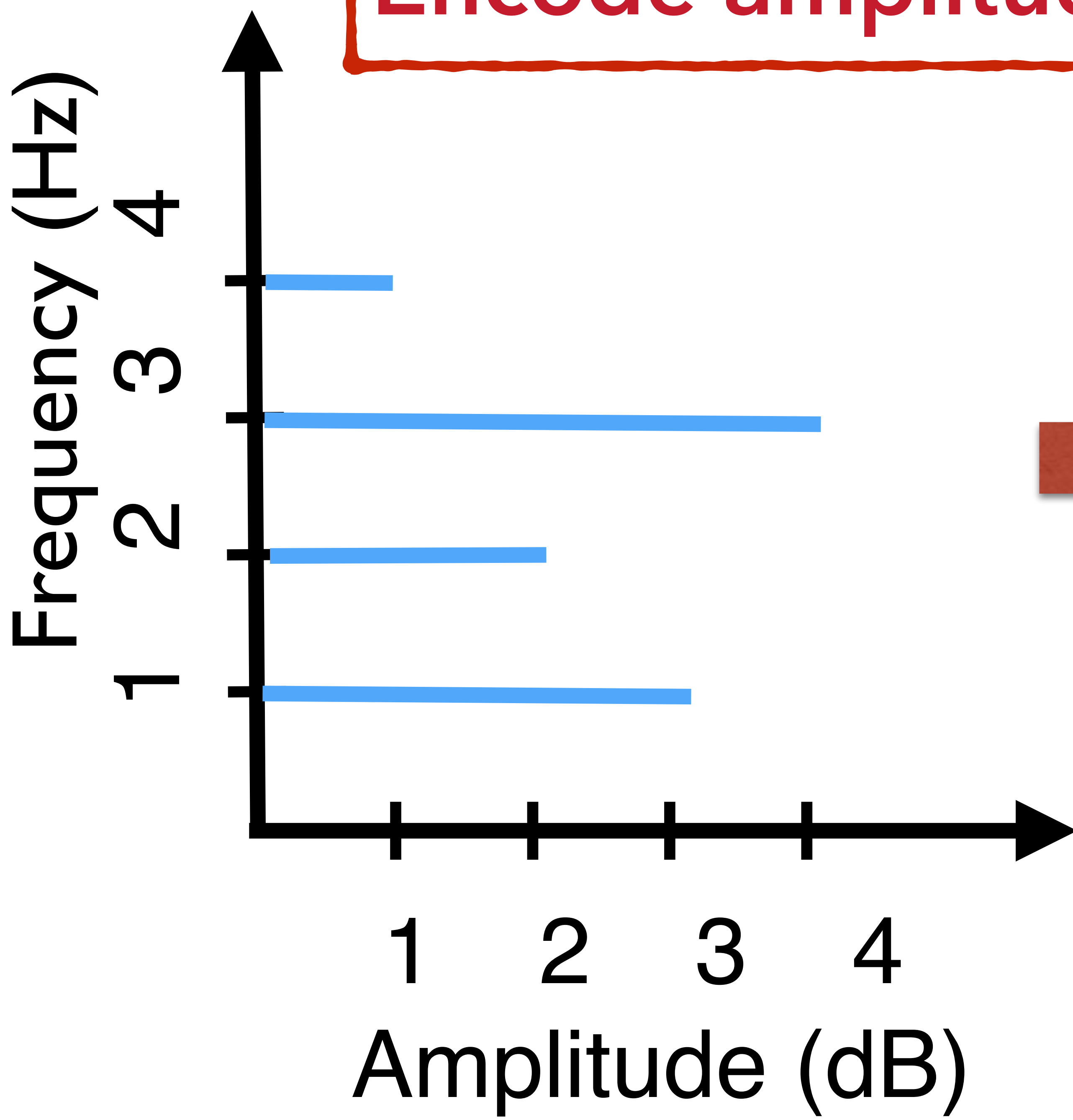
Encode amplitude levels (the height/ length of the blue spikes) in grayscale*

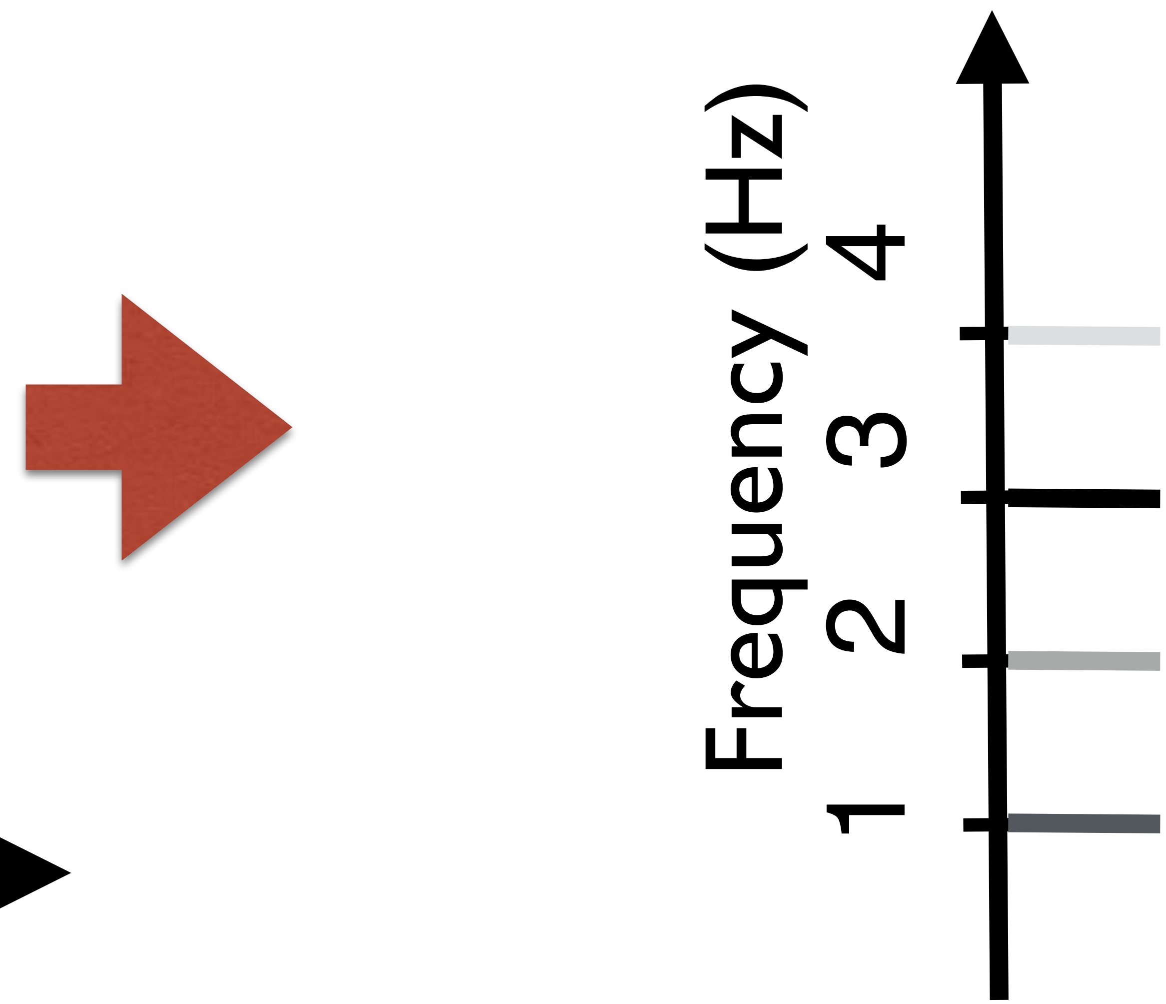


*or with heat map

FROM SPECTRUM TO SPECTROGRAM: STEP 4

Encode amplitudes in grayscale…

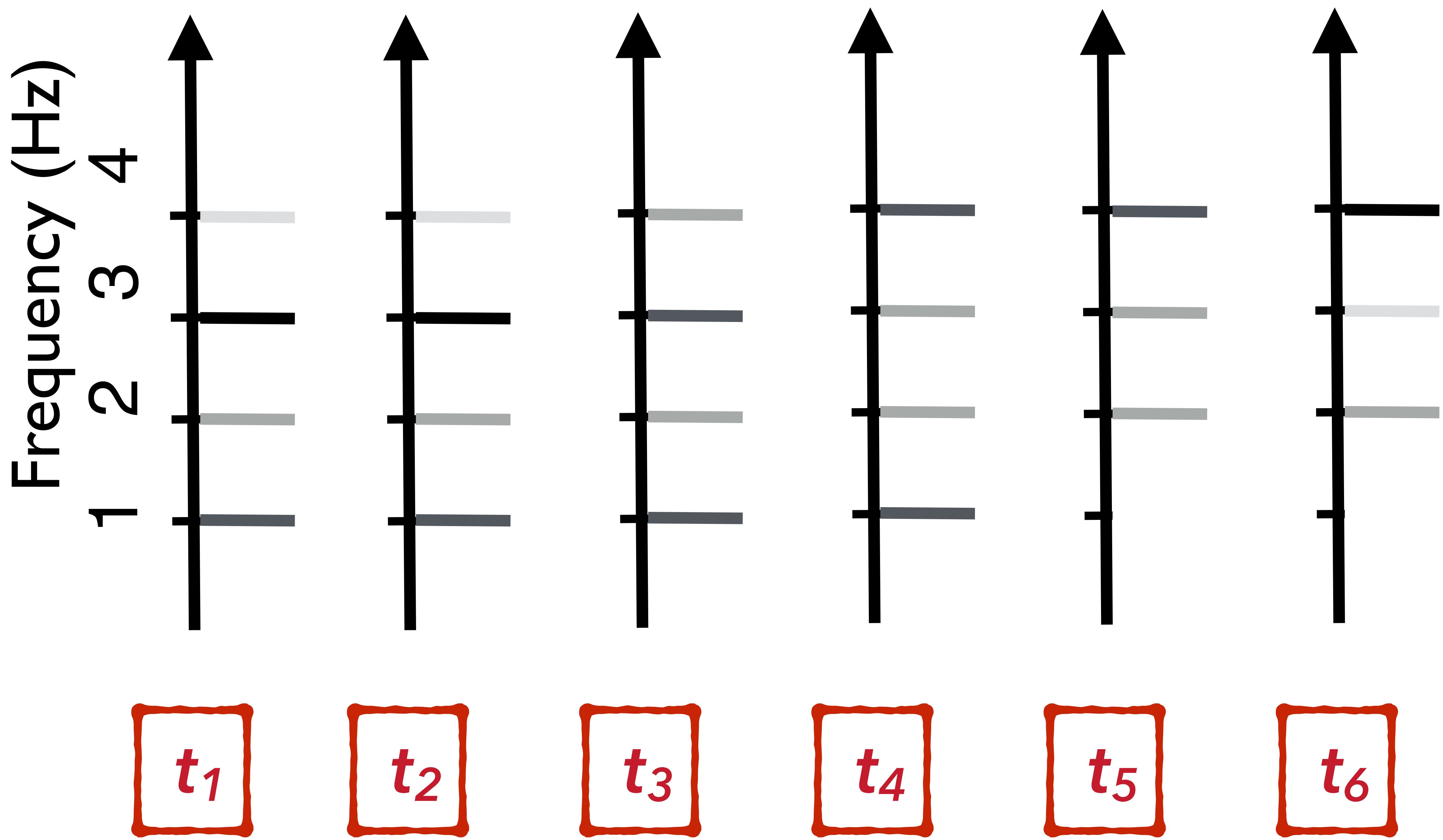




…at time t_1

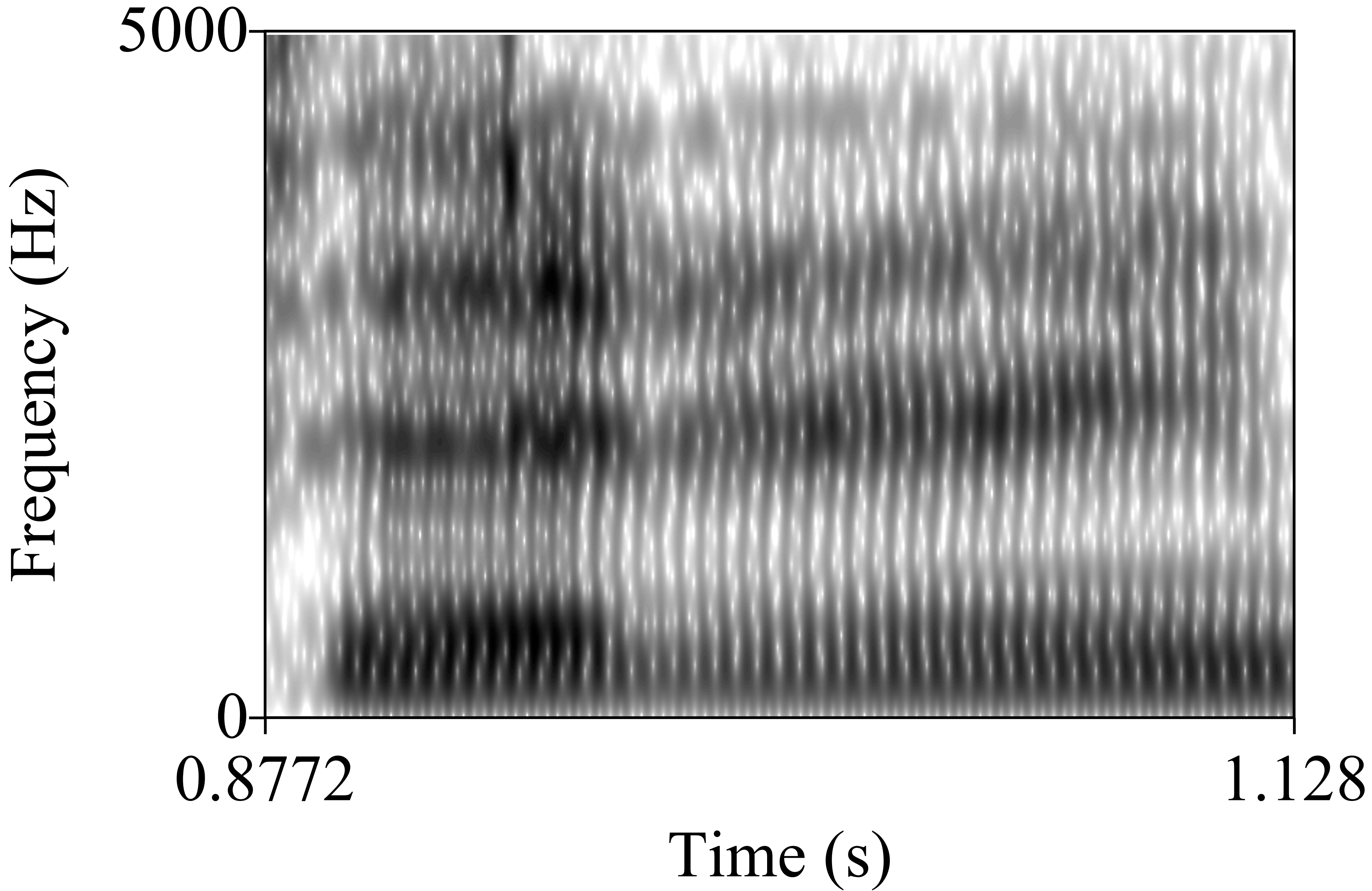
FROM SPECTRUM TO SPECTROGRAM: STEP 5

Line up spectral slices taken over time…



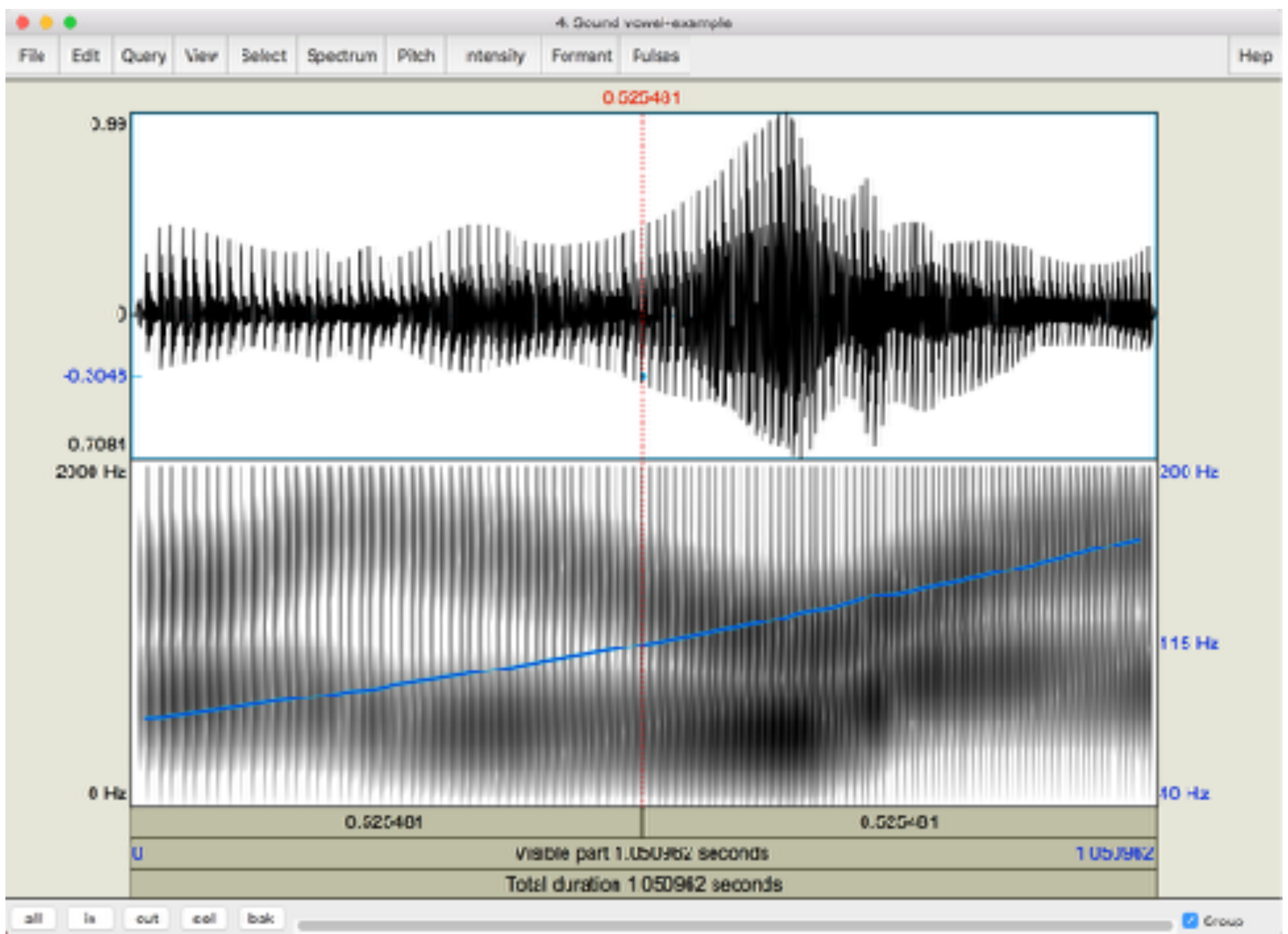
FROM SPECTRUM TO SPECTROGRAM: FINIS

…and you have a spectrogram



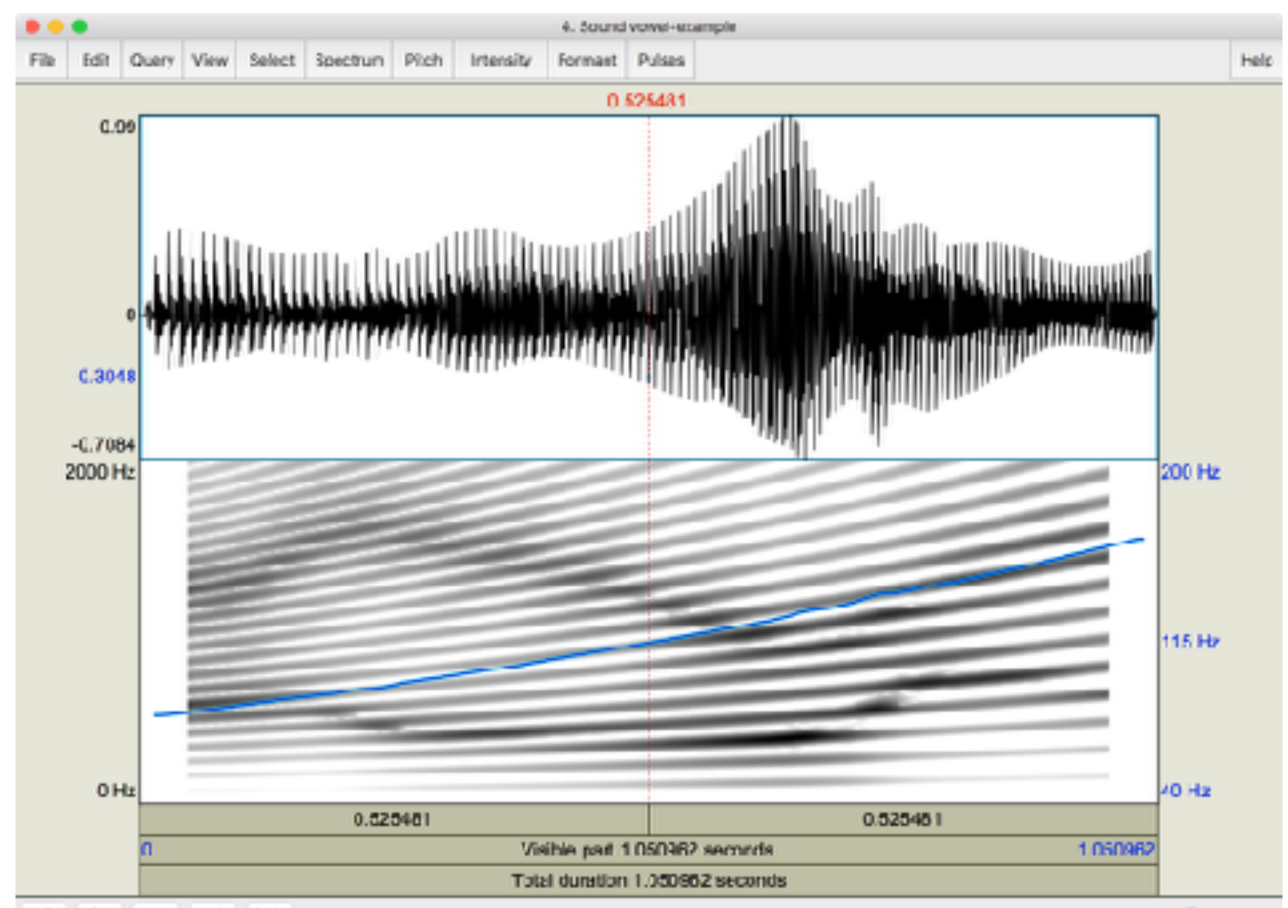
WIDE VS. NARROW-BAND SPECTROGRAMS

Wide: high temporal resolution, low frequency resolution
Narrow: low temporal resolution, high frequency resolution



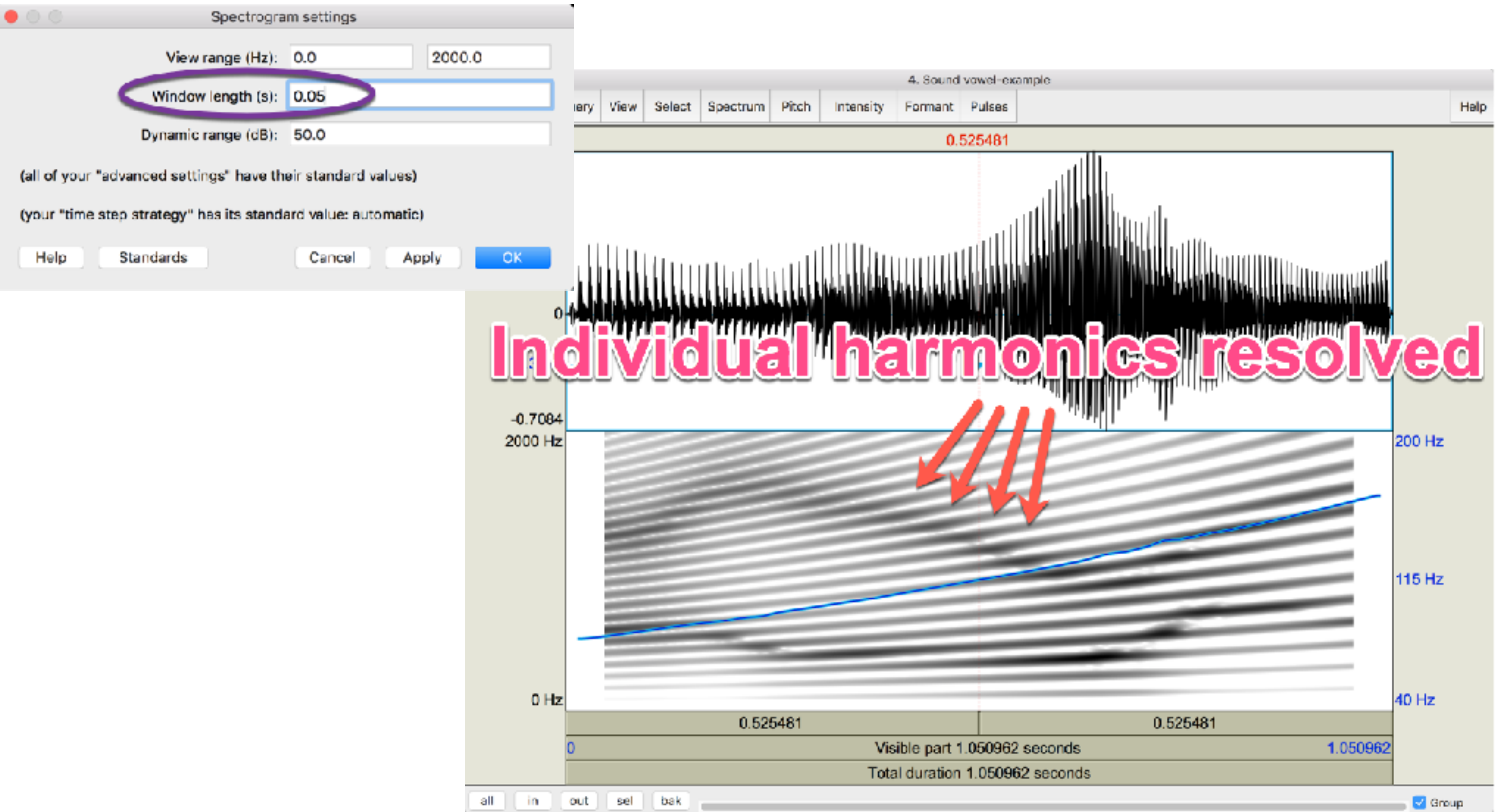
WIDE

NARROW



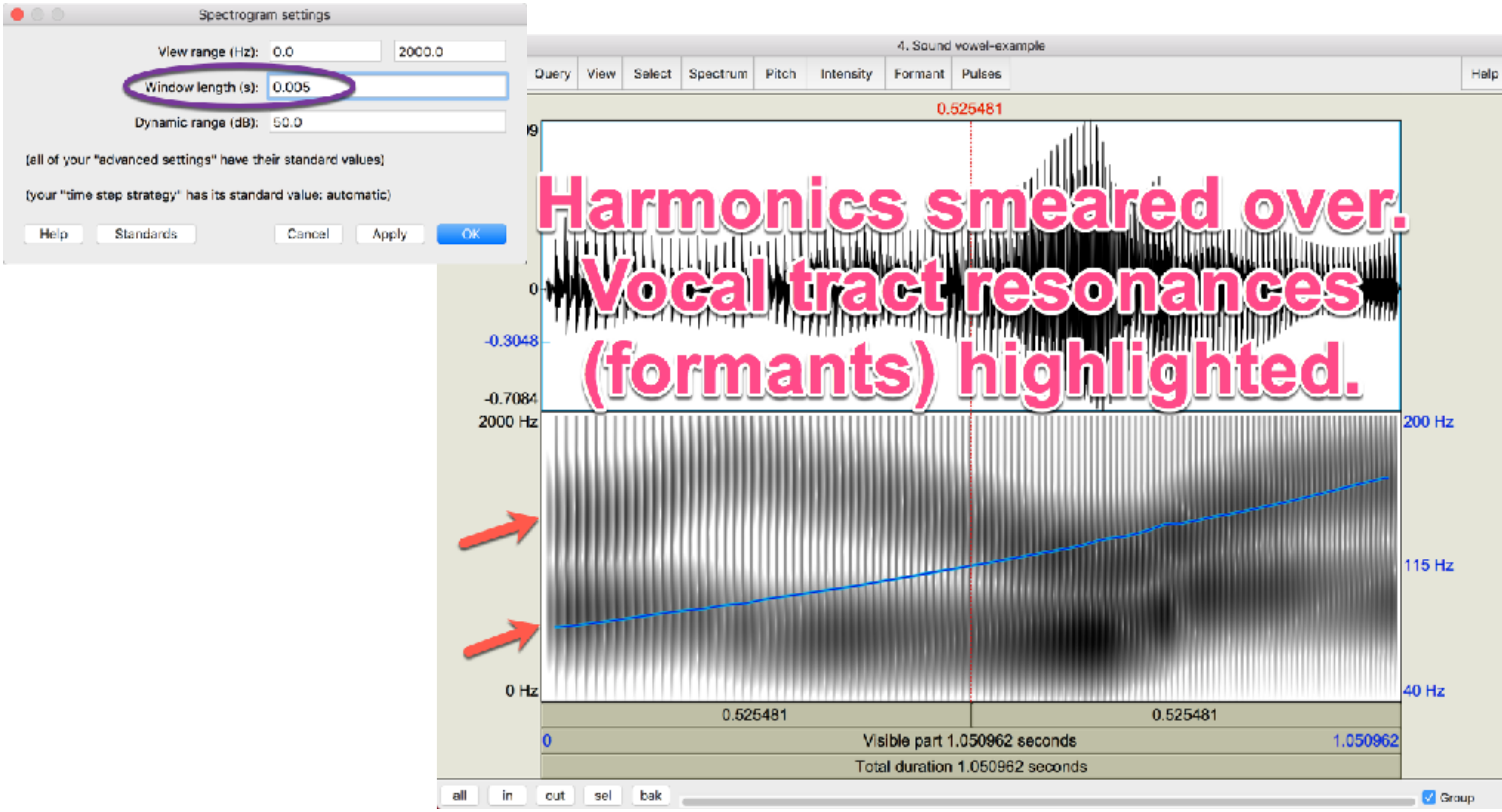
NARROW-BAND SPECTROGRAM

Narrow: low temporal resolution, high frequency resolution



WIDE-BAND SPECTROGRAM

Wide: high temporal resolution, low frequency resolution

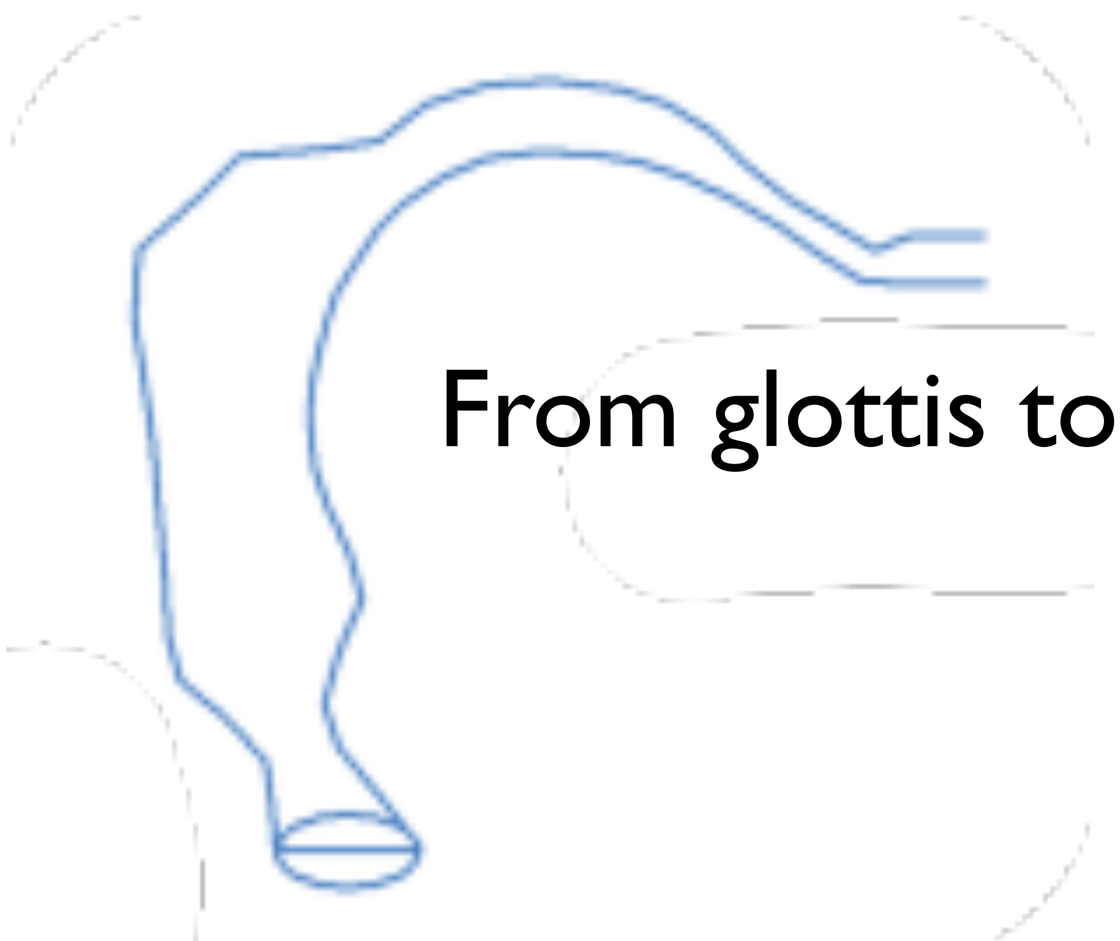


SOURCE-FILTER THEORY

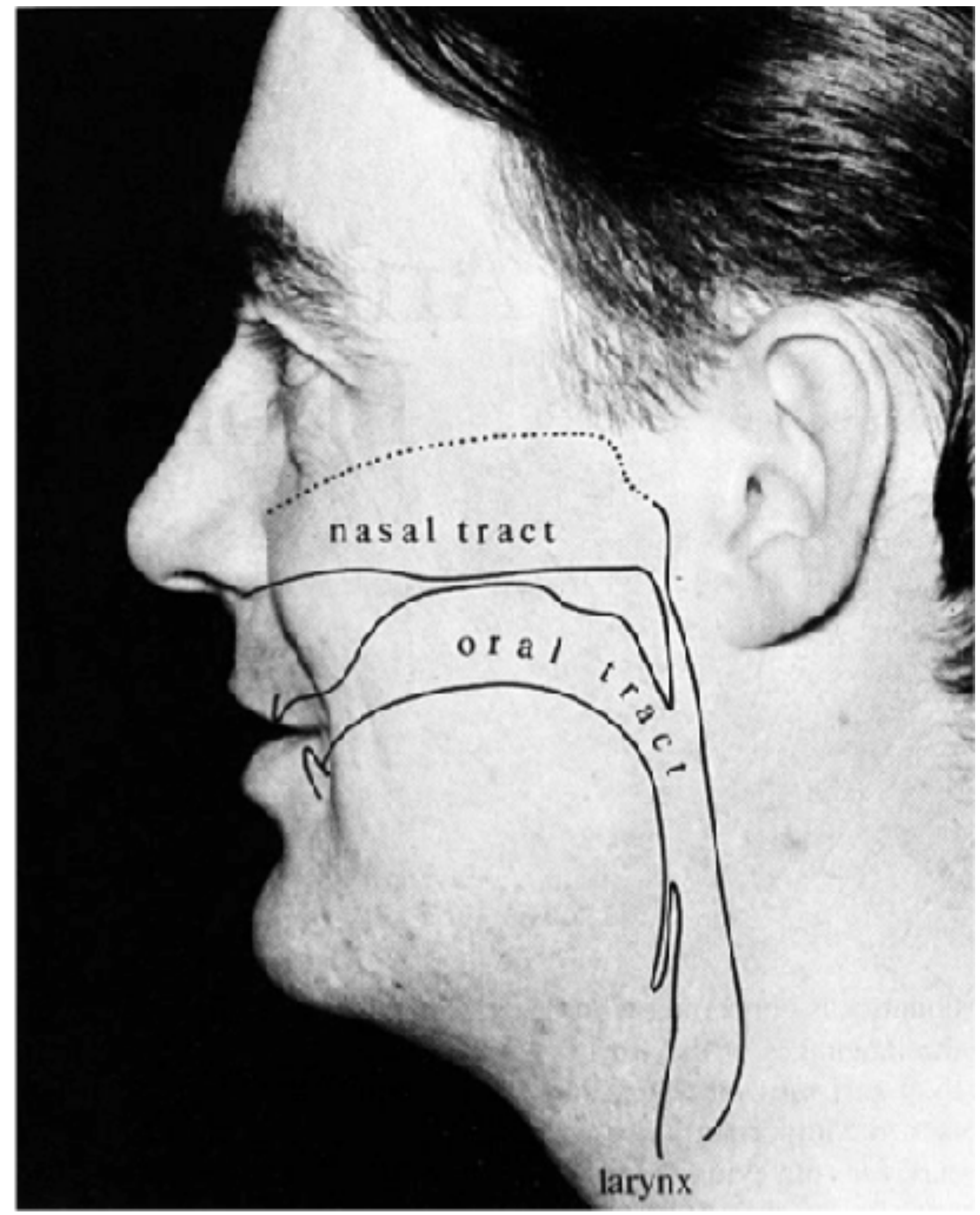
SOURCE-FILTER THEORY

- **Input = (Voice) source**
  - Result: harmonics of voice source
- **Filter = Vocal tract**
  - Result: harmonic amplitudes get modulated
- **Output = Speech**
  - Result: combined effects of source and filter

VOCAL TRACT

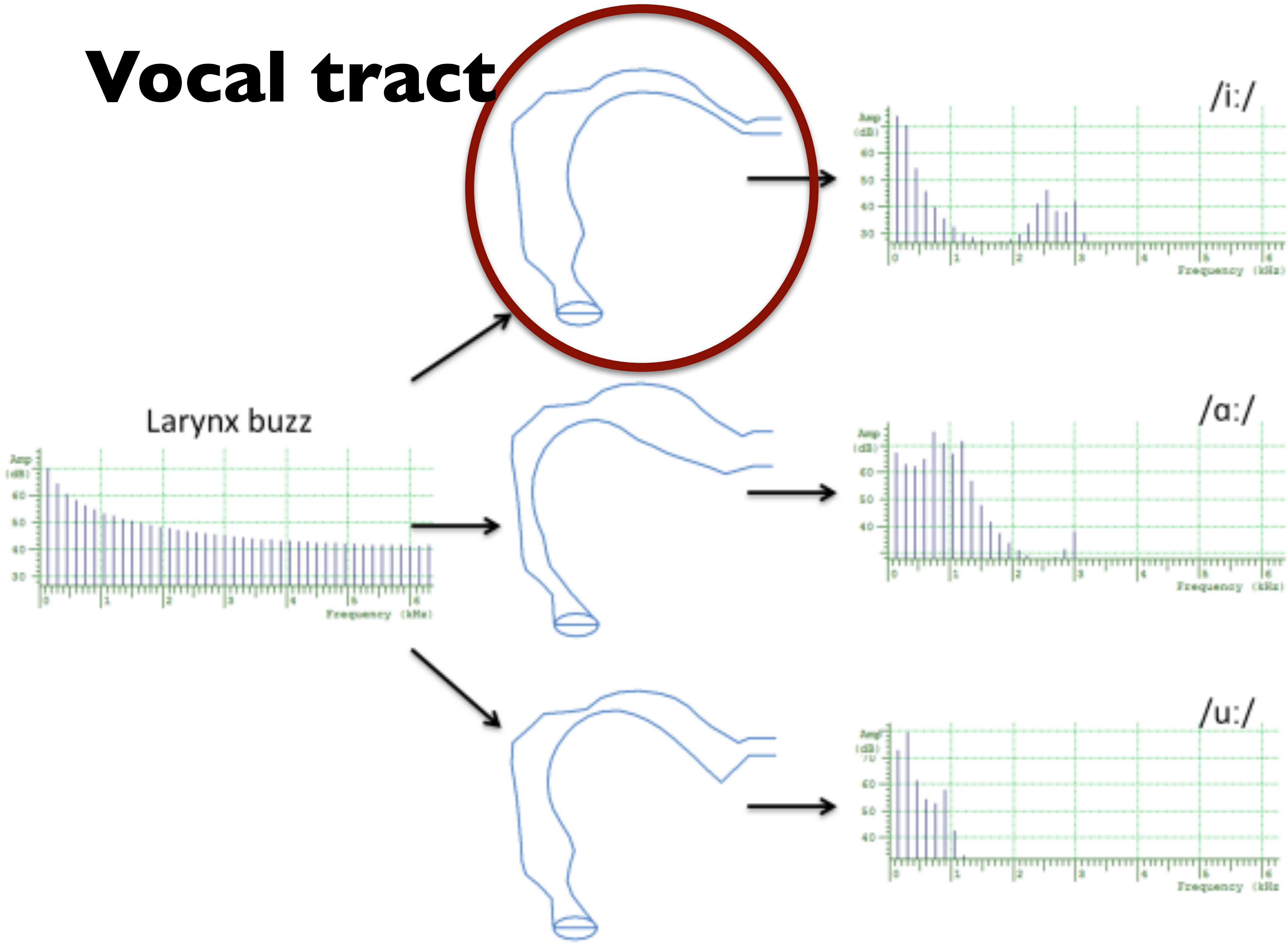


From glottis to lips!



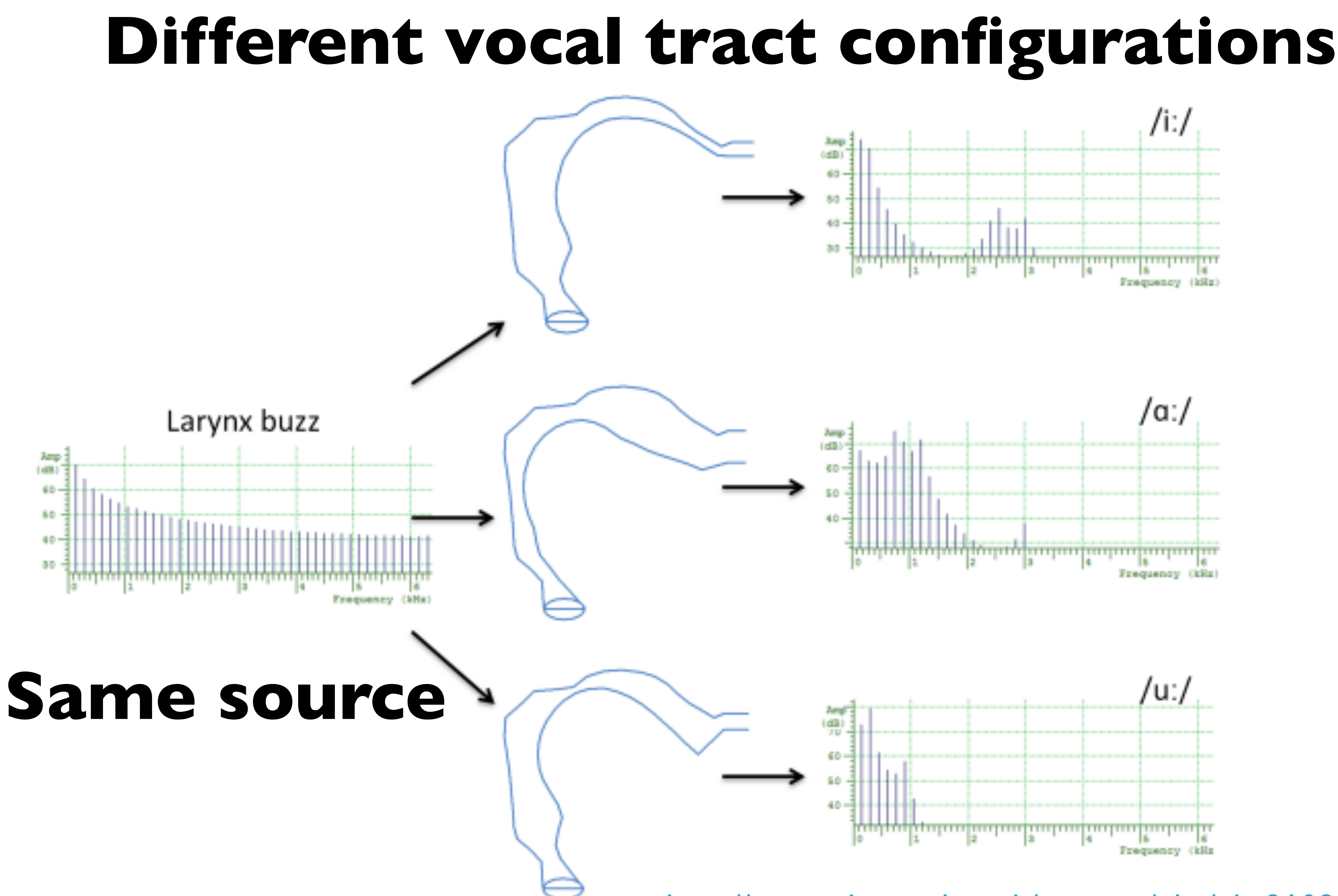
Ladefoged and Johnson (2010), p. 5

SOURCE–FILTER THEORY: A DIAGRAM



http://www.phon.ucl.ac.uk/courses/plin/plin2108/week5.php

DIFFERENT SPEECH SOUNDS HAVE DIFFERENT VOCAL TRACT CONFIGURATIONS



http://www.phon.ucl.ac.uk/courses/plin/plin2108/week5.php

SOURCE–FILTER THEORY: A SCHEMATIC

source **filter** **output**

waveform

spectrum

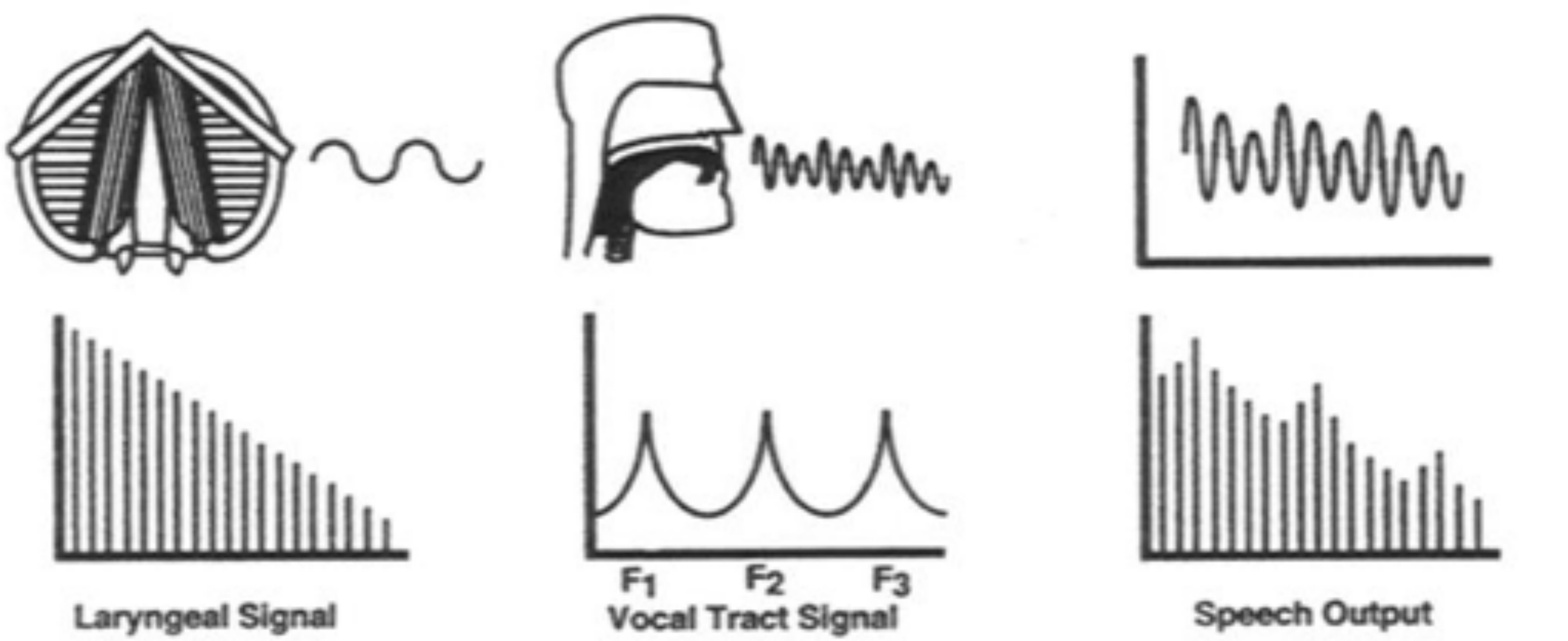


FIGURE 4-12 Effects of Acoustic Filtering

Fucci and Lass

SOURCE-FILTER THEORY: A SCHEMATIC

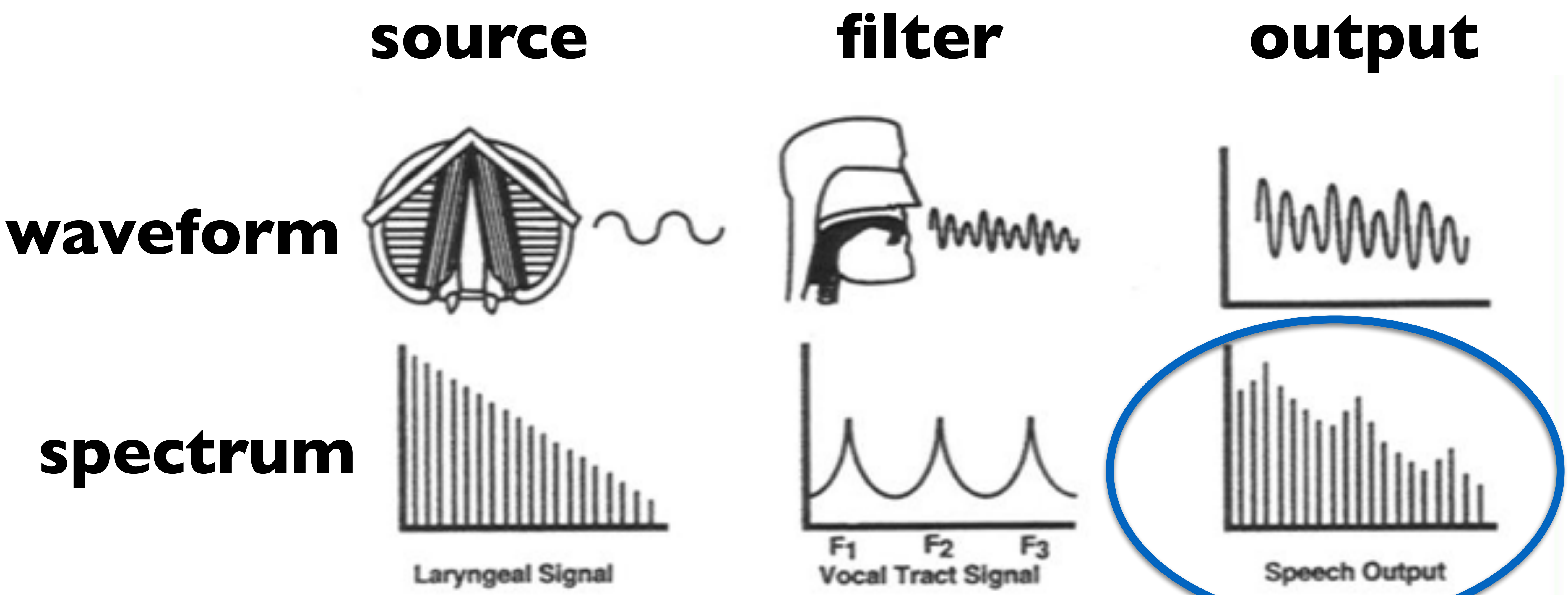


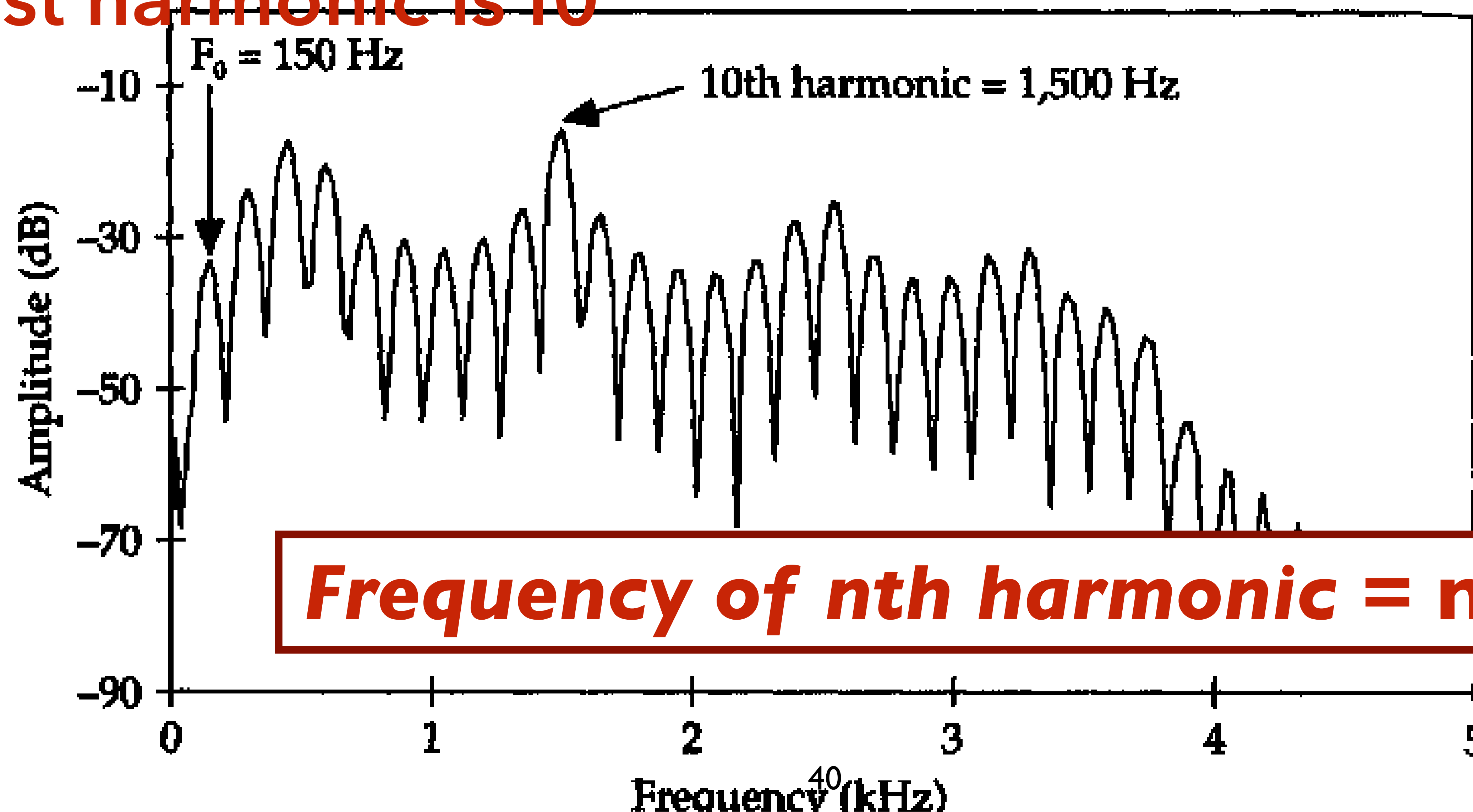
FIGURE 4-12 Effects of Acoustic Filtering

Fucci and Lass

OUTPUT SPECTRUM

- Output = Speech
- Result: combined effects of source and filter

1st harmonic is f0



Frequency of nth harmonic = n * f0

THE SOURCE:
THE VIBRATING LARYNX

SOURCE–FILTER THEORY: A SCHEMATIC

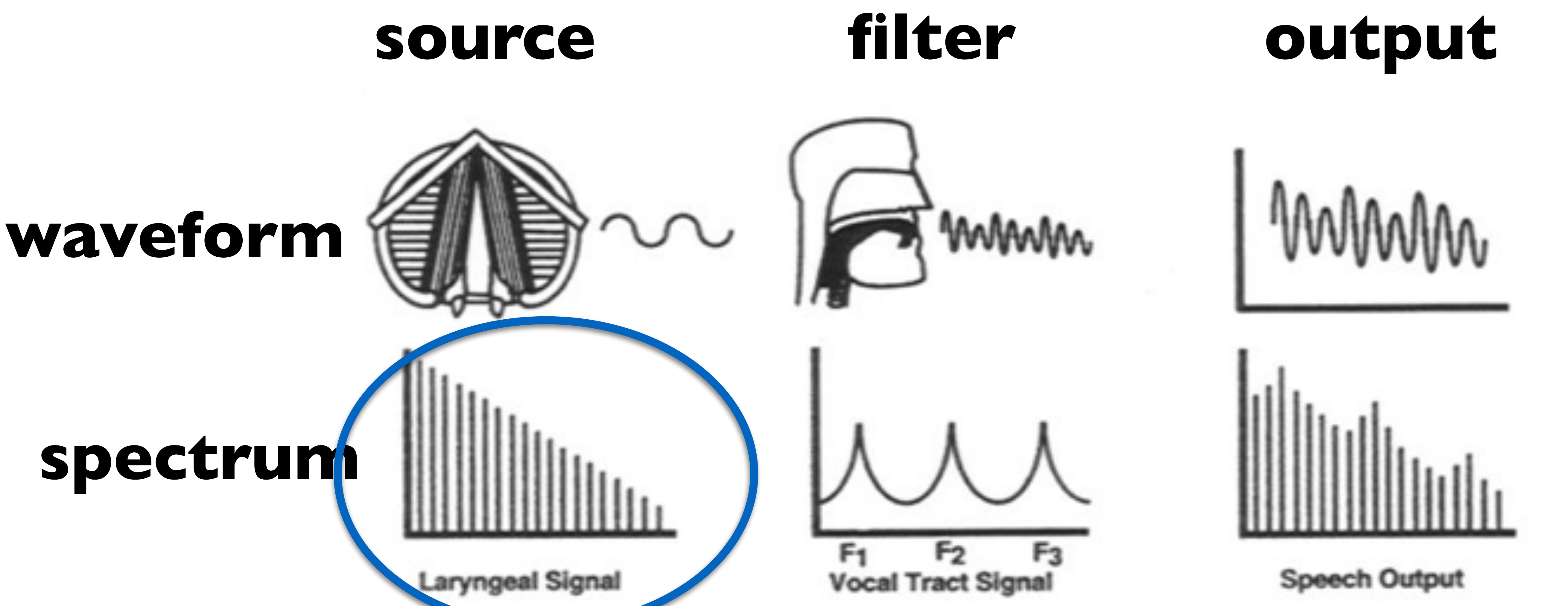


FIGURE 4-12 Effects of Acoustic Filtering

Fucci and Lass

AMPLITUDE OF HARMONICS IN SOURCE

- Can get by recording at the larynx or inverse filtering of speech
- Amplitudes of source harmonics generally decrease as frequency goes up (about 3 dB fall per octave)
- Rate of decrease depends on phonation quality (creaky, breathy, etc.)

ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

http://www.linguistics.ucla.edu/faciliti/facilities/physiology/EGG.htm

ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Glottal waveform

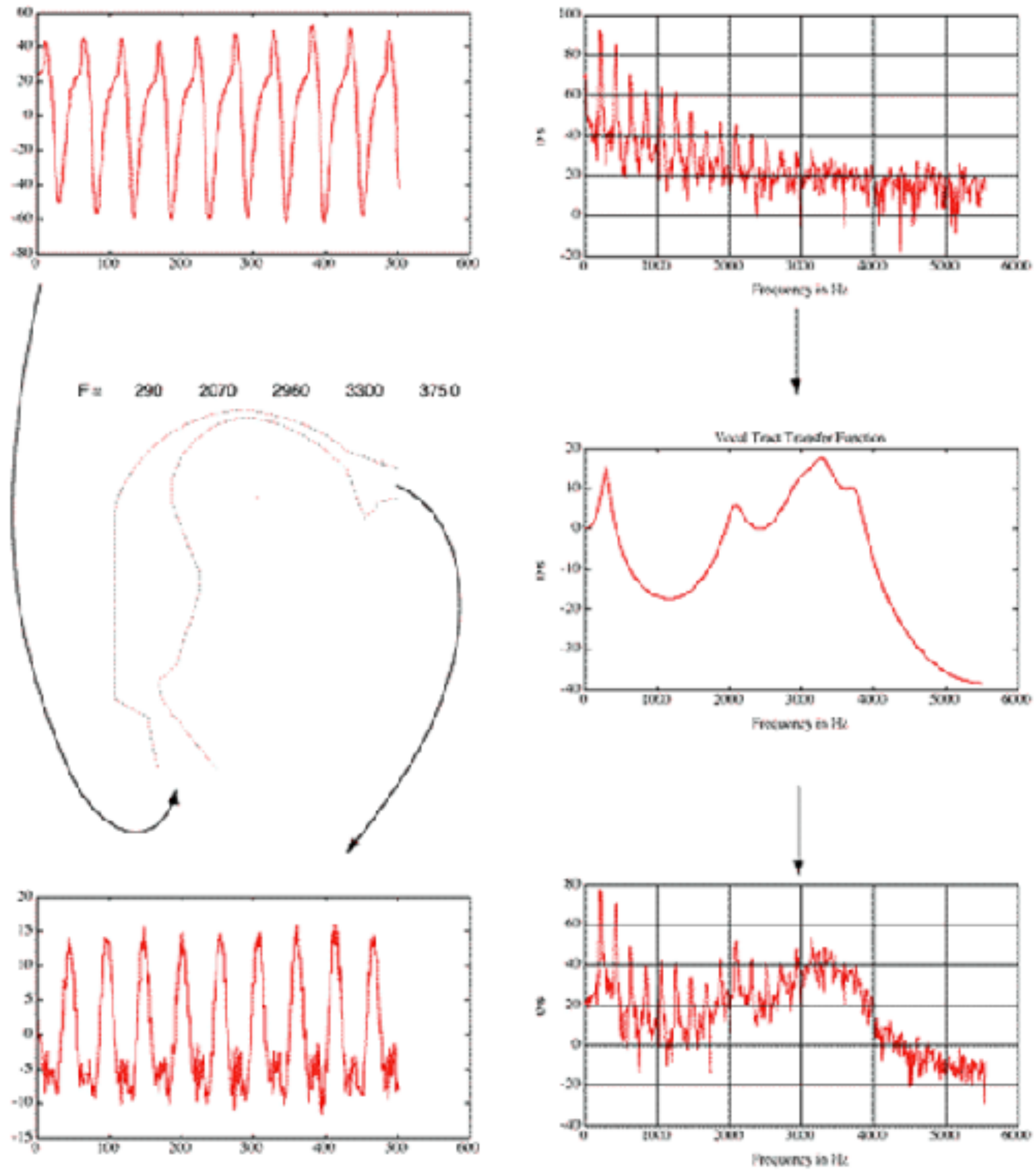
Glottal spectrum

Vocal tract shape for [i]

Vocal tract transfer function

Filtered waveform: [i]

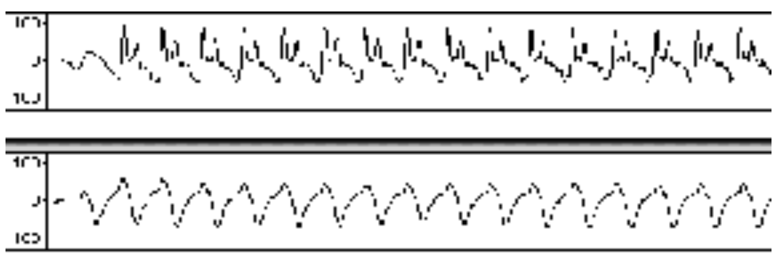
Filtered spectrum: [i]



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

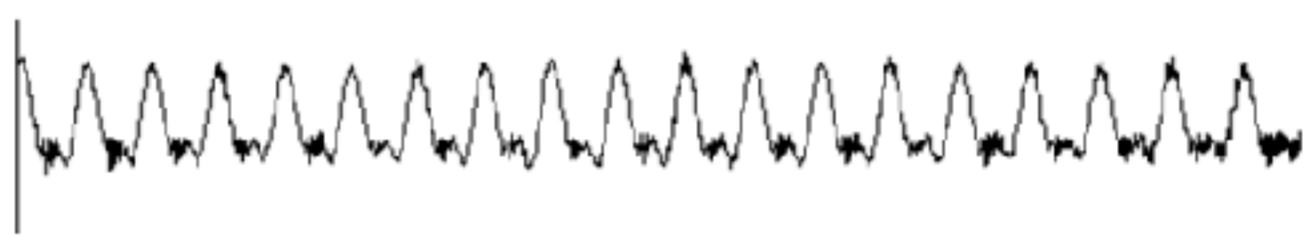
ELECTROGLOTTOGRAPHY: RECORDING AT THE LARYNX

Mic

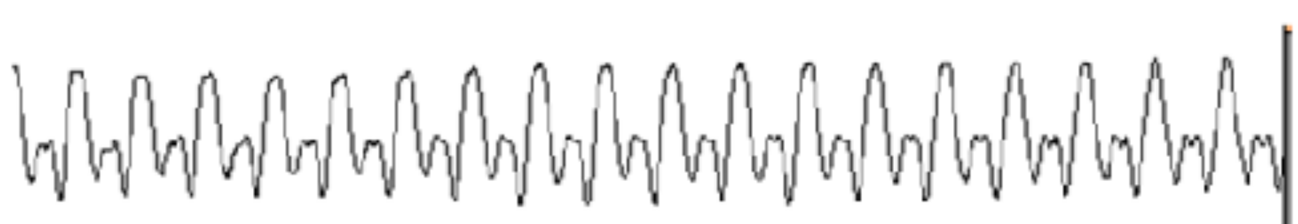


EGG source

[a]-filtered



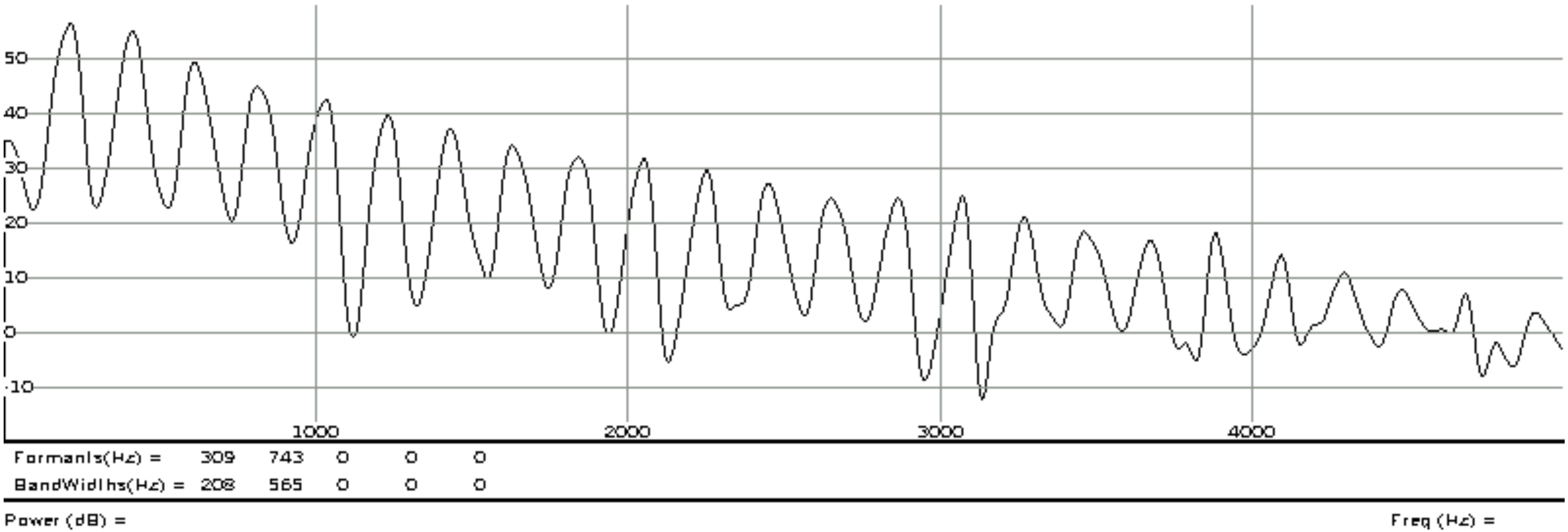
[i]-filtered



[u]-filtered

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

INVERSE FILTERED SPECTRUM EXAMPLE



THE FILTER:
THE VOCAL TRACT

THE VOCAL TRACT AS A FILTER

- Configuration of vocal tract acts on amplitude of harmonics from voice source
- **No new harmonics are added nor or their frequencies changed!**
- Some harmonics get stronger, some get weaker
  - Particular vocal tract configuration has particular resonance frequencies (formants); if these are close to the frequencies of some harmonics, those harmonics get strengthened

SYMPATHETIC VIBRATION: TUNING FORKS

https://youtu.be/zWKiWaiM3Pw

Other fun resonance videos at:
http://blog.prosig.com/2011/09/20/5-videos-that-explain-resonance/

EXAMPLE: RESONANCE

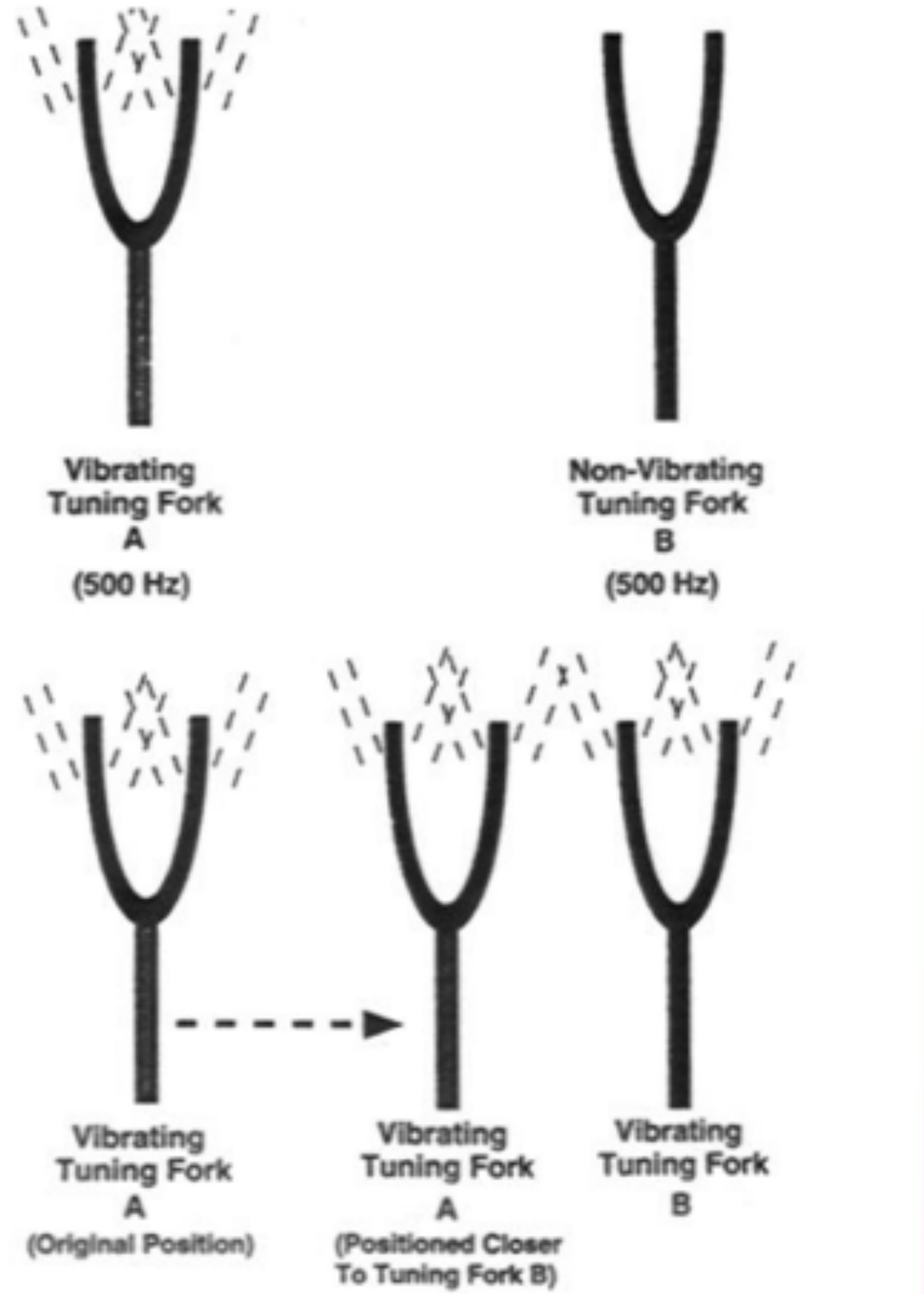


FIGURE 4-2 Sympathetic Vibration

B resonates to A if B's vibrations make A vibrate too.

Effect: B has more energy than it did

The closer B's natural frequency is to A's the stronger the vibrations of B

RESONANCES FOR DIFFERENT VOCAL TRACT CONFIGURATIONS

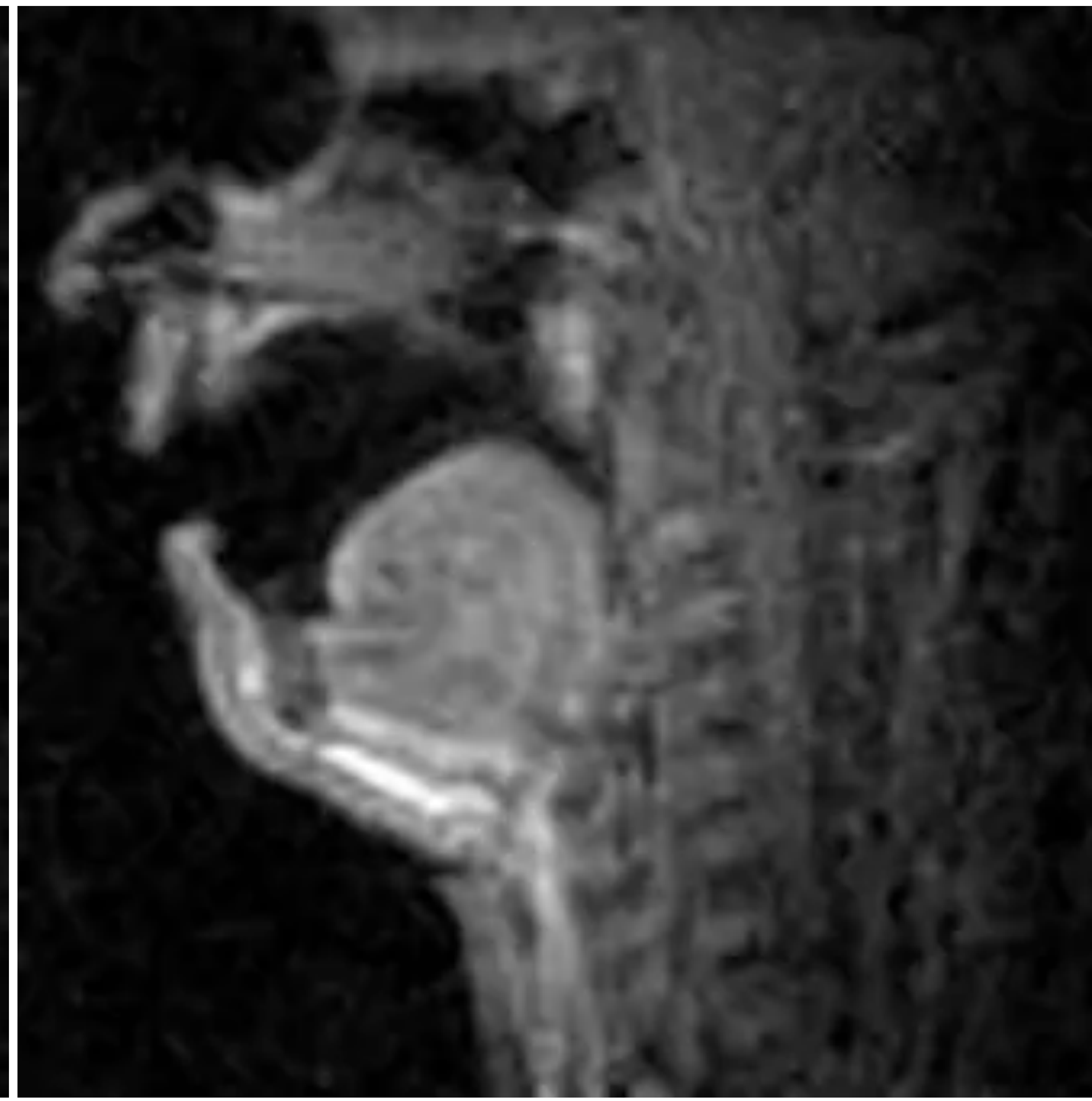
- Resonances depend on size and shape of airway
  - Can be approximated as multitube models, with connected Helmholtz resonators
  - Helmholtz resonances are the formants
- See article by Sandberg on *The Acoustics of the Singing Voice* for further reading

http://faculty.washington.edu/losterho/Sundberg.pdf

MRI VIDEOS OF VOWEL PRODUCTIONS!

front close unrounded vowel

back open unrounded vowel

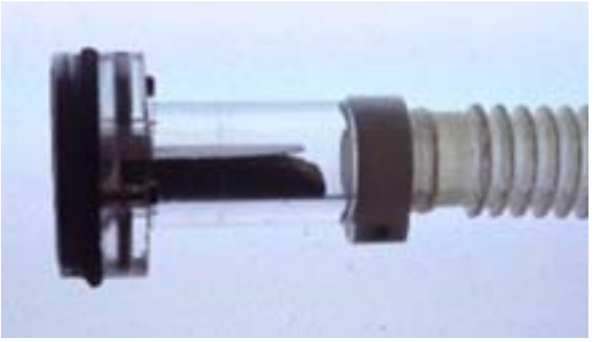


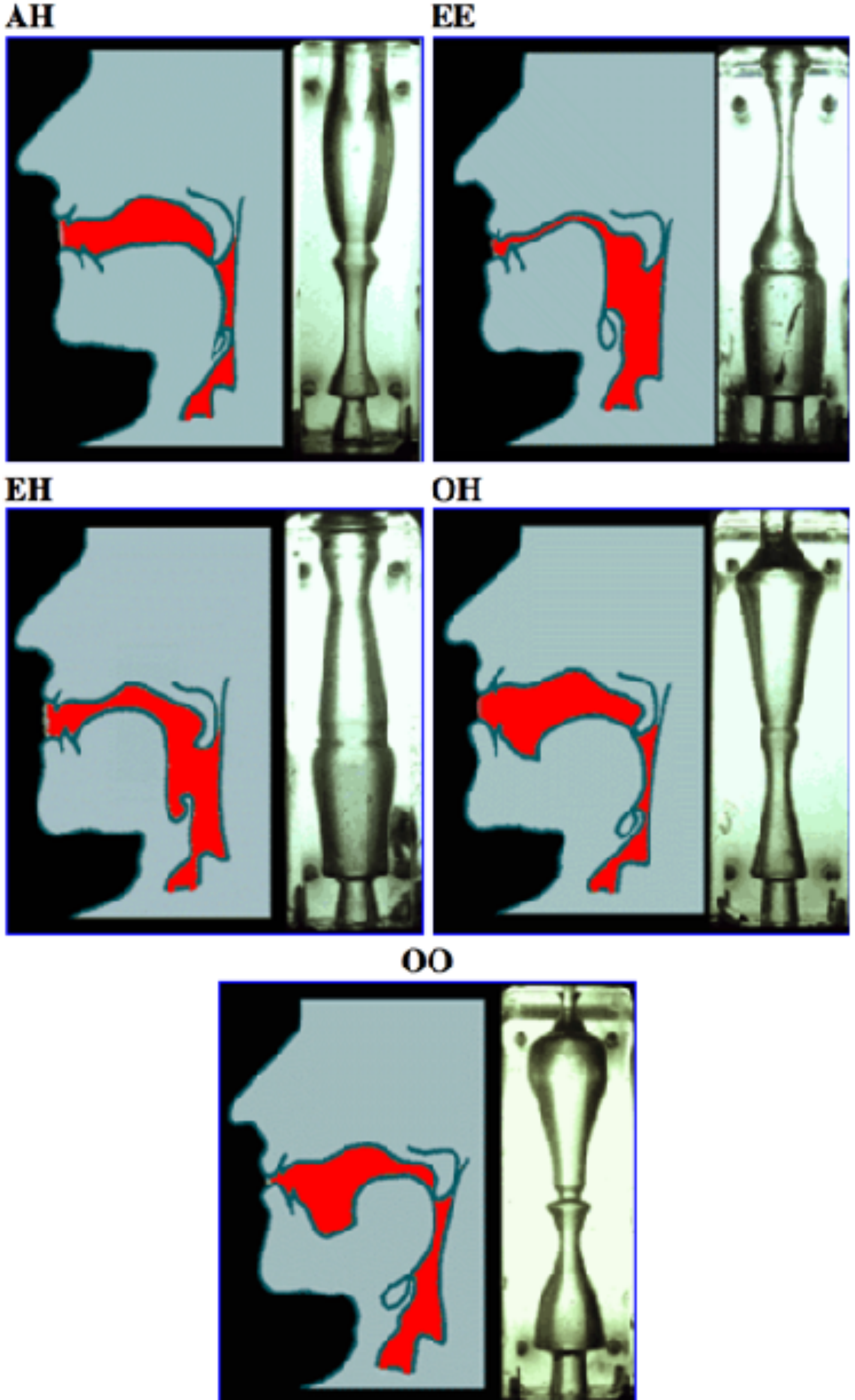
http://sail.usc.edu/span/rtmri_ipa/pk_2015.html

HELMHOLTZ RESONATORS AND FORMANTS

http://www.exploratorium.edu/exhibits/vocal_vowels/vocal_vowels.html

duck call sound source: vibrating reed



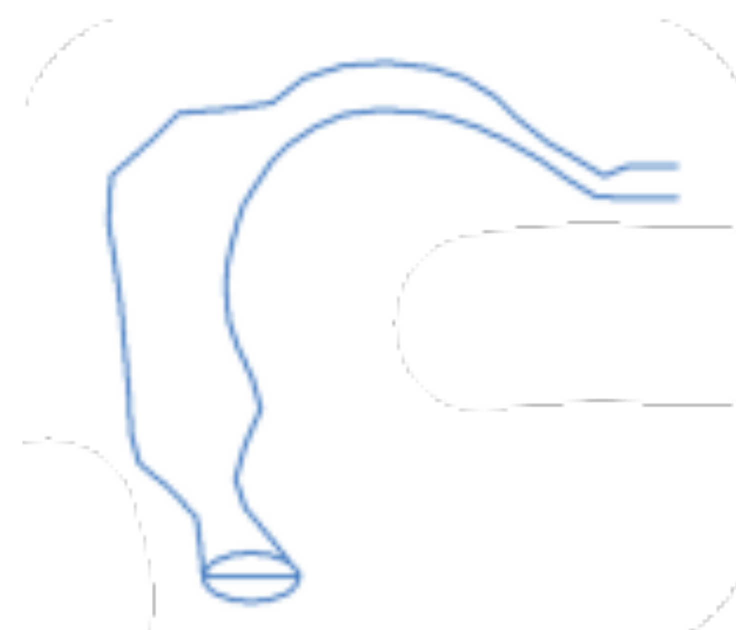


SOME POINTS OF CONFUSION

POINT OF CONFUSION: VOCAL FOLD VS. VOCAL TRACT LENGTH

- **Vocal tract length** has a systematic effect on formant frequencies (vocal tract resonances)
  - As vocal tract length increases, formant frequencies go down: inverse relation

Who has higher formant frequencies?
A baby or an adult?



POINT OF CONFUSION: VOCAL FOLD VS. VOCAL TRACT LENGTH

- **Vocal fold length** has a systematic effect on fundamental frequency (property of voice source)
  - Natural vocal fold length accounts for some portion of individual differences in fundamental frequencies, i.e. differences in f0 between individuals

http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/influence.html

POINT OF CONFUSION: VOCAL FOLD VS. VOCAL TRACT LENGTH

Vocal fold length

If we assume that the vocal folds are 'ideal strings' with uniform properties, their Fo is governed by this equation:

Longitudinal stress

Fundamental frequency

$$F_0 = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}$$

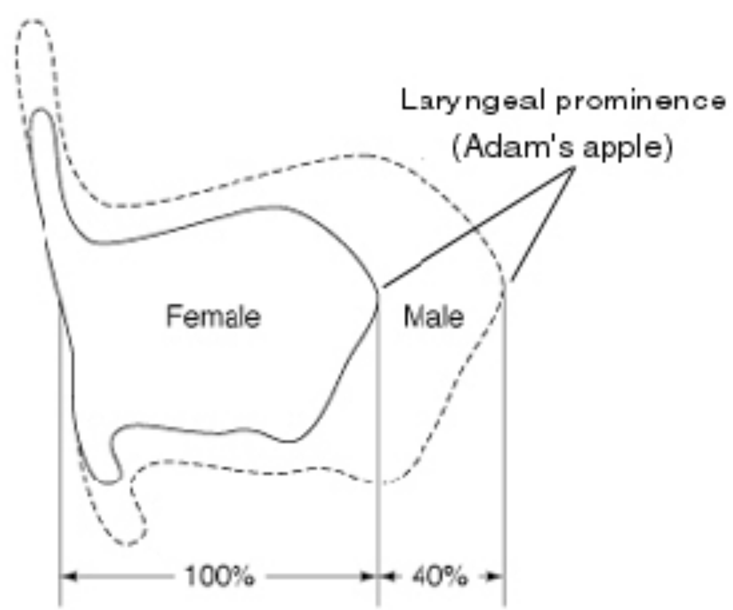
Length of vocal folds

Tissue density

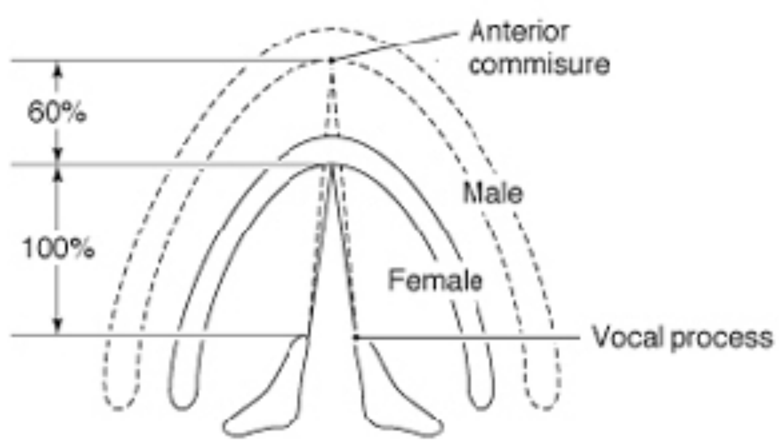
The key variable here is the **length** of the part of the vocal folds that is actually in vibration, which we call **effective vocal fold length**. If we examine this quantity for men and women, we find that men have a 60% longer effective fold length than women, on average, which fully accounts for the difference we see in Fo between the sexes.

http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/influence.html

POINT OF CONFUSION: VOCAL FOLD VS. VOCAL TRACT LENGTH



(a)



(b)

On average, men have 60% longer effective vocal fold length than women.

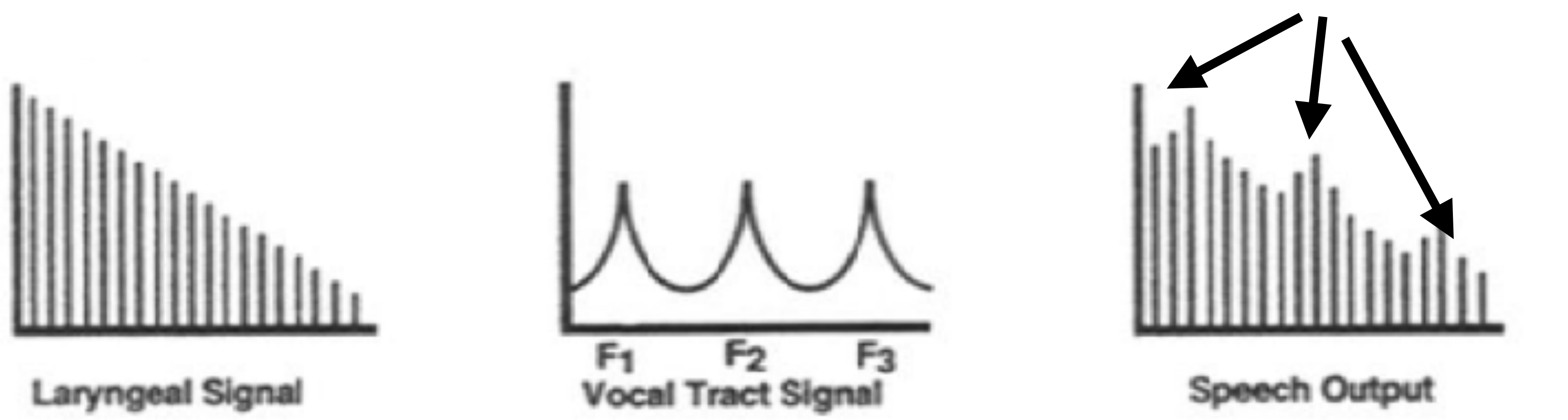
What does this tell you about f0 on average, compared for men vs. women?

http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/influence.html

POINT OF CONFUSION: FORMANTS VS. F0

Formants: vocal tract resonances that "filter" harmonic amplitudes from the voice source: we indirectly see what the formants are from their effects on the harmonics in speech

*Harmonics **boosted** around F1, F2, and F3*

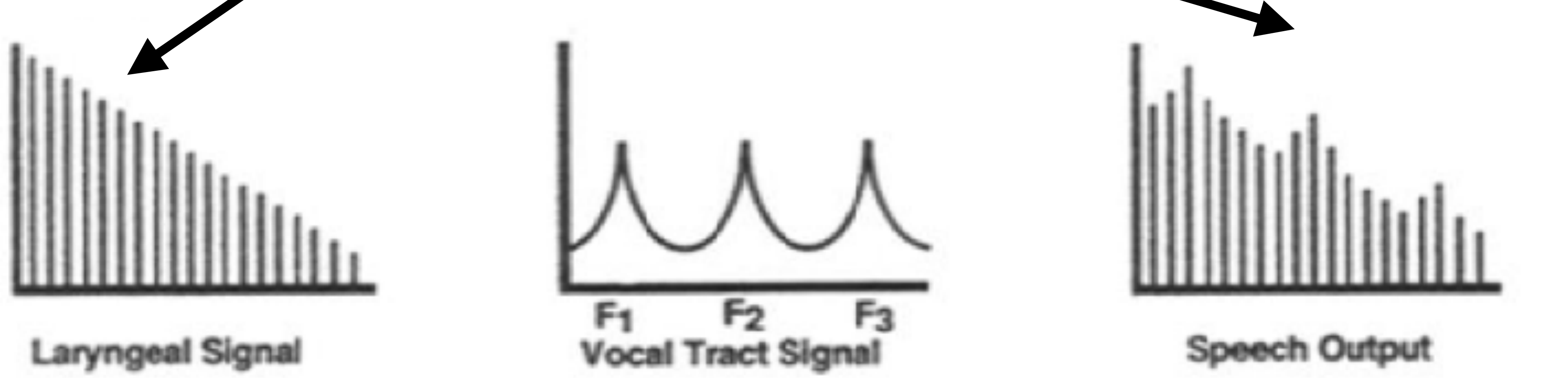


formants F1, F2, F3

POINT OF CONFUSION: FORMANTS VS. F0

Fundamental frequency: a property of the voice source, the rate of vocal fold vibration, the lowest harmonic (the "first" harmonic), also the spacing between harmonics

f0 not affected by vocal tract configuration: spacing between harmonics unaffected by vocal tract filtering*



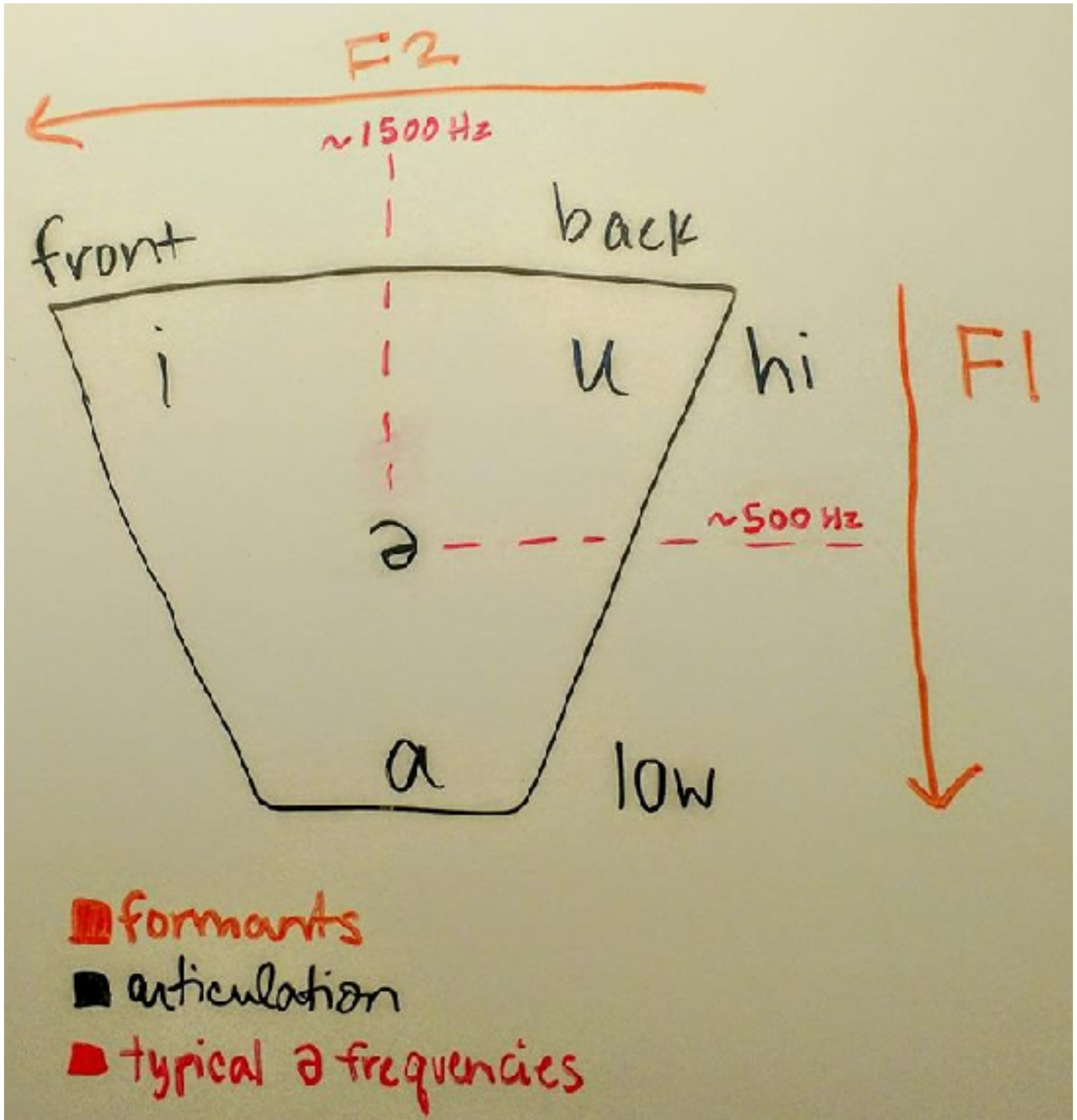
*to a first approximation

EXERCISE: INDEPENDENCE OF F0, FORMANTS

- Demonstrate that you can keep f0 constant, while changing formants

- Demonstrate that you can keep formants constant, while changing f0

HANDY FORMANT CHART!



http://www.facebook.com/groups/ling5

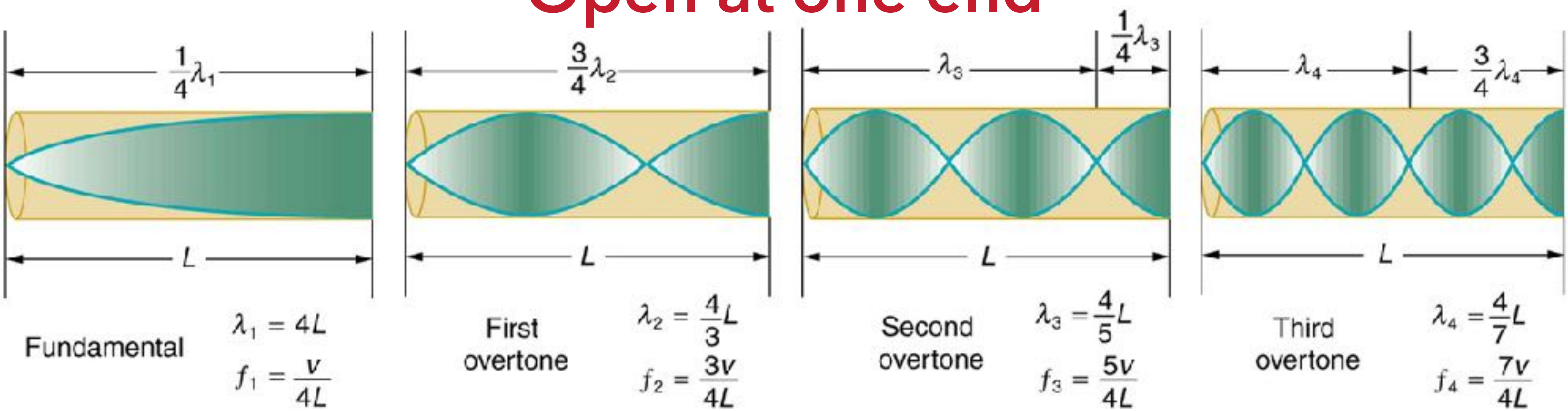
Ivy Hauser

https://blogs.umass.edu/ihauser/

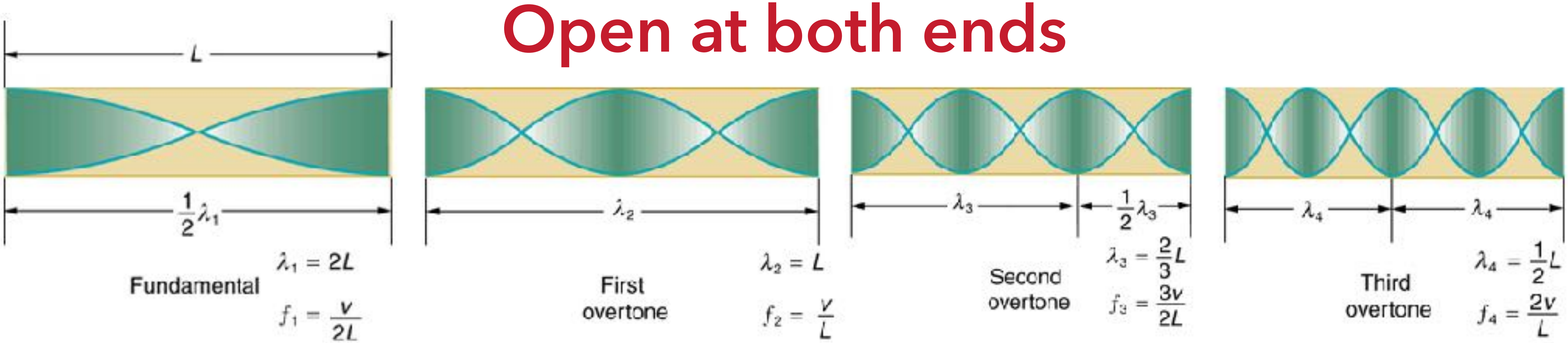
TUBE RESONANCE IN THE VOCAL TRACT

TUBE RESONANCE: OPEN AT ONE END

Open at one end

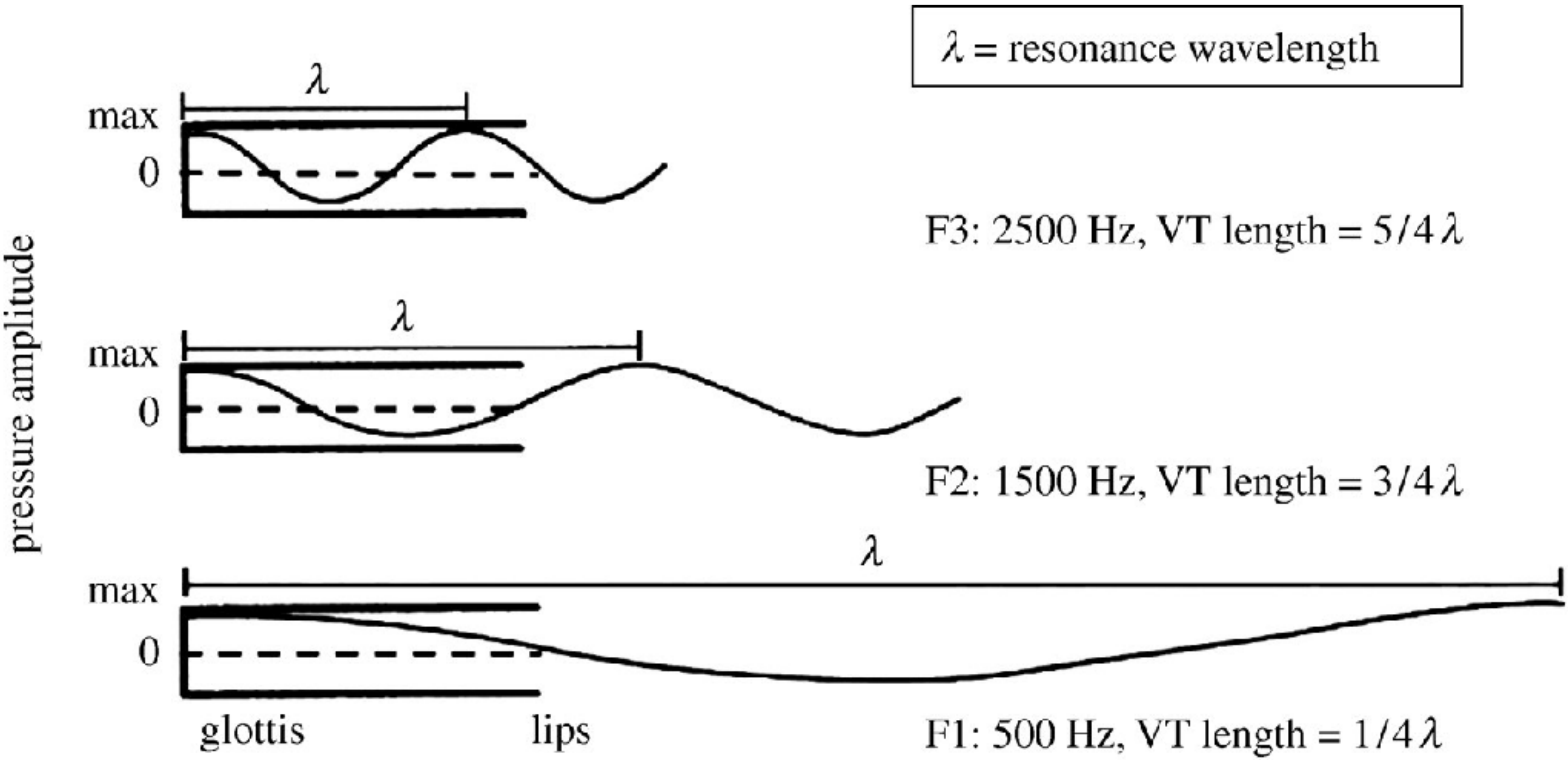


Open at both ends



http://philschatz.com/physics-book/contents/m42296.html#import-auto-id1379919

NATURAL RESONANCES FOR SCHWA



http://rstb.royalsocietypublishing.org/content/363/1493/965

ESTIMATING YOUR VOCAL TRACT LENGTH!

The resonances of an unconstricted tube or pipe are a function of the length of the tube.

$$f = \frac{nc}{4L} \text{ for } n = 1, 3, 5, \ldots$$

f = formant frequency in Hz
c = speed of sound 34,000 cm/s
L = length of vocal tract in cm

So the lowest formant frequency in a 17 cm. vocal tract is:

$$f = \frac{c}{4L}$$

= *34,000 / 4 * 17*
= *500 Hz*

And the spacing between formants is:

$$\Delta f = \frac{c}{2L}$$ (always twice the lowest f)

= *1000 Hz*

http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter/SFb.html#VTL

ESTIMATING YOUR VOCAL TRACT LENGTH!

- Record yourself producing a schwa-type vowel /ə/, and while continuing to phonate, slowly raise the jaw a bit to a higher vowel, then lower again to schwa. Now glide smoothly to an /ɛ/-type vowel (as in "head'), and back to schwa. Save this recorded file as schwa_YOURINITIALS.wav, e.g., schwa_KY.wav

- Create a textgrid. Examine the spectrogram of your recording, and select a moment in time for labeling where the formants appear to be fairly equally spaced in frequency. Measure the values of F1-F3 as in Part I and record their values in the textgrid. Calculate the F2-F1 and F3-F2 at this point. Take the average of these as the inter-formant distance. Save this TextGrid as schwa_YOURINITIALS.TextGrid, e.g., schwa_KY.TextGrid.

- Upload both files to this folder: https://drive.google.com/open?id=1icPQn8vZ214lV70i8BJI_-YeAvucSYUG (you may need to be signed into your UMass account to access the folder)

http://sail.usc.edu/~lgoldste/General_Phonetics/Week10/Formant_Analysis/index.html