

Subjective Dimensions of Sound

- Location
- Duration
- Timbre
- Loudness
- Density
- Dissonance
- PITCH

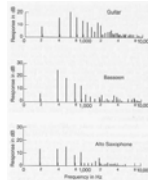
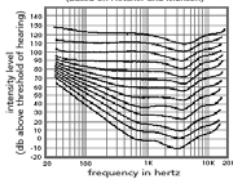
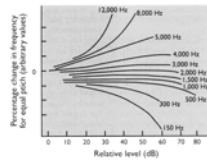


Figure 12.8
Relative amplitudes of the harmonics of a guitar, a horn, and an alto saxophone, along with the relative amplitudes of the harmonics of a sine wave. The position of the lines on the horizontal axis indicates the frequency of the harmonics and their relative amplitudes.

Pitch

- Frequency (Hz)?
- Equal Pitch Contours
- Equal Loudness Contours
- Missing Fundamental



Theories

- Place Principle
 - traveling wave
 - tonotopic organization
- Frequency Principle
 - refractory period limitation
 - volley principle

Which theory is correct?

- Both?
- Place Principle (500 – 20,000Hz)
- Frequency Principle (< 4,000Hz)

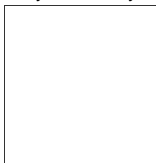
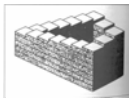
Music Perception

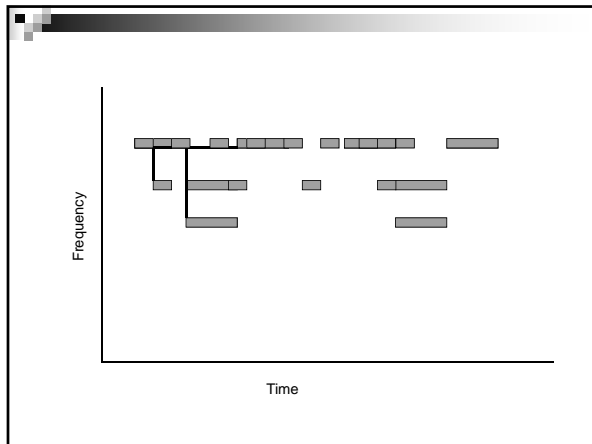
- height – perceived pitch
- chroma - identity within octave

Music Perception

- similar to visual perception

- depiction in space
- Tonal Staircase Illusion
- Grouping principles
 - good continuation – masking
 - proximity – auditory stream segmentation





Speech Recognition

- computers
 - filtering signal from noise
 - sloppy pronunciation
 - accents
 - gender
 - age
 - speed
- how far have we come?
 - Friends, Romans, countryman, lend me your ears
 - Friends, Romans, countryman, linear years

Units of Speech

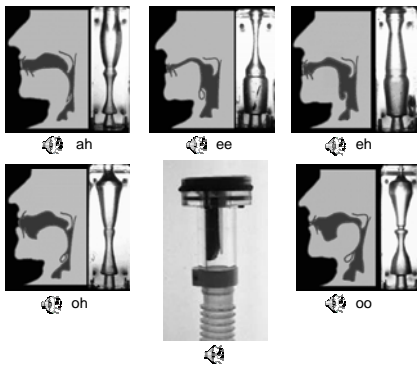
- sentence?
- letter?
- phoneme?
 - shortest segment of speech that if changed, changes the meaning of a word
 - /b/, /i/, /t/
 - approx. 47 in English

Analyzing Speech

- patterns of pressure changes
- shape of vocal tract
- articulators – tongue, lips, jaw, teeth, soft palate
- two main types of phonemes
 - vowels and consonants

Vowels

- vibrations of vocal chords
- shape of vocal tract > different resonant frequencies



Vowels

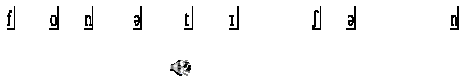
- vibrations of vocal chords
- shape of vocal tract > different resonant frequencies
- frequency peaks called formants
- first formant lowest frequency, etc.

Consonants

- formant transitions
- rapid shifts in frequency preceding or following formants

Oscillogram

Pressure

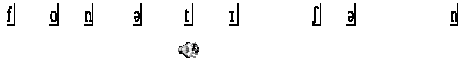


Pitch Analysis

- fundamental (dominant) frequency (f_0)
- Males : 80 -200Hz
- Females: 150-350Hz

Fundamental Frequency

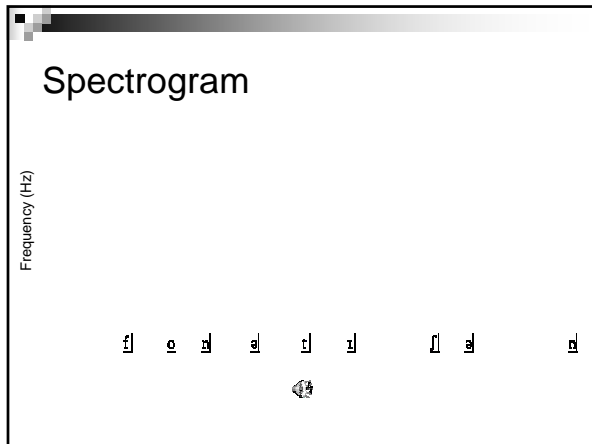
Frequency (Hz)

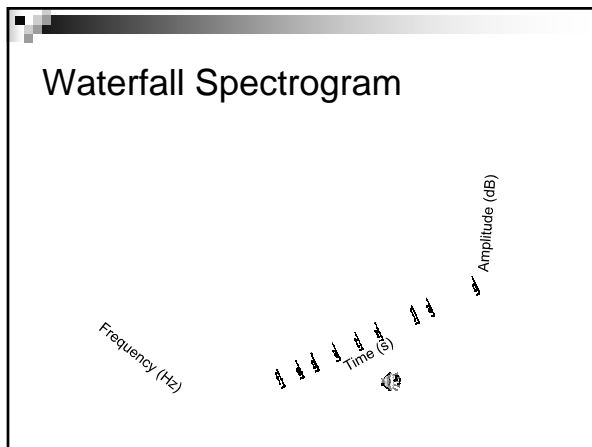


Spectrum

Amplitude (dB)

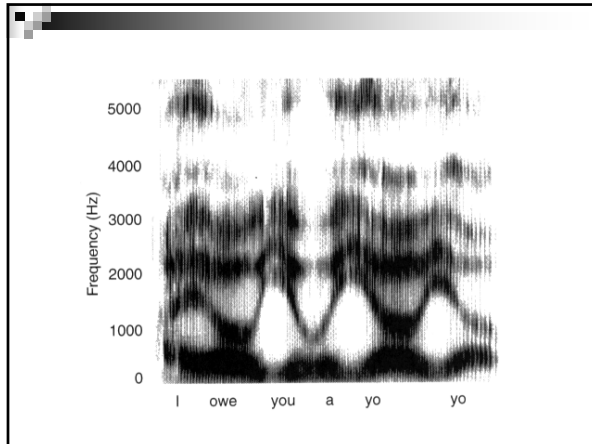
Frequency (Hz)

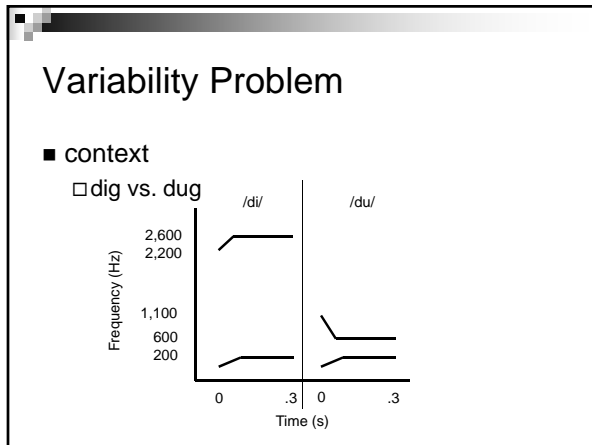




Segmentation Problem

- how do we segment sound into words?
 - look for breaks in sound stimulus?





Coarticulation

- when we say the word "happy"
- before you say anything tongue has moved into position to make the "a" sound
- "h" will sound a little like an "a"
- while saying "a" closing lips for "pp"
- spreads out vowel and consonant information to aid understanding
- allows us to communicate at a rate of about 5 syllables/second

Variability Problem

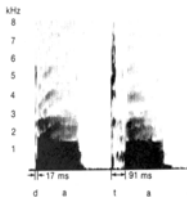
- different speakers
 - fast vs. slow speakers
 - accents
 - sloppy pronunciation
 - "This was the best buy."
 - "Did you go to the store?"
 - "What are you doing?"

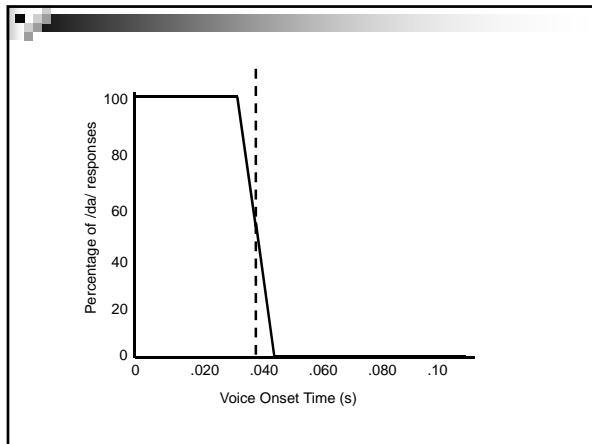
Acoustic-Phonetic Invariance

- there must be some constant set of acoustic features associated with each perceived phoneme

Categorical Perception

- creates two categories of sounds from a wide range of acoustic signals
- Voice Onset Time (VOT) – /da/ (17 ms) vs. /ta/ (91 ms)
 - phonemic boundary 35-40 ms





Multimodal Information

- McGurk effect
 - information from visual domain is integrated with information from auditory domain to assist speech perception

Top-Down Influences

- Read the following sentences:
 M*R* H*D * L*TTL* L*MB I*S FL**C* W*S
 WH*T* *S SN*W

 S*M* W**DS *R* EA*I*R T* U*D*R*T*N* T*A*
 *T*E*S

Meaning and Segmentation

- What do you hear?

Meaning and Phoneme Perception

- What do you hear?
- phonemic restoration effect
- "There was time to *ave...
- rave? save? wave? shave?

Knowledge of Language

- if things are hard to make out (noise, accents)
 - meaningful grammatical sentences > non-meaningful grammatical sentences > ungrammatical strings of words
 - permissible word structures (e.g. TAN) > non-permissible word structures (e.g., TQN)

Theories of Speech Perception

- passive (data-driven) vs. active (conceptually-driven)
- passive – feature detectors/template matching (Pandemonium-like models)
- active – cohort theory – passive used to establish cohort, fit with meaning, etc. used to eliminate possibilities
 - trace theory – nodes activate all nodes connected higher and lower in network
