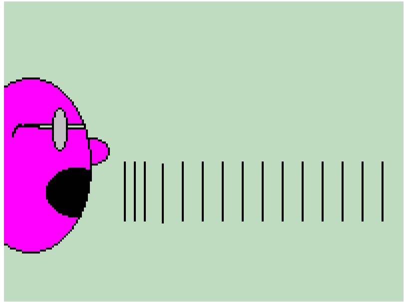
Acoustic Phonetics

How speech sounds are physically represented

Chapters 12 and 13

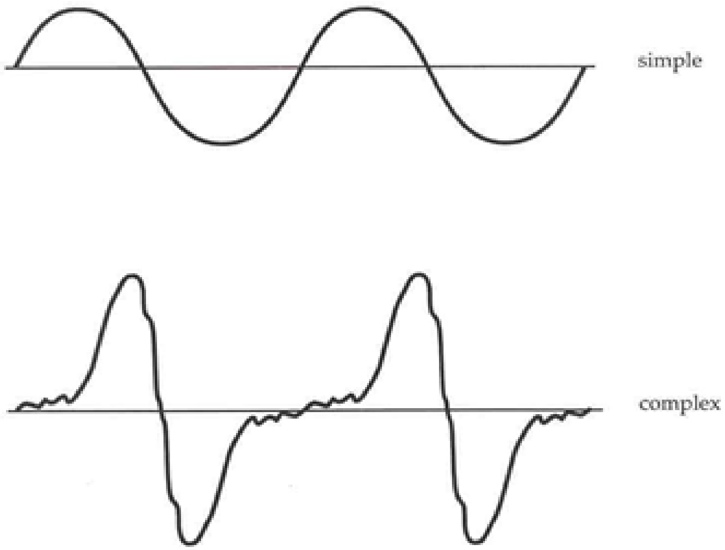
Sound

- Energy
- Travels through a medium to reach the ear
- Compression waves



Information from *Phonetics for Dummies.* William F. Katz. “Making Waves: An Overview of Sound.” 2013.

Periodic waves



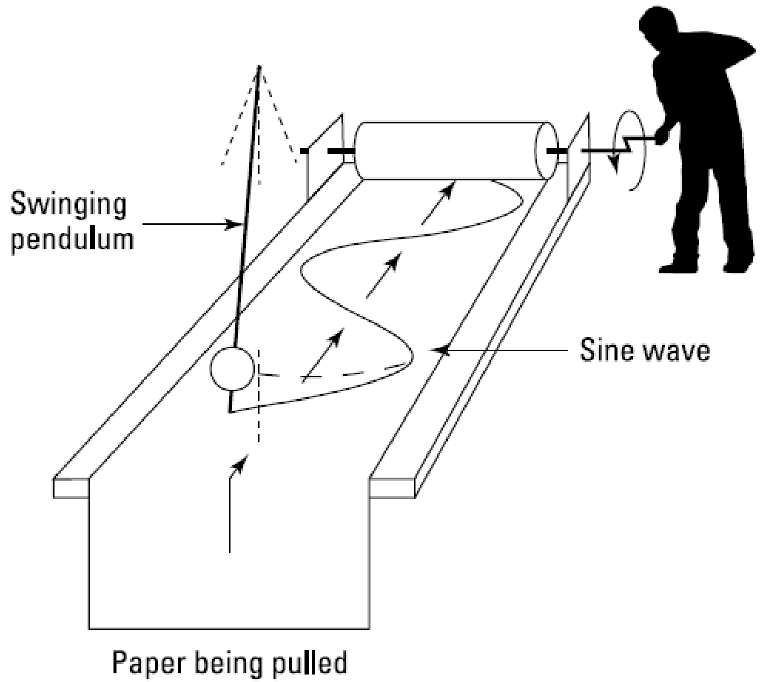
- Simple (sine; sinusoid)
- Complex (actually a composite of many overlapping simple waves)

Sinusoid waves

- Simple periodic motion from perfectly oscillating bodies
- Found in in nature (e.g., swinging pendulum, sidewinder snake trail, airflow when you whistle)
- Sinusoids sound ‘cold’ (e.g. flute)

Information from *Phonetics for Dummies.* William F. Katz. “Making Waves: An Overview of Sound.” 2013.

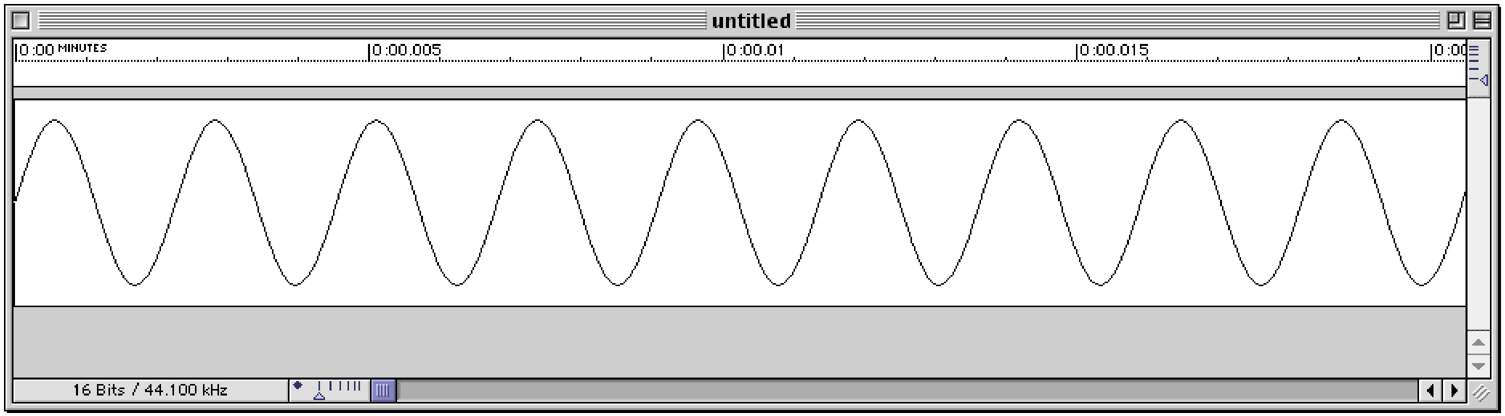
Let’s crank one out!

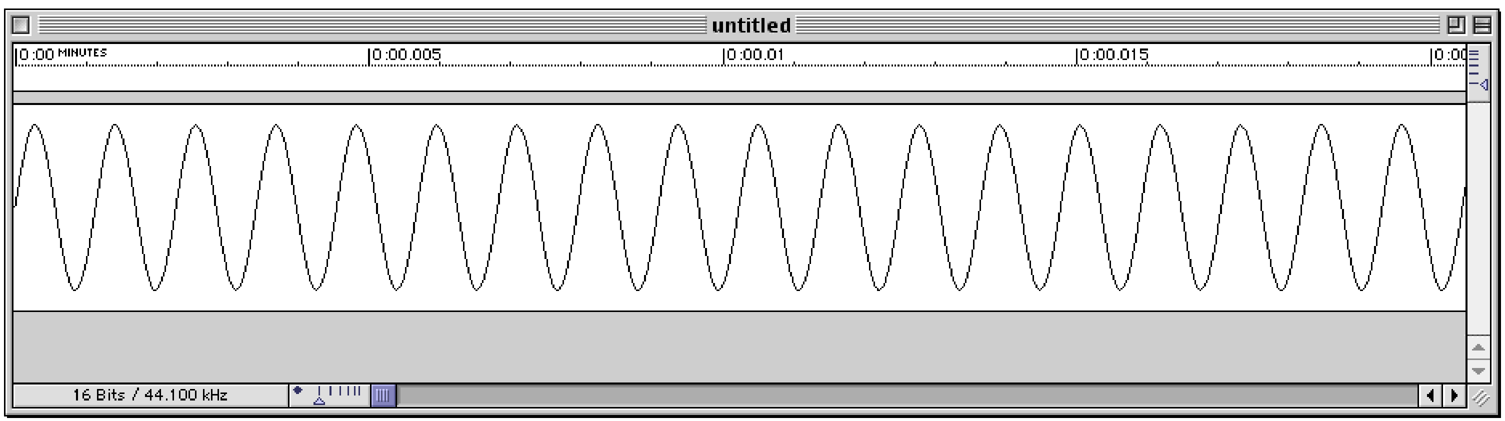


Pg. 175

Information from *Phonetics for Dummies*. William F. Katz. “Making Waves: An Overview of Sound.” 2013.

Frequency - Tones





Simple waves - key properties

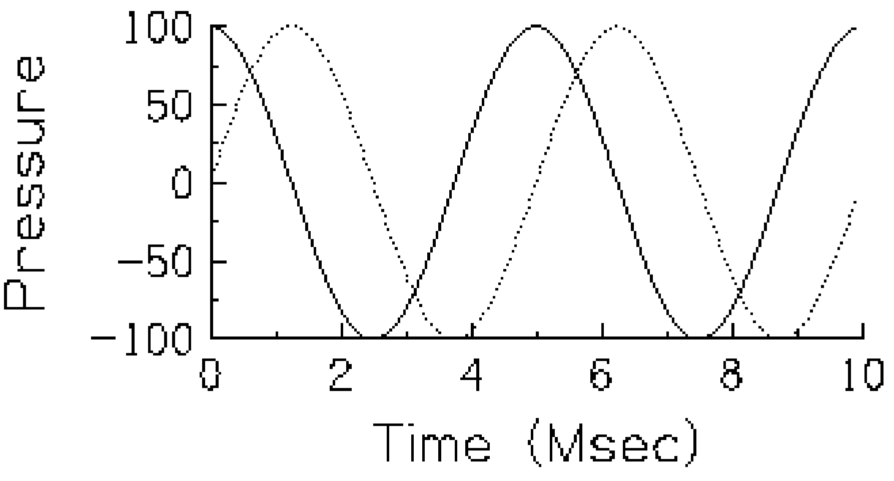
- Frequency = cycles per sec (cps) = Hz
- Amplitude – measured in decibels (dB), 1/10 of a Bell

(Note: dB is on a log scale, increases by powers of 10)

Information from *Phonetics for Dummies*. William F. Katz. "Making Waves: An Overview of Sound." 2013.

Phase

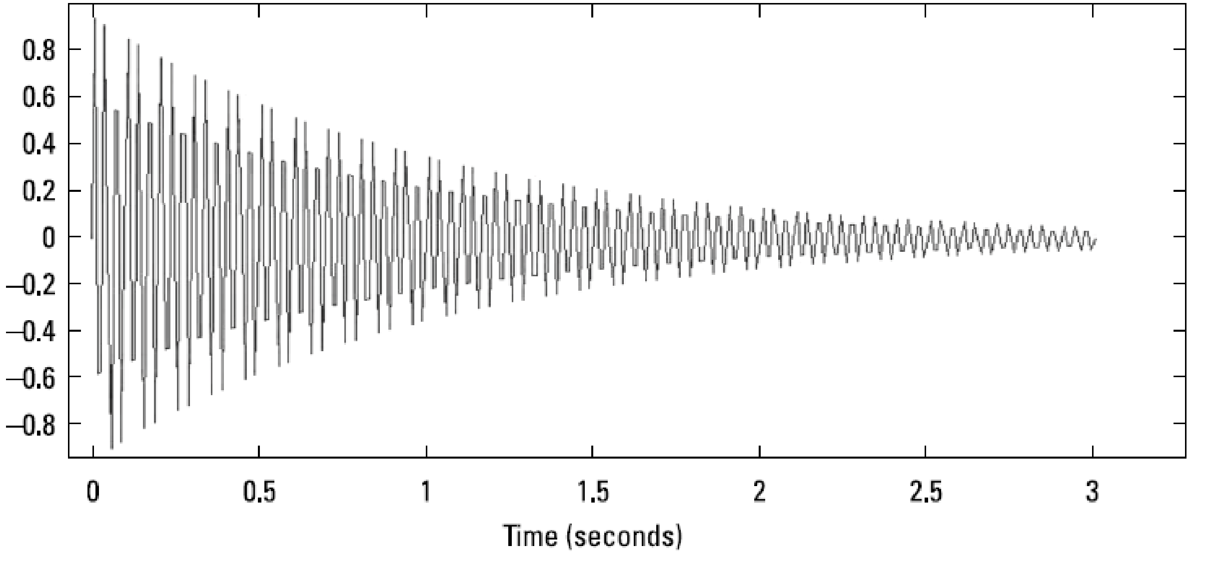
- A measure of the position along the sinusoidal vibration
- These two waveforms are slightly out of phase (approx. 90^0 difference)
- Used in sound localization



Information from *Phonetics for Dummies.* William F. Katz. "Making Waves: An Overview of Sound." 2013.

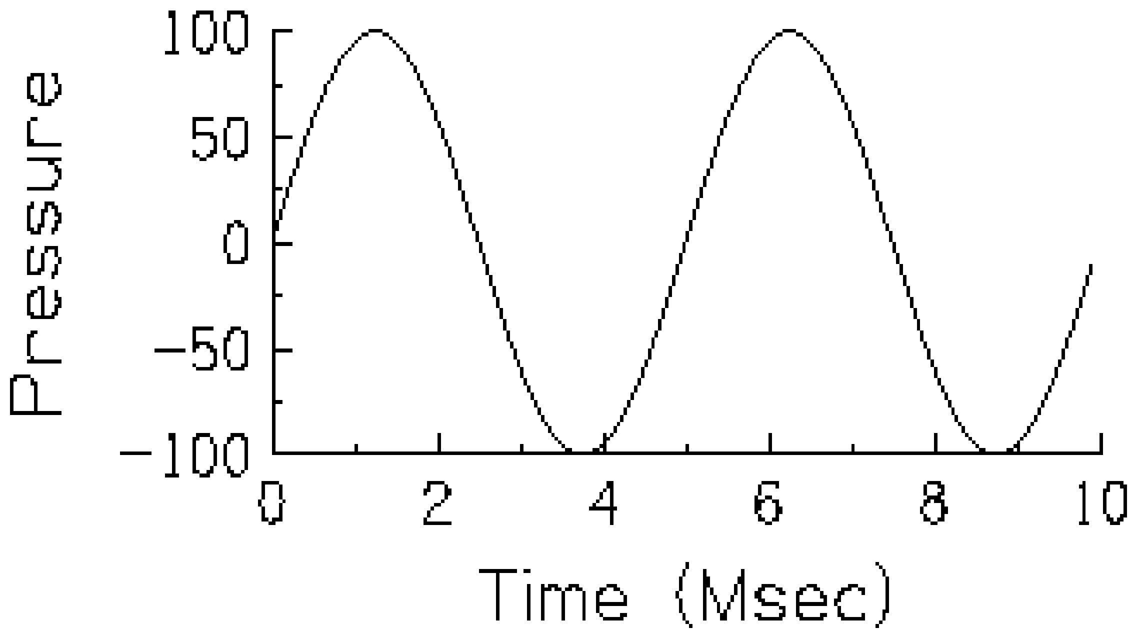
Damping

- Loss of vibration due to friction



Information from *Phonetics for Dummies.* William F. Katz. "Making Waves: An Overview of Sound." 2013.

Quickie Quiz!



Q: What is the frequency of this wave ?

HINT: It repeats twice in 10 msec

Answer:

- 200 Hz!

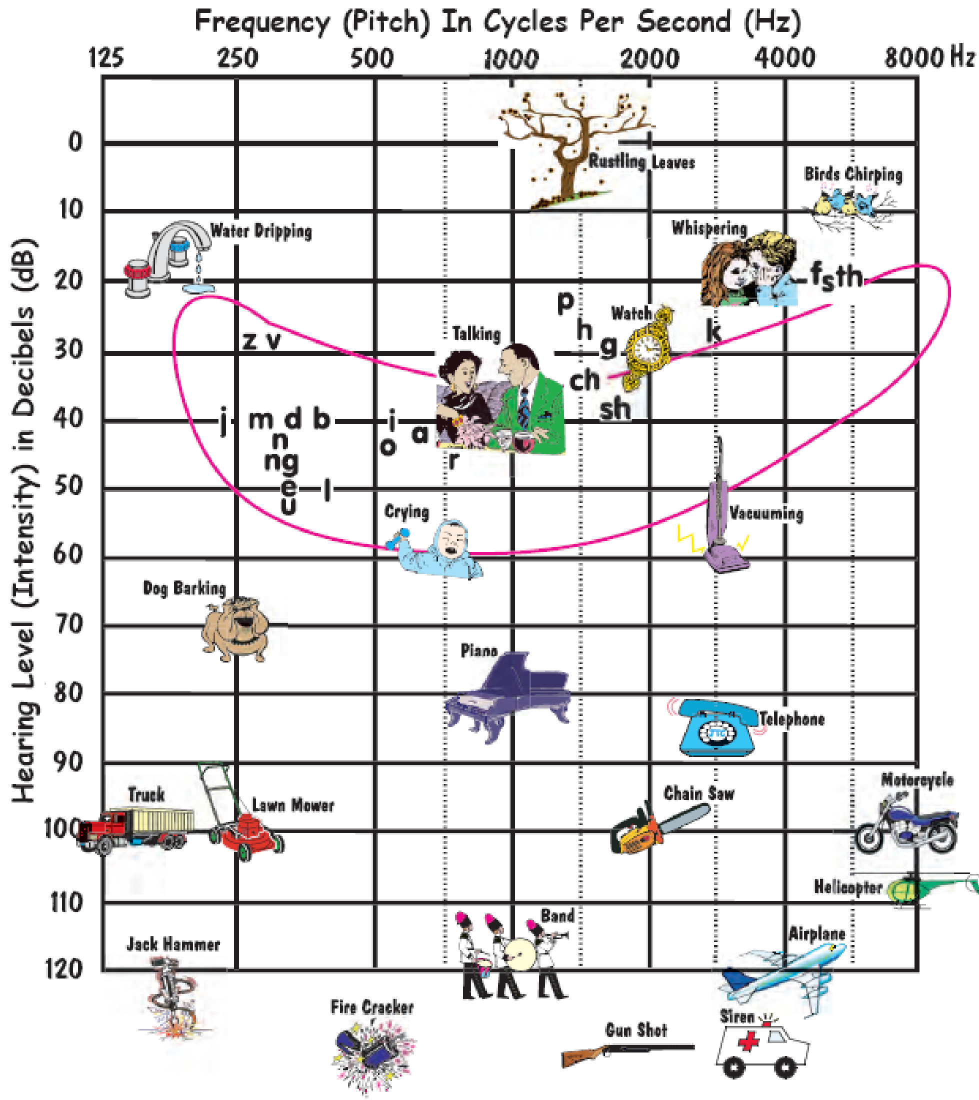
(2 cycles in .01 sec = 200 cps)

Physical vs. perceptual

PHYSICAL	PERCEPTUAL
• Fundamental frequency (F_0) →	"Pitch"
• Amplitude/ Intensity →	"Loudness"
• Duration →	"Length"

Information from *Phonetics for Dummies*. William F. Katz. "Making Waves: An Overview of Sound." 2013.

Frequency Spectrum of Familiar Sounds

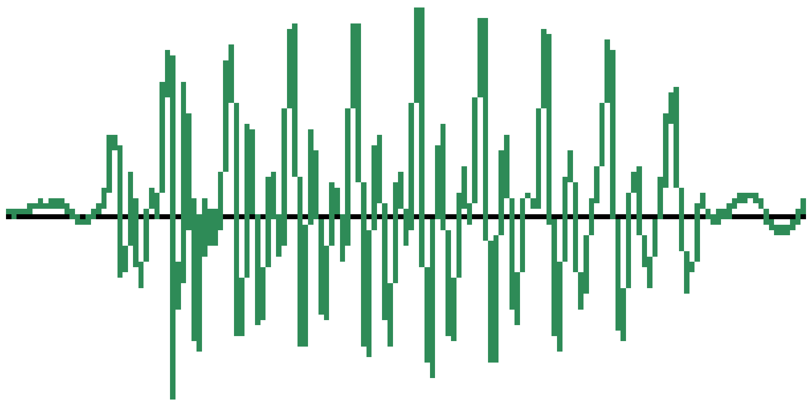


Speech is here

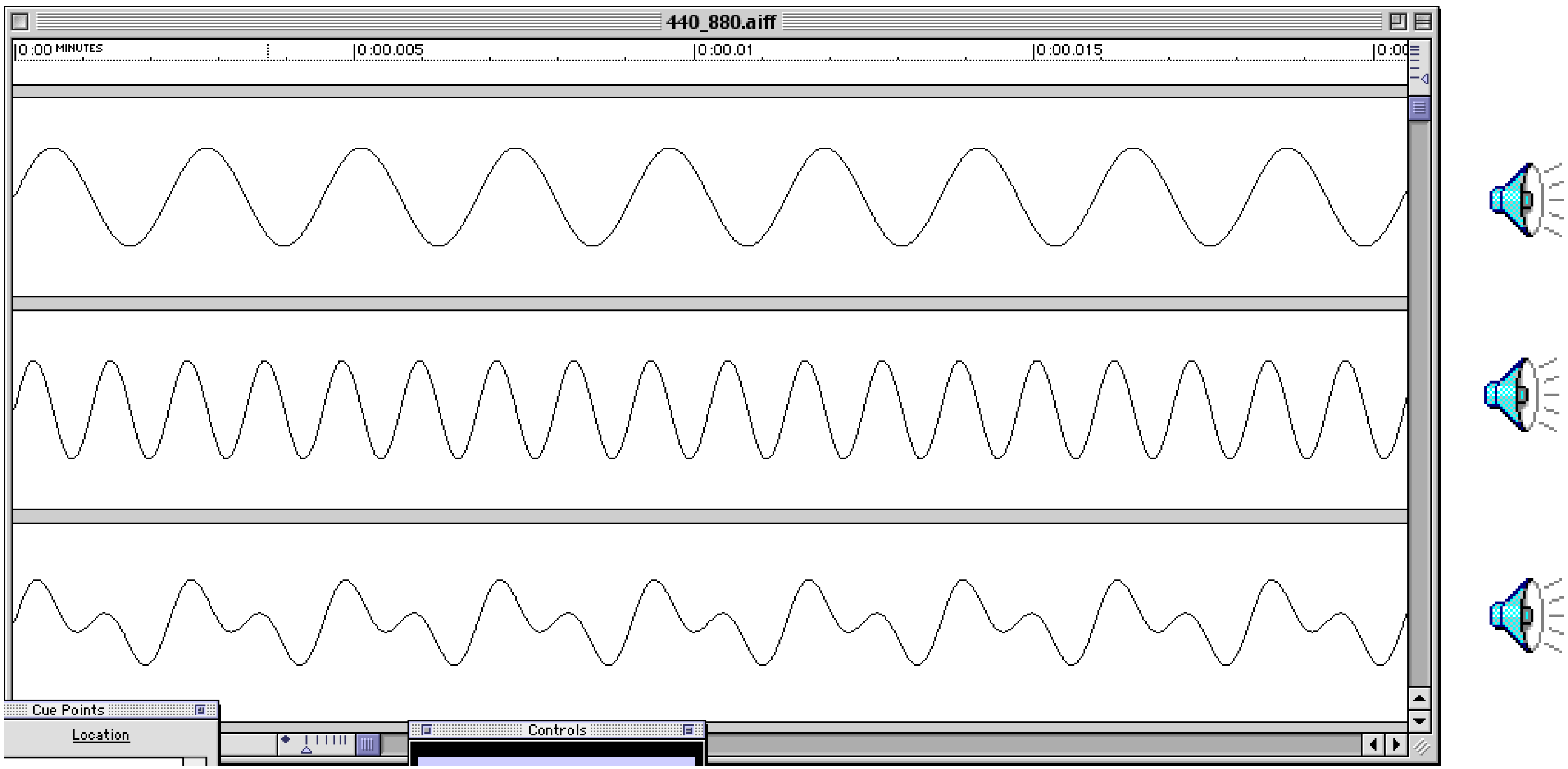
Image from *Fetal Hydrocephalus.* The Amazing Owen. "Great News from the Audiologist." March 23, 2009. Accessed June 13, 2016. http://fetalhydrocephalus.com/hydro/SIblog/default.aspx?id=35&t=Great-news-from-the-Audiologist

Complex periodic waves

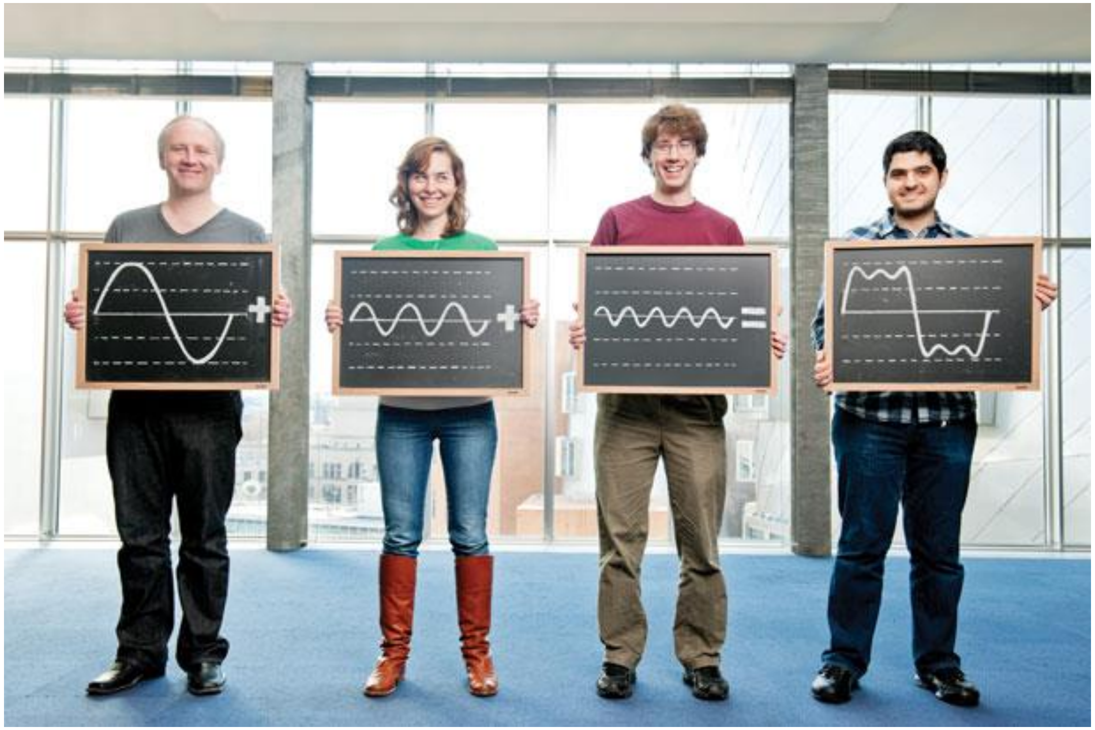
- Results from imperfectly oscillating bodies
- Demonstrate simple harmonic motion
- Examples - a vibrating string, the vocal folds



Frequency – Tones/ Adding

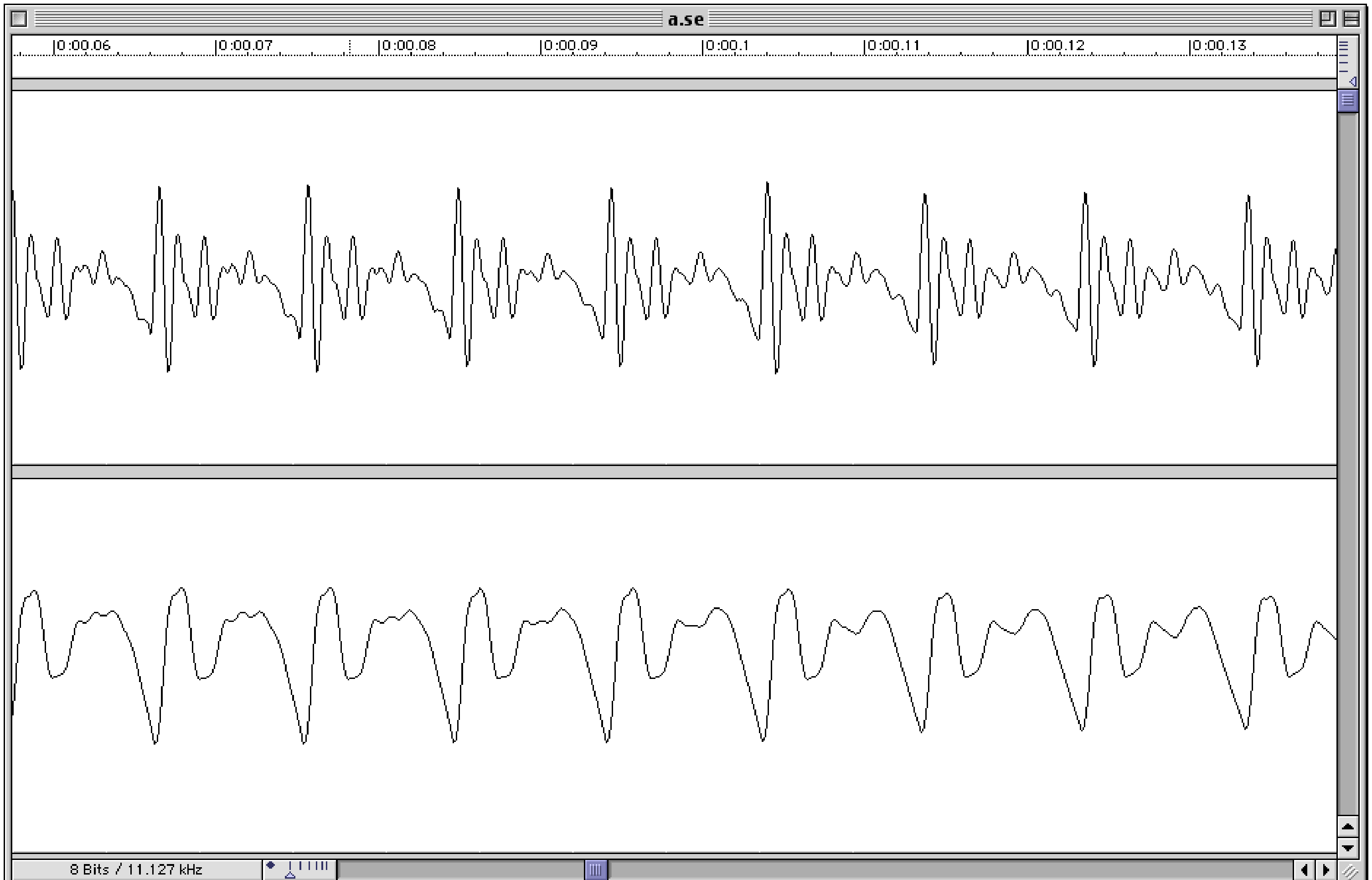


Another example…..

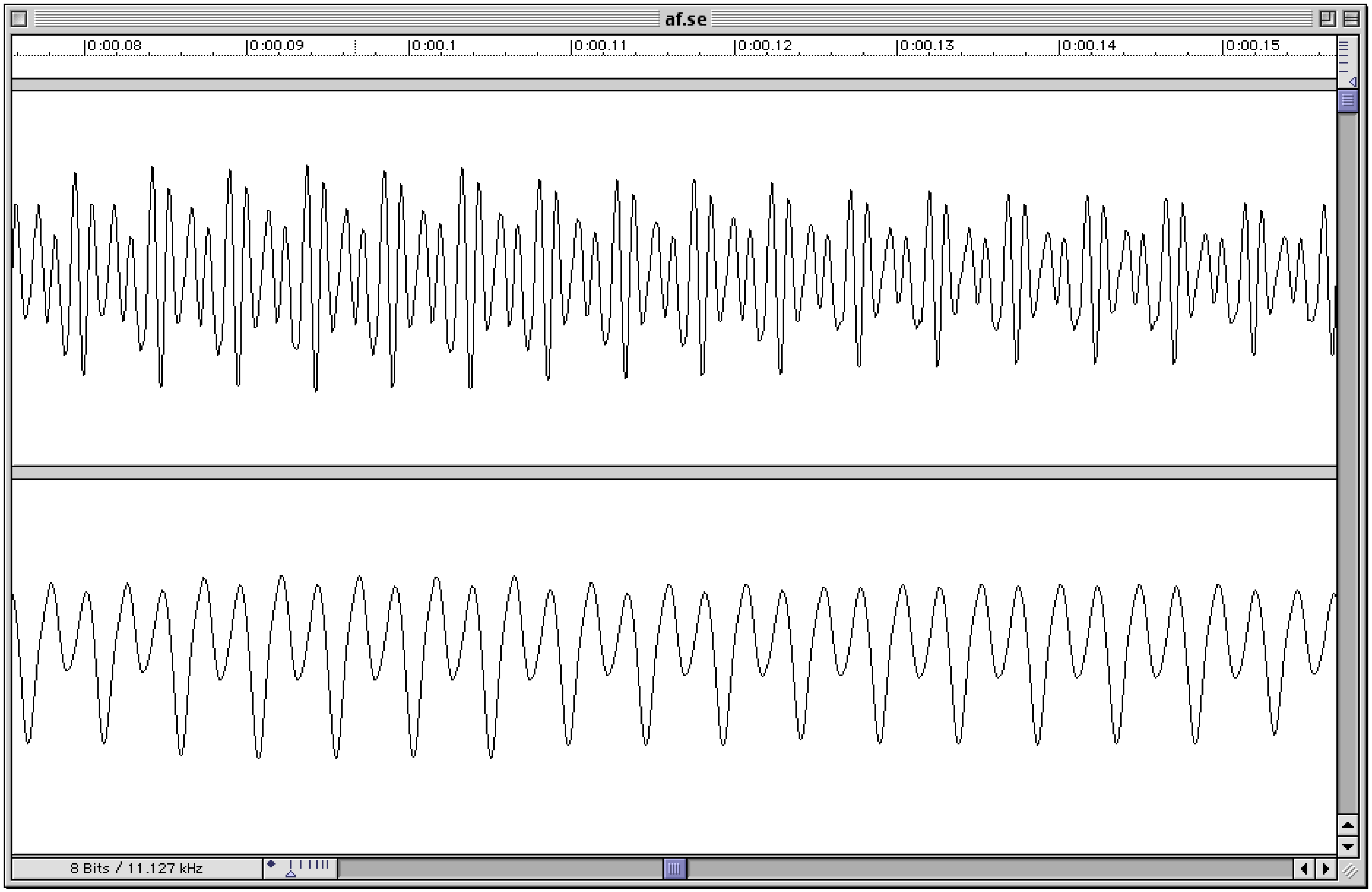


"http://www.askamathematician.com/wpcontent/uploads/2012/09/IndykKatabiPriceHassanieh.jpg">

Waveforms - Male Vowels



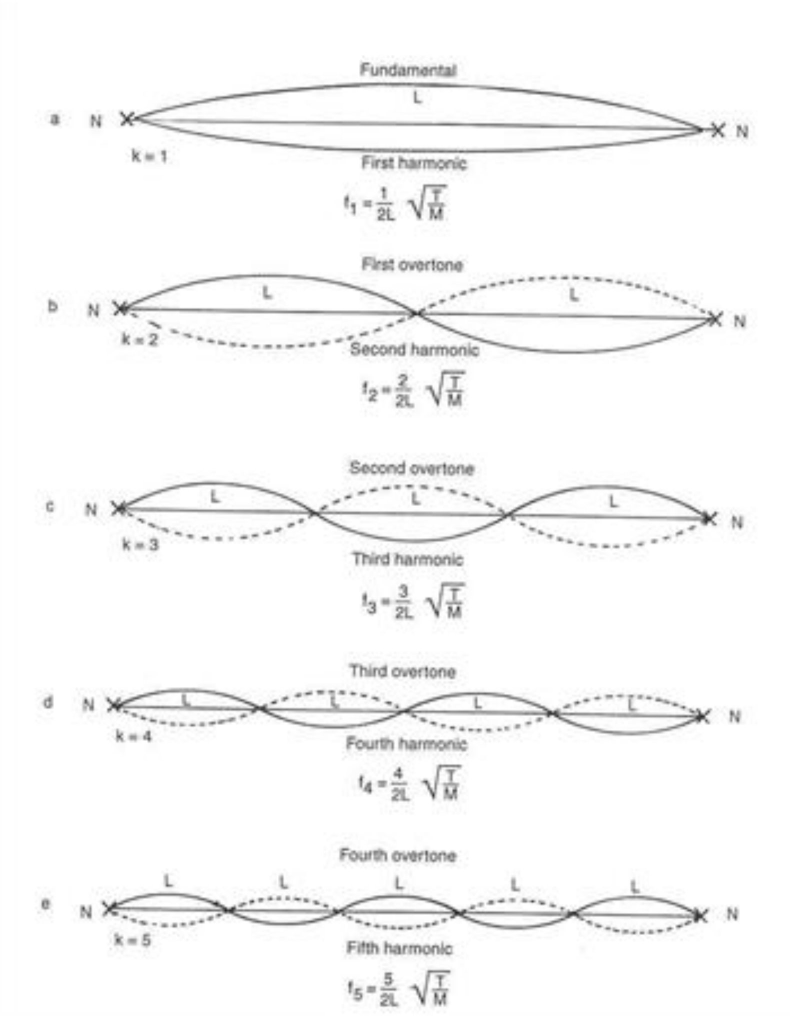
Waveforms - Female Vowels



Complex periodic waves – cont'd

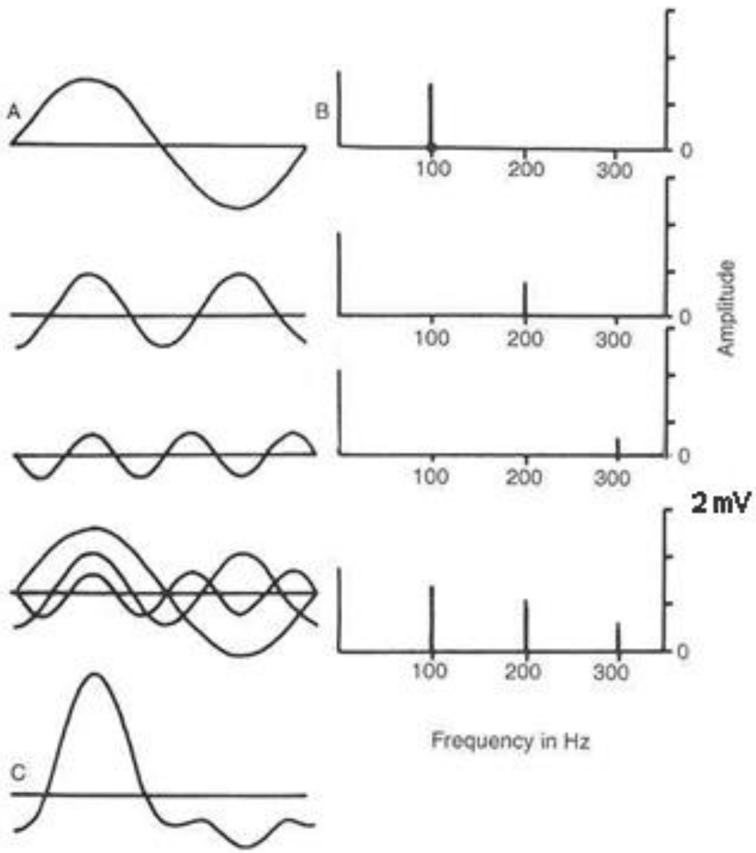
- Consists of a fundamental (F_0) and harmonics
- Harmonics ("overtones") consist of energy at integer multiples of the fundamental (x2, x3, x4 etc…)

Harmonic series



- Imagine you pluck a guitar string and could look at it with a really precise strobe light
- Here is what its vibration will look like

From complex wave to its components… and the frequency spectrum



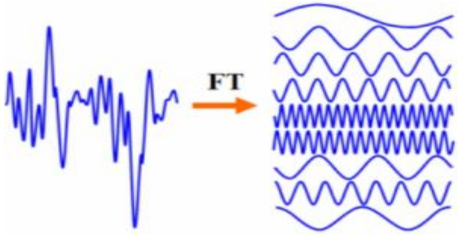
(complex wave)

- Also known as a "line spectrum"
- Here, complex wave at the bottom…
- ..is broken into its component sin waves shown at the top

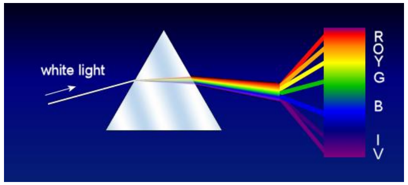
Fourier analysis

1768-1830

Complex wave → component sinusoids



Sound



Light

Review of source characteristics

- Simple waves are a good way to learn about basic properties of frequency, amplitude, and phase.
- Examples include whistling; not really found much in speech

- Complex waves are found in nature for oscillating bodies that show simple harmonic motion (e.g., the vocal folds)

Information from *Phonetics for Dummies.* William F. Katz. "Making Waves: An Overview of Sound." 2013.

Now let's look at the filter

- In speech, the filter is the *supralaryngeal vocal tract* (SLVT)
- The shape of the oral/pharyngeal cavity determines vowel quality
- SLVT shape is chiefly determined by tongue movement, but lips, velum and (indirectly) jaw also play a role

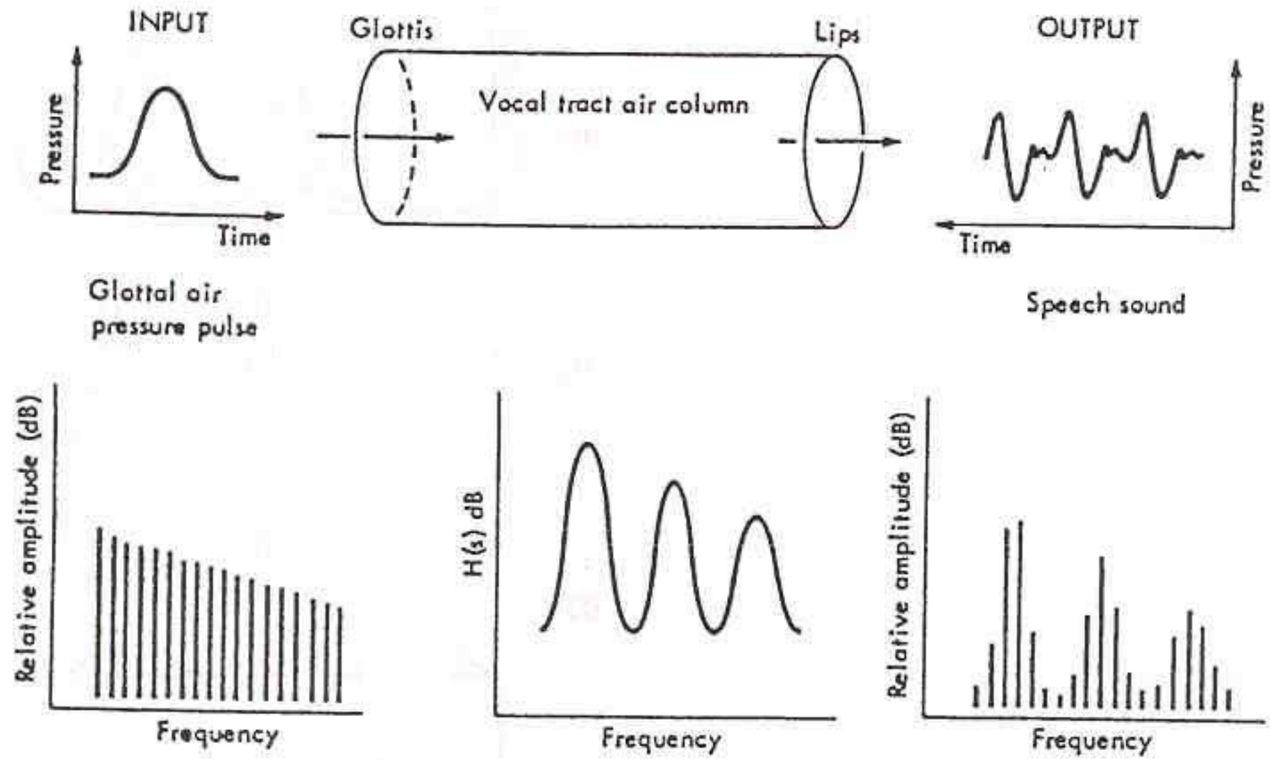
Resonance

- Reinforcement or shaping of frequencies as a function of the boundary conditions through which sound is passed
- FUN: Try producing a vowel with a paper towel roll placed over your mouth!
- The 'extra tube' changes the resonance properties

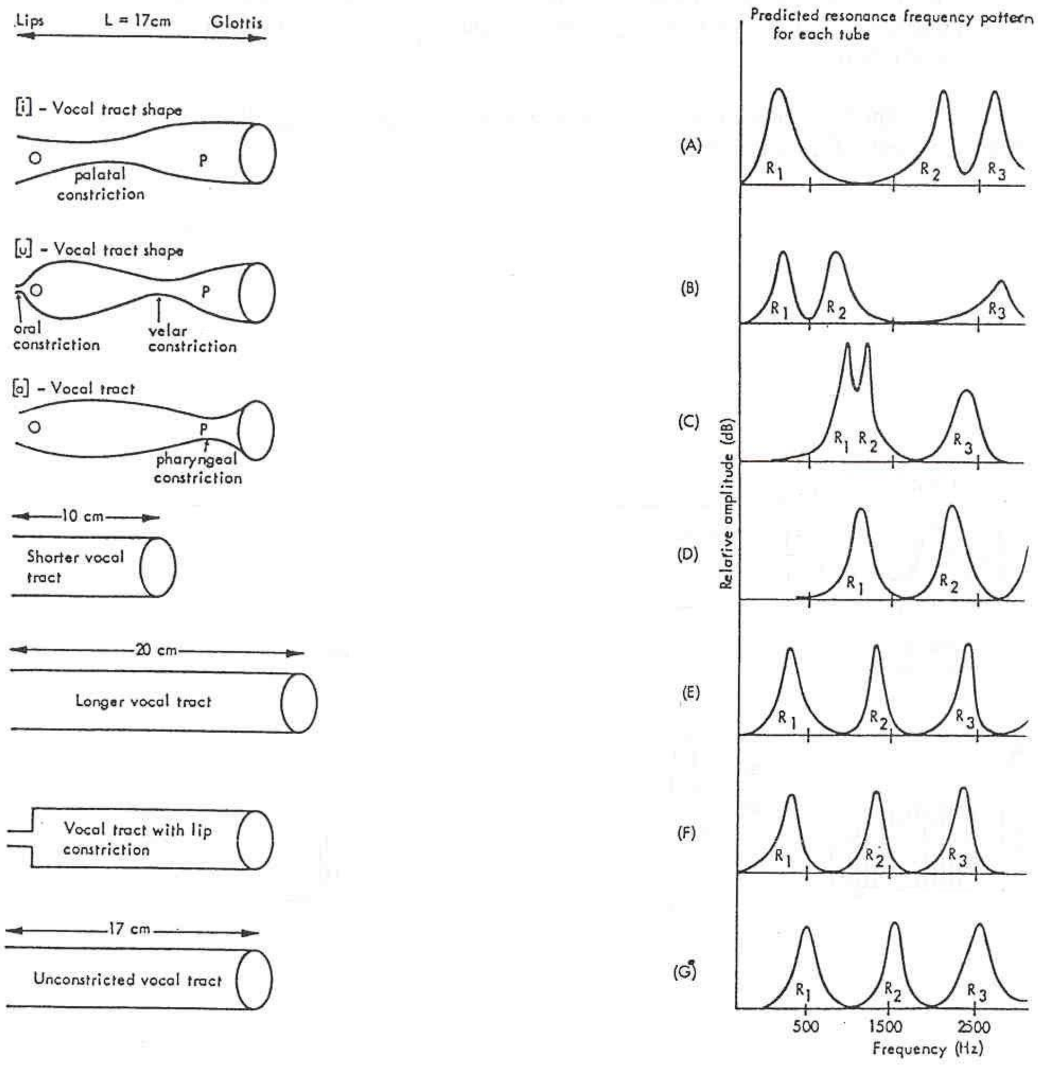
Resonance / Formants

- The SLVT can be modeled as a kind of bottle with different shapes… as sound passes through this chamber it achieves different sound qualities
- The resonant peaks of speech that relate to vowel quality are called formants.
- Thus, R1 = F1 (“first formant). R2 = F2 (“second formant”) etc.
- F1 and F2 are critical determinants of vowel quality

Input → SLVT → final output



Vocal tract shape → formant frequencies



Resonance – FOUR basic rules

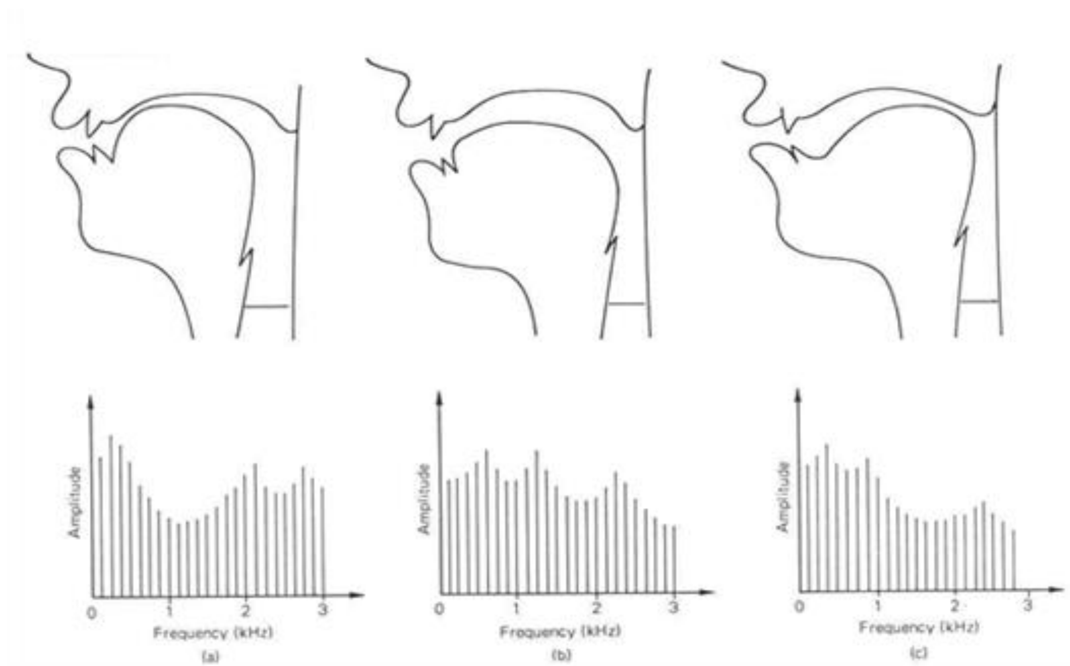
- **F1 rule** – inversely related to jaw height. As the jaw goes down, F1 goes up, etc.
- **F2 rule** – directly related to tongue fronting. As the tongue moves forward, F2 increases.
- **F3 rule** – F3 drops with r-coloring
- **Lip rounding rule** – All formants are lowered by liprounding (because lip protrusion lengthens the vocal tract 'tube')

Information from *Phonetics for Dummies.* William F. Katz. "Making Waves: An Overview of Sound." 2013.

Examples of resonance for /i/, /ɑ/, /u/

- /i/ is made with the tongue high (thus, low F1) and fronted (high F2)
- /ɑ/ is made with the tongue low (high F1) and back (low F2)

/i/ /ɑ/ /u/



Download a (free) cool, interactive demo: https://www.phon.ucl.ac.uk/resource/vtdemo/

American English Vowels

Table 12-1 Mean Formant Frequencies for Men

Vowels	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ʌ/	/ɝ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/
F3	3003	2654	2557	2643	2580	2539	1686	2468	2564	2390	2364	2321
F2	2345	1974	1982	1855	1809	1455	1457	1214	1081	1182	1376	1373
F1	300	445	497	534	694	638	523	754	654	523	426	353

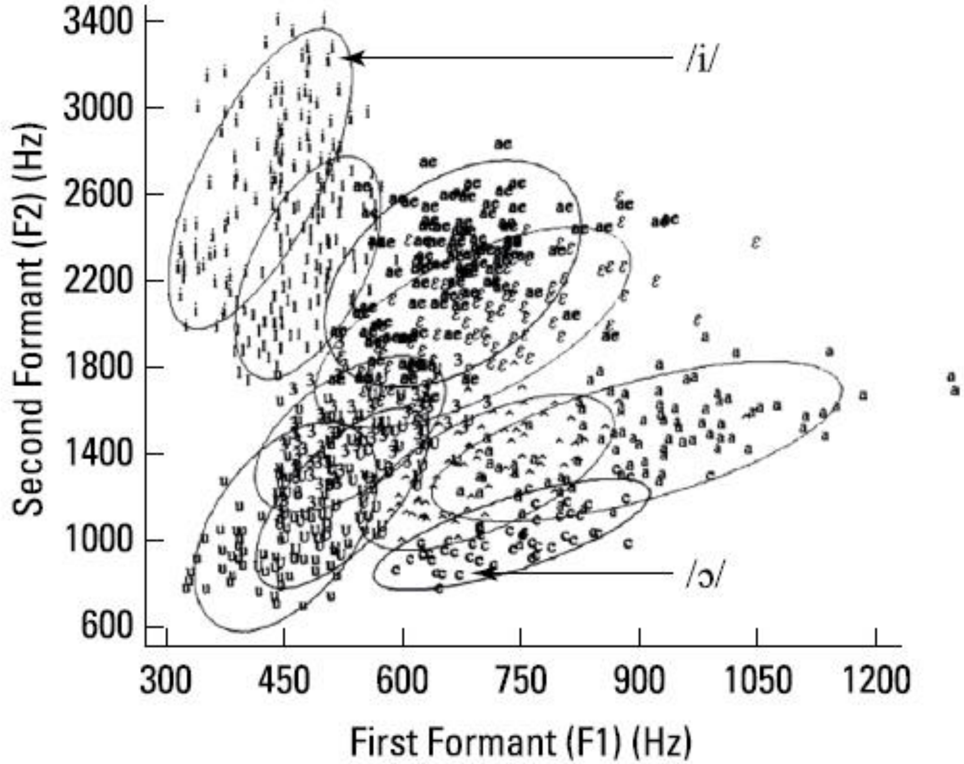
Table 12-2 Mean Formant Frequencies for Children

Vowels	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ʌ/	/ɝ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/
F3	3256	2965	2990	2929	2875	2887	1870	2966	2947	2634	2734	2636
F2	2588	2161	2309	2144	2051	1751	1508	1273	1203	1470	1685	1755
F1	429	522	572	586	836	767	640	688	816	636	516	430

(Assmann & Katz, 2000)

Tables from *Phonetics for Dummies.* William F. Katz. “Making Waves: An Overview of Sound.” 2013.

F2 x F1 plot
American English Vowels



- Peterson & Barney, 1952

Figure from *Phonetics for Dummies.* William F. Katz. “Making Waves: An Overview of Sound.” 2013.

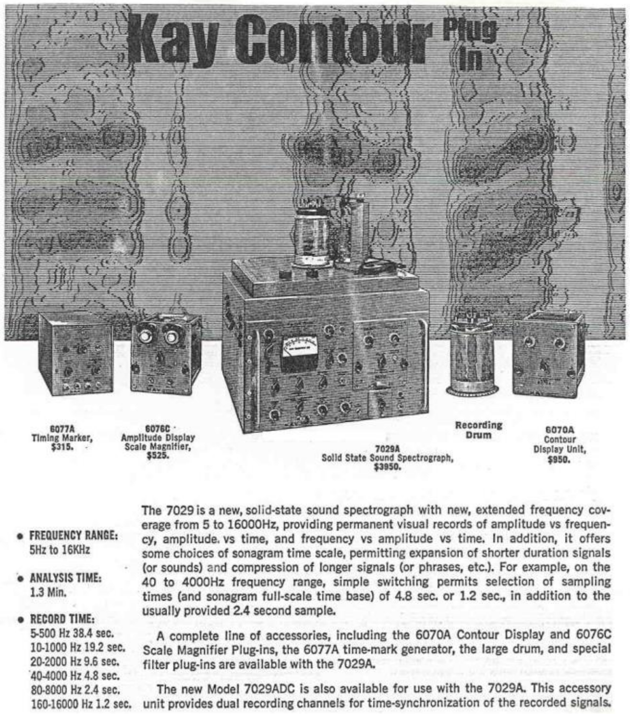
Chap 13

- Reading a sound spectrogram

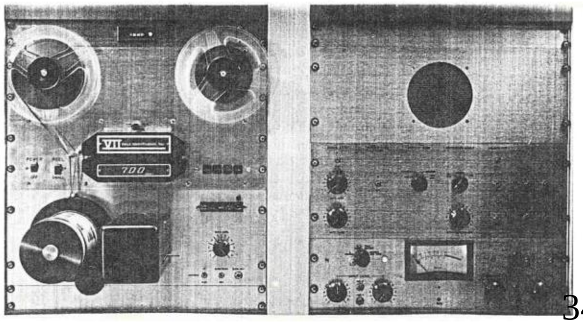
The sound spectrograph

- Invented in the 1940s
- First called ‘visible speech’
- Originally thought to produce a *“speech fingerprint” (?)*
- We now know speech perception is far more complicated and ambiguous..

ca. 1960's



1970's



Basics of spectrogram operation

- Original systems used bandpass filters
- Accumulated energy was represented by a dark image <u>burned</u> onto specially-treated paper
- Nowadays, performed digitally using variety of algorithms (e.g., LPC = *linear predictive coding*)

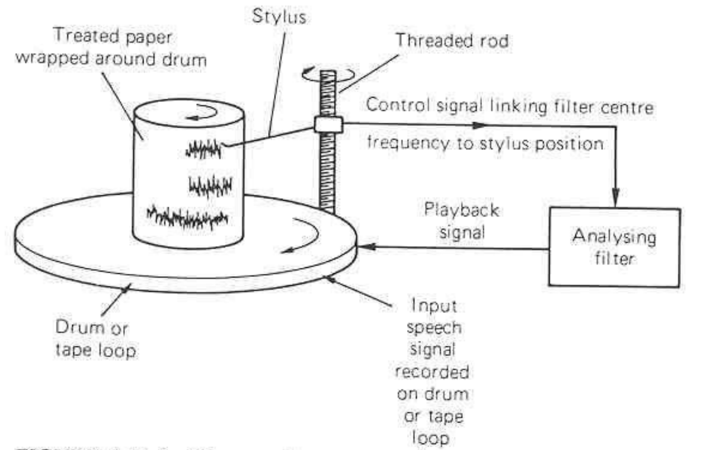


FIGURE 7.14.2 The speech spectrograph

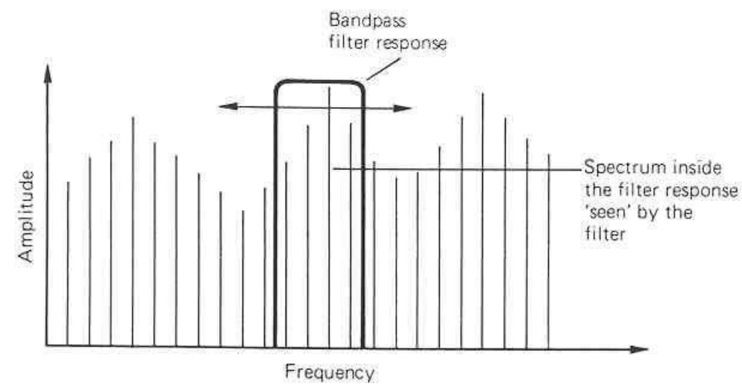
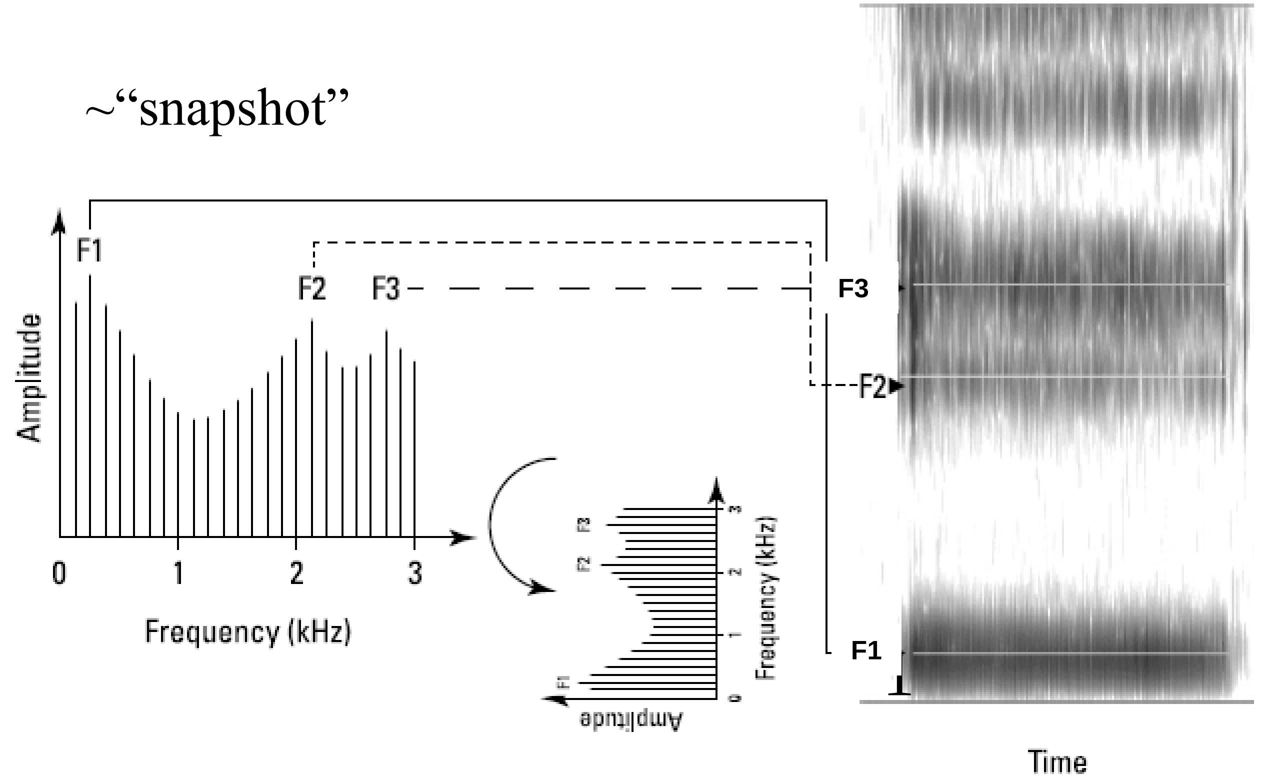


FIGURE 7.14.3 Bandpass filtering

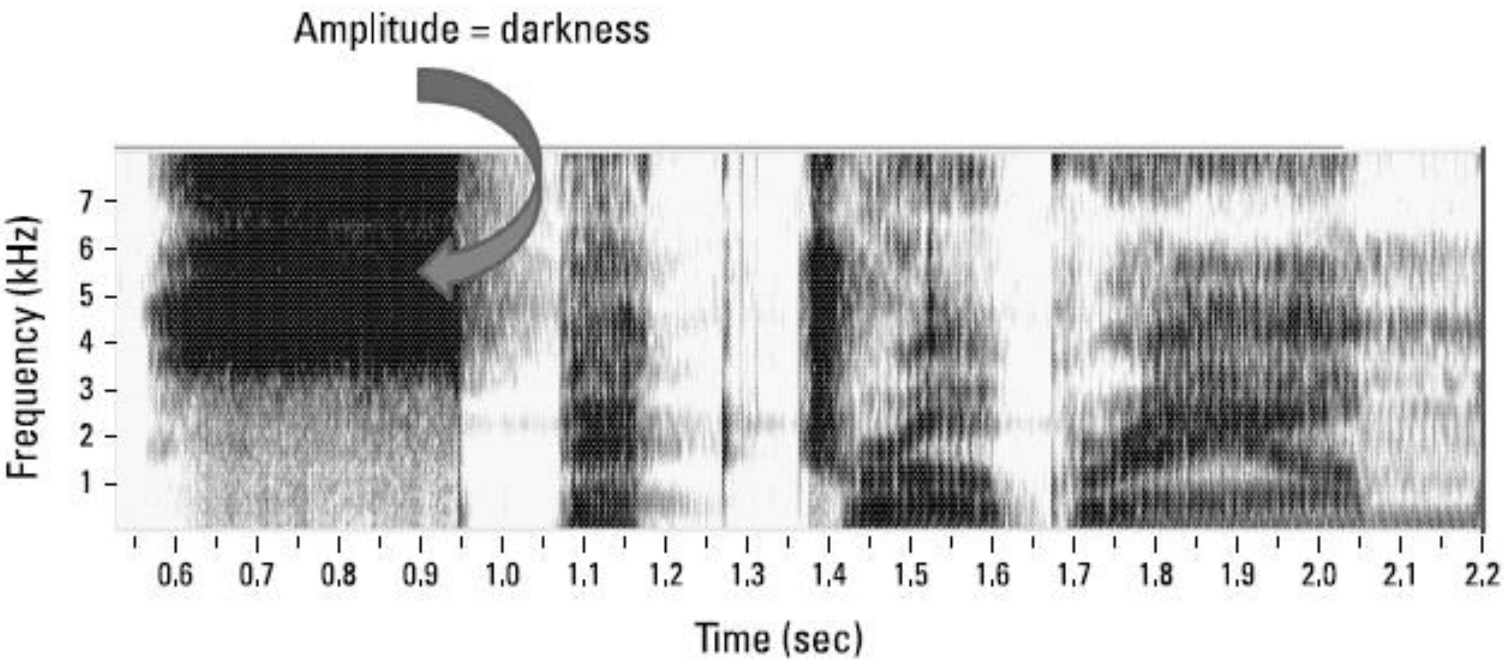
Relating line spectrum to spectrogram

~ “video”

~“snapshot”



Sample of word "*spectrogram*"

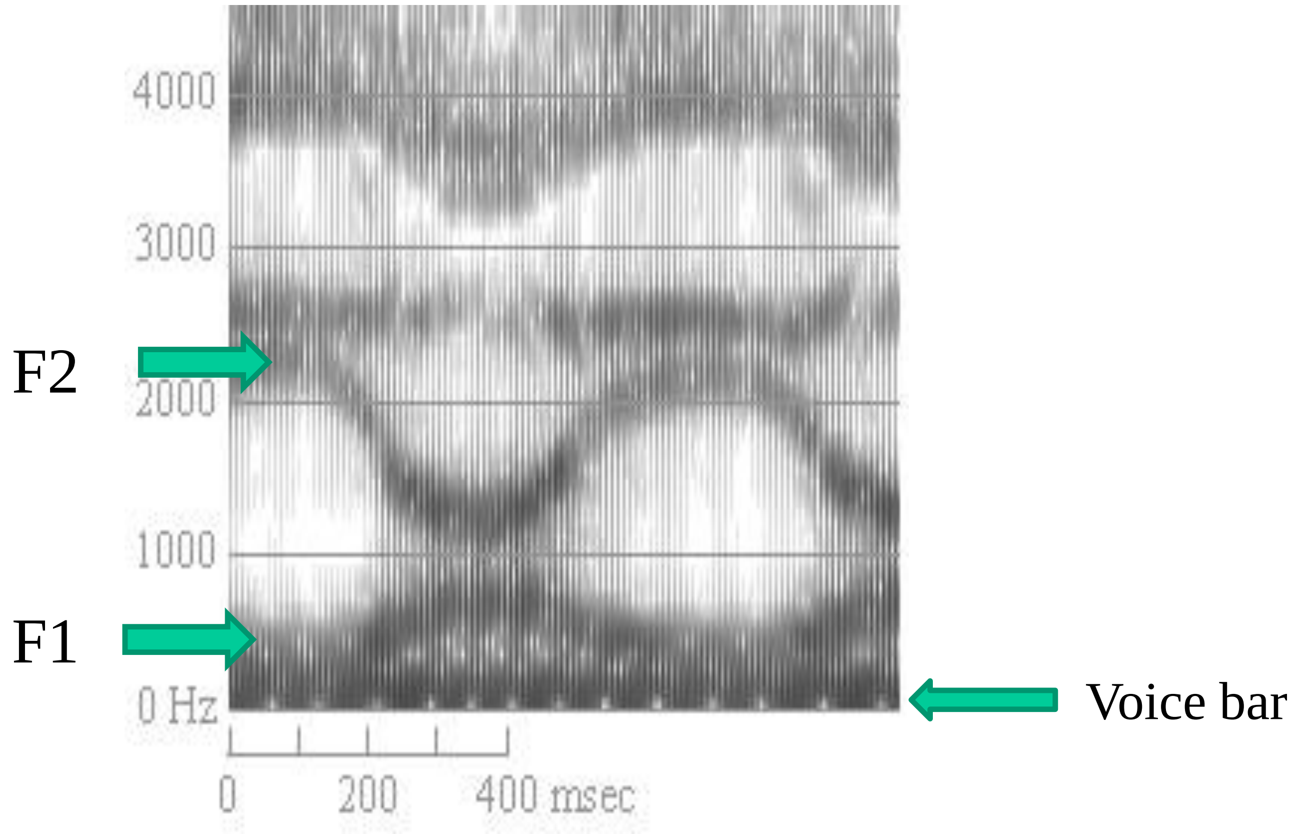


- Pg. 192

Figure from *Phonetics for Dummies.* William F. Katz. "Reading a Sound Spectrogram." 2013.

Vowel basics

- Here is /i ɑ i ɑ / produced with level pitch



Spectrogram from Ladefoged and Johnson, A course in phonetics

Let’s find some vowels!

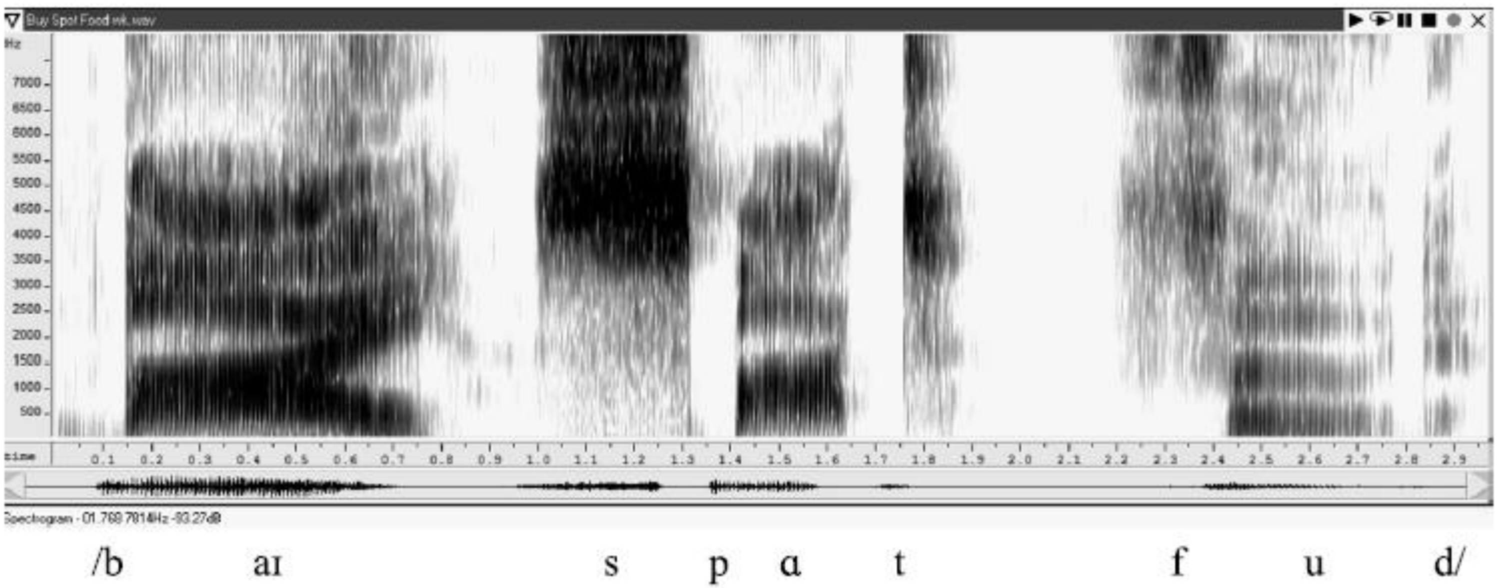


Figure from *Phonetics for Dummies.* William F. Katz. “Reading a Sound Spectrogram.” 2013.

Here they are:

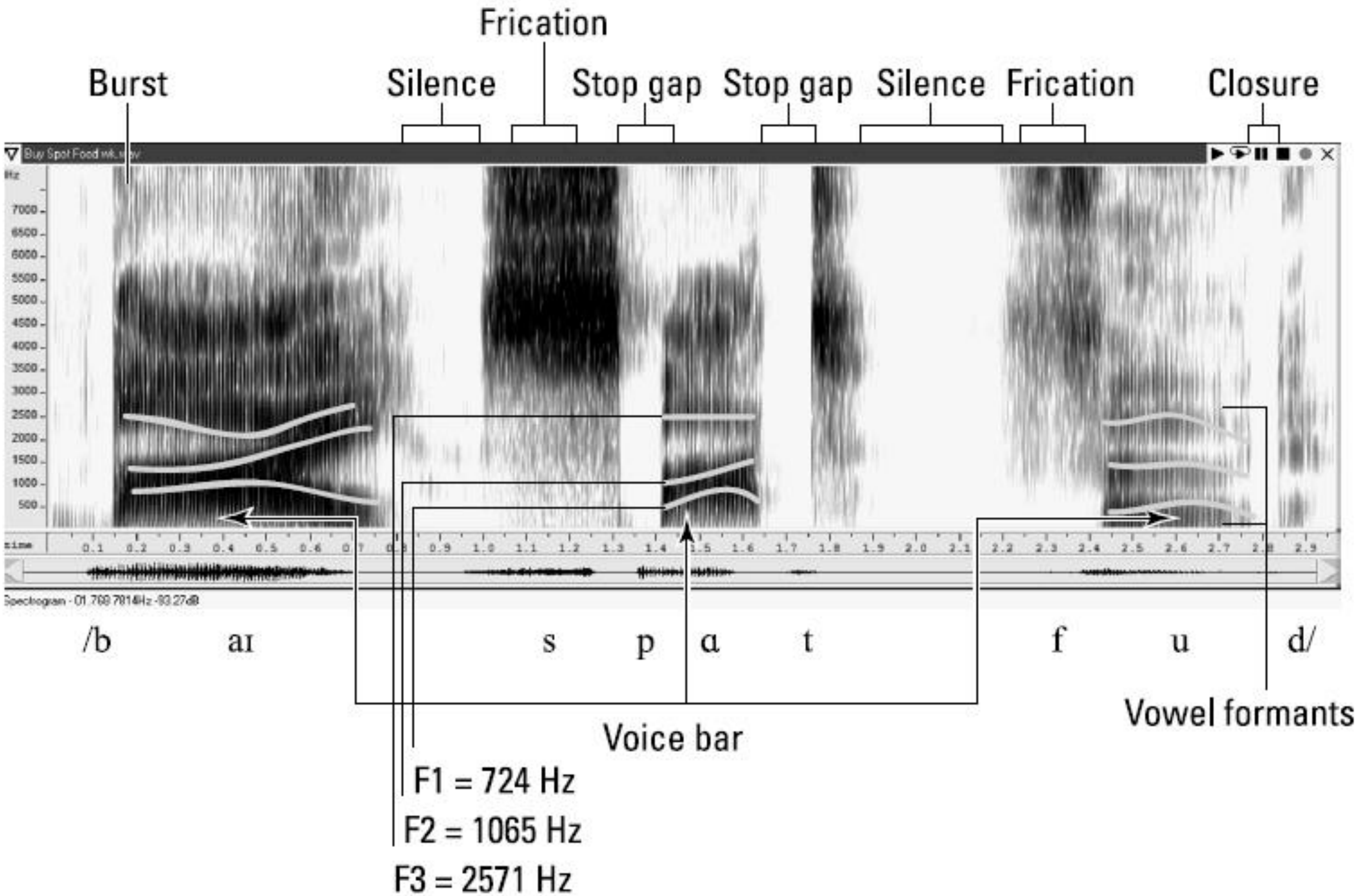


Figure from *Phonetics for Dummies.* William F. Katz. "Reading a Sound Spectrogram." 2013.

Consonants – formant transitions

- An example of an F1 transition for the syllable /da/

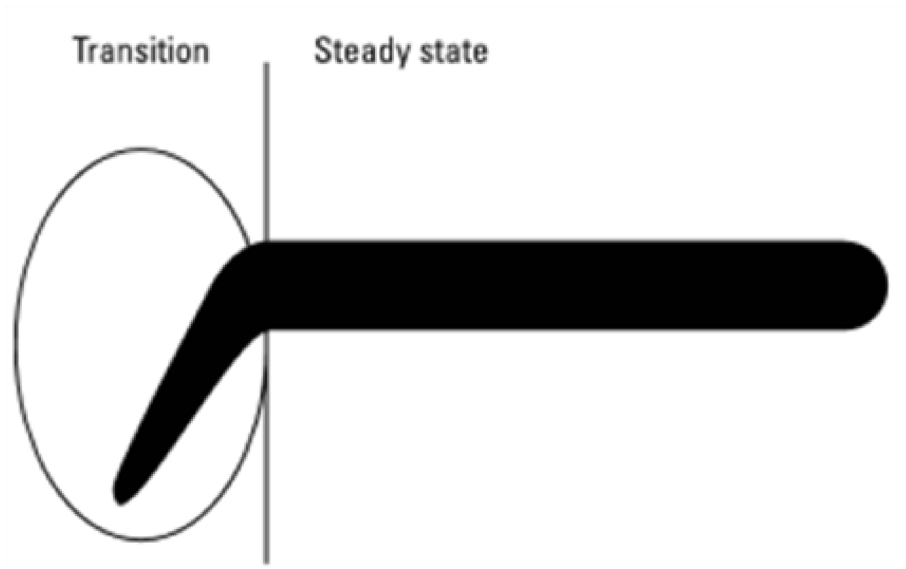
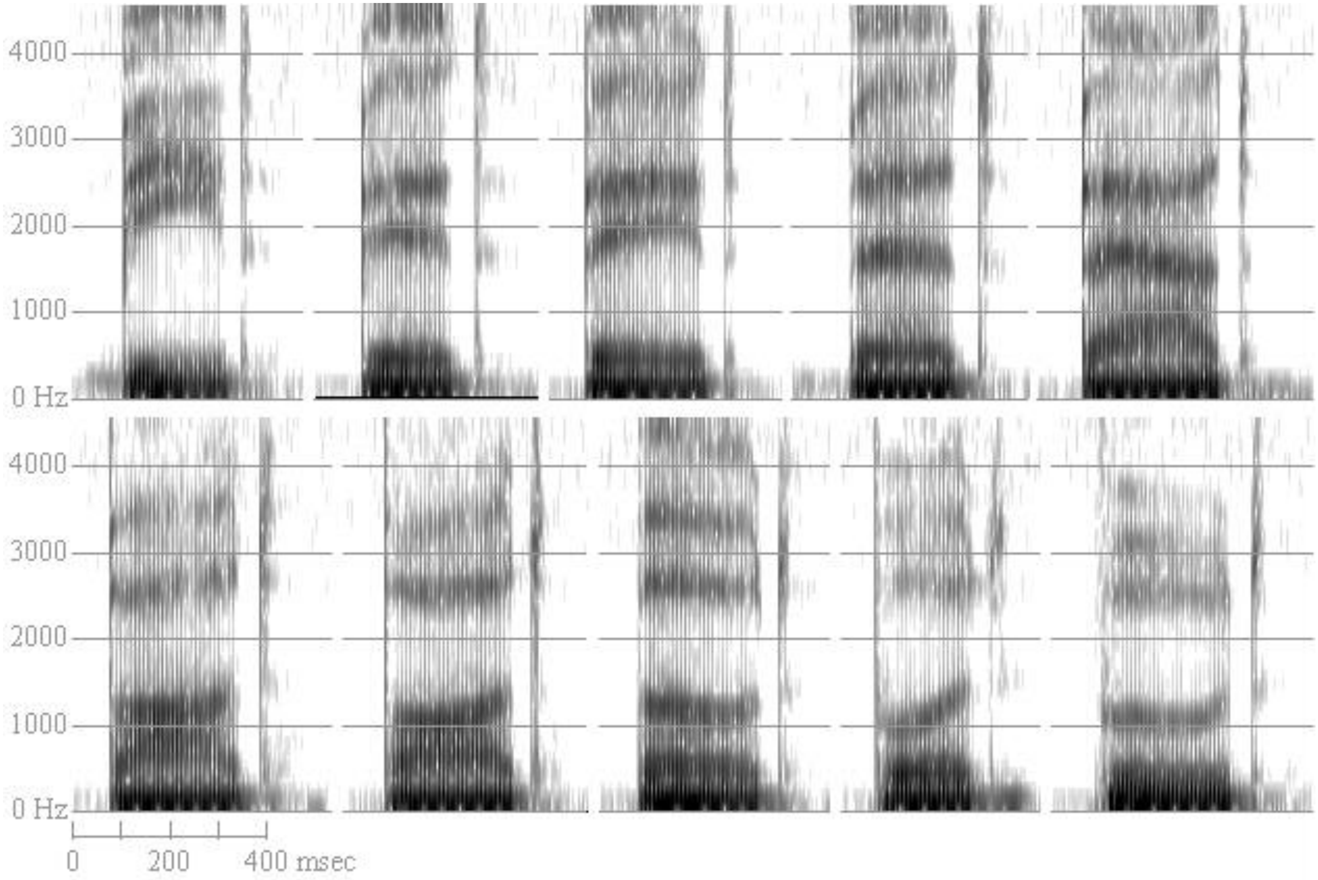


Figure from *Phonetics for Dummies.* William F. Katz. “Reading a Sound Spectrogram.” 2013.

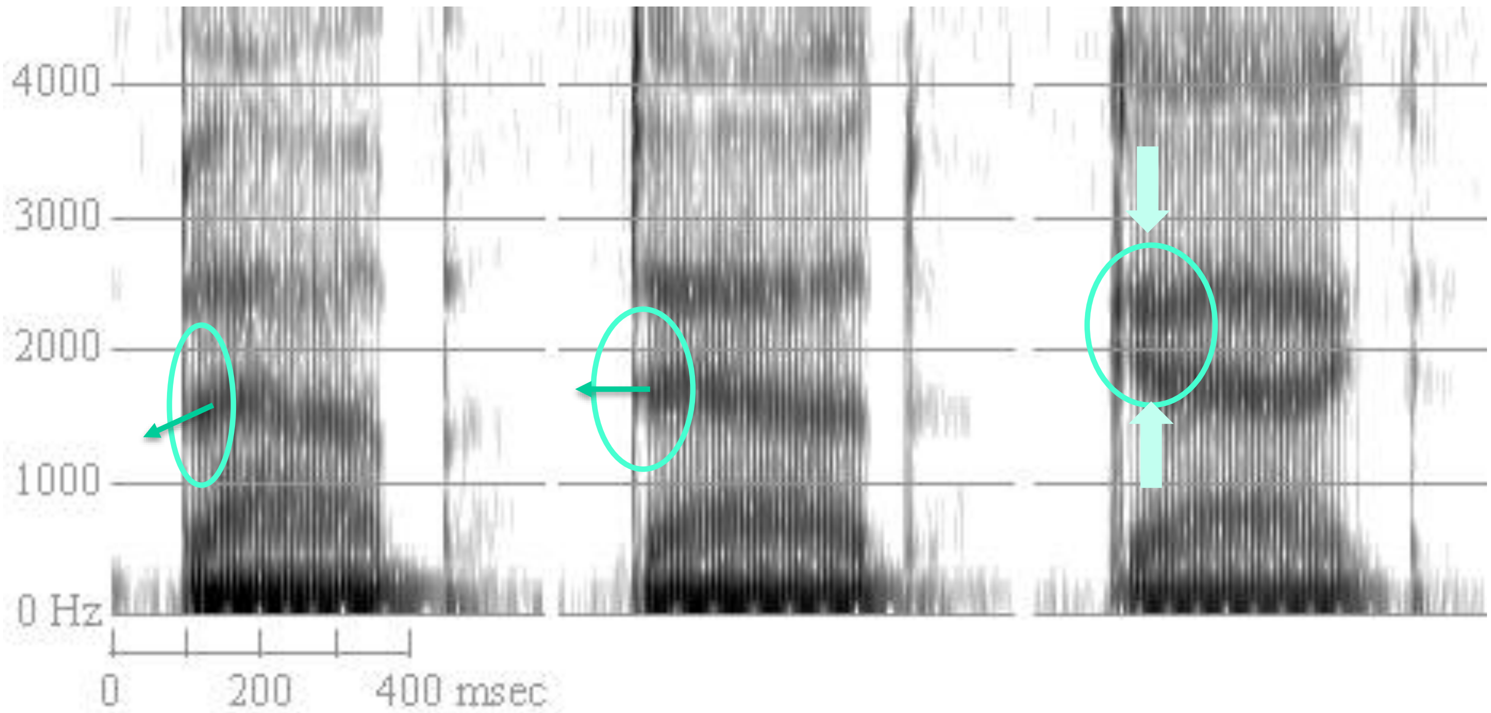
American English vowels in /b_d/ context



- TOP ROW (front vowels): "bead bid bade bed bad"
- BOTTOM ROW (back vowels) "bod bawd bode buhd booed"

Spectrograms from Ladefoged and Johnson, A course in phonetics

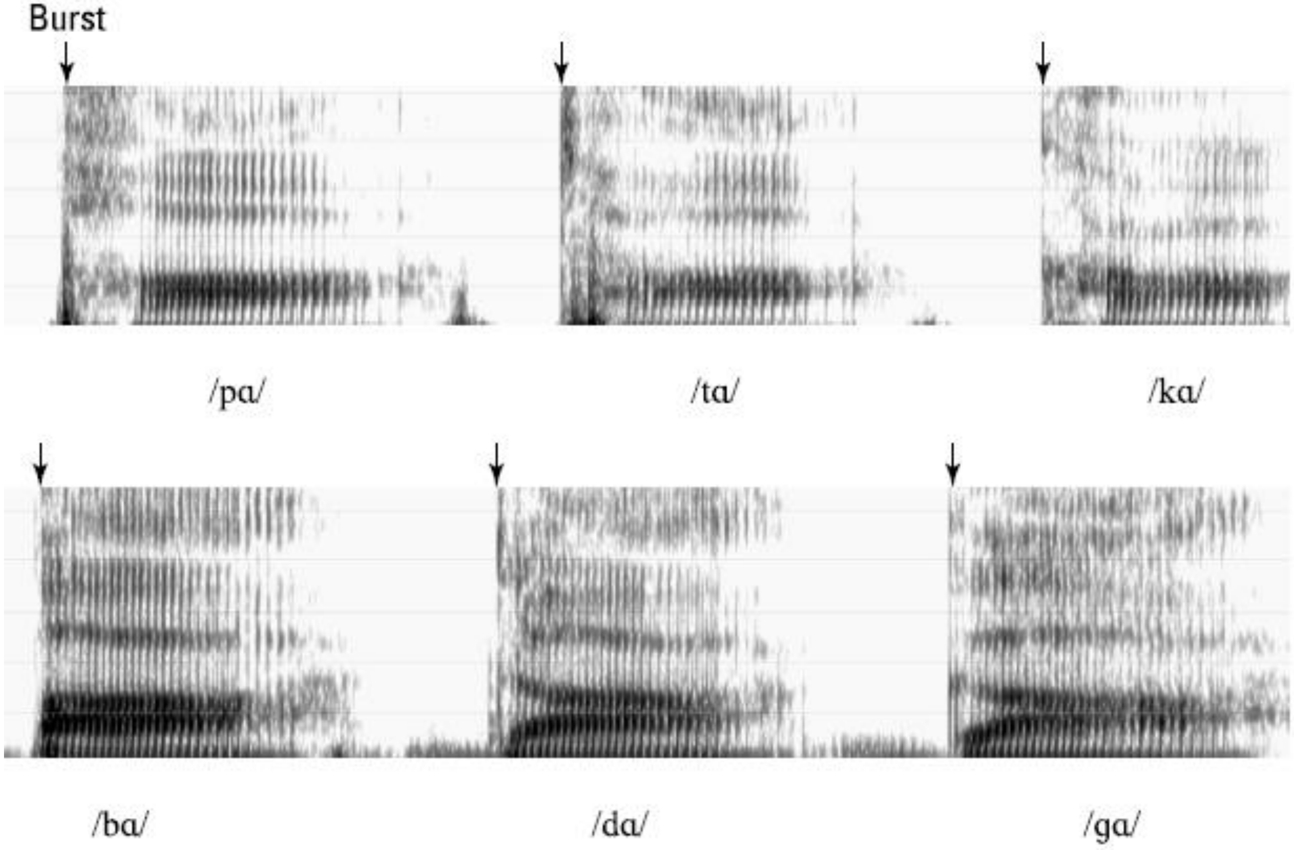
Stops/ formant transitions



- Spectrograms of "bab" "dad" and "gag"
- Labials – F2 point down, alveolars F2 point to ~1700-1800 Hz, velars "pinch" F2 and F3 together
- Note: bottom-most fuzzy is the voice bar!

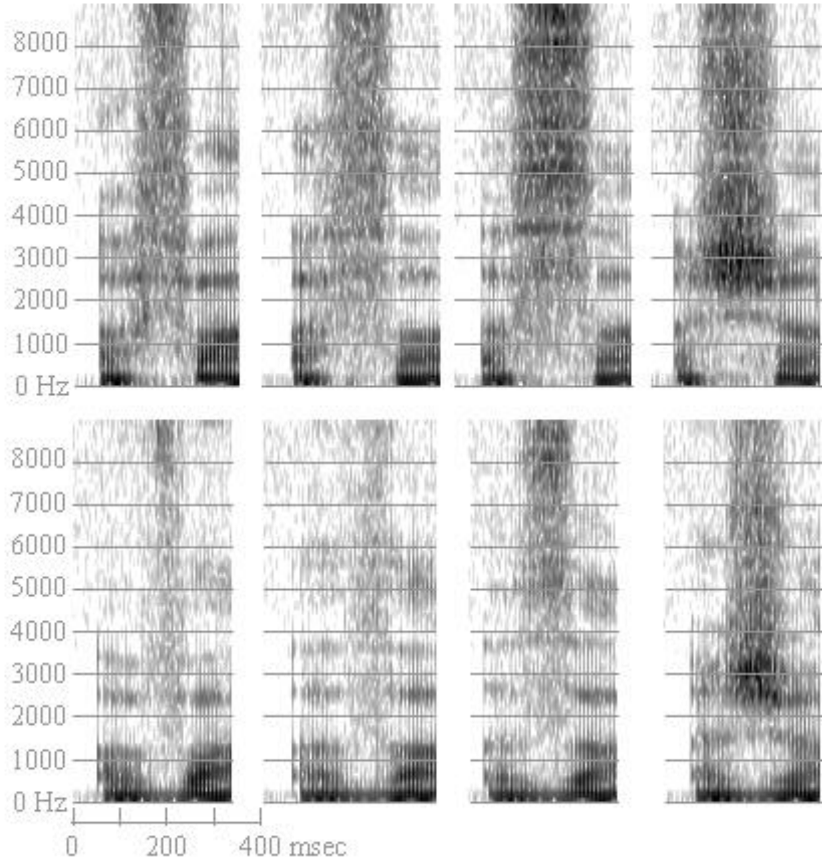
Spectrogram from Ladefoged and Johnson, A course in phonetics

Voicing



(voice of WK)

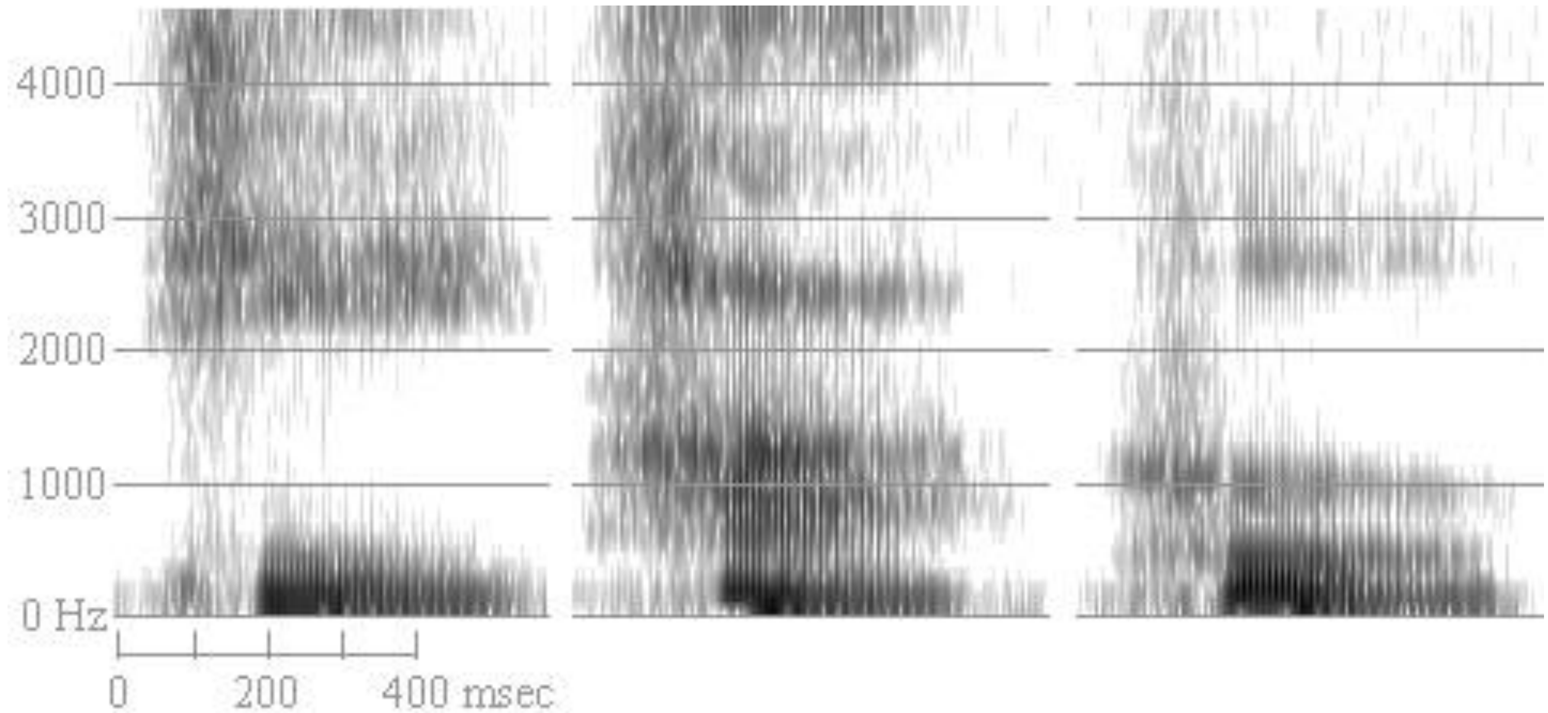
Fricatives



- Top row: /f/, theta, s, esh,
- Bottom row: /v/, ethe, z, long z
- Distribution of the spectral noise is the key here!

Spectrogram from Ladefoged and Johnson, A course in phonetics

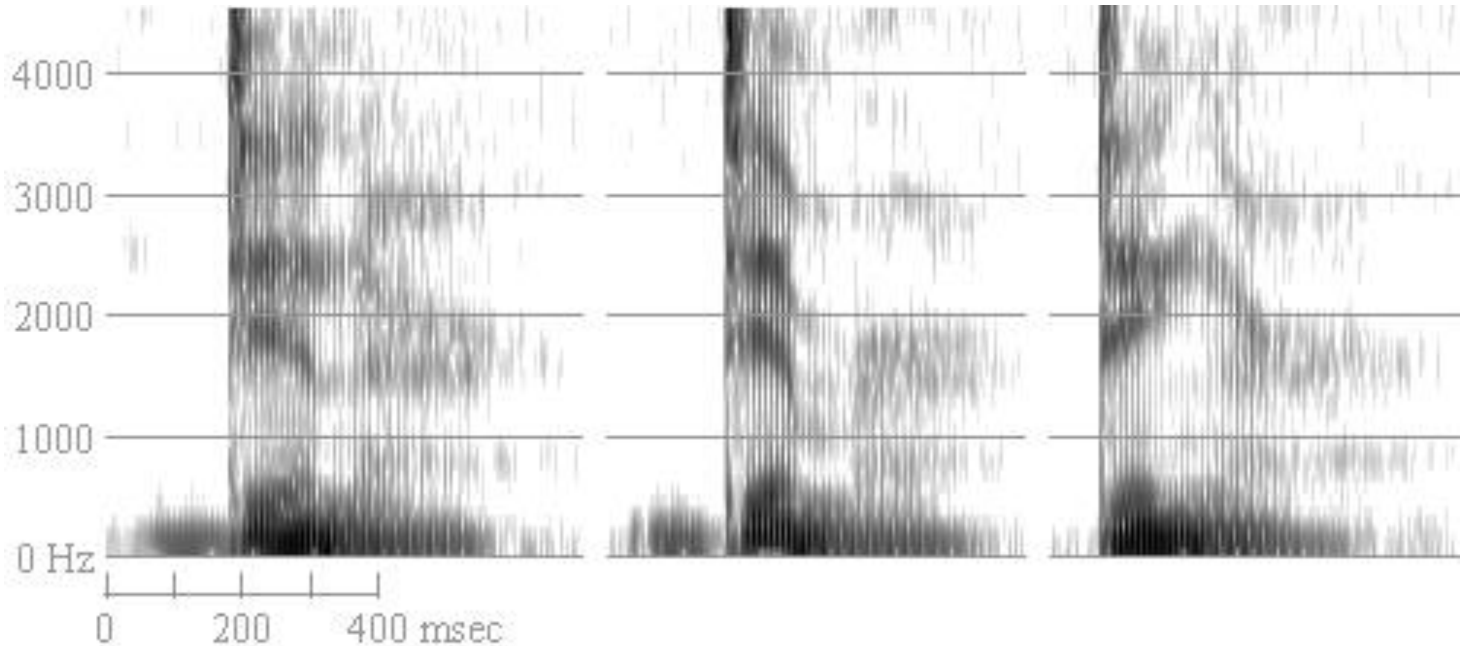
The fricative /h/



- Commonly excites all the formant cavities
- May look slightly different in varying vowel contexts

Spectrogram from Ladefoged and Johnson, A course in phonetics

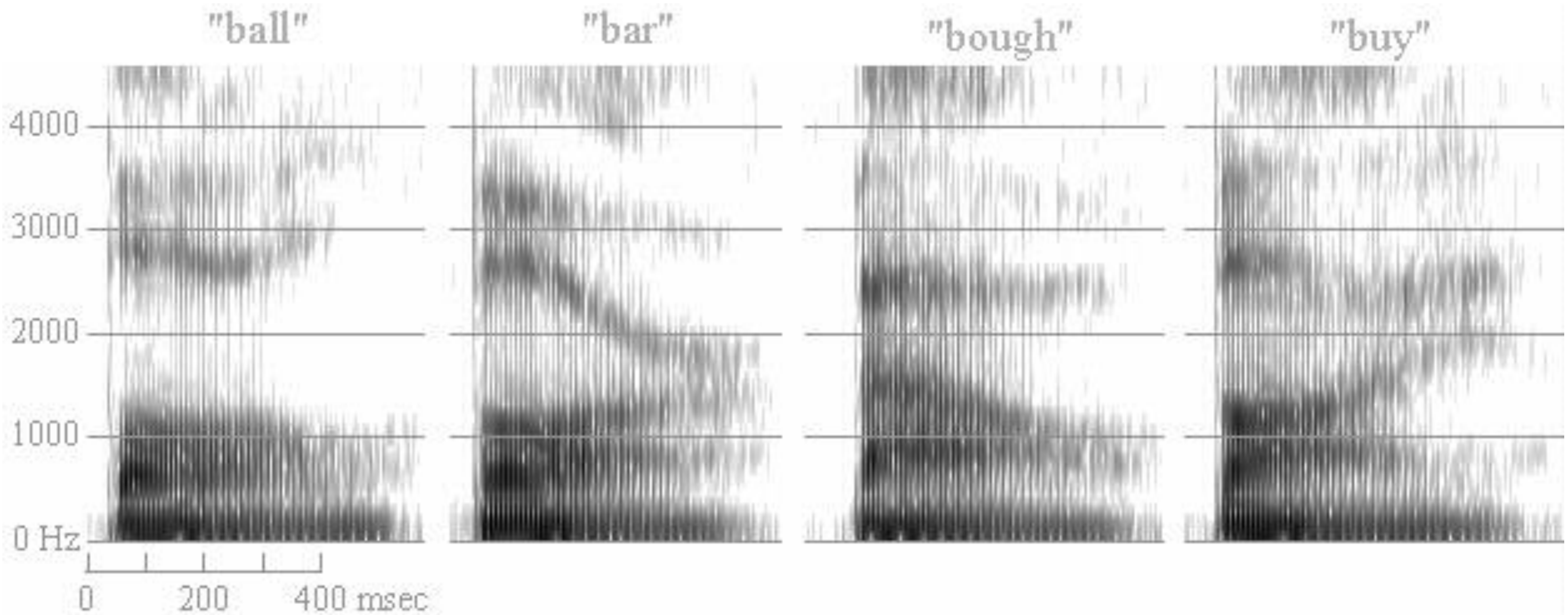
Nasal stops



- Spectrograms of *"dinner dimmer dinger"*
- Marked by "zeroes" or formant regions with little energy
- Can also result in broadening of formant bandwidths (fuzzying the edges)

Spectrogram from Ladefoged and Johnson, A course in phonetics

Approximants



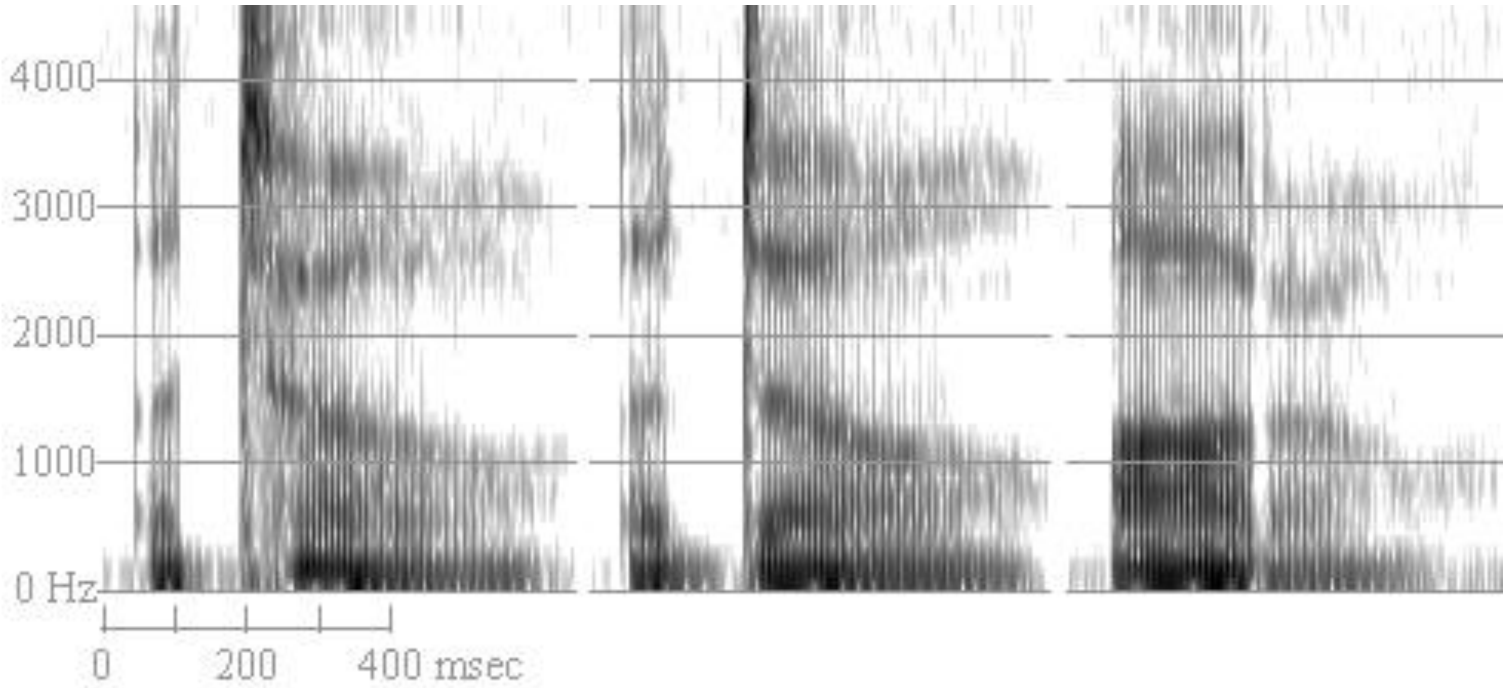
/ɹ/ - very low third formant, just above F2

/l/ - formants in the neighborhood of 250, 1200, and 2400 Hz; less apparent in final position.

Higher formants considerable reduced in intensity

Spectrogram from Ladefoged and Johnson, A course in phonetics

Stops versus tap/flap



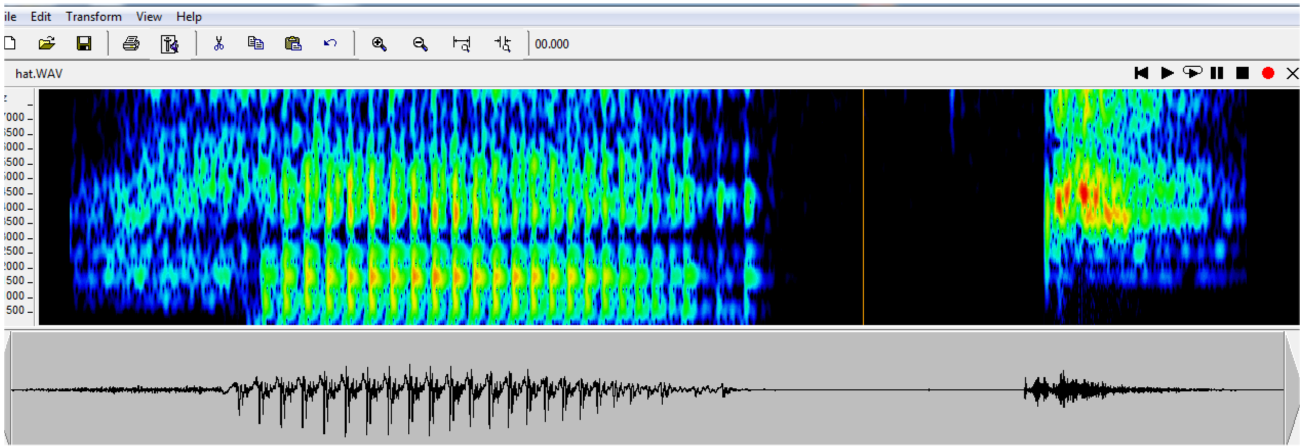
“a toe” “a doe” “otto”

- For full stops, there is about 100 ms of silence
- For tap, only about 10-30 ms

Spectrogram from Ladefoged and Johnson, A course in phonetics

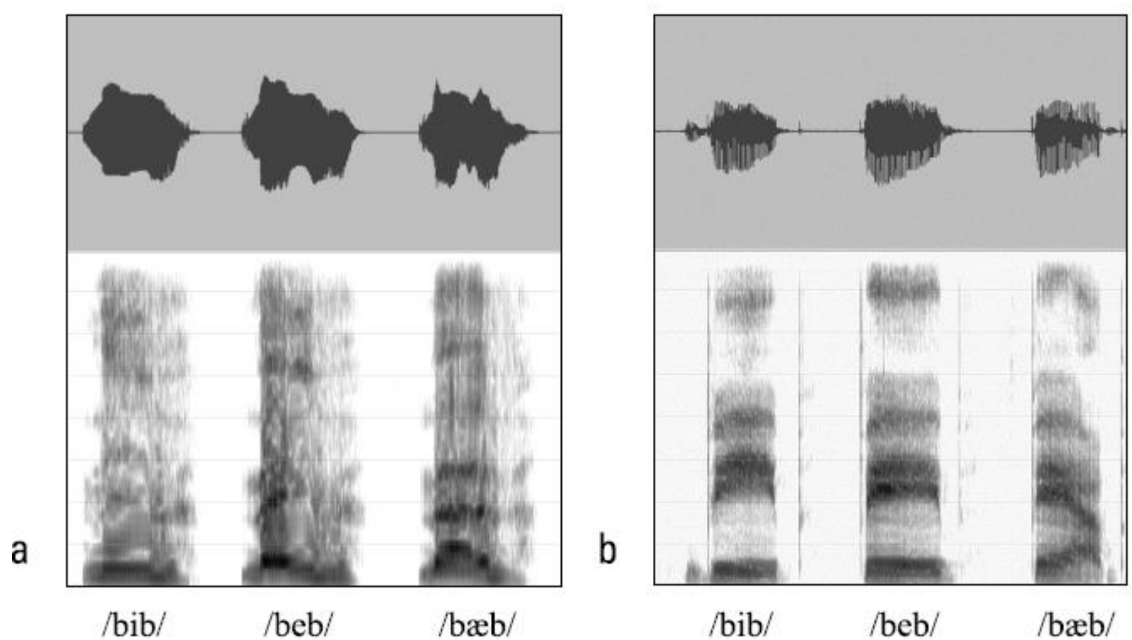
Pseudo-colored example

- Here is an American English /æ/ (male)
- “Hot” areas (in green/yellow/red) have more energy



Wavesurfer

Some “tough cases”….



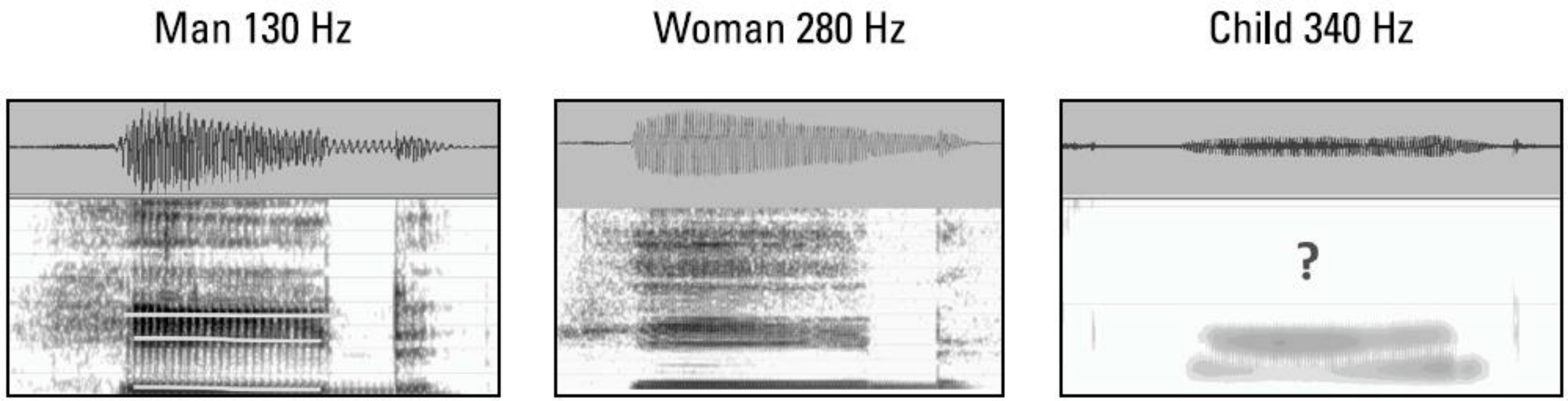
ALS-

Amotryophic lateral sclerosis

(notice loss of formant frequency quality)

(Healthy male control)

Women and children



(High F_0 can cause problems estimating formants)