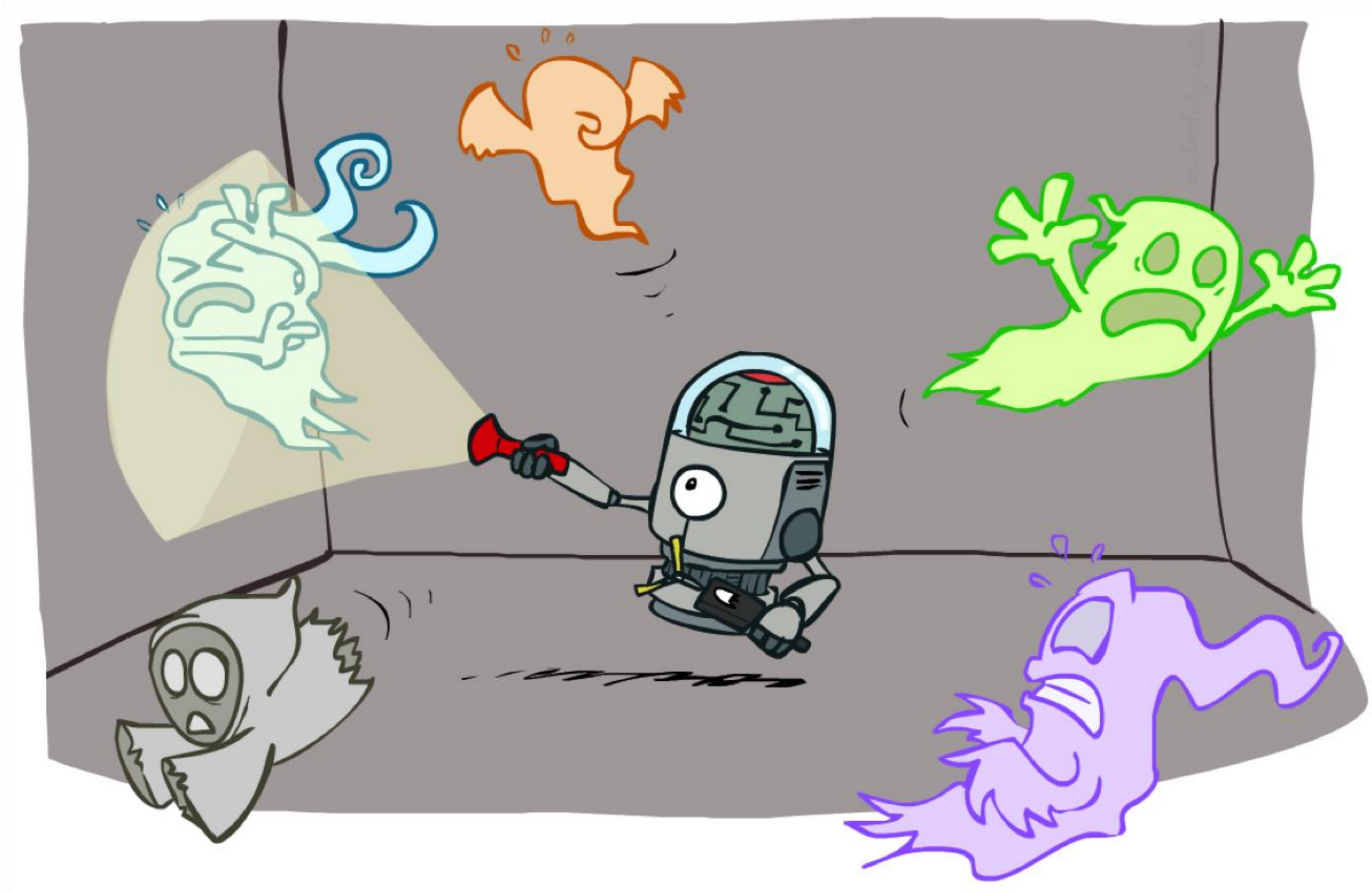
Applications of HMMs



[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]

Today

- Applications:
  - “I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis”
  - Speech recognition

AI in the News

I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis
Brad Miller, Ling Huang, A. D. Joseph, J. D. Tygar (UC Berkeley)

Challenge

- Setting
  - User we want to spy on use HTTPS to browse the internet
- Measurements
  - IP address
  - Sizes of packets coming in
- Goal
  - Infer browsing sequence of that user

- E.g.: medical, financial, legal, …

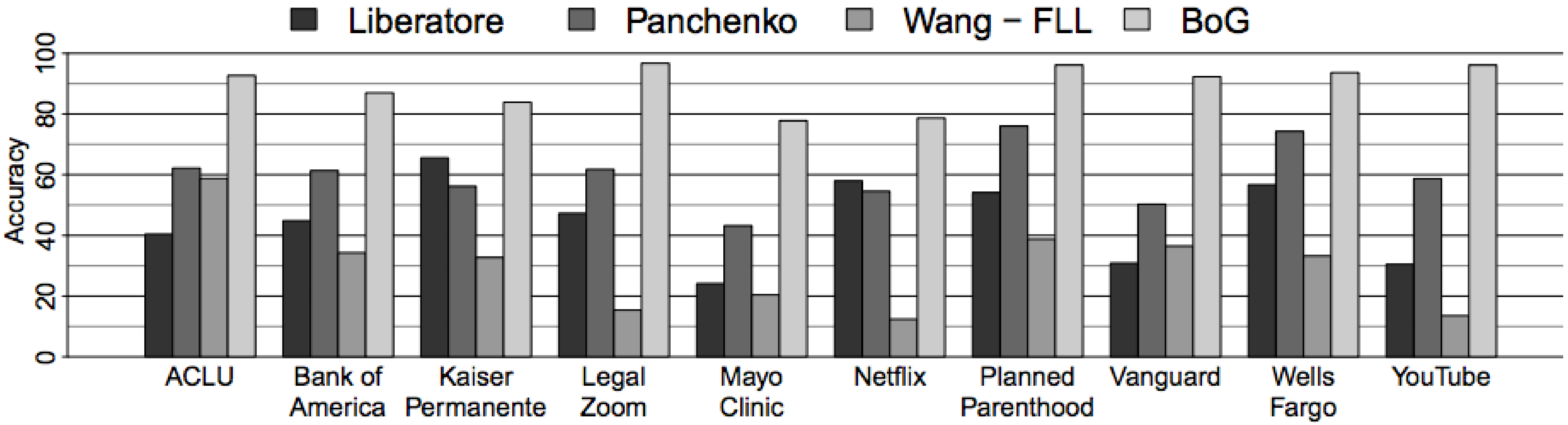
HMM

- Transition model
  - Probability distribution over links on the current page + some probability to navigate to any other page on the site

- Noisy observation model due to traffic variations
  - Caching
  - Dynamically generated content
  - User-specific content, including cookies

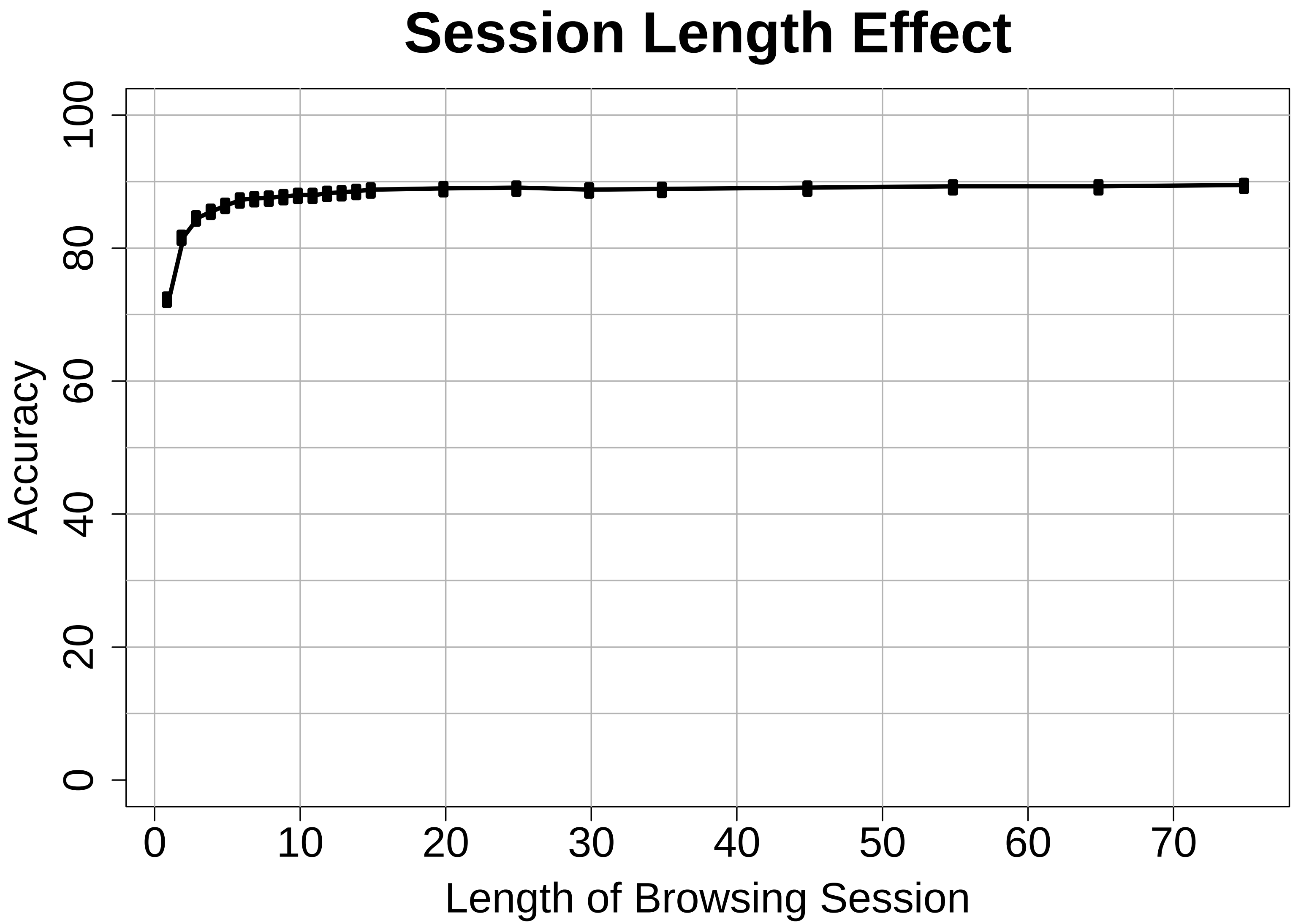
  → Probability distribution P(packet size | page)

Results

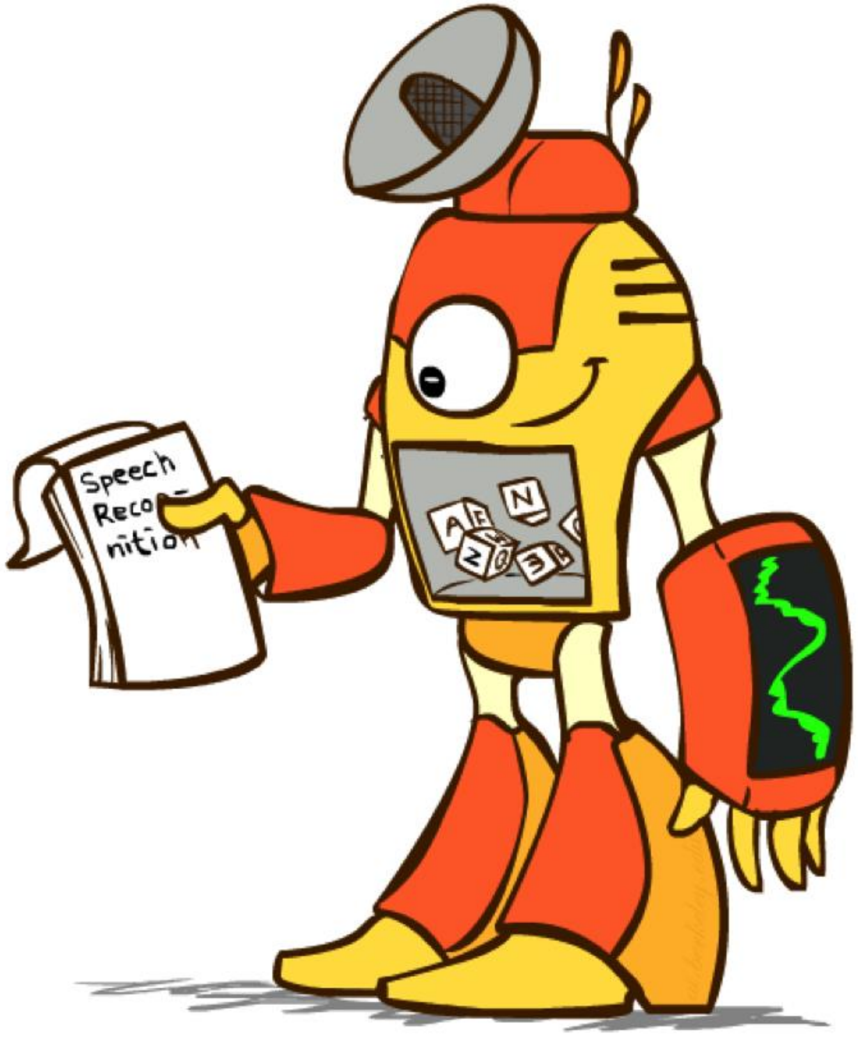


BoG = described approach, others are prior work

Results


Session Length Effect
Accuracy
0
20
40
60
80
100
0
10
20
30
40
50
60
70
Length of Browsing Session

Speech Recognition



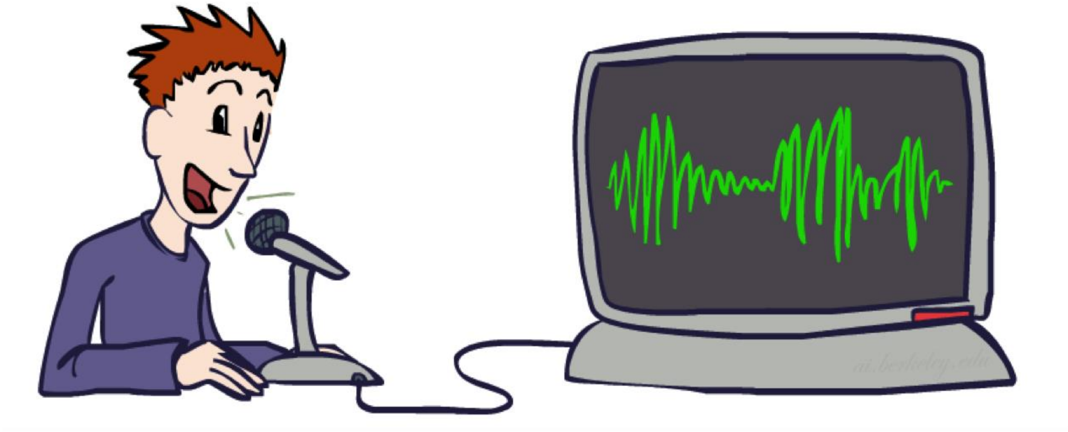
Hello, this is charter. We're trying to reach you to confirm your appointment scheduled tomorrow. Saturday. Yep, timber 20th, between 5 PM, and 7 PM, service calls typically take up to 1 hour to complete. After our technician arrives during the scheduled appointment window if you will receive another call from us tomorrow. 45 minutes prior to your scheduled appointment. If you do not confirm your appointment. At that time, your service call may be cancelled. Please call us at 1(800) 747-8134 if you wish to make any changes to your appointment. Thank you. We look forward to resolving your service issue, goodbye.

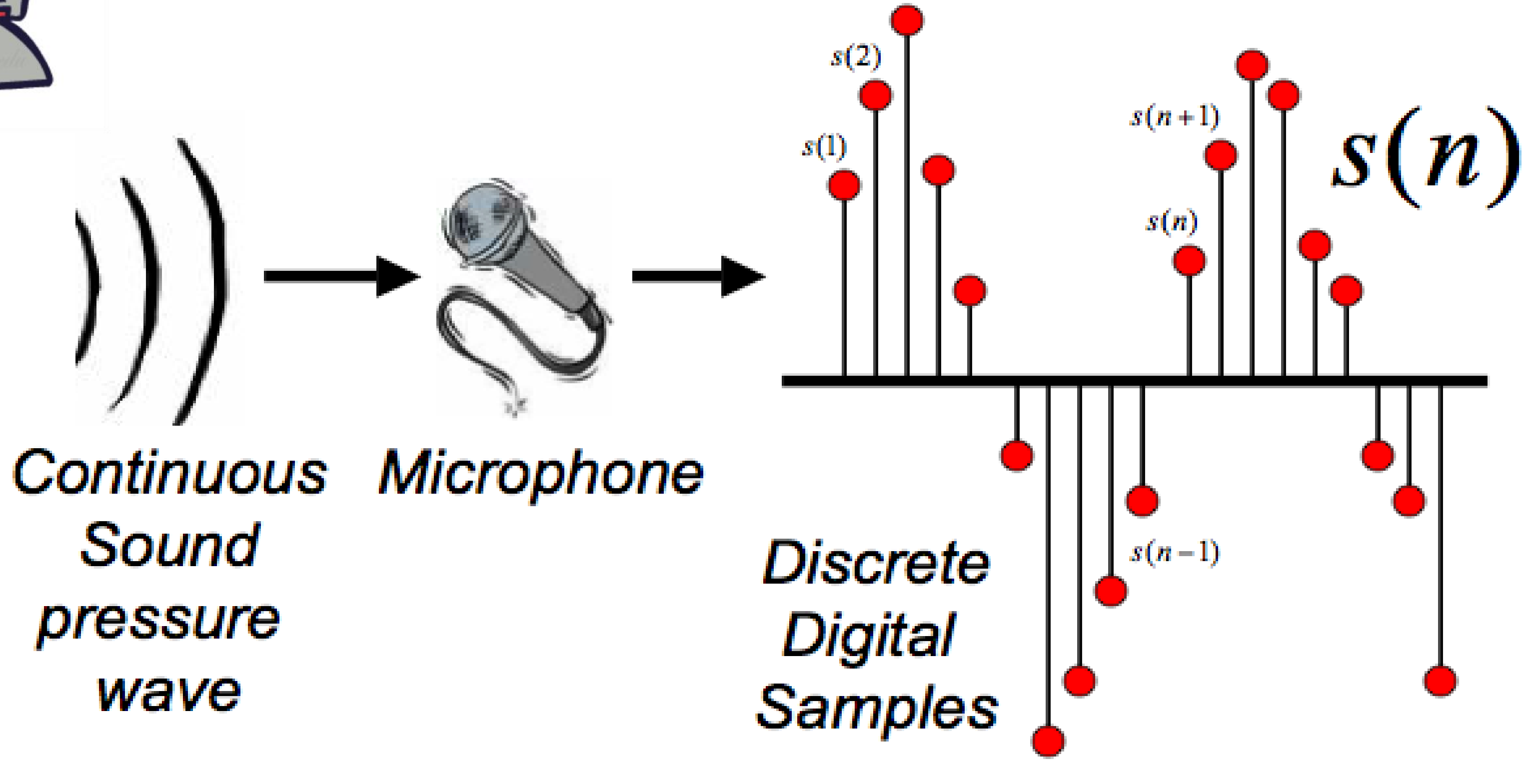
Play message

And you're a hard died in mail bone. But you'd be able more excitable listening Stone A at my place soul. Hot no ready for you to test it all here. One of these days ago but.

Play message

Digitizing Speech





Thanks to Bryan Pellom for this slide!

Speech in an Hour

- Speech input is an acoustic waveform

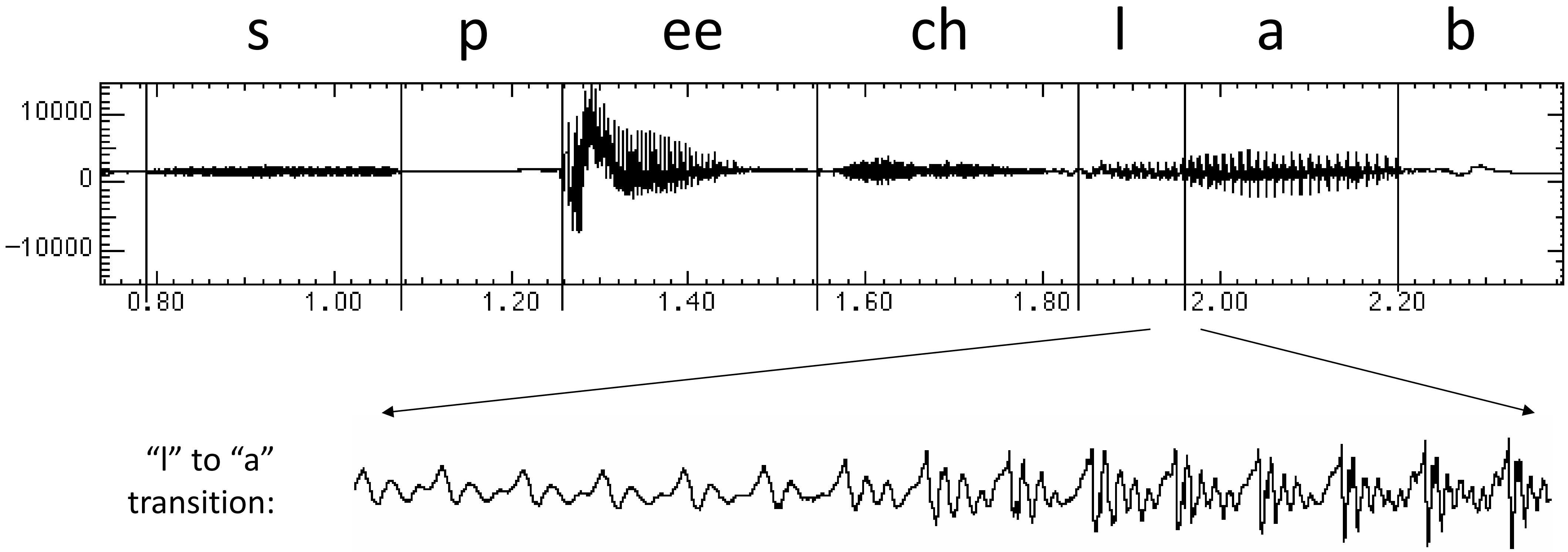
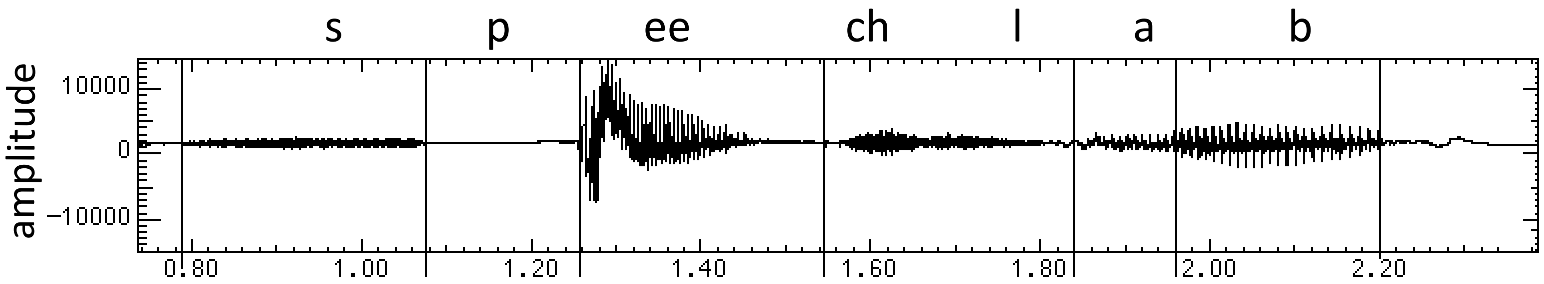
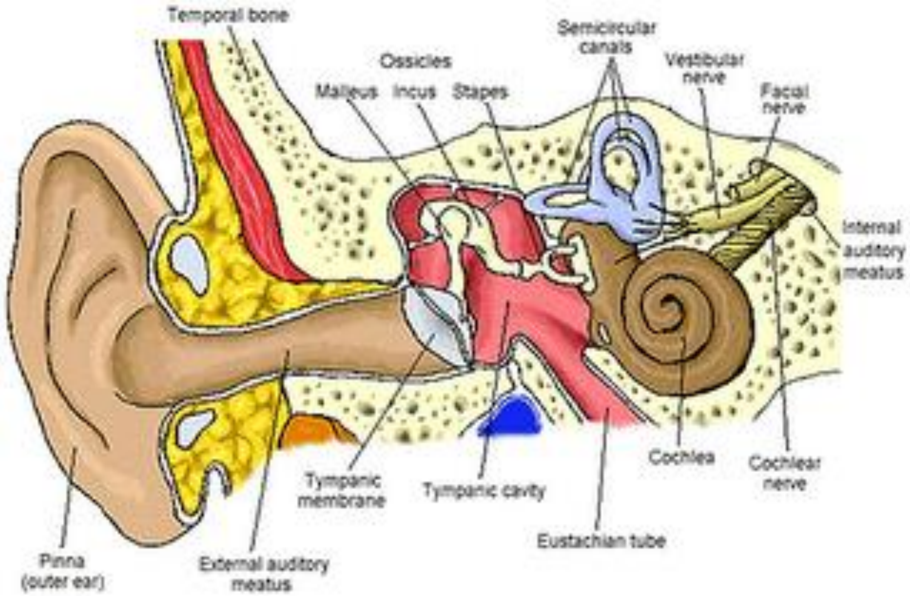


Figure: Simon Arnfield, http://www.psyc.leeds.ac.uk/research/cogn/speech/tutorial/

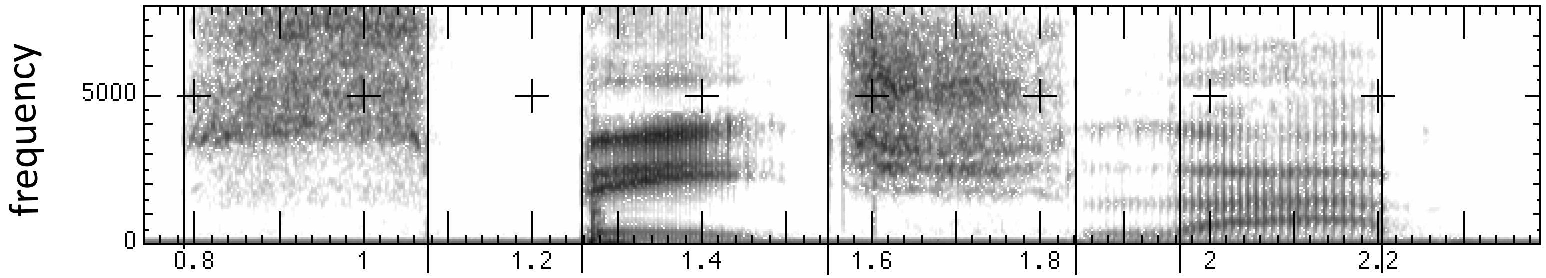
Spectral Analysis

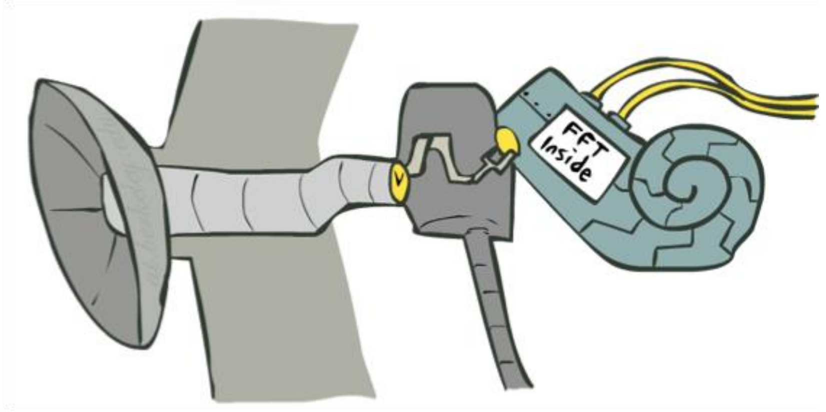
- Frequency gives pitch; amplitude gives volume
  - Sampling at ~8 kHz (phone), ~16 kHz (mic) (kHz=1000 cycles/sec)





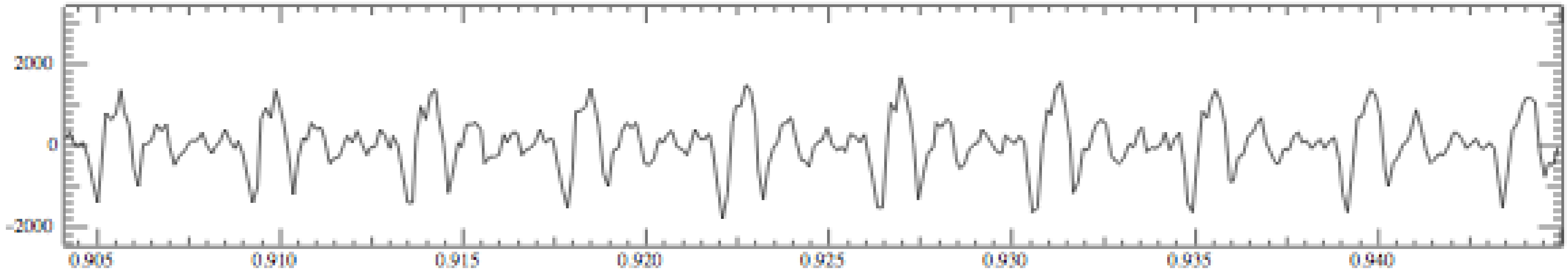
- Fourier transform of wave displayed as a spectrogram
  - Darkness indicates energy at each frequency



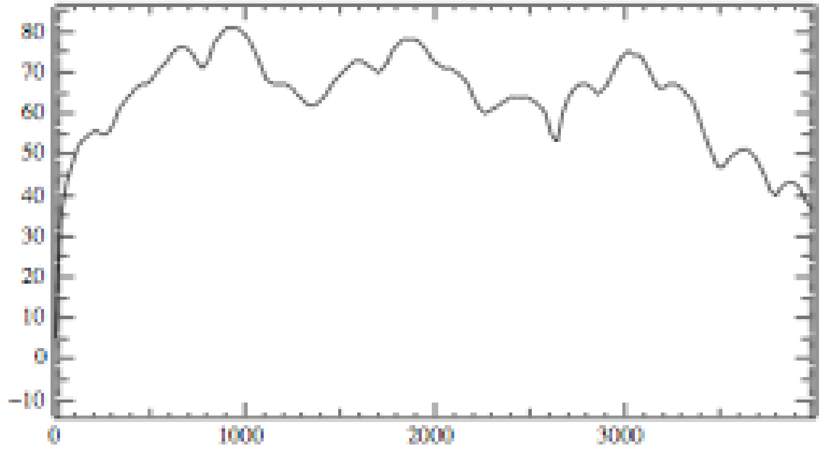


Human ear figure: depion.blogspot.com

Part of [ae] from "lab"

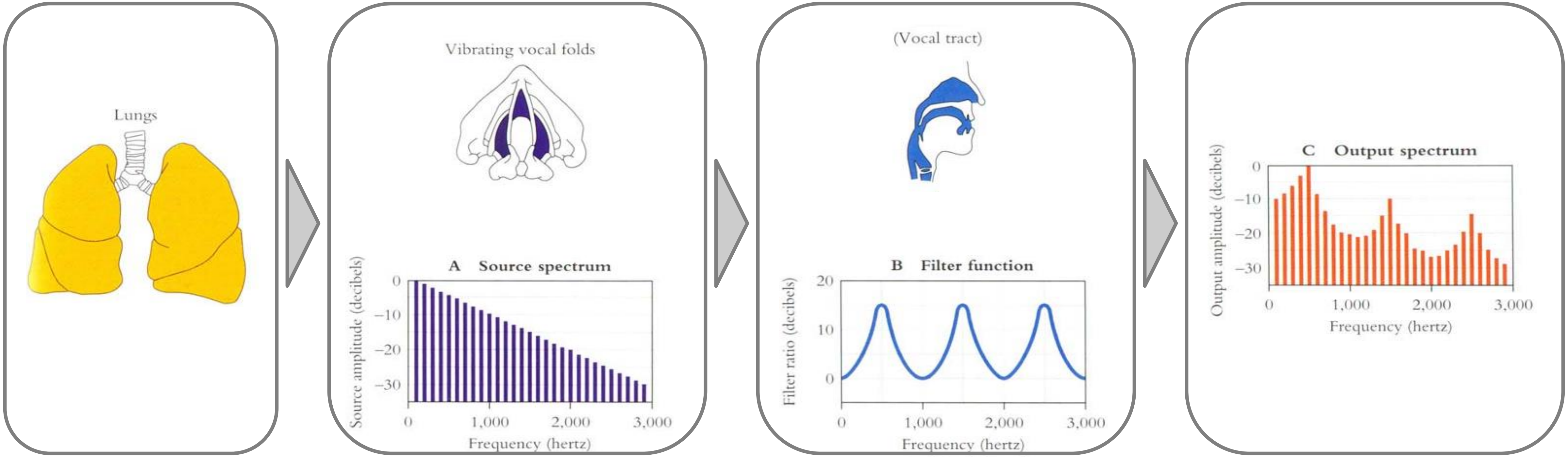


- Complex wave repeating nine times
  - Plus smaller wave that repeats 4x for every large cycle
  - Large wave: freq of 250 Hz (9 times in .036 seconds)
  - Small wave roughly 4 times this, or roughly 1000 Hz



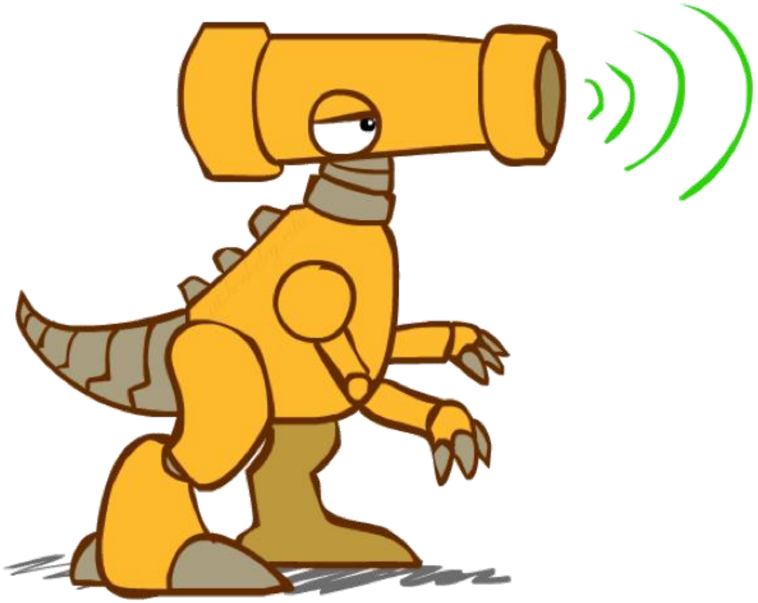
Why These Peaks?

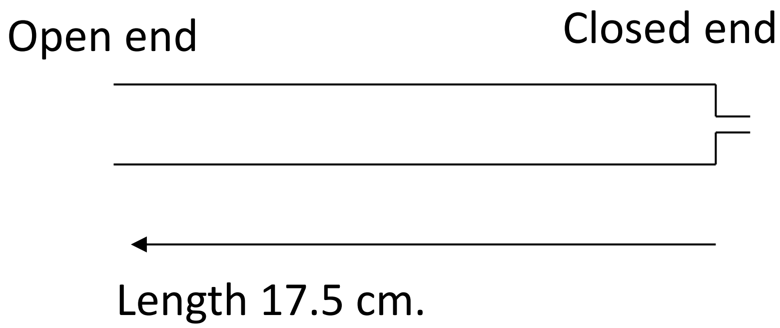
- Articulator process:
  - Vocal cord vibrations create harmonics
  - The mouth is an amplifier
  - Depending on shape of mouth, some harmonics are amplified more than others

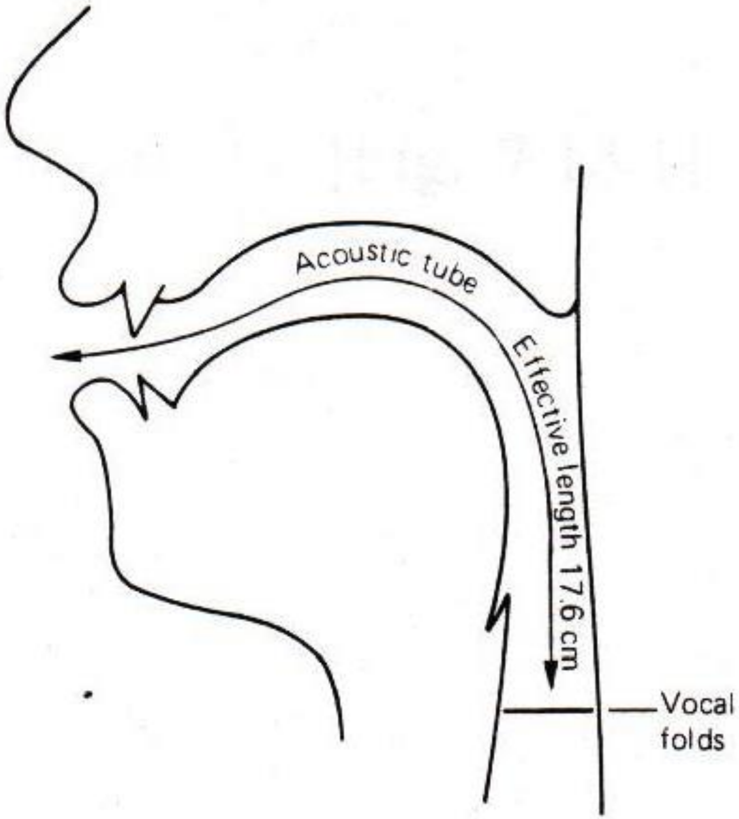


Resonances of the Vocal Tract

- The human vocal tract as an open tube







- Air in a tube of a given length will tend to vibrate at resonance frequency of tube
- Constraint: Pressure differential should be maximal at (closed) glottal end and minimal at (open) lip end

Figure: W. Barry Speech Science slides

Spectrum Shapes

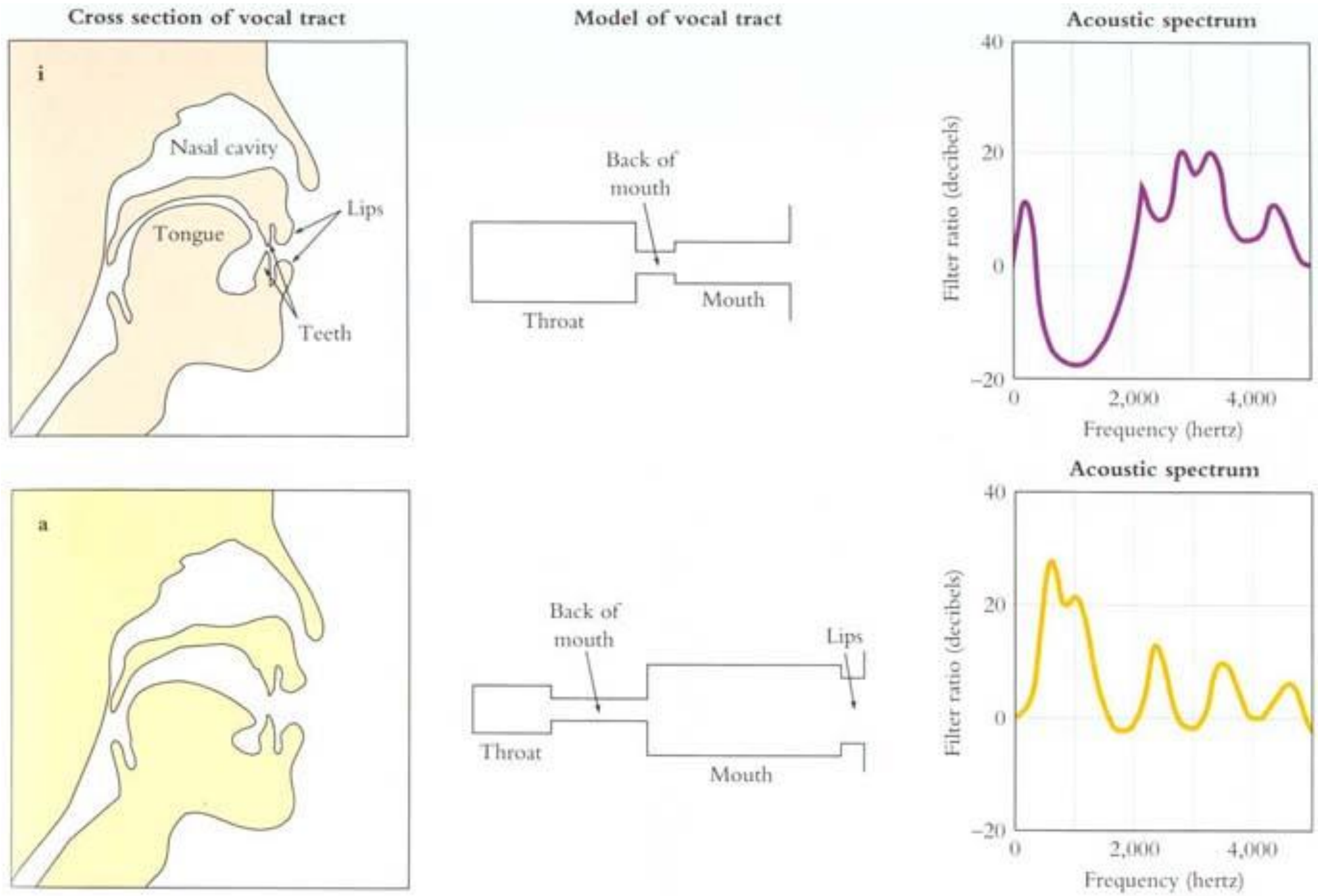
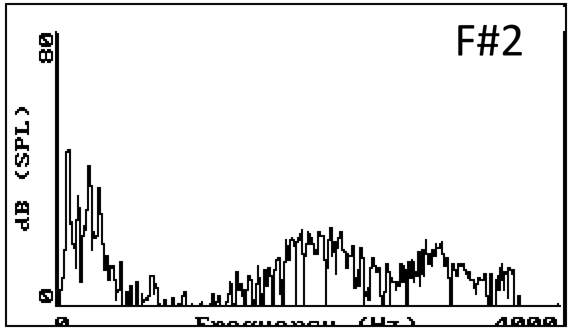
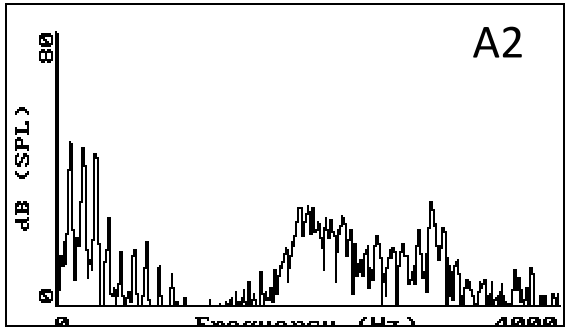


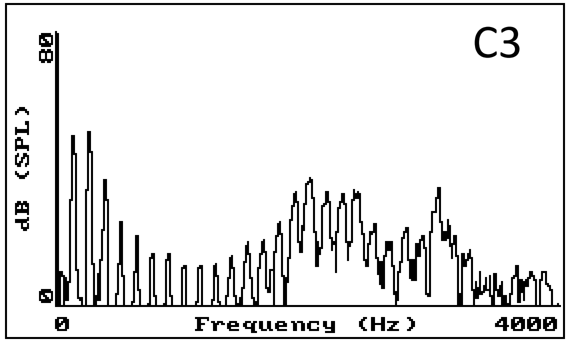
Figure: Mark Liberman

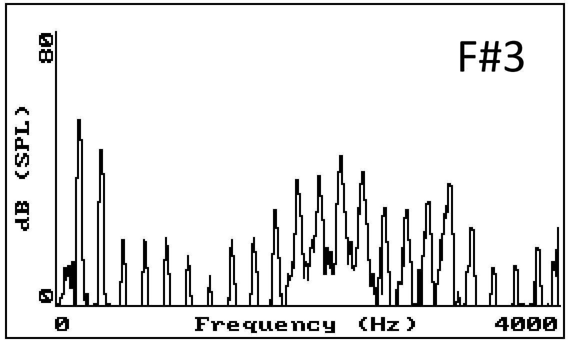
[Demo: speech synthesis]

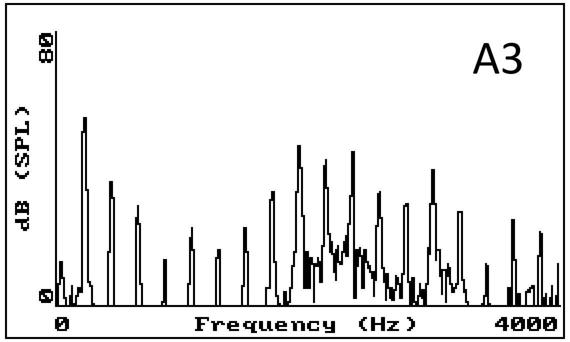
Vowel [i] sung at successively higher pitches

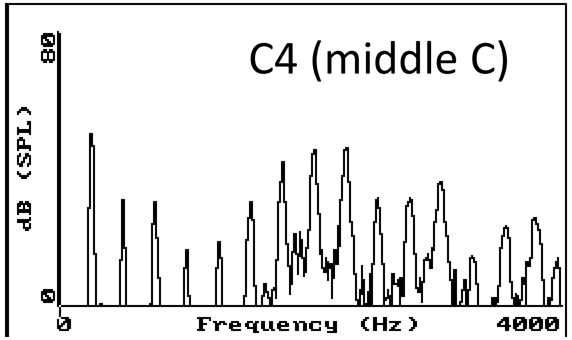


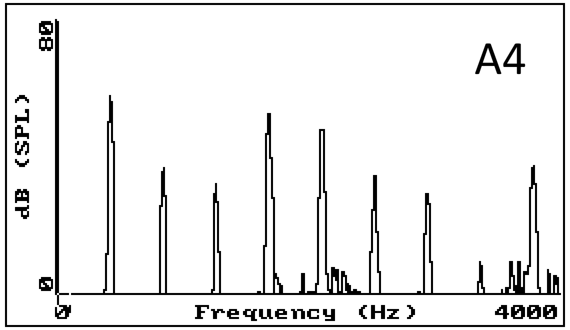


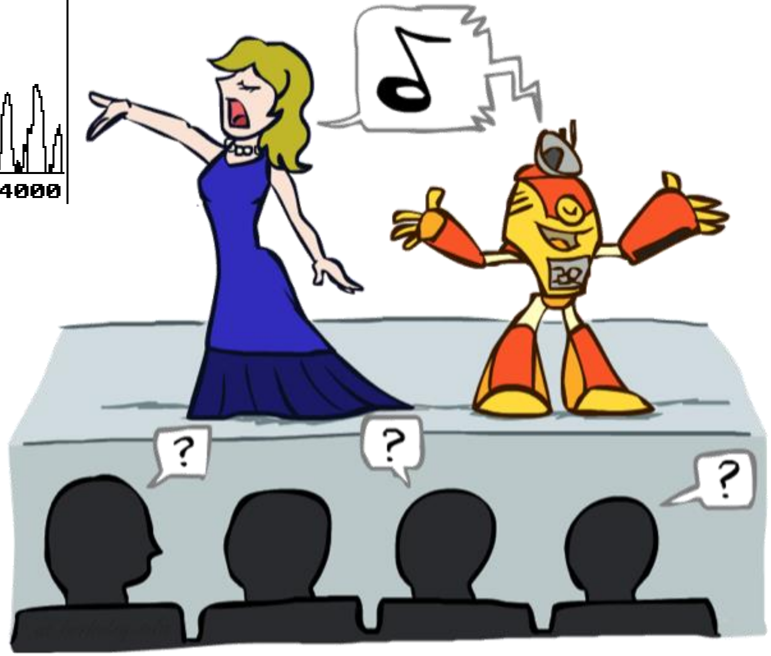








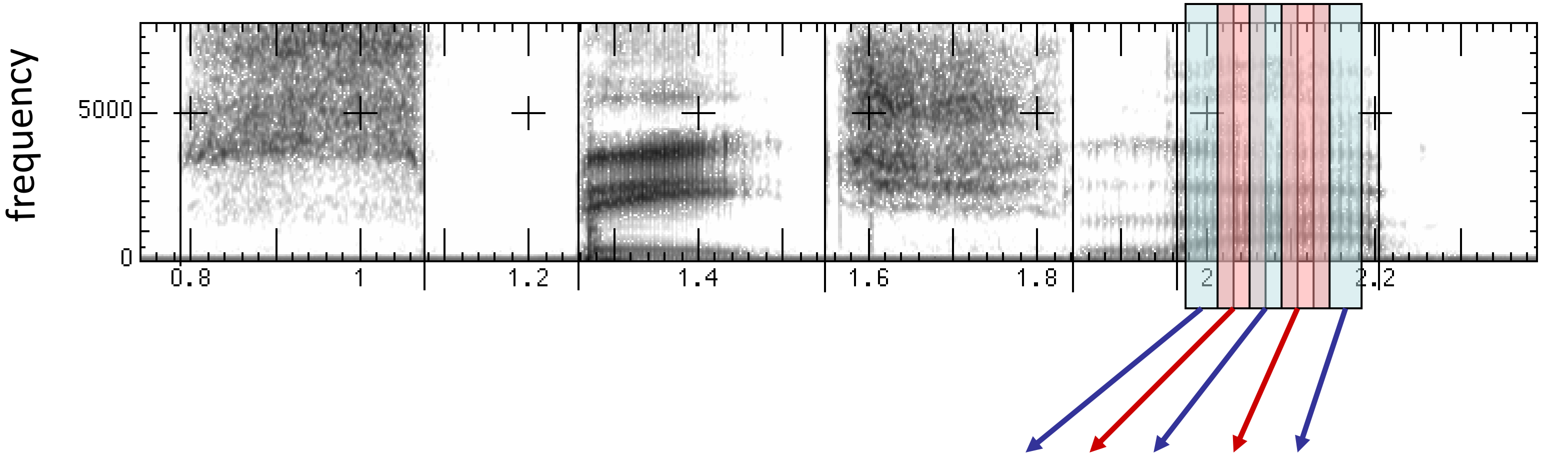




Graphs: Ratree Wayland

Acoustic Feature Sequence

- Time slices are translated into acoustic feature vectors (~39 real numbers per slice)



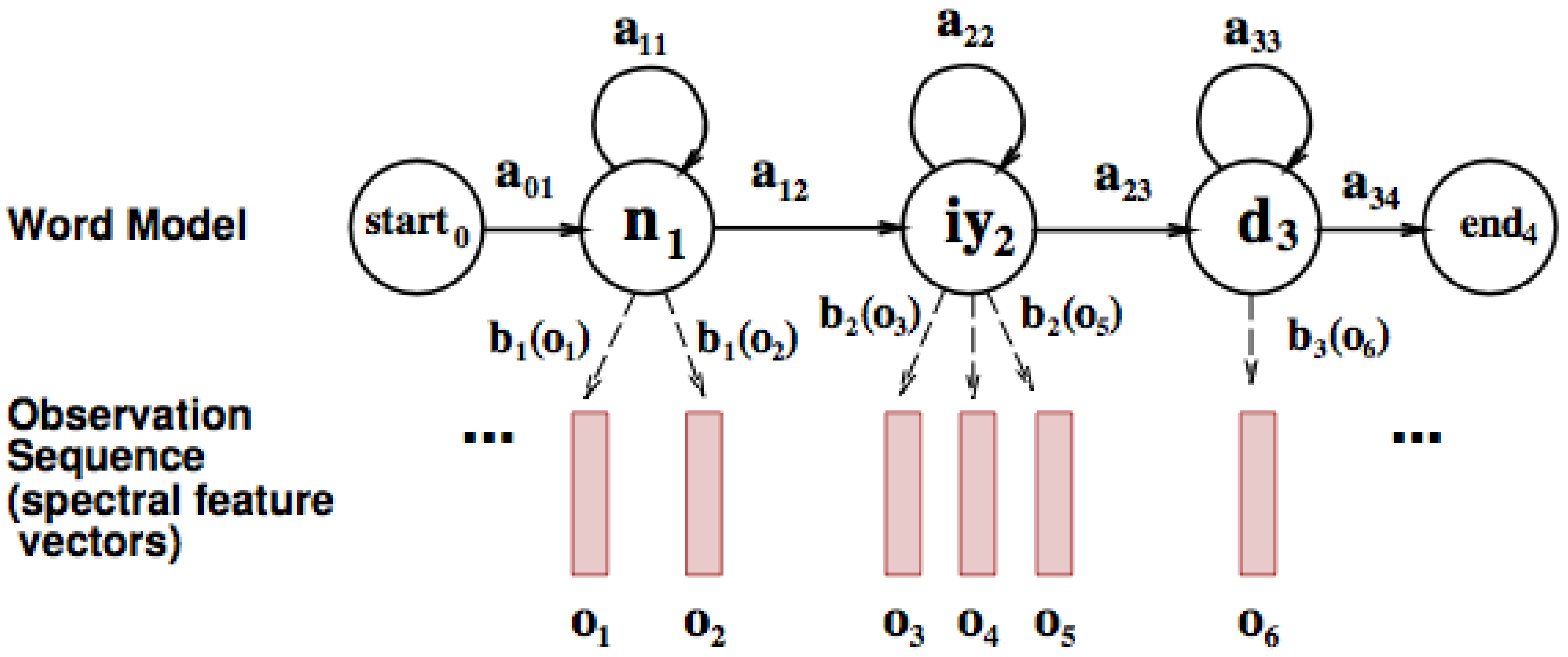
……………………………………………… $e_{12}e_{13}e_{14}e_{15}e_{16}$ ………..

- These are the observations E, now we need the hidden states X

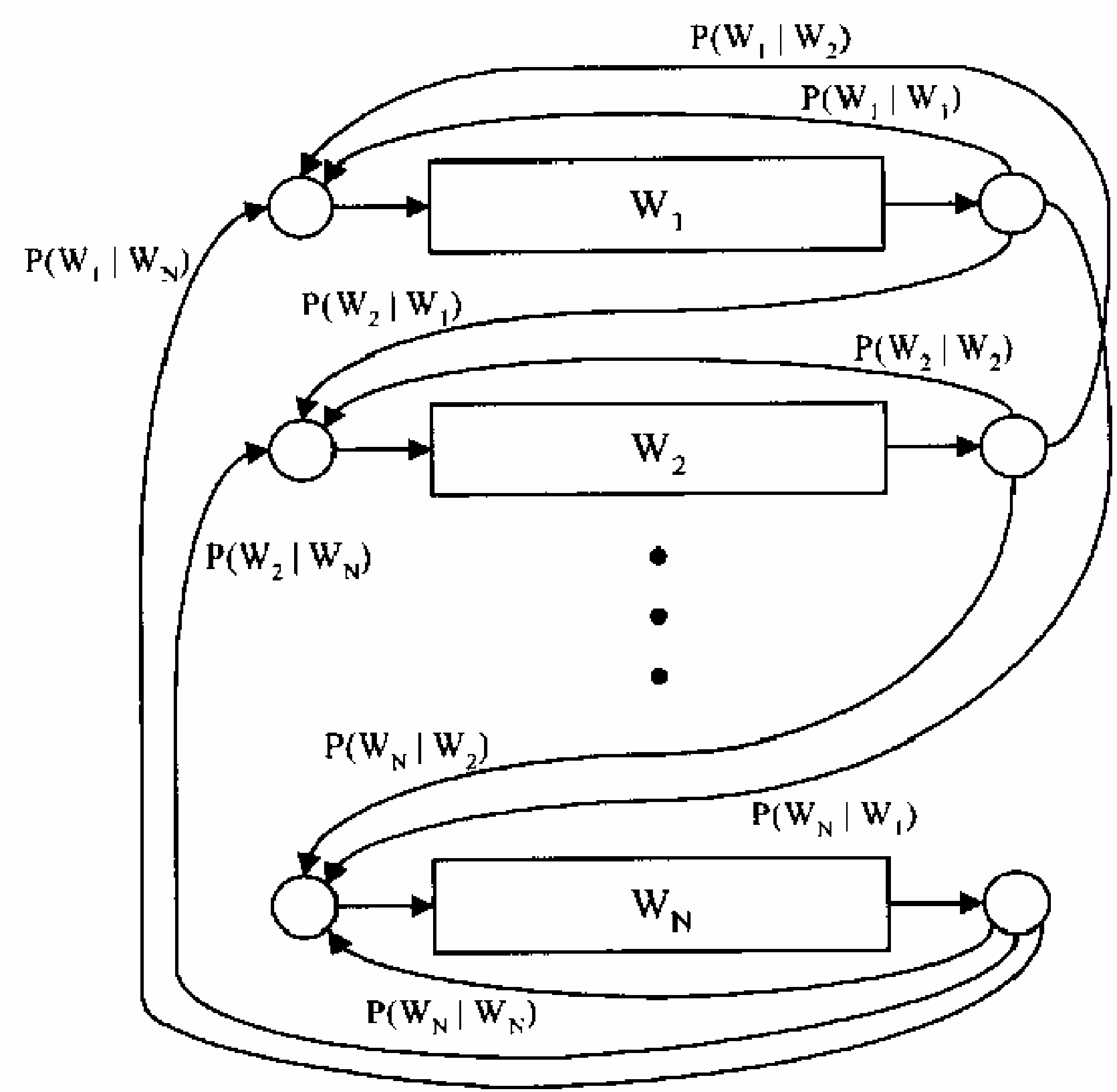
Speech State Space

- HMM Specification
  - P(E|X) encodes which acoustic vectors are appropriate for each phoneme (each kind of sound)
  - P(X|X’) encodes how sounds can be strung together
- State Space
  - We will have one state for each sound in each word
  - Mostly, states advance sound by sound
  - Build a little state graph for each word and chain them together to form the state space X

States in a Word



Transitions with a Bigram Model



Training Counts

198015222	the first
194623024	the same
168504105	the following
158562063	the world
…	
14112454	the door

23135851162	the *

$$\hat{P}(\text{door}|\text{the}) = \frac{14112454}{23135851162}$$

$$= 0.0006$$

Figure: Huang et al, p. 618

Decoding

- Finding the words given the acoustics is an HMM inference problem
- Which state sequence $x_{1:T}$ is most likely given the evidence $e_{1:T}$?

$$x^*_{1:T} = \arg\max_{x_{1:T}} P(x_{1:T}|e_{1:T}) = \arg\max_{x_{1:T}} P(x_{1:T}, e_{1:T})$$

- From the sequence x, we can simply read off the words