


HCS 7367  
Speech Perception Lab



Dr. Peter Assmann  
Fall 2011

## Class web page

<http://www.utdallas.edu/~assmann/hcs7367/>

- Course information
- Lab details
- Speech demos
- Outside readings
- Additional resources

## Software requirements

✓ *MATLAB & Simulink Student Version 2011a.*  
*The MathWorks* (required software).



✓ *Praat: doing phonetics by computer.* Boersma &  
Weenink, *University of Amsterdam* (free download).



✓ *Wavesurfer.* Kåre Sjölander and Jonas Beskow,  
*KTH Sweden* (free download).



## Matlab Background

Kermit Sigmon, *MATLAB Primer 2nd Edition.*

<http://www.fi.uib.no/Fysisk/Teori/KURS/WRK/mat/singlemat.html>

Getting started with Matlab (The MathWorks):

[http://www.mathworks.com/help/techdoc/learn\\_matlab/bqr\\_2pl.html](http://www.mathworks.com/help/techdoc/learn_matlab/bqr_2pl.html)

UTD IR – Matlab and Simulink: Resources for Getting Started

[http://www.utdallas.edu/ir/how-to/ml\\_help/index.html](http://www.utdallas.edu/ir/how-to/ml_help/index.html)

## Praat : doing phonetics by computer

- Download Praat:  
<http://www.fon.hum.uva.nl/praat/>
- Praat tutorial:  
<http://www.fon.hum.uva.nl/praat/manual/Intro.html>

## Wavesurfer

- Download Wavesurfer:  
[www.speech.kth.se/wavesurfer](http://www.speech.kth.se/wavesurfer)
- Wavesurfer User Manual  
[www.speech.kth.se/wavesurfer/man.html](http://www.speech.kth.se/wavesurfer/man.html)

## Starting with Matlab

- **Interactive MATLAB Tutorial**  
[http://www.mathworks.com/help/techdoc/learn\\_matlab/f0-11759.html](http://www.mathworks.com/help/techdoc/learn_matlab/f0-11759.html)  
[http://www.mathworks.com/academia/student\\_center/tutorials/ml\\_onramp/player.html?slide=1](http://www.mathworks.com/academia/student_center/tutorials/ml_onramp/player.html?slide=1)
- **Start Matlab**
  - doc Matlab
  - Click on "Getting Started"
  - This launches a video in your browser

## Recommended Books

- ✓ W.M. Hartmann (1996). **Signals, sound and sensation.** (Springer-Verlag).
- ✓ Kent, R.D. & Read, C. (2001). **The Acoustic Analysis of Speech.** (Singular Press).



## Recommended Books

- ✓ B. Gold & N. Morgan (2000). **Speech and Audio Signal Processing: Processing and perception of speech and music.** Wiley (ISBN: 0-471-35154-7).



## Recommended Books

- ✓ Ian McLoughlin (2009). **Applied Speech and Audio Processing: With Matlab Examples.** Cambridge University Press (ISBN: 978-0521519540).



## Course Requirements

- Matlab problems & lab assignments (40%)
- Oral report on term project (10%)
- Term project paper (50%)

## Dates for lab assignments

- Lab assignment 1: **Sept 15**
- Lab assignment 2: **Oct 6**
- Lab assignment 3: **Oct 27**
- Lab assignment 4: **Nov 17**
- 3-5 page written reports on lab projects

## Term project: important dates

**Aug 20-27:** Pick a topic

**Sep 22:** Preliminary project presentation

**Sep 29:** Turn in project outline

**Nov 17/24:** Oral presentations

**Dec 10:** Final project paper due

## Examples of topics

- Acoustic analysis and intelligibility of children's speech
- Neural network models of vowel recognition
- Simulating distortions introduced by hearing loss
- Noise reduction algorithms for hearing aid processors
- Production and perception of foreign accents
- Contribution of prosody to connected speech intelligibility
- Effects of noise, reverberation on speech communication
- Monaural vs. binaural speech understanding in noise
- Development of speech perception in infants
- Models of speech coding in the auditory cortex

## Project grading

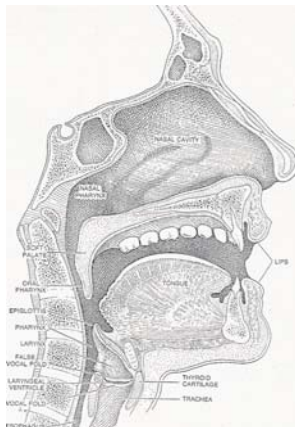
- **Project organization** (35%)
- **Technical content** (35%)
- **Communication of ideas** (30%)

## Initial stages

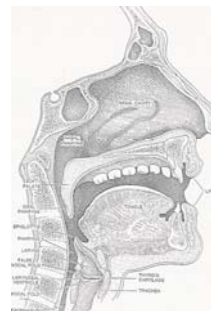
- Identify a topic area and read relevant papers
- Refine your topic; choose a manageable problem
- Set specific goals and define evaluation metric
- Identify the approach to solve the problem
- Start right away.

## Acoustics of speech

- Phonation
- Articulation

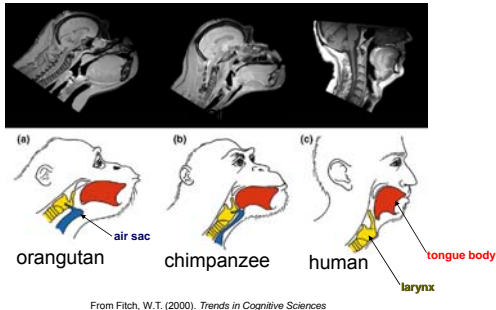


## Organs of speech



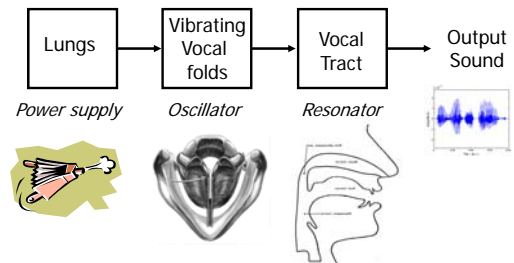
- **Lungs:** apply pressure to generate air stream (power supply)
- **Larynx:** air forced through the glottis, a small opening between the vocal folds (sound source)
- **Vocal tract:** pharynx, oral and nasal cavities serve as complex resonators (filter)

## Human vocal tract



From Fitch, W.T. (2000). *Trends in Cognitive Sciences*

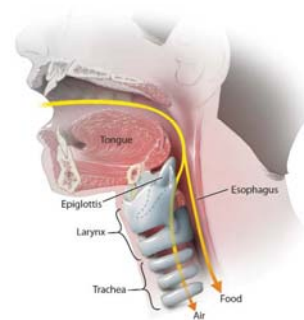
## Source-filter theory of speech production Fant (1960)



## Source-Filter Theory: Vowels

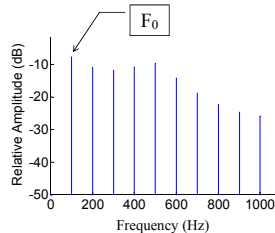
- G. Fant (1960). *Acoustic Theory of Speech Production*
- Linear systems theory
- Assumptions: (1) linearity (2) time-invariance
- Vowels can be decomposed into two primary components: a **source** (input signal) and a **filter** (modulates the input).

## Human vocal tract



## Source properties

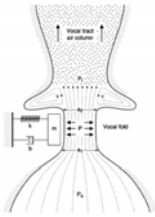
- In **voiced** sounds the glottal source spectrum contains a series of lines called **harmonics**.
- The lowest one is called the **fundamental frequency** ( $F_0$ ).



## Source properties: Pitch

- **Fundamental frequency** ( $F_0$ ) is determined by the rate of vocal fold vibration, and is responsible for the perceived voice **pitch**.

## Vocal fold oscillation

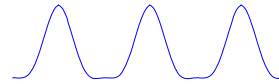


- One-mass model
  - Air flow through the glottis during the closing phase travels at the same speed because of inertia, producing lowered air pressure above the glottis.

Source: <http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/model.html>

## Audio demo: the source signal

- Source signal for an adult male voice 🗣️
- Source signal for an adult female voice 🗣️
- Source signal for a 10-year child 🗣️

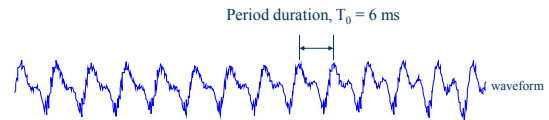


## Source properties: Pitch

- $F_0$  can be removed by **filtering** (as in telephone circuits) and the pitch remains the same.
- This is the **problem of the missing fundamental**, one of the oldest problems in hearing science.
- Pitch is determined by the frequency pattern of the harmonics (or their equivalent in the time domain, the periodicities in the signal).

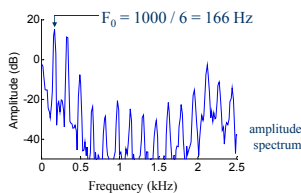
## Harmonicity and Periodicity

- **Period**: regularly repeating pattern in the waveform



## Harmonicity and Periodicity

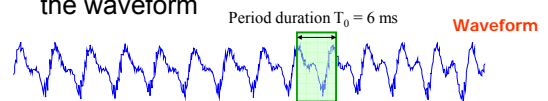
- **Harmonic**: regularly repeating peak in the amplitude spectrum



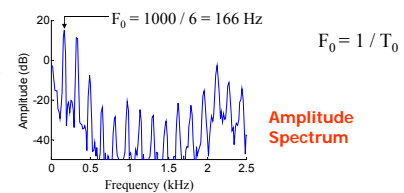
$$F_0 = 1 / T_0$$

## Harmonicity and Periodicity

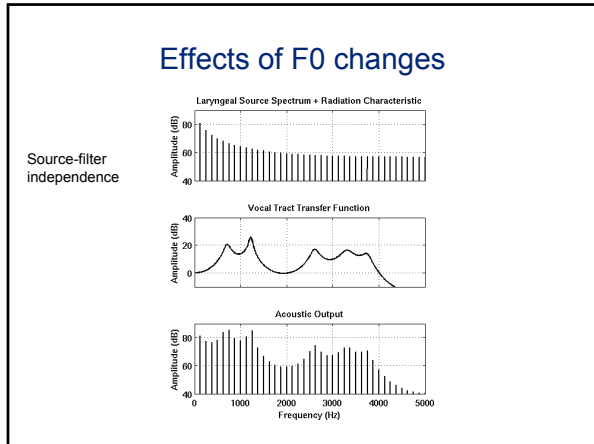
- **Period**: regularly repeating pattern in the waveform



Harmonics are integer multiples of  $F_0$  and are evenly spaced in frequency

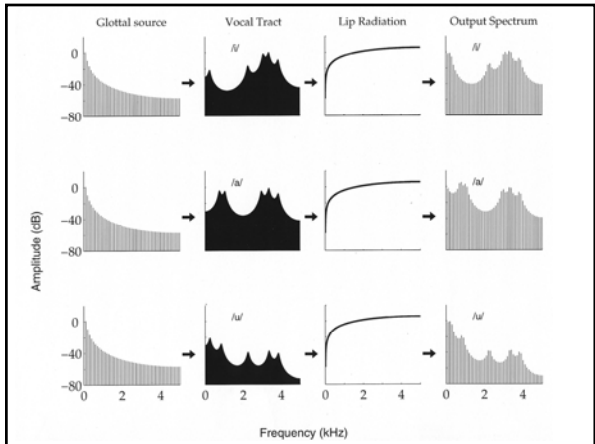
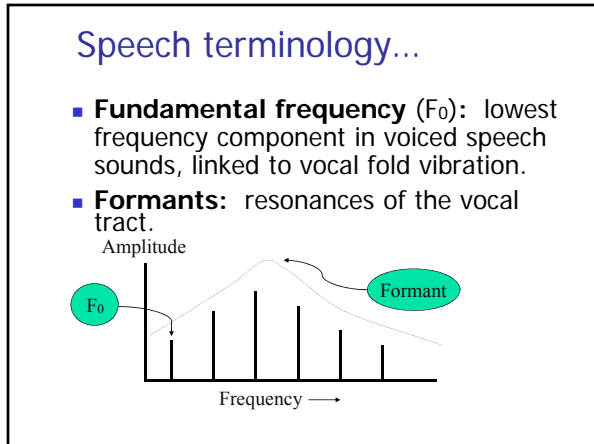
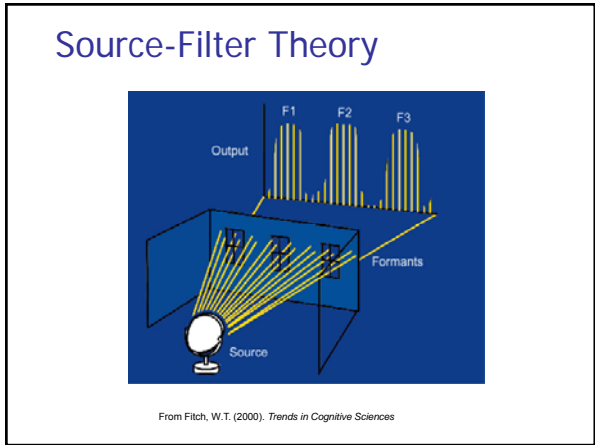


$$F_0 = 1 / T_0$$

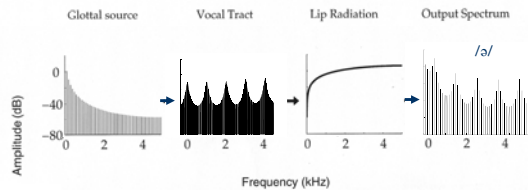


- ### Voicing irregularities
- **Shimmer**: variation in amplitude from one cycle to the next.
  - **Jitter**: variation in frequency (period duration) from one cycle to the next.

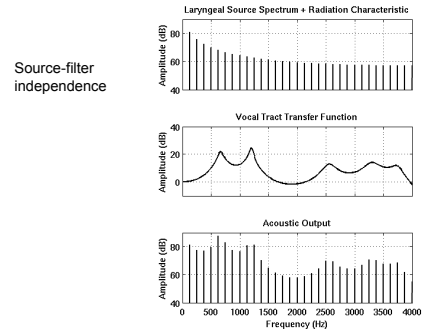
- ### Voicing irregularities
- **Breathy voice** is associated with a glottal waveform with a steeper roll-off than modal voice. As a result there is less energy in the higher harmonics (steeper slope in the spectrum).



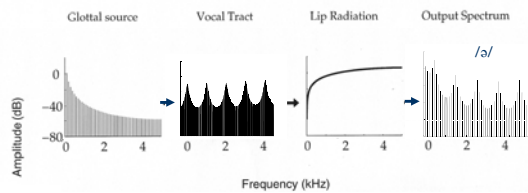
## Uniform tube model (schwa)



## Effects of formant frequency changes



## Uniform tube model (schwa)



## Vocal tract model

### • Quarter-wave resonator:

$$F_n = (2n - 1) c / 4 L$$

- $F_n$  is the frequency of formant  $n$  in Hz
- $c$  is the velocity of sound (about 35000 cm/sec)
- $L$  is the length of the vocal tract (17.5 for adult male)

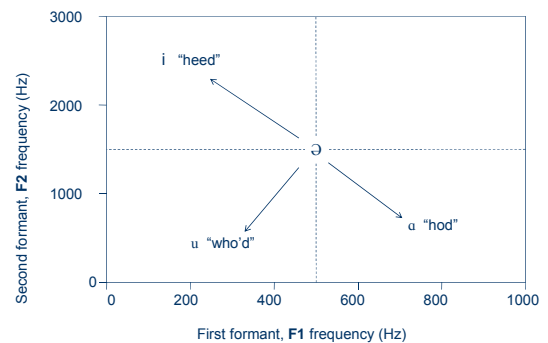
## Vocal tract model

### • Quarter-wave resonator:

$$F_n = (2n - 1) c / 4 L$$

- $F_1 = (2(1) - 1) * 35000 / (4 * 17.5) = 500$  Hz
- $F_2 = (2(2) - 1) * 35000 / (4 * 17.5) = 1500$  Hz
- $F_3 = (2(3) - 1) * 35000 / (4 * 17.5) = 2500$  Hz

## Acoustic vowel space

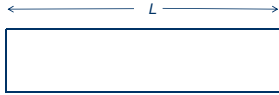


## Vocal tract model

- Quarter-wave resonator:

$$F_n = (2n - 1) c / 4 L$$

- $F_n$  is the frequency of formant  $n$  in Hz
- $c$  is the velocity of sound in air (about 35000 cm/sec)
- $L$  is the length of the vocal tract (17.5 for adult male)



## Vocal tract model

- Quarter-wave resonator:

$$F_n = (2n - 1) c / 4 L$$

- $F_1 = (2(1) - 1) * 35000 / (4 * 17.5) = 500$  Hz
- $F_2 = (2(2) - 1) * 35000 / (4 * 17.5) = 1500$  Hz
- $F_3 = (2(3) - 1) * 35000 / (4 * 17.5) = 2500$  Hz



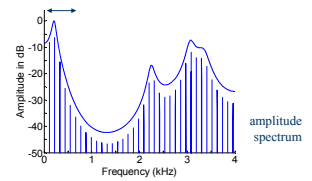
## Perturbation Theory

- The **first formant (F1)** frequency is lowered by a constriction in the front half of the vocal tract (*/u/* and */i/*), and raised when the constriction is in the back of the vocal tract, as in */a/*.



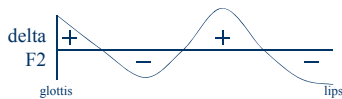
## Perturbation Theory

- F1 frequency is correlated with jaw opening (and inversely related to tongue height).



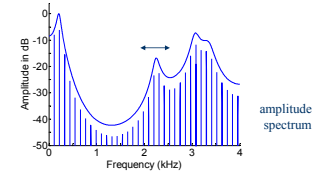
## Perturbation Theory

- The **second formant (F2)** is lowered by a constriction near the lips or just above the pharynx; in */u/* both of these regions are constricted. F2 is raised when the constriction is behind the lips and teeth, as in the vowel */i/*.



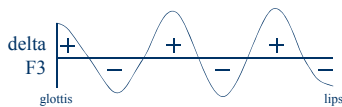
## Perturbation Theory

- F2 frequency is correlated with tongue advancement (front-back dimension)



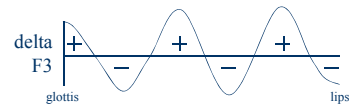
## Perturbation Theory

- The **third formant (F3)** is lowered by a constriction at the lips **or** at the back of the mouth **or** in the upper pharynx. This occurs in /r/ and /r/-colored vowels like American English /æ/.



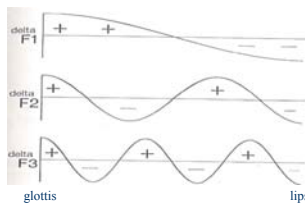
## Perturbation Theory

- F3 is raised when the constriction is behind the lips and teeth or near the upper pharynx.



## Perturbation Theory

- All formants** tend to drop in frequency when the vocal tract length is increased **or** when a constriction is formed at the lips.



## Helium speech

- The speed of sound in a helium/oxygen mixture at 20°C is about 93000 cm/s, compared to 35000 cm/s in air. This increases the resonance frequencies but has relatively little effect on F<sub>0</sub>. In helium speech, the formants are shifted up but the pitch stays the same.

## Helium speech

- Using Matlab as a calculator, find the frequencies of F<sub>1</sub>, F<sub>2</sub> and F<sub>3</sub> for a 17.5 cm vocal tract producing the vowel /ə/ in a helium/air mixture (velocity  $c \approx 93000$  cm/s)

$$F_n = (2n - 1) c / 4 L$$

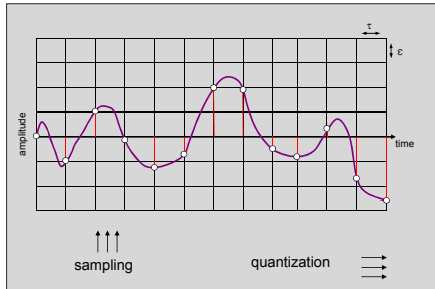
- >  $F_1 = (2(1) - 1) * 93000 / (4 * 17.5) =$
- >  $F_2 = (2(2) - 1) * 93000 / (4 * 17.5) =$
- >  $F_3 = (2(3) - 1) * 93000 / (4 * 17.5) =$

## Helium speech

- Audio demos**
  - Speech in air 🗣️
  - Speech in helium 🗣️
  - Pitch in air 🗣️
  - Pitch in helium 🗣️

[http://phys.unsw.edu.au/phys\\_about/PHYSICS!/SPEECH\\_HELIUM/speech.html](http://phys.unsw.edu.au/phys_about/PHYSICS!/SPEECH_HELIUM/speech.html)

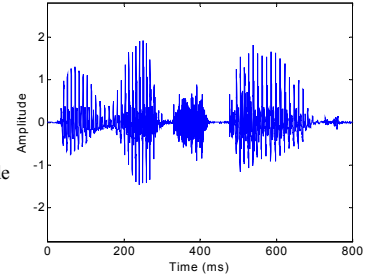
## Digital representations of signals



## Digital representations of speech

- Speech waveform: representation of the amplitude variations in the signal as a function of time.

- **Amplitude quantization**
- **Time sampling**
- In the example, the sample rate is **8 kHz** (8000 samples/s) with **16-bit** amplitude quantization.



## Vector representation of speech

- In Matlab speech signals are represented as **row** or **column vectors** (e.g., N rows x 1 columns, where N is the number of samples in the waveform).

```

» load sent1.mat
» size( sent1 )
ans =
    15043     1
» x=(1:length(y))./(rate/1000);
» plot(x,y);
» axis( [ 0 800 -28000 28000 ] );
» xlabel('Time (ms)');
» ylabel('Amplitude');
    
```

## Fourier analysis and synthesis

- D.P.W. Ellis (2009). **An introduction to signal processing for speech**. In *The Handbook of Phonetic Science*, , 2<sup>nd</sup> edition, edited by Hardcastle, Laver, and Gibbon. chapter 22, pp. 757-780, Blackwell.

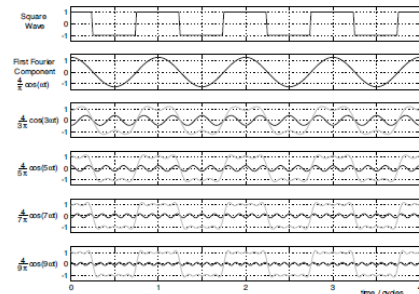


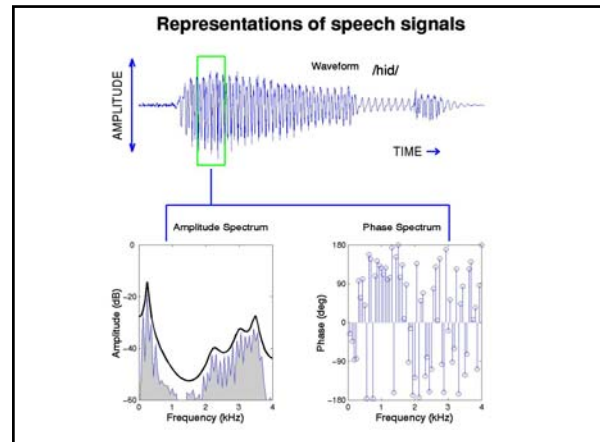
Figure 6: Illustration of how any periodic function can be approximated by a sum of harmonics i.e. sinusoids at integer multiples of the fundamental frequency of the original waveform. Top pane shows the target waveform, a square wave. Next panes show the first five harmonics; in each pane, the dark curve is the sinusoid, and the light curve is the cumulative sum of all harmonics so far, showing how the approximation comes increasingly close to the target signal. Ellis (2009, p. 12)

## Spectral analysis

- **Amplitude spectrum**: sound pressure levels associated with different frequency components of a signal
  - Power or intensity
  - Amplitude or magnitude
  - Log units and decibels (dB)
- **Phase spectrum**: relative phases associated with different frequency components
  - Degrees or radians

## Frequency domain representation

- Why perform spectral analyses of speech?
  - The ear+brain carry out a form of frequency analysis
  - The relevant features of speech are more readily visible in the amplitude spectrum than in the raw waveform
  - BUT: the ear is not a Fourier analyzer.
  - Fourier analysis provides **amplitude** and **phase spectrum**; speech cues have been mainly associated with the amplitude spectrum. However, the ear is not "phase-deaf"; many phase changes are clearly audible.
  - Frequency selectivity is greatest for frequencies below 1 kHz and declines at higher frequencies



## Spectral analysis in Matlab

- Amplitude spectrum of a vector:
  - » `X= fft (y);`
  - » `help fft`

FFT Discrete Fourier transform.

FFT(X) is the discrete Fourier transform (DFT) of vector X. If the length of X is a power of two, a fast radix-2 fast-Fourier transform algorithm is used. If the length of X is not a power of two, a slower non-power-of-two algorithm is employed. For matrices, the FFT operation is applied to each column.

FFT(X,N) is the N-point FFT, padded with zeros if X has less than N points and truncated if it has more.

## Spectral analysis in Matlab

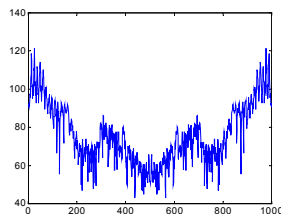
- Log magnitude (amplitude) spectrum:
  - » `X= fft (y);`
  - » `m = 20 * log10 ( abs ( X ) );`
  - » `help abs`

ABS Absolute value.

ABS(X) is the absolute value of the elements of X. When X is complex, ABS(X) is the complex modulus (magnitude) of the elements of X.

## Spectral analysis in Matlab

- Log magnitude (amplitude) spectrum:
  - » `plot(20*log10(abs(fft(y))))`



## Plotting amplitude spectra

- » `help fp`

FP: function to compute & plot amplitude spectrum

Usage: `[a,f]=fp(wave,n,rate,string>window);`

**wave**: input waveform

**n**: if unspecified, `n=length(wave)`; padded with zeros if necessary

**rate**: sample rate in Hz (default 10000 Hz)

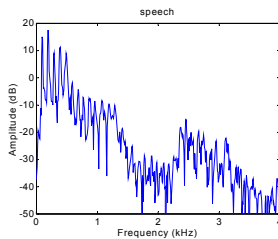
**string**: title for graph (default none)

**window** options: 'hann', 'hamm', 'kaiser', or 'rect' (default=hanning)

**[a,f]**: FFT log magnitude, frequency

## Plotting amplitude spectra

```
» [a,f]=fp(wave,n,rate,string>window);
» [a,f]=fp(y,1000,rate,'speech','hann');
```



## Speech spectrograms

- What is a speech spectrogram?
  - Display of amplitude spectrum at successive instants in time ("running spectra")
  - How can 3 dimensions be represented on a two-dimensional display?
    - Gray-scale spectrogram
    - Waterfall plots
    - Animation
- Why are speech spectrograms useful?
  - Shows dynamic properties of speech
  - Includes frequency analysis

## Speech spectrograms in Matlab

```
» help specgram
```

SPECGRAM Calculate spectrogram from signal.

**B = SPECGRAM(A,NFFT,Fs,WINDOW,NOVERLAP)** calculates the spectrogram for the signal in vector A.

SPECGRAM splits the signal into overlapping segments, windows each with the WINDOW vector and forms the columns of B with their zero-padded, length NFFT discrete Fourier transforms.

## Speech spectrograms in Matlab

```
» help sp
```

**sp:** create gray-scale spectrogram

Usage: **h=sp(wave,rate,nfft,nsamp,nhop,pre,drng);**

**wave:** input waveform

**rate:** sample rate in Hz (default 8000 Hz)

**nfft:** FFT window length (default: 256 samples)

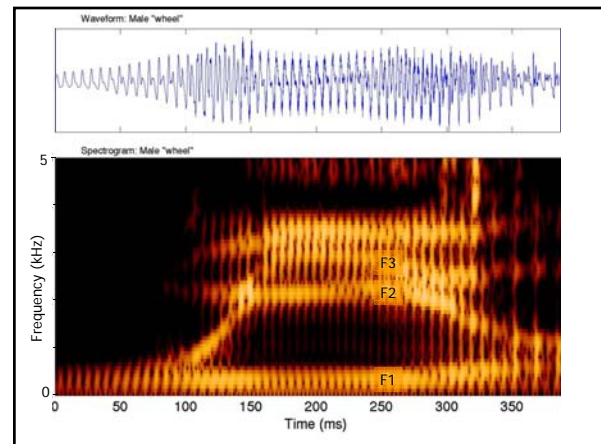
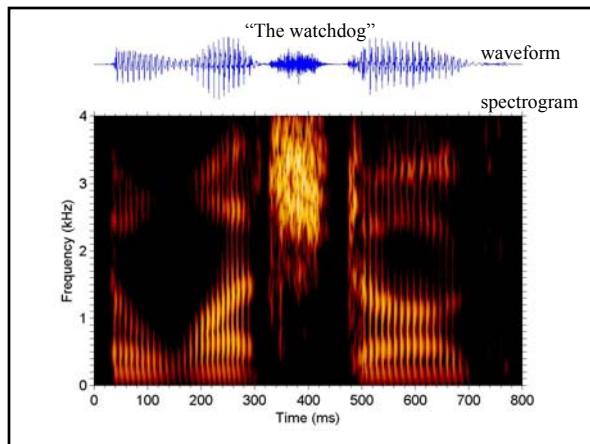
**nsampf:** number of samples per frame (default: 60)

**nhop:** number of samples to hop to next frame (default: 5 samples)

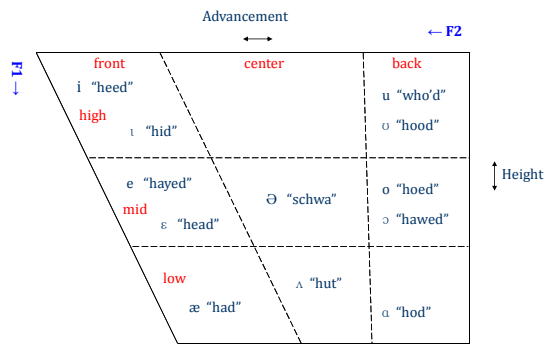
**pre:** preemphasis factor (0-1) (default: 1)

**drng:** dynamic range in dB (default: 80)

**title:** title for graph (default: none)



## American English vowel space



## Assignment 1

### Part 1: (Matlab code, plots, brief summary)

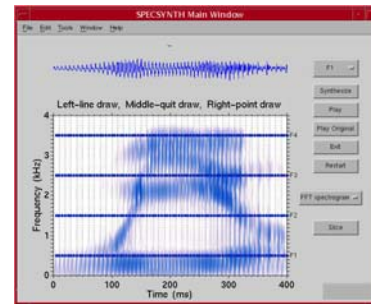
- Make a set of digital recordings (WAV files) of the 12 vowels of American English:

/i/ "heed"	/ɪ/ "hid"	/e/ "hayed"	/ɛ/ "head"
/æ/ "had"	/ʌ/ "hud"	/ɑ/ "hod"	/ɔ/ "hawed"
/o/ "hoed"	/u/ "hood"	/u/ "who'd"	/ɜ/ "herd"

## Assignment 1

- Load waveforms into Matlab; make 12 subplots of the amplitude spectra of the vowels, sampled near the midpoint.
  - » `[ y, fs ] = wavread ('heed.wav');`
  - » `subplot (4,3,1);`
  - » `start = ( length ( y ) / 2 ) - 256;`
  - » `stop = ( length ( y ) / 2 ) + 256;`
  - » `fp ( y ( start : stop ) , 512 , fs, 'heed.wav', 'Hamming');`

### TrackDraw: a graphical speech synthesizer



### TrackDraw: a graphical speech synthesizer

