

# CS 224S / LINGUIST 236 Speech Recognition and Synthesis

Dan Jurafsky

## Lecture 2: Acoustic Phonetics

1/6/05

CS 224S Winter 2005

1

## Today, Jan 6, Week 1

- Acoustic Phonetics
  - Waves, sound waves, and spectra
  - Speech waveforms
  - Deriving schwa
  - Formants
  - Spectrograms
  - Reading spectrograms
  - PRAAT

1/6/05

CS 224S Winter 2005

2

## Acoustic Phonetics

- Sound Waves

- <http://www.kettering.edu/~drussell/Demos/waves-intro/waves-intro.html>
- <http://www.kettering.edu/~drussell/Demos/waves/Lwave.gif>

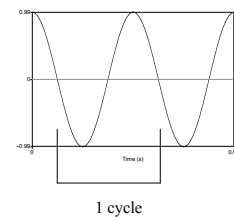
1/6/05

CS 224S Winter 2005

3

## Simple Period Waves (sine waves)

- Characterized by:
  - period:  $T$
  - amplitude  $A$
  - phase  $\phi$
- Fundamental frequency in cycles per second, or Hz
  - $F_0 = 1/T$

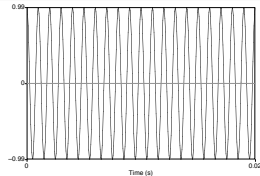


1/6/05

CS 224S Winter 2005

4

## Simple periodic waves of sound



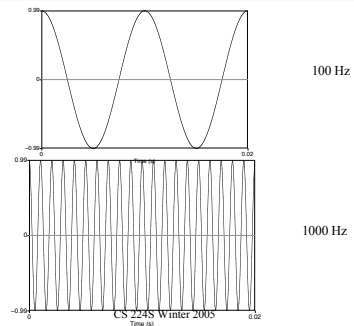
- Y axis: Amplitude = amount of air pressure at that point in time
- Zero is normal air pressure, negative is rarefaction
- X axis: time. Frequency = number of cycles per second.
- Frequency =  $1/\text{Period}$
- 20 cycles in .02 seconds = 1000 cycles/second = 1000 Hz

1/6/05

CS 224S Winter 2005

5

## Waves have different frequencies

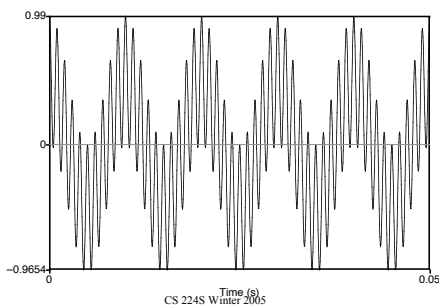


1/6/05

CS 224S Winter 2005

6

## Complex waves: Adding a 100 Hz and 1000 Hz wave together



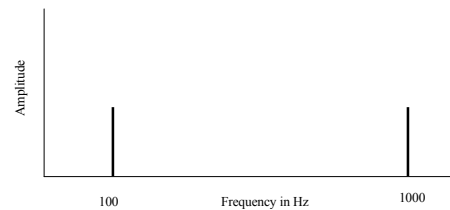
1/6/05

CS 224S Winter 2005

7

## Spectrum

Frequency components (100 and 1000 Hz) on x-axis



1/6/05

CS 224S Winter 2005

8

## Spectra continued

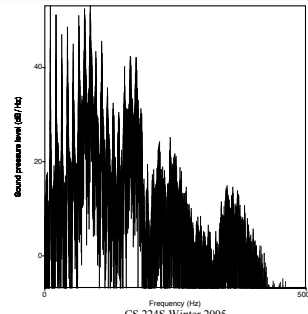
- **Fourier analysis: any wave can be represented as the (infinite) sum of sine waves of different frequencies (amplitude, phase)**

1/6/05

CS 224S Winter 2005

9

## Spectrum of one instant in an actual soundwave: many components across frequency range



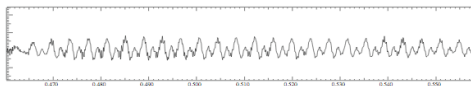
1/6/05

CS 224S Winter 2005

10

## Waveforms for speech

- **Waveform of the vowel [iy]**

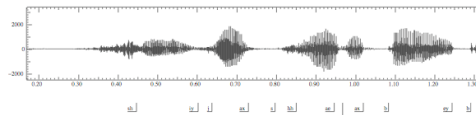


- **Frequency: repetitions/second of a wave**
- Above vowel has 28 reps in .11 secs
- So freq is  $28 / .11 = 255$  Hz
- This is speed that vocal folds move, hence voicing
- **Amplitude: y axis: amount of air pressure at that point in time**

Zero is normal air pressure, negative is rarefaction

11

## She just had a baby



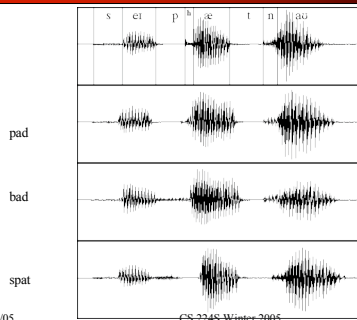
- **What can we learn from a wavefile?**
  - Vowels are voiced, long, loud
  - Length in time = length in space in waveform picture
  - Voicing: regular peaks in amplitude
  - When stops closed: no peaks: silence.
  - Peaks = voicing: .46 to .58 (vowel [iy], from second .65 to .74 (vowel [ax]) and so on
  - Silence of stop closure (1.06 to 1.08 for first [b], or 1.26 to 1.28 for second [b])
  - Fricatives like [sh] intense irregular pattern; see .33 to .46

1/6/05

CS 224S Winter 2005

12

## Examples from Ladefoged

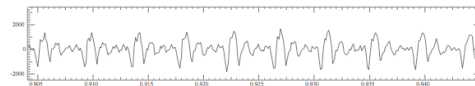


1/6/05

CS 224S Winter 2005

13

## Part of [ae] waveform from "had"



- Note complex wave repeating nine times in figure
- Plus smaller waves which repeats 4 times for every large pattern
- Large wave has frequency of 250 Hz (9 times in .036 seconds)
- Small wave roughly 4 times this, or roughly 1000 Hz
- Two little tiny waves on top of peak of 1000 Hz waves

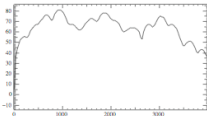
1/6/05

CS 224S Winter 2005

14

## Back to spectrum

- Spectrum represents these freq components
- Computed by Fourier transform, algorithm which separates out each frequency component of wave.



- x-axis shows frequency, y-axis shows magnitude (in decibels, a log measure of amplitude)
- Peaks at 930 Hz, 1860 Hz, and 3020 Hz.

1/6/05

CS 224S Winter 2005

15

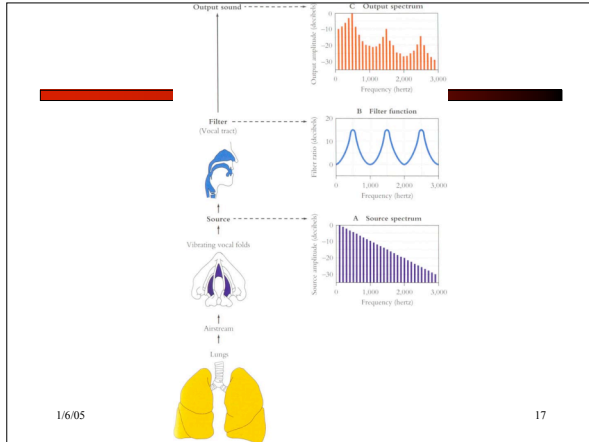
## Why is a speech sound wave composed of these peaks?

- Articulatory facts:
  - The vocal cord vibrations create **harmonics**
  - The mouth is an amplifier
  - Depending on shape of mouth, some harmonics are amplified more than others

1/6/05

CS 224S Winter 2005

16



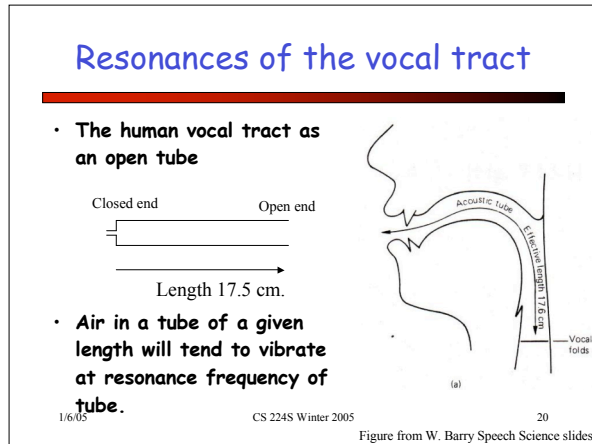
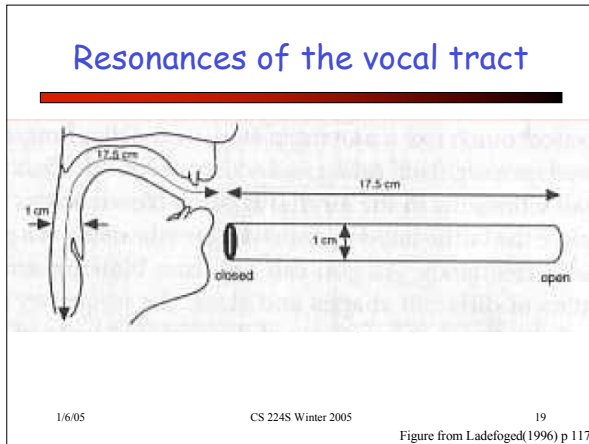
## Deriving schwa: how shape of mouth (filter function) creates peaks!

- Reminder of basic facts about sound waves
- $f = c/\lambda$
- $c =$  speed of sound (approx 35,000 cm/sec)
- A sound with  $\lambda=10$  meters has low frequency  $f = 35$  Hz (35,000/1000)
- A sound with  $\lambda=2$  centimeters has high frequency  $f = 17,500$  Hz (35,000/2)

1/6/05

CS 224S Winter 2005

18



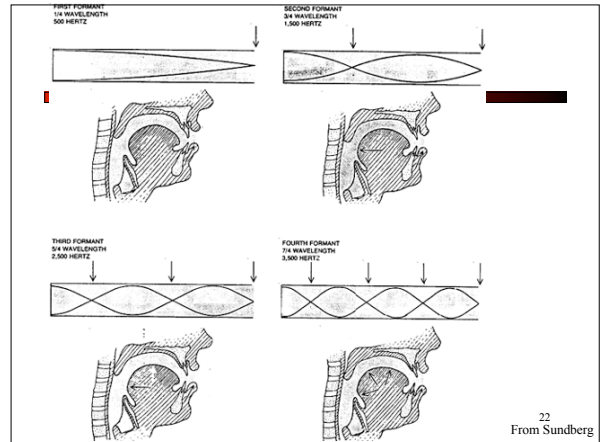
## Resonances of the vocal tract

- If vocal tract is cylindrical tube open at one end
- Standing waves form in tubes
- Waves will resonate if their wavelength corresponds to dimensions of tube
- Constraint: Pressure differential should be maximal at (closed) glottal end and minimal at (open) lip end.
- Next slide shows what kind of length of waves can fit into a tube with this constraint

1/6/05

CS 224S Winter 2005

21



22  
From Sundberg

## Computing the 3 formants of schwa

- Let the length of the tube be  $L$
- $F_1 = c/\lambda_1 = c/(4L) = 35,000/4 \cdot 17.5 = 500\text{Hz}$
- $F_2 = c/\lambda_2 = c/(4/3L) = 3c/4L = 3 \cdot 35,000/4 \cdot 17.5 = 1500\text{Hz}$
- $F_3 = c/\lambda_3 = c/(4/5L) = 5c/4L = 5 \cdot 35,000/4 \cdot 17.5 = 2500\text{Hz}$
- So we expect a neutral vowel to have 3 resonances at 500, 1500, and 2500 Hz
- These vowel resonances are called **formants**

1/6/05

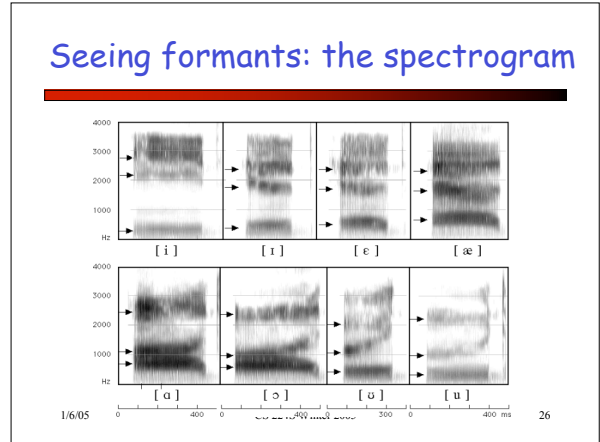
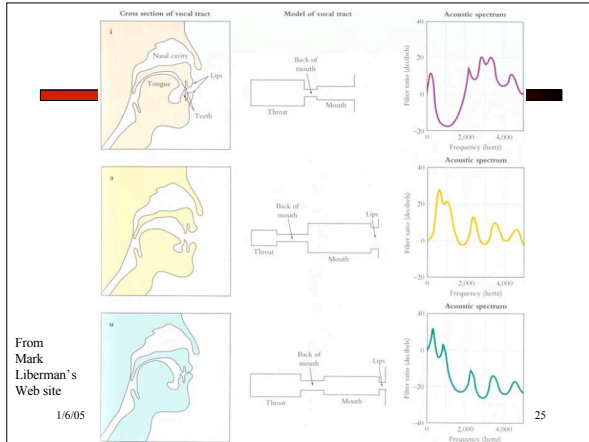
CS 224S Winter 2005

23

## Different vowels have different formants

- Vocal tract as "amplifier"; amplifies different frequencies
- Formants are result of different shapes of vocal tract.
- Any body of air will vibrate in a way that depends on its size and shape.
- Air in vocal tract is set in vibration by action of vocal cords.
- Every time the vocal cords open and close, pulse of air from the lungs, acting like sharp taps on air in vocal tract,
- Setting resonating cavities into vibration so produce a number of different frequencies.

24



## Formants

- Vowels largely distinguished by 2 characteristic pitches.
- One of them (the higher of the two) goes downward throughout the series iy ih eh ae aa ao ou u (whisper iy eh uw)
- The other goes up for the first four vowels and then down for the next four.
  - creaky voice iy ih eh ae (goes up)
  - creaky voice aa ow uh uw (goes down)
- These are called "formants" of the vowels, lower is 1st formant, higher is 2nd formant.

1/6/05 CS 224S Winter 2005 27

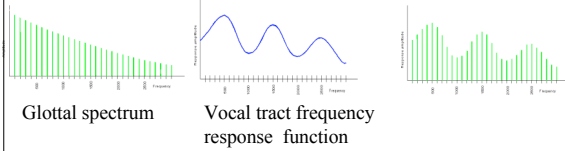
## How formants are produced

- Q: Why do vowels have different pitches if the vocal cords are same rate?
- A: This is a confusion of frequencies of SOURCE and frequencies of FILTER!

1/6/05 CS 224S Winter 2005 28

## Remember source-filter model of speech production

Input → Filter → Output



Source and filter are independent, so:  
 Different vowels can have same pitch  
 The same vowel can have different pitch

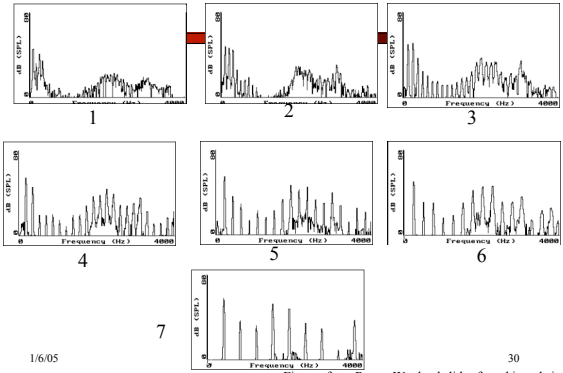
1/6/05

CS 224S Winter 2005

29

Figures and text from Ratreay Wayland slide from his website

Vowel [i] sung at successively higher pitch.

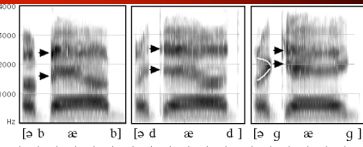


1/6/05

30

Figures from Ratreay Wayland slides from his website

## How to read spectrograms



- **bab**: closure of lips lowers all formants: so rapid increase in all formants at beginning of "bab"
- **dad**: first formant increases, but F2 and F3 slight fall
- **gag**: F2 and F3 come together: this is a characteristic of velars. Formant transitions take longer in velars than in alveolars or labials

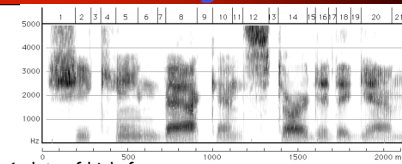
1/6/05

CS 224S Winter 2005

31

From Ladefoged "A Course in Phonetics"

## She came back and started again



1. lots of high-freq energy
3. closure for k
4. burst of aspiration for k
5. ey vowel; faint 1100 Hz formant is nasalization
6. bilabial nasal
7. short b closure, voicing barely visible.
8. ae; note upward transitions after bilabial stop at beginning
9. note F2 and F3 coming together for "k"

1/6/05

CS 224S Winter 2005

32

From Ladefoged "A Course in Phonetics"

