T-61.184
Automatic Speech Recognition: From Theory to Practice

http://www.cis.hut.fi/Opinnot/T-61.184/
September 20, 2004

Prof. Bryan Pellom
Department of Computer Science
Center for Spoken Language Research
University of Colorado

pellom@cslr.colorado.edu

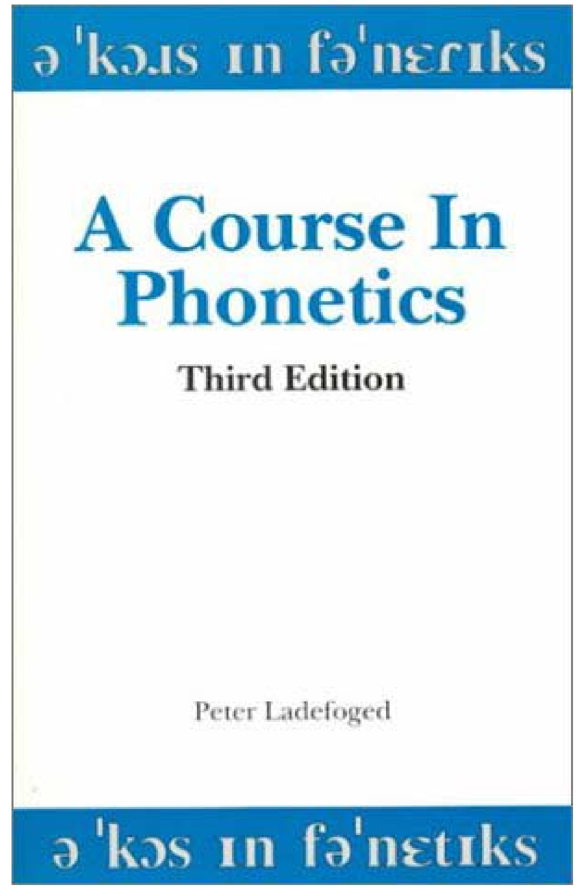
Today

- Introduction to Speech Production and Phonetics
- Short Video on Spectrogram Reading
- Quick Review of Probability and Statistics
- Formulation of the Speech Recognition Problem
- Talk about Homework #1

Speech Production and Phonetics

- **Peter Ladefoged, "A Course In Phonetics," Harcourt Brace Jovanovich, ISBN 0-15-500173-6**

- **Excellent introductory reference to this material**



Speech Production Anatomy

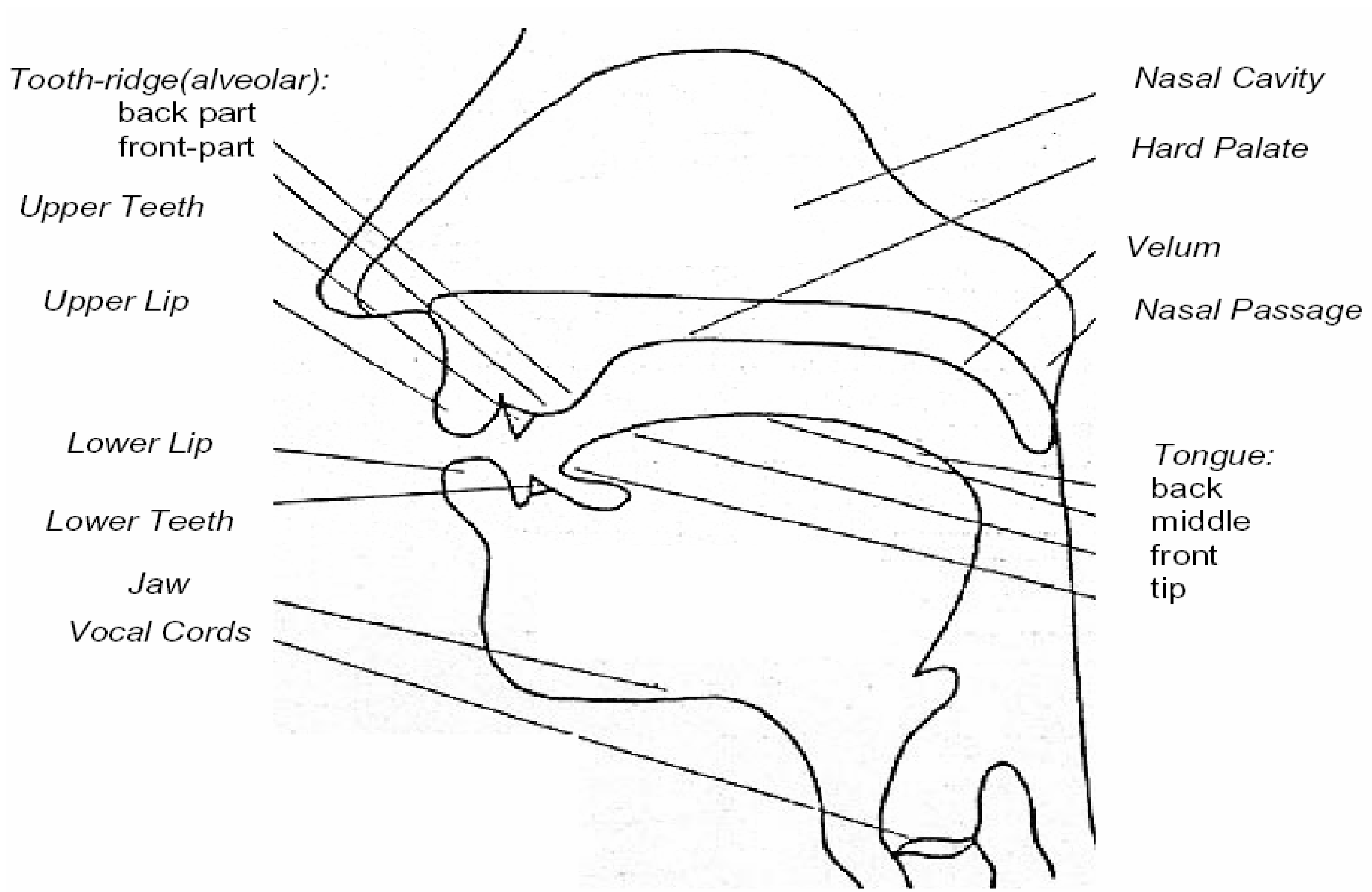


Figure from Spoken Language Processing (Huang, Acero, Hon)

Speech Production Anatomy

- **Vocal Tract**
  - Consists of the pharyngeal and oral cavities
- **Articulators**
  - Components of the vocal tract which move to produce various speech sounds
  - Include: vocal folds, velum, lips, tongue, teeth

Source-Filter Representation of Speech Production

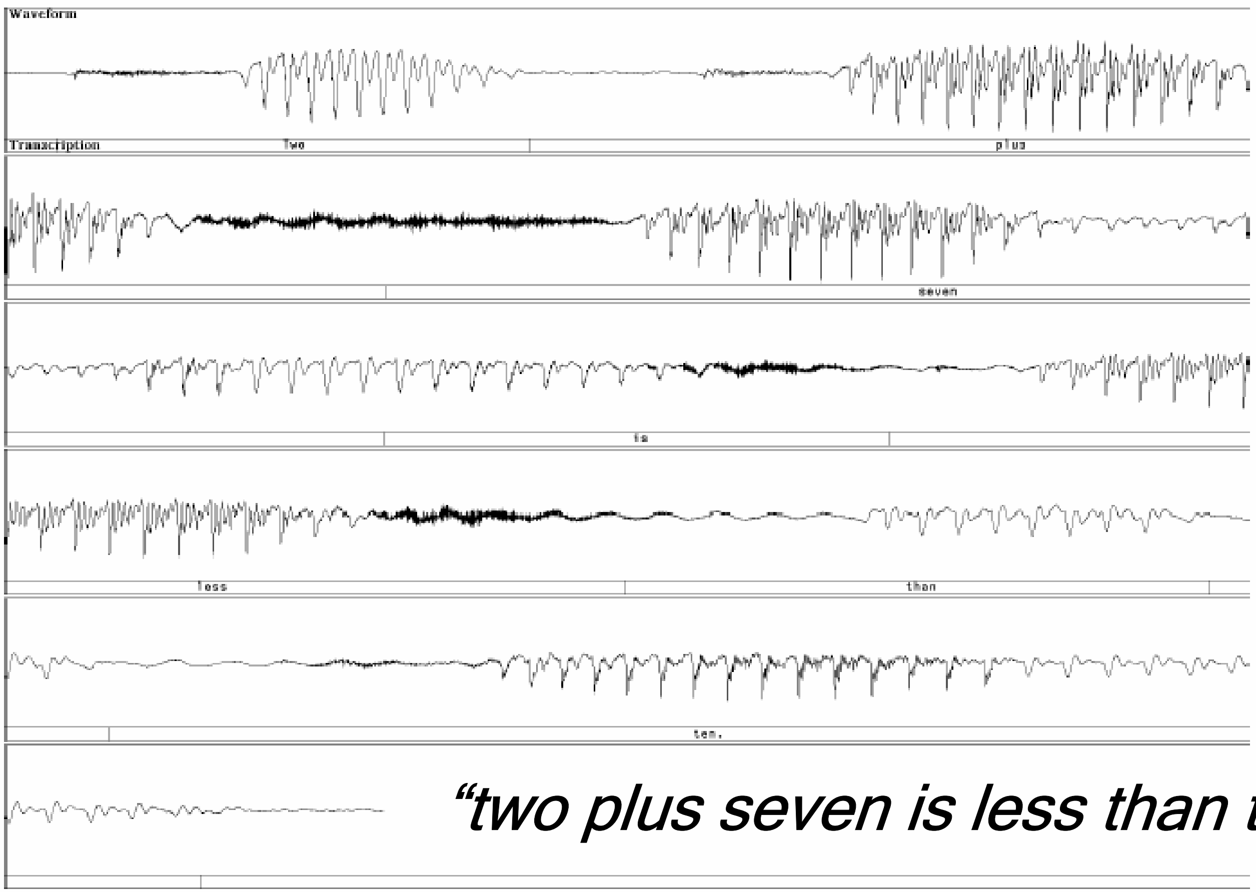
- **Production viewed as an acoustic filtering operation**
- **Larynx and lungs provide input or <u>source excitation</u>**
- **Vocal and nasal tracts act as <u>filter</u>. Shape the spectrum of the resulting signal**

T-61.184

Describing Sounds

- **The study of speech sounds and their production, classification and transcription is known as <u>phonetics</u>**
- **A <u>phoneme</u> is an abstract unit that can be used for writing a language down in a systematic or unambiguous way**
- **Sub-classifications of phonemes**
  - <u>Vowels</u> – air passes freely through resonators
  - <u>Consonants</u> – air passes partially or totally obstructed in one or more places as it passes through the resonators

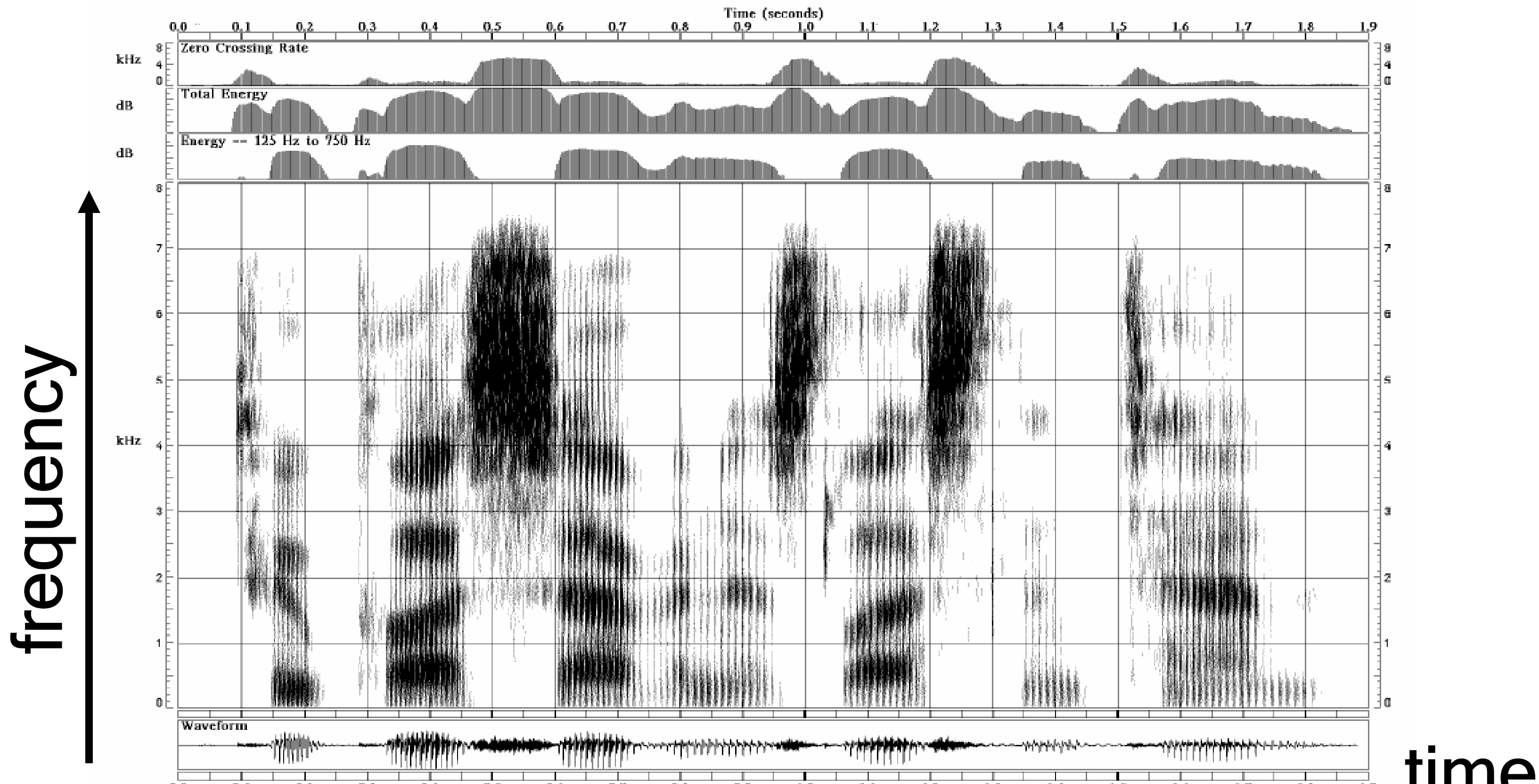
Time-Domain Waveform Example



"two plus seven is less than ten"

Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

Wide-Band Spectrogram



"two plus seven is less than ten"

Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

T-61.184

Phonetic Alphabets

- Allow us to describe the primitive sounds that make up a language
- Each language will have a unique set of phonemes
- Useful for speech recognition since words can be represented by sequences of phonemes as described by a phonetic alphabet.

International Phonetic Alphabet (IPA)

- **Phonetic representation standard which describes sounds in most/all world languages**
- **IPA last published in 1993 and updated in 1996**
- **Issue: character set difficult to manipulate on a computer…**
- **`http://www2.arts.gla.ac.uk/IPA/ipa.html`**

IPA Consonants

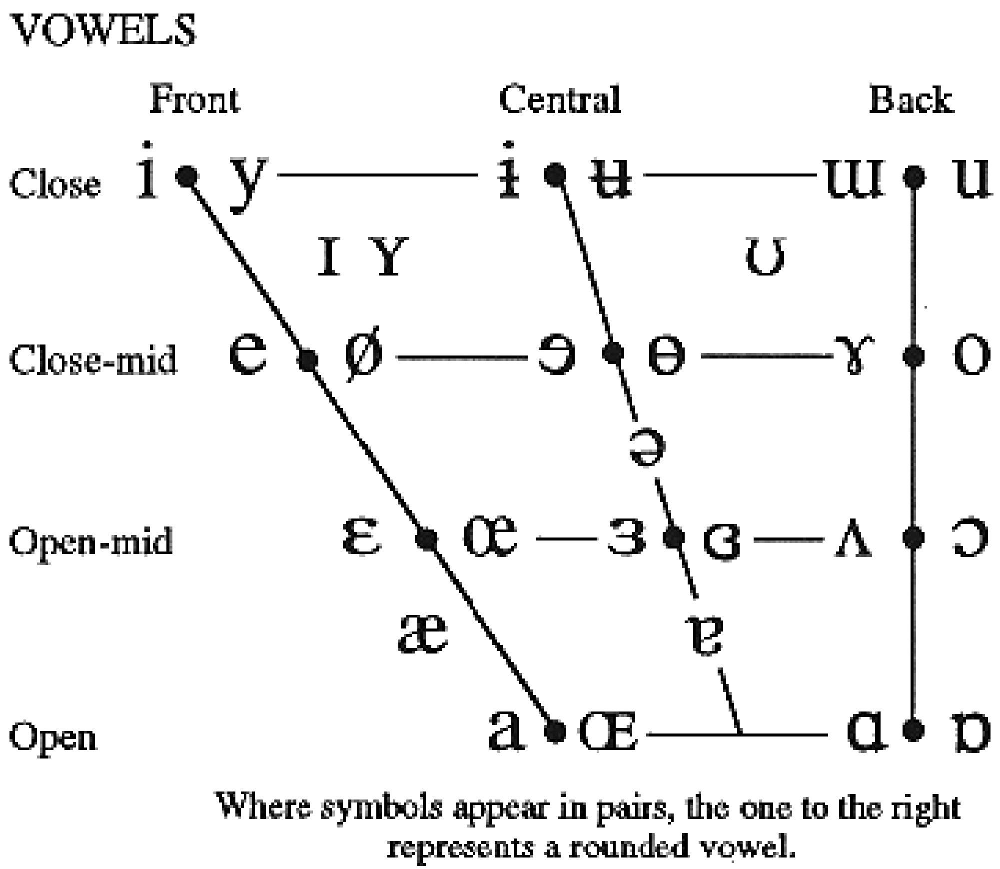
THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993, updated 1996)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

IPA Vowels



T-61.184

American English Phonemes (IPA)

PHONEME	*EXAMPLE*	*PHONEME*	*EXAMPLE*	*PHONEME*	*EXAMPLE*
/i^{y}/	beat	/s/	see	/w/	wet
/ɪ/	bit	/š/	she	/r/	red
/e^{y}/	bait	/f/	fee	/l/	let
/ɛ/	bet	/θ/	thief	/y/	yet
/æ/	bat	/z/	z	/m/	meet
/ɑ/	Bob	/ž/	Gigi	/n/	neat
/ɔ/	bought	/v/	v	/ŋ/	sing
/ʌ/	but	/ð/	thee	/č/	church
/o^{w}/	boat	/p/	pea	/ǰ/	judge
/ʊ/	book	/t/	tea	/h/	heat
/u^{w}/	boot	/k/	key		
/ɝ/	Burt	/b/	bee		
/ɑ^{y}/	bite	/d/	Dee		
/ɔ^{y}/	Boyd	/g/	geese		
/ɑ^{w}/	bout				
/ə/	about				

Table from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

Alternative Phonetic Alphabets

- **ARPAbet**
  - English only ASCII representation
  - Phoneme units represented by 1-2 letters
  - Similar representation used by CMU Sphinx-II recognizer, `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`

- **SAMPA**
  - Speech Assessment Methods Phonetic Alphabet
  - Computer Readable representation
  - Maps symbols of the IPA into ASCII codes

CMU Sphinx-II Phonetic Symbols

Phone	Example	Phone	Example	Phone	Example
AA	odd	**EY**	ate	**P**	pee
AE	at	**F**	fee	**PD**	lip
AH	hut	**G**	green	**R**	read
AO	ought	**GD**	bag	**S**	sea
AW	cow	**HH**	he	**SH**	she
AX	abide	**IH**	it	**T**	tea
AXR	user	**IX**	acid	**TD**	lit
AY	hide	**IY**	eat	**TH**	theta
B	be	**JH**	gee	**TS**	bits
BD	Dub	**K**	key	**UH**	hood
CH	cheese	**KD**	lick	**UW**	two
D	dee	**L**	lee	**V**	vee
DD	dud	**M**	me	**W**	we
DH	thee	**N**	note	**Y**	yield
DX	matter	**NG**	ping	**Z**	zee
EH	ed	**OW**	oat	**ZH**	seizure
ER	hurt	**OY**	toy	**SIL**	(silence)

Example Words and Corresponding CMU Dictionary Transcriptions

▪ basement	B EY S M AX N TD
▪ Bryan	B R AY AX N
▪ perfect	P AXR F EH KD TD
▪ speech	S P IY CH
▪ recognize	R EH K AX G N AY Z

Classifications of Speech Sounds

- **Voiced vs. voiceless**
  - Voiced if vocal chords vibrate
- **Nasal vs. Oral**
  - Nasal if air travels through nasal cavity and oral cavity closed
- **Consonant vs. Vowel**
  - Consonants: obstruction in air stream above the glottis. The Glottis is defined as the space between the vocal cords.
- **Lateral vs. Non-lateral**
  - Non-lateral If the air stream passes through the middle of the oral cavity (compared to along side the oral cavity)

Consonants and Vowels

- **Consonants are characterized by:**
  - Place of articulation
  - Manner of articulation
  - Voicing
- **Vowels are characterized by:**
  - lip position
  - tongue height
  - tongue advancement

Places of Articulation

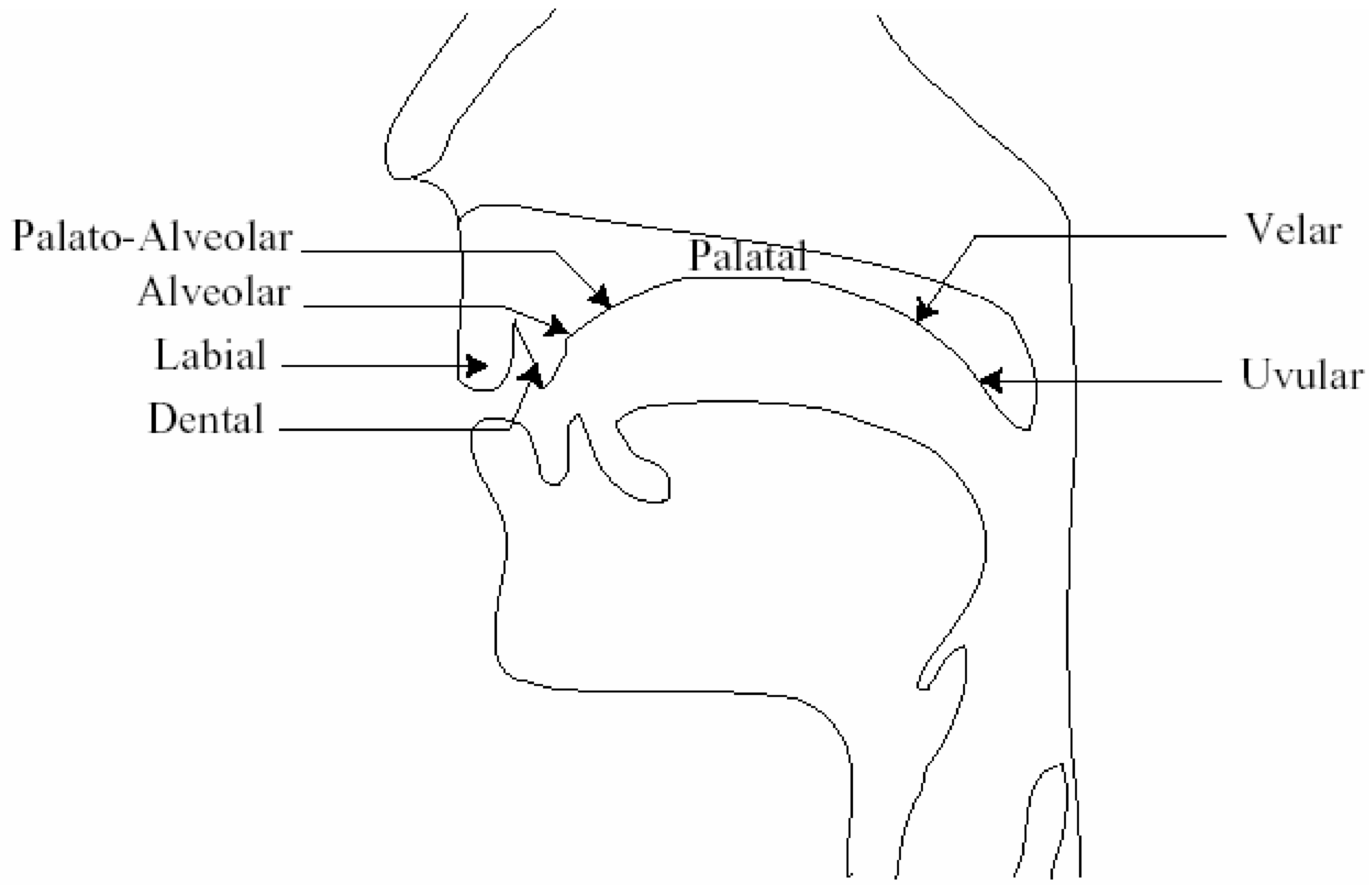


Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

Places of Articulation

■	**Biliabial**	Made with the two lips {P,B,M}
■	**Labio-dental**	Lower lip & upper front teeth {F,V}
■	**Dental**	Tongue tip or blade & upper front teeth {TH,DH}
■	**Alveolar**	Tongue tip or blade & alveolar ridge {T,D,N,…}
■	**Retroflex**	Tongue tip & back of the alveolar ridge {R}
■	**Palato-Alveolar**	Tongue blade & back of the alveolar ridge {SH}
■	**Palatal**	Front of the tongue & hard palate {Y,ZH}
■	**Velar**	Back of the tongue & soft palate {K,G,NG}

Manners of Articulation

- **Stop**
  - complete obstruction with sudden (explosive) release
- **Nasal**
  - Airflow stopped in the oral cavity, soft palate down, airflow is through the nasal tract
- **Fricative**
  - Articulators close together, turbulent airflow produced

Manners of Articulation

- **Retroflex (Liquid)**
  - Tip of the tongue is curled back slightly (/r/)
- **Lateral (Liquid)**
  - Obstruction of the air stream at a point along the center of the oral tract, with incomplete closure between one or both sides of the tongue and the roof of the mouth (/l/)
- **Glide**
  - Vowel-like, but initial position within a syllable (/y/, /w/)

American English Consonants by Place and Manner of Articulation

Place

Manner

	Labial	Labio-dental	Dental	Alveolar	Palatal	Velar	Glottal
Plosive	*p b*			*t d*		*k g*	*?*
Nasal	*m*			*n*		*ng*	
Fricative		*f v*	*th dh*	*s z*	*sh zh*		*h*
Retroflex Sonorant				*r*			
Lateral sonorant				*l*			
Glide	*w*				*y*		

American English Unvoiced Fricatives

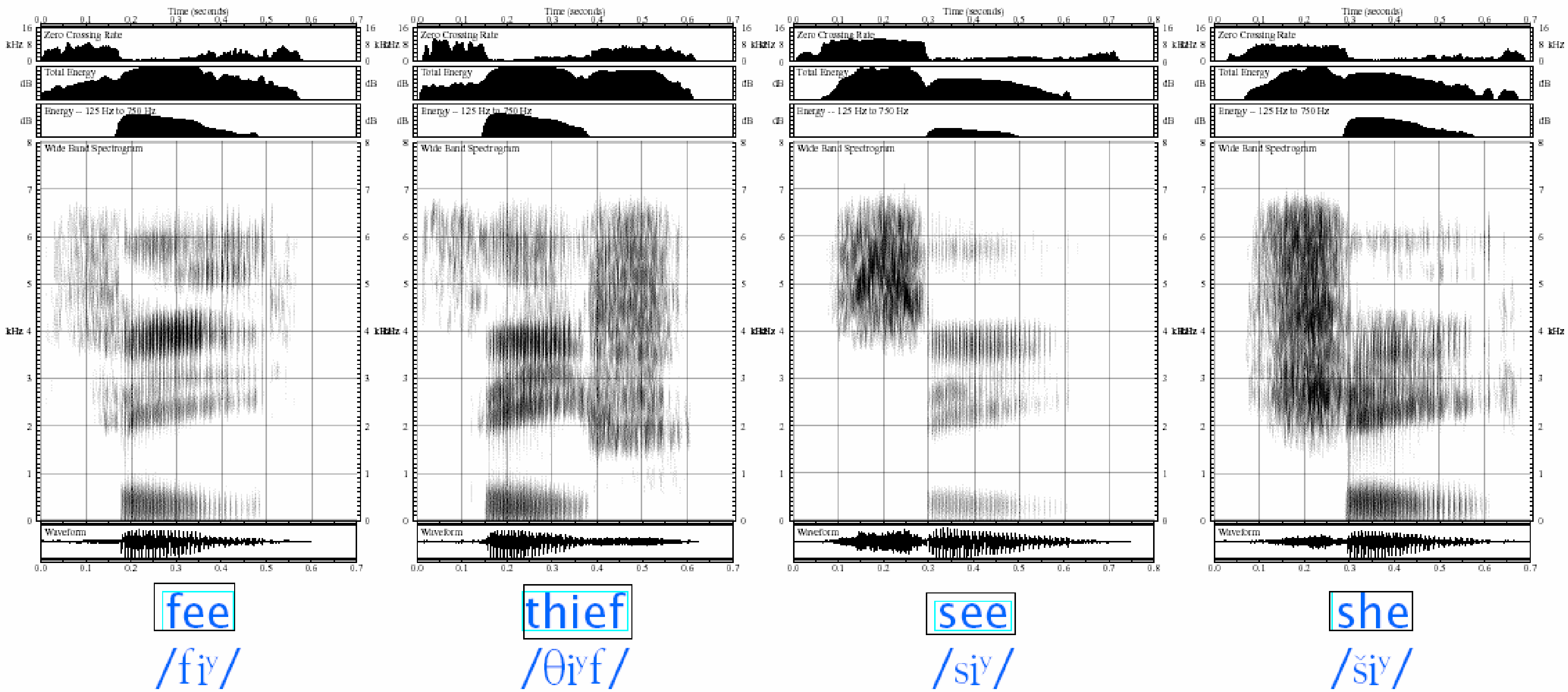
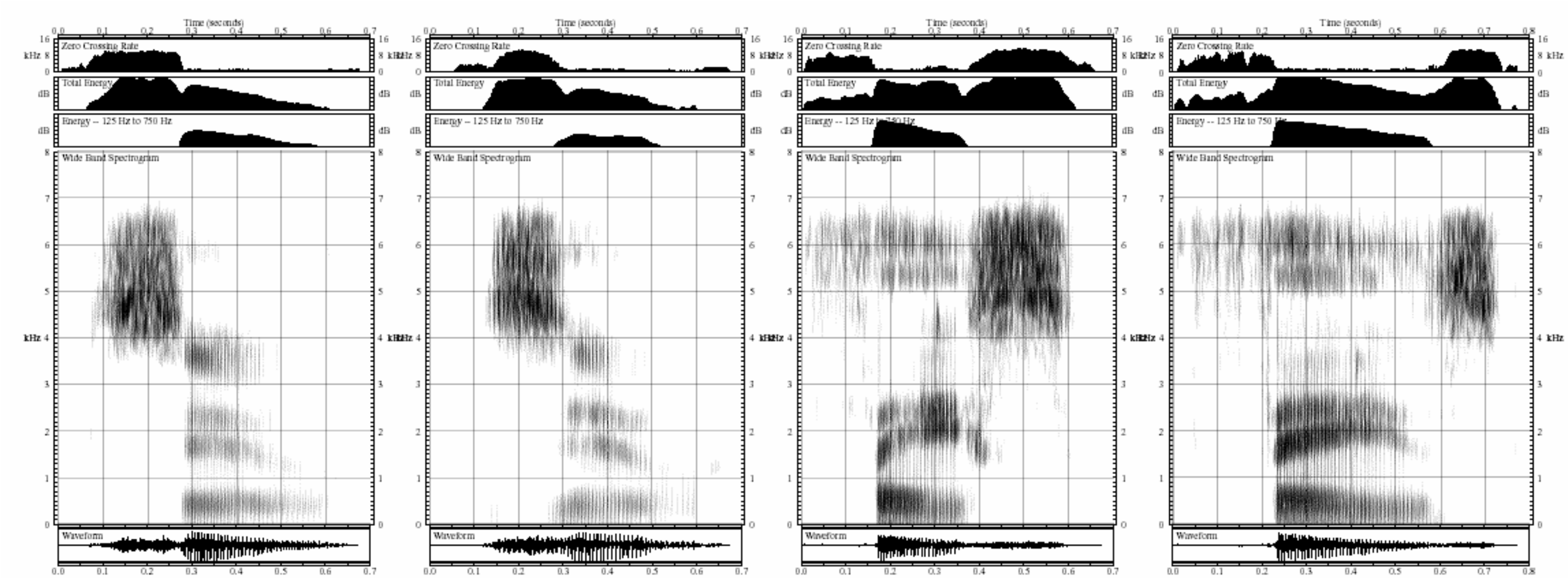


Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

Voiced vs. Unvoiced Fricatives



sue /su/ zoo /zu/ face /fe^{y}s/ phase /fe^{y}z/

- **Voiced fricatives tend to be shorter than unvoiced fricatives**

Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

American English Unvoiced Stops

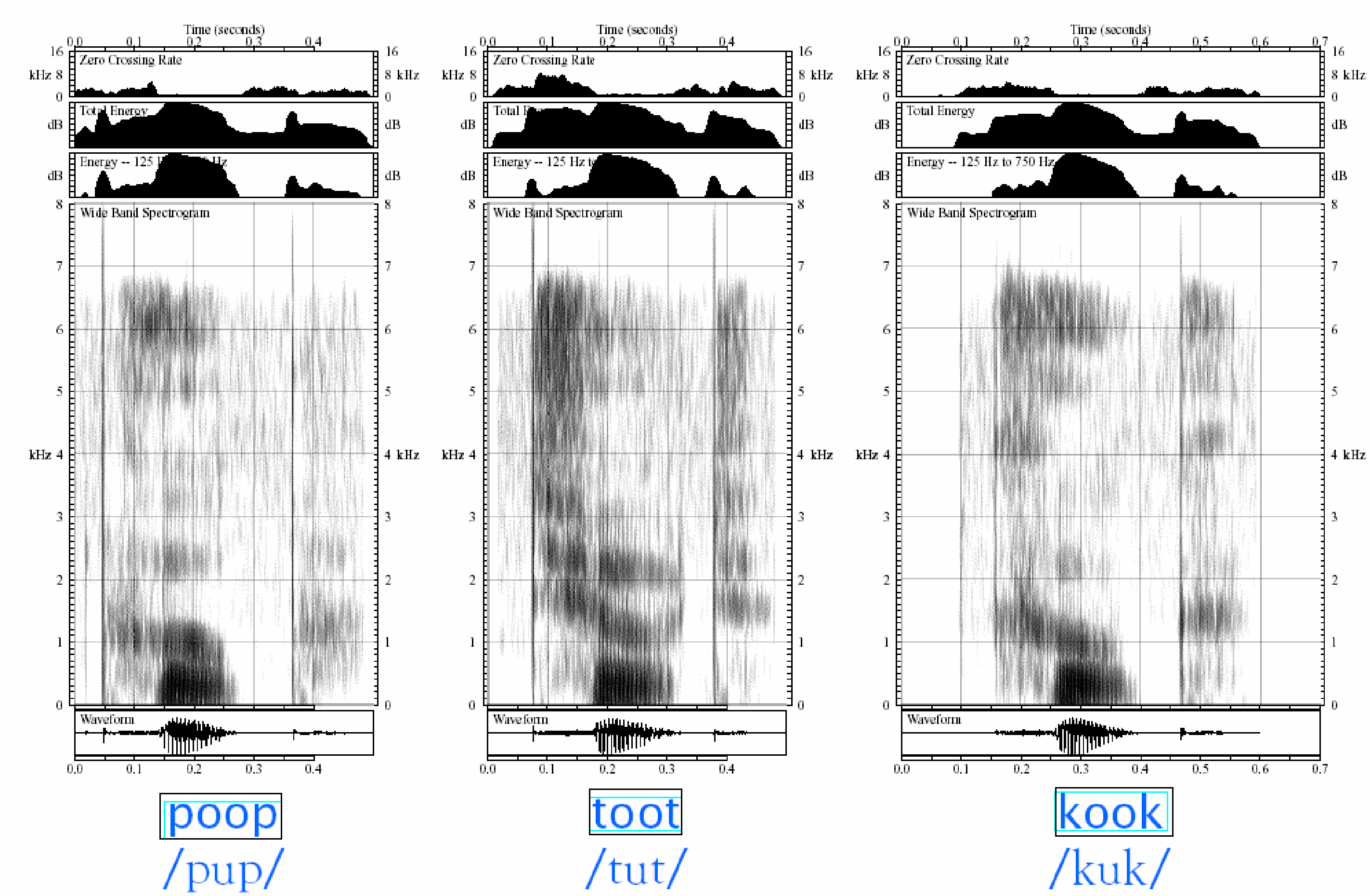


Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

T-61.184

Voiced vs. Unvoiced Stops

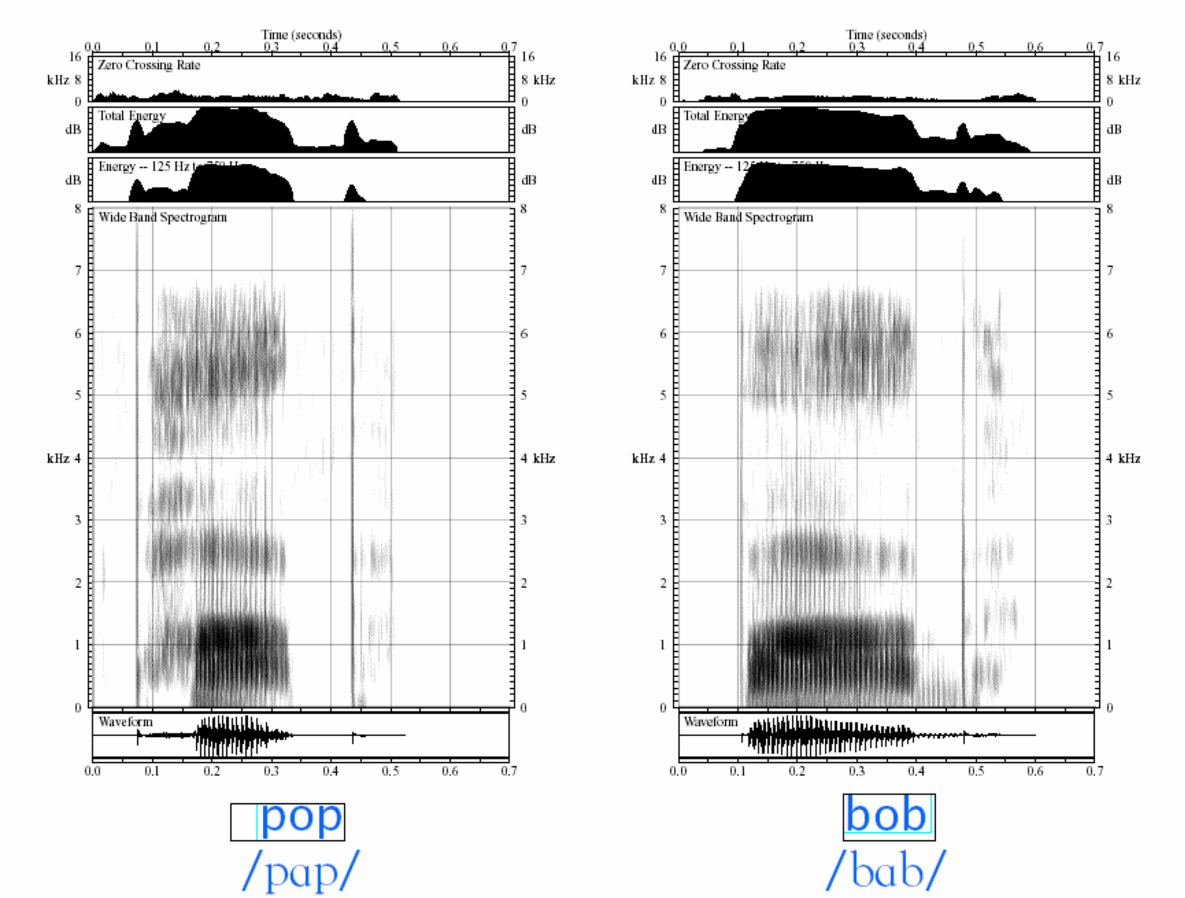


Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

Stop Consonant Formant Transitions

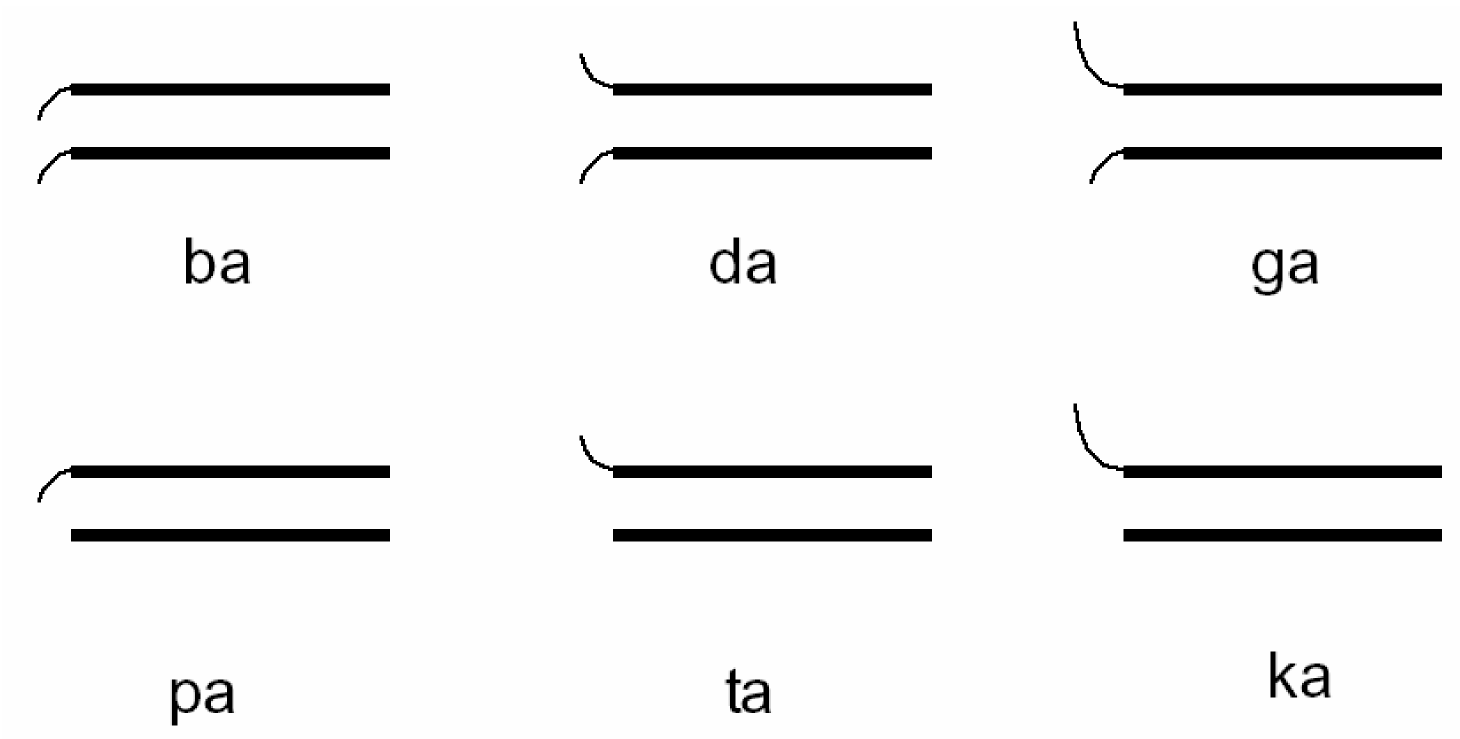
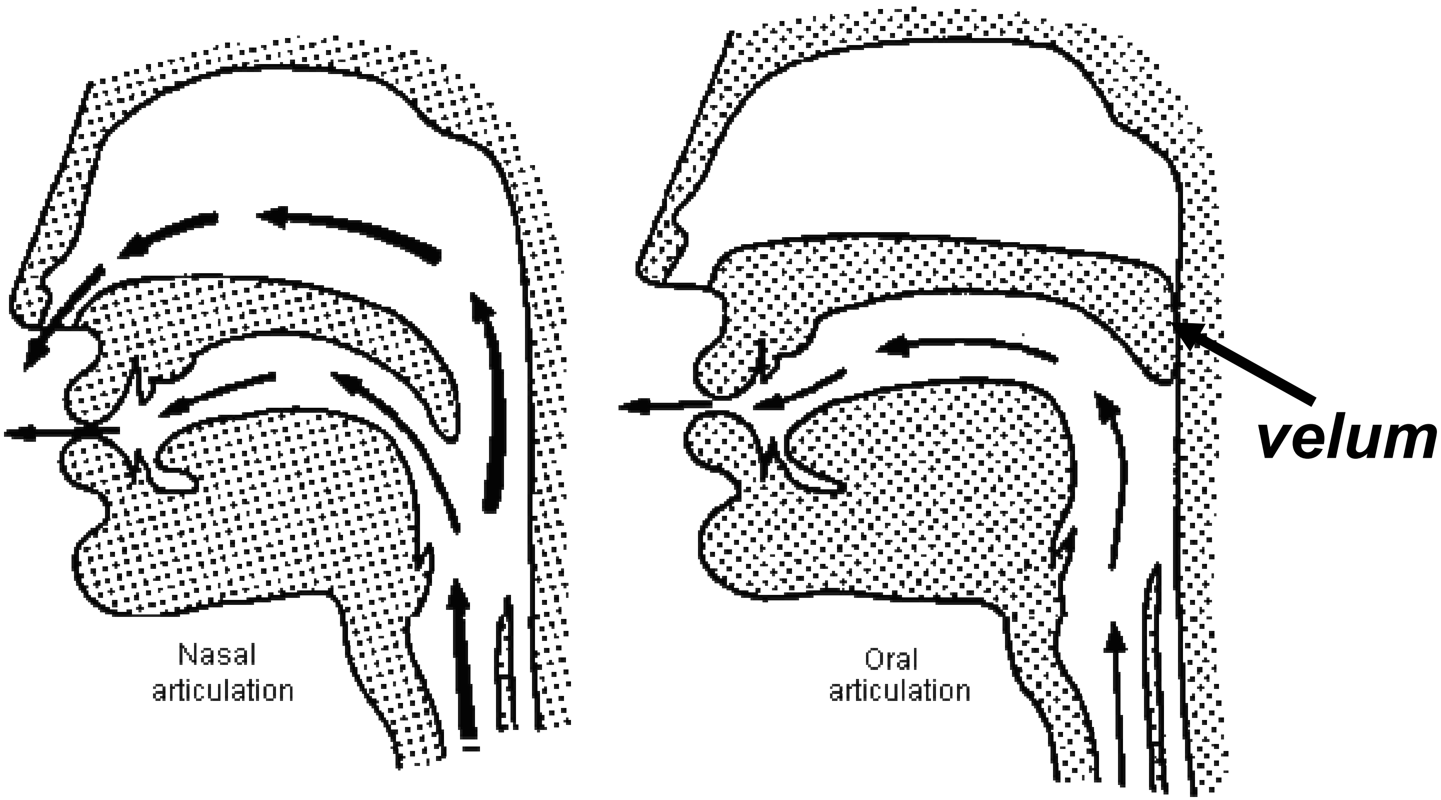


Figure from http://hitchcock.dlt.asu.edu/media5/a_spanias/speech-recognition/real-lectures/PDF/LECT04-5.PDF

Nasal vs. Oral Articulation



/M/, /N/, /NG/

T-61.184

Spectrogram of Nasals

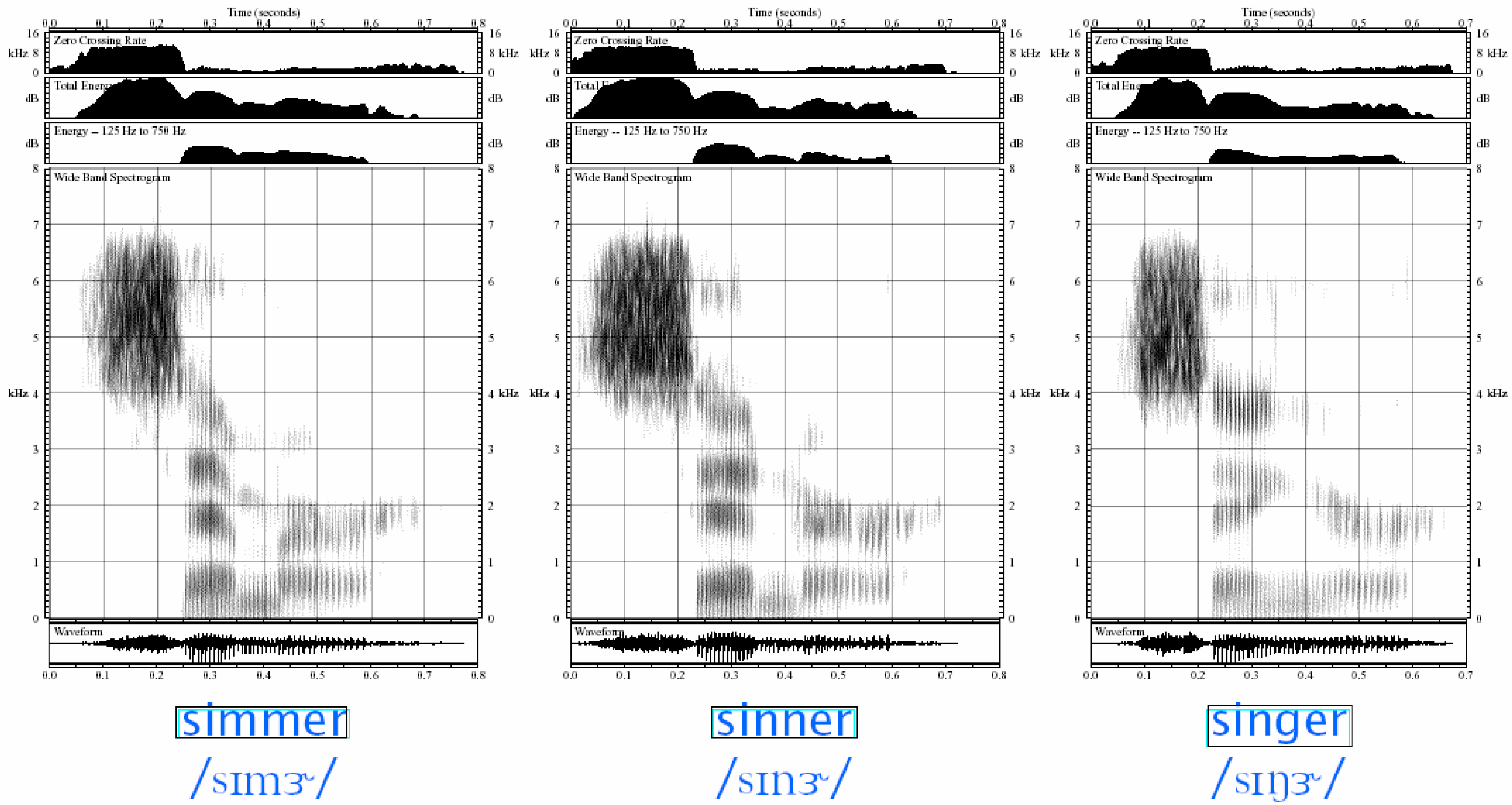


Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

American English Semivowels

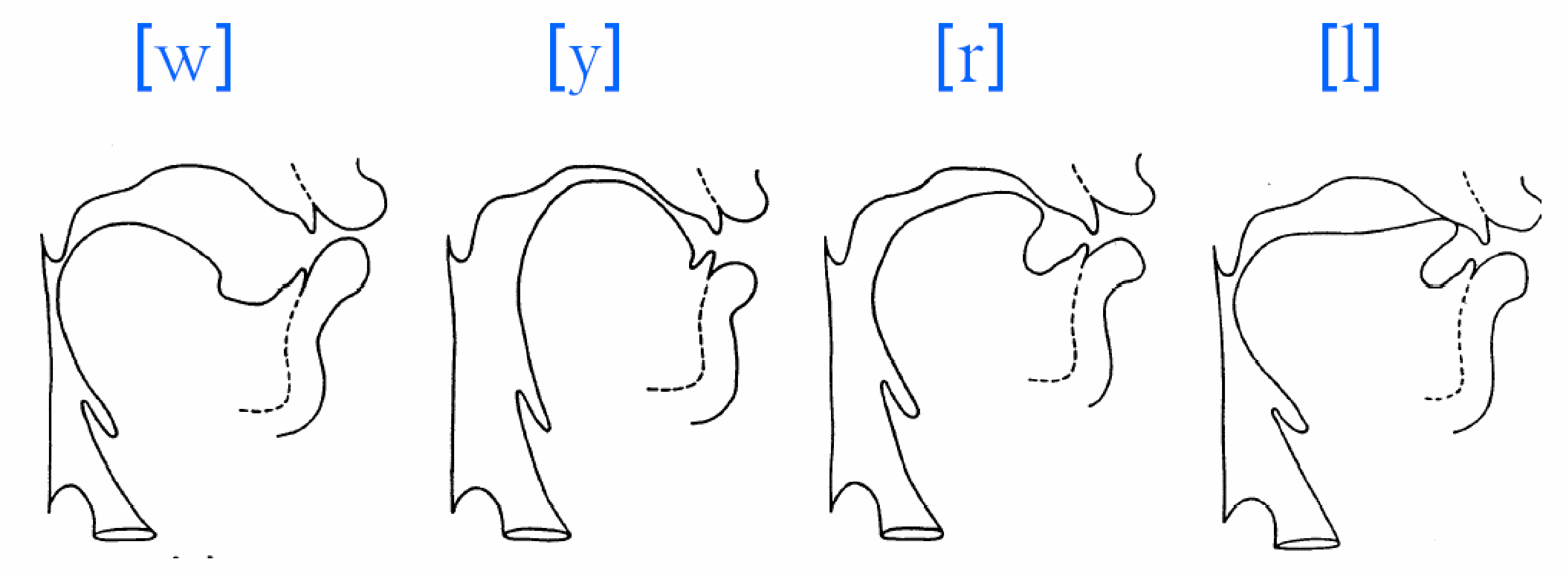
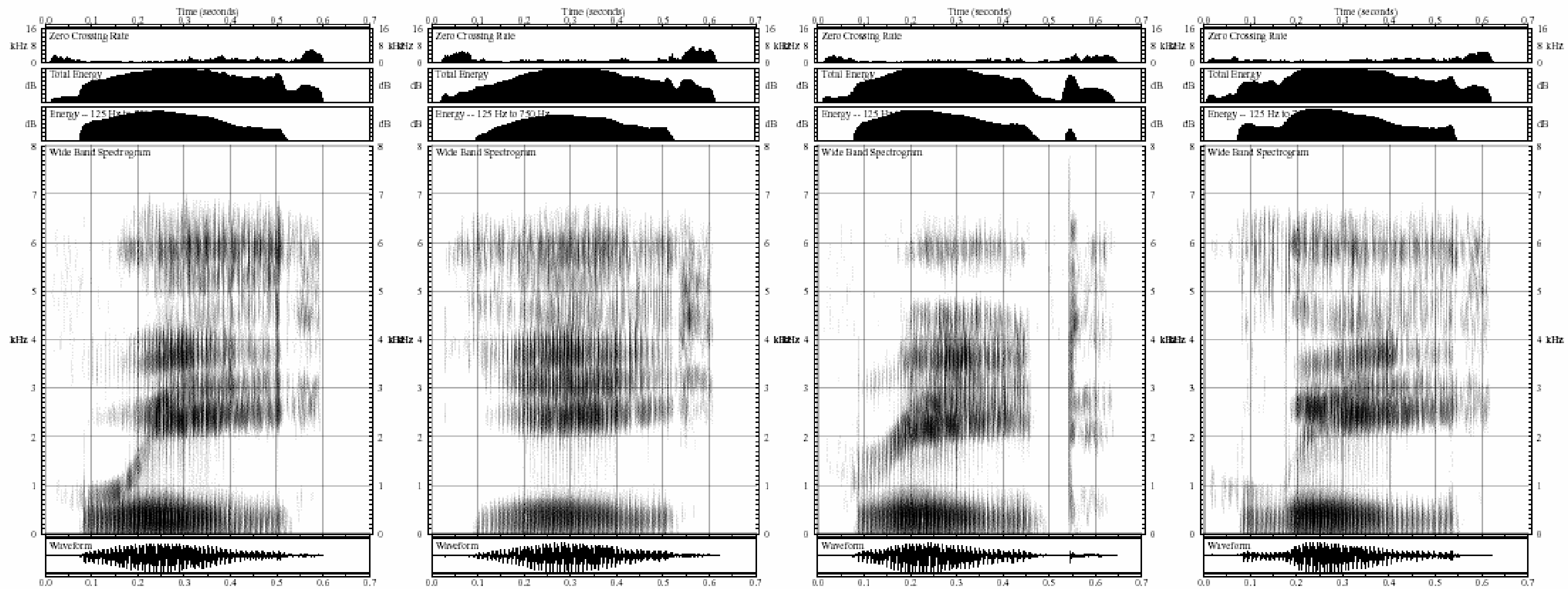


Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

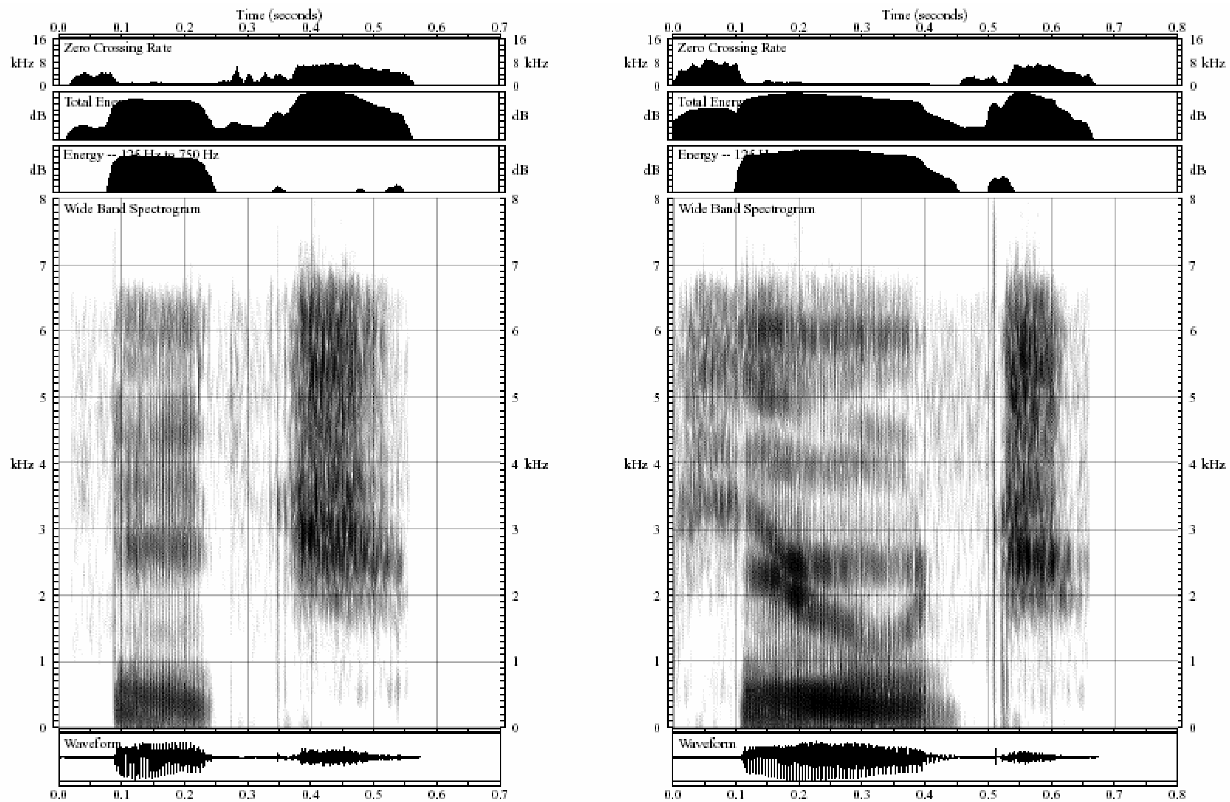
Semivowels



we	ye	reed	lee
$/wi^y/$	$/yi^y/$	$/ri^yd/$	$/li^y/$

Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

Affricates



each

/ɪʸč/

huge

/hyuǰ/

Figure from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

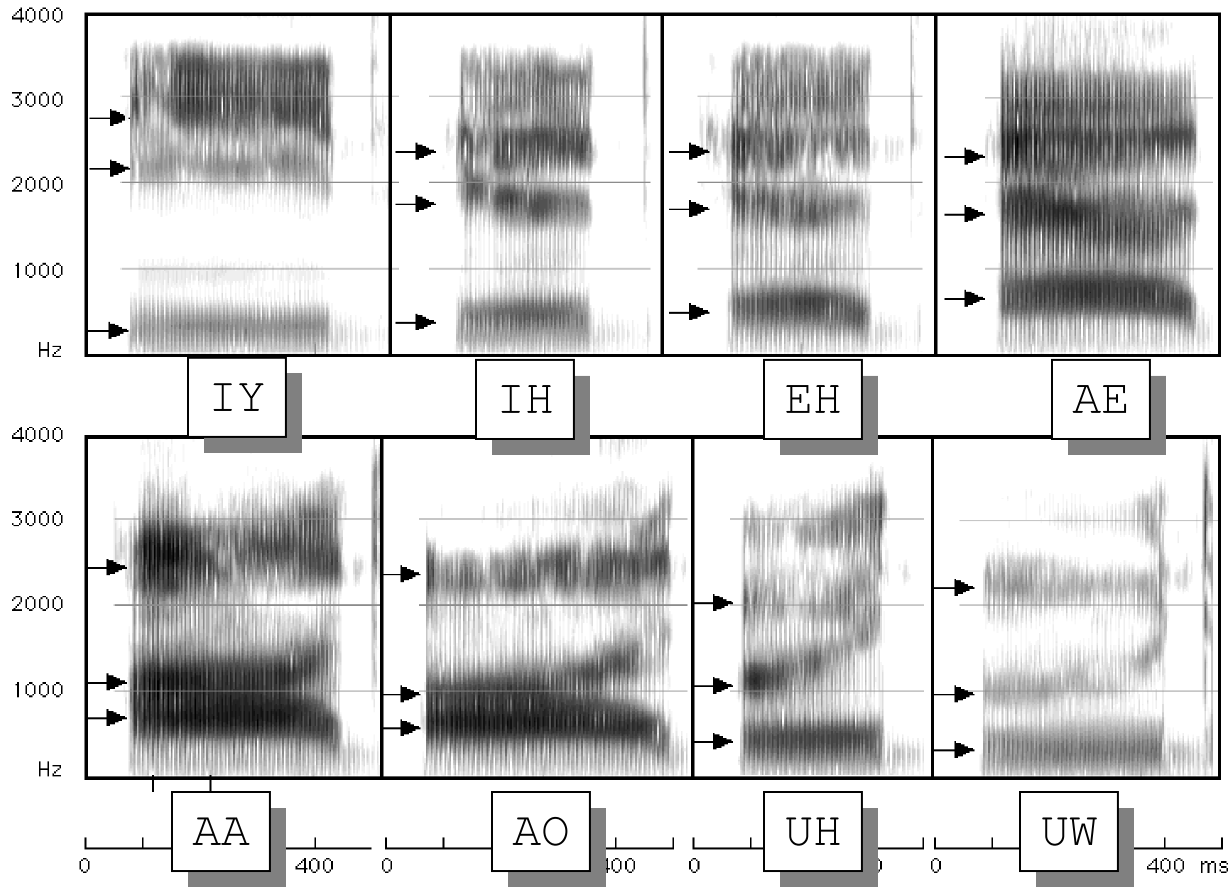
Describing Vowels

- **Velum Position**
  - nasal vs. non-nasal
- **Lip Shape**
  - Rounded vs. unrounded
- **Tongue height**
  - High, mid, low
  - Correlated to first formant position
- **Tongue advancement**
  - Front, central, back
  - Correlated to second formant position

American English Vowels

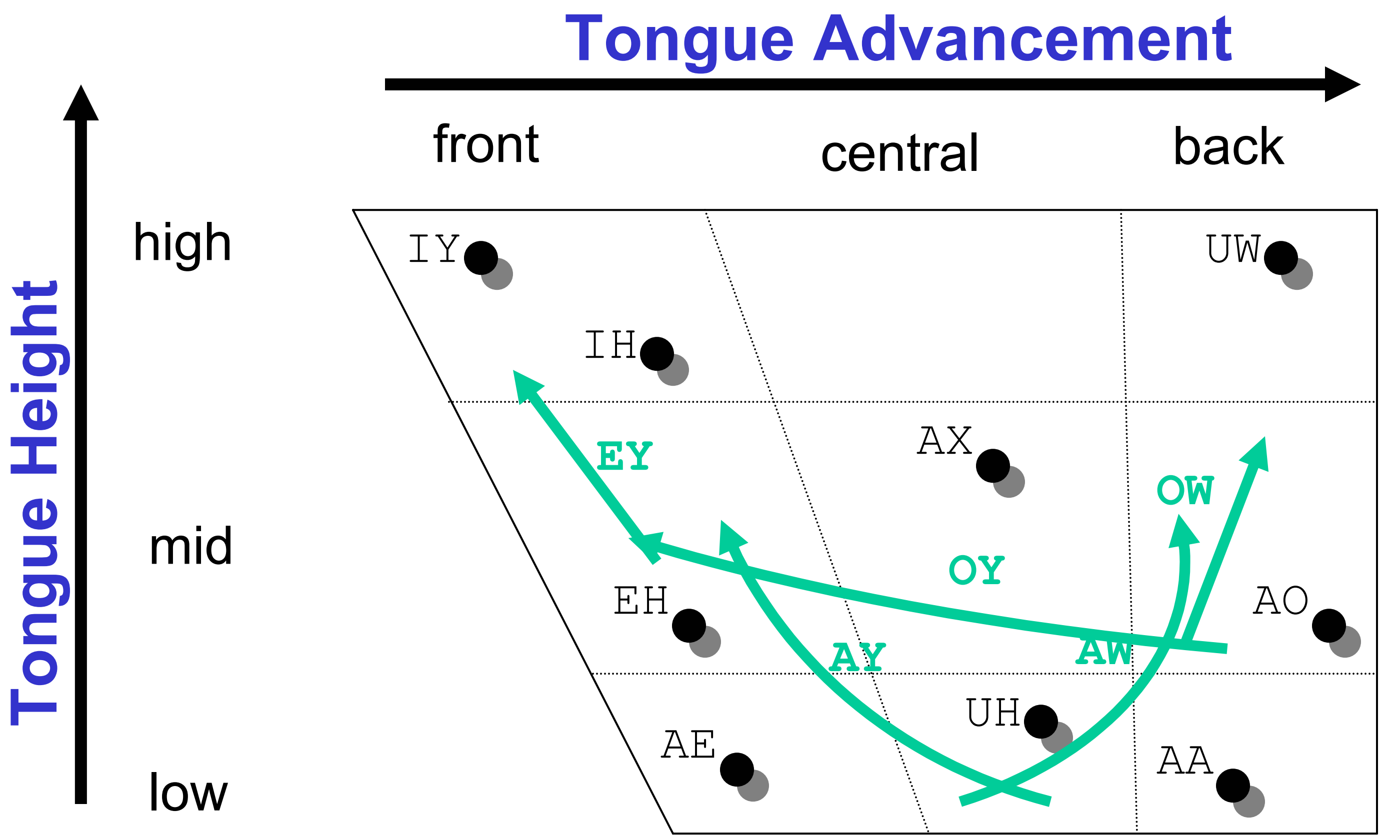
"heed"
"hid"
"head"
"had"

"hod"
"hawed"
"hood"
"who'd"



T-61.184

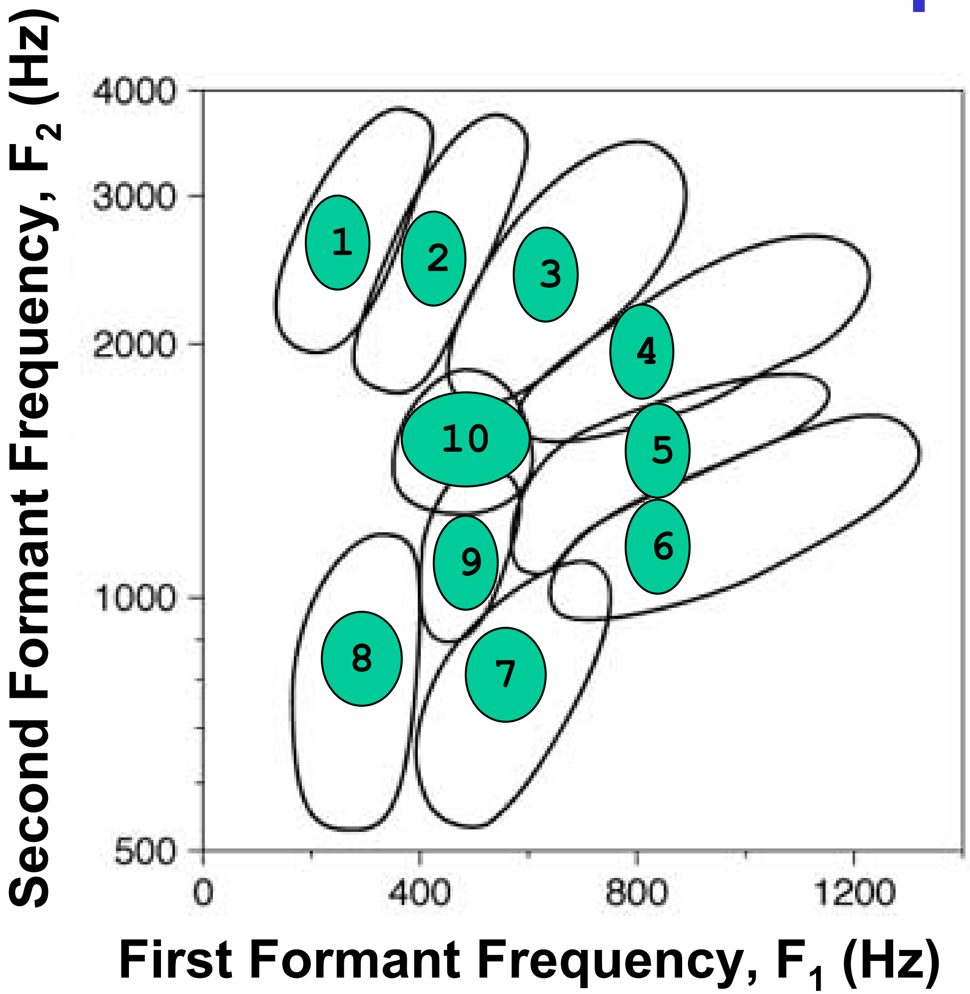
Vowel Chart



Diphthongs shown in green

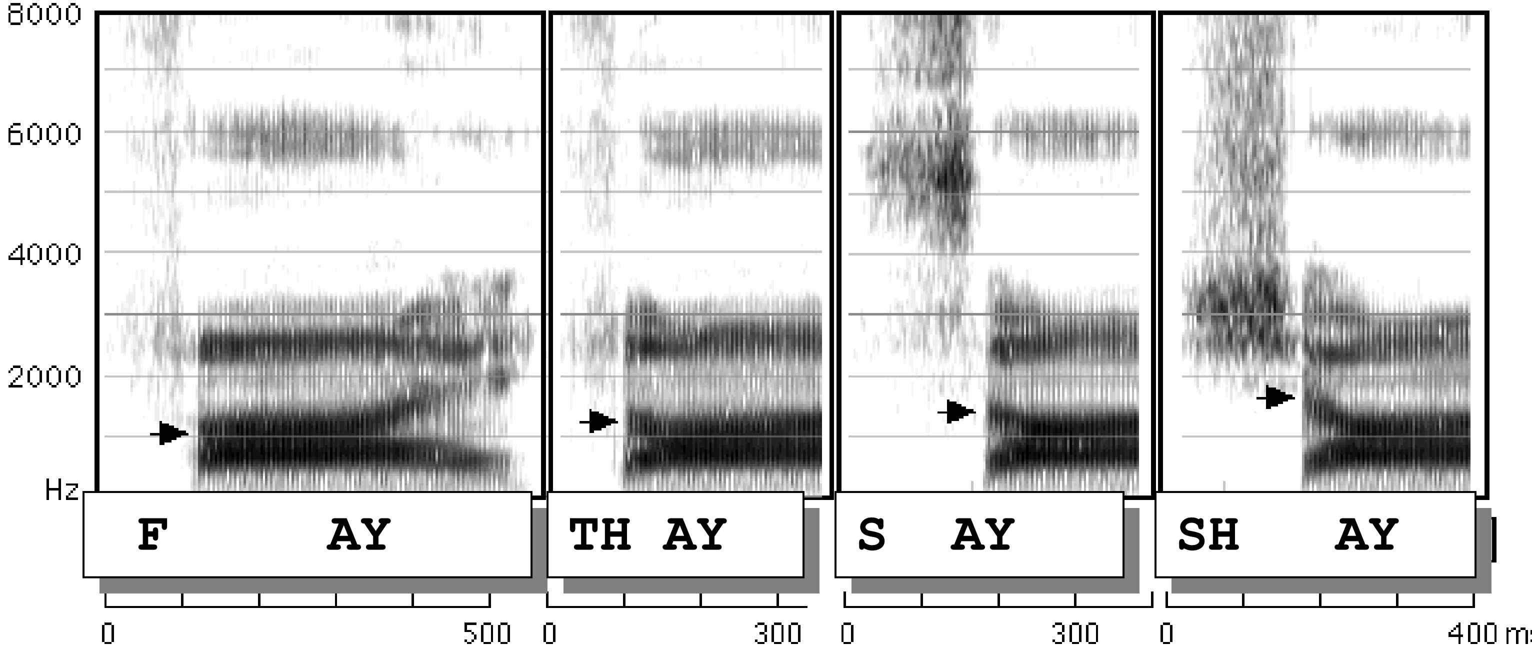
T-61.184

The Vowel Space



1. IY
2. IH
3. EH
4. AE
5. UH
6. AA
7. AO
8. UW
9. *(non English "u")*
10. AX

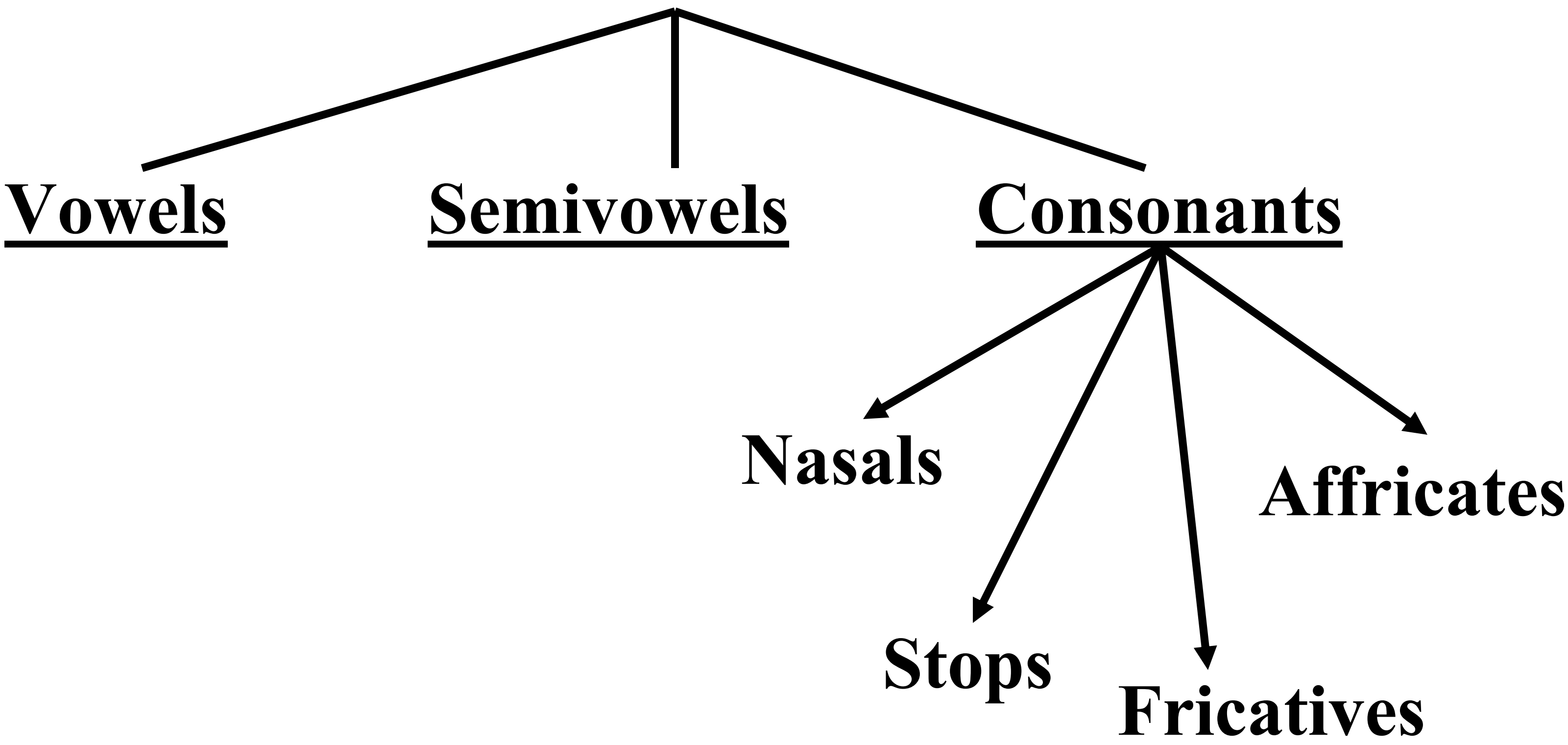
Coarticulation



Notice position of 2nd formant onset for these words: "fie", "thigh", "sigh", "shy"

http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants

Sound Classification Summary



Spectrogram Reading Video

- "Speech as Eyes See It" (12 minute video)
- 1977-1978 video by Ron Cole and Victor Zue
- After 2000-3000 hours of training: phonemes and words can be transcribed from a spectrogram alone
- 80-90% agreement on segments
- Provided insight into the speech recognition problem during the 1970's

Review: Probability & Statistics for Speech Recognition

Relative-Frequency and Probability

- **Relative Frequency of "A":**
  - Experiment is performed ***N*** times
  - With 4 possible outcomes: ***A, B, C, D***
  - N_A is number of times event ***A*** occurs:

$$N_A + N_B + N_C + N_D = N$$

$$\boxed{r(A) = \frac{N_A}{N}}$$

- **P(A)**
  - Defined as "the probability of event ***A***"
  - *Relative Frequency* that event ***A*** would occur if an experiment was repeated many times

Probability

- **Example (con't)**

$$N_A + N_B + N_C + N_D = N$$

$$r(A) + r(B) + r(C) + r(D) = 1$$

$$\boxed{P(A) = \lim_{N \to \infty} r(A)}$$

$$P(A) + P(B) + P(C) + P(D) = 1$$

Mutually Exclusive Events

- For <u>mutually exclusive</u> events A, B, C, …, M:

$$0 \le P(A) \le 1$$

$$P(A) + P(B) + P(C) + \cdots + P(M) = 1$$

$$P(A) = 0 \ \text{ represents impossible event}$$
$$P(A) = 1 \ \text{ represents certain event}$$

Joint and Conditional Probability

- P(AB) is the <u>joint probability</u> of event A and B both occurring

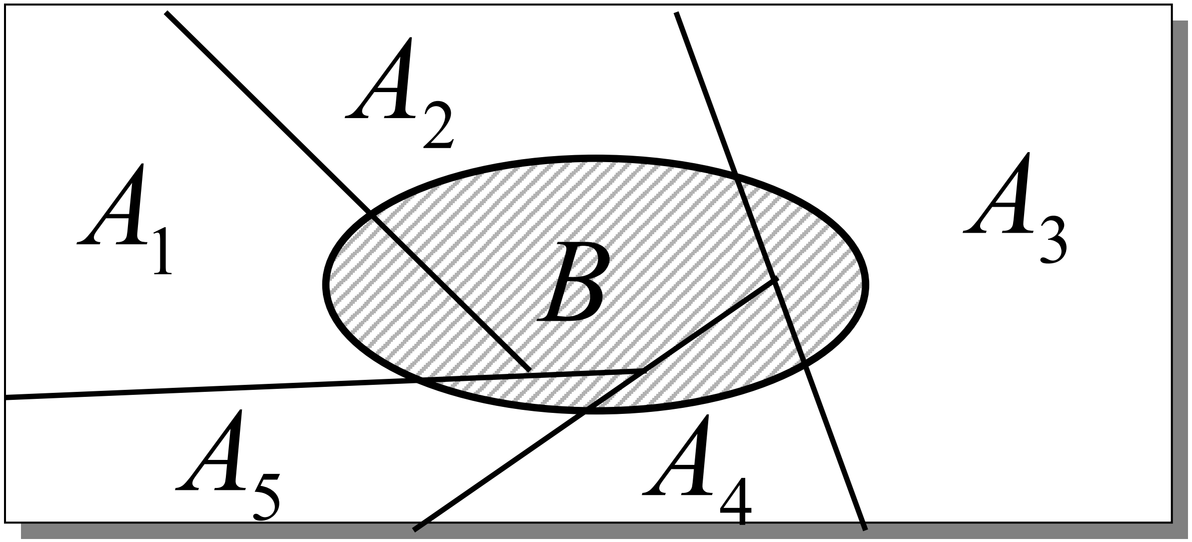
$$P(AB) = \lim_{N \to \infty} \frac{N_{AB}}{N}$$

- P(A|B) is the <u>conditional probability</u> of event A given that event B has occurred:

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \lim_{N \to \infty} \frac{N_{AB} / N}{N_B / N}$$

Marginal Probability

Probability of an event occurring across all conditions



$$P(B) = \sum_{k=1}^{n} P(A_k B) = \sum_{k=1}^{n} P(B \mid A_k) P(A_k)$$

Bayes' Theorem

$$P(AB) = P(B \mid A)P(A)$$

$$P(B) = \sum_{k=1}^{n} P(A_k B) = \sum_{k=1}^{n} P(B \mid A_k)P(A_k)$$

$$\boxed{P(A_i \mid B) = \frac{P(A_i B)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{P(B)}}$$

Bayes' Rule

- **Bayes' Rule allows us to update prior beliefs with new evidence,**

$$\underbrace{P(W_i)}_{\textit{Prior Probability}} \left[\frac{\overbrace{P(O \mid W_i)}^{\textit{Likelihood}}}{\underbrace{\sum_j P(O \mid W_j)P(W_j)}_{\textit{Evidence}\ \mathbf{P(O)}}} \right] = \underbrace{P(W_i \mid O)}_{\textit{Posterior Probability}}$$

Statistically Independent Events

- **Occurrence of event A does not influence occurrence of event B:**

$$P(AB) = P(A)P(B)$$

$$P(A \mid B) = P(A)$$

$$P(B \mid A) = P(B)$$

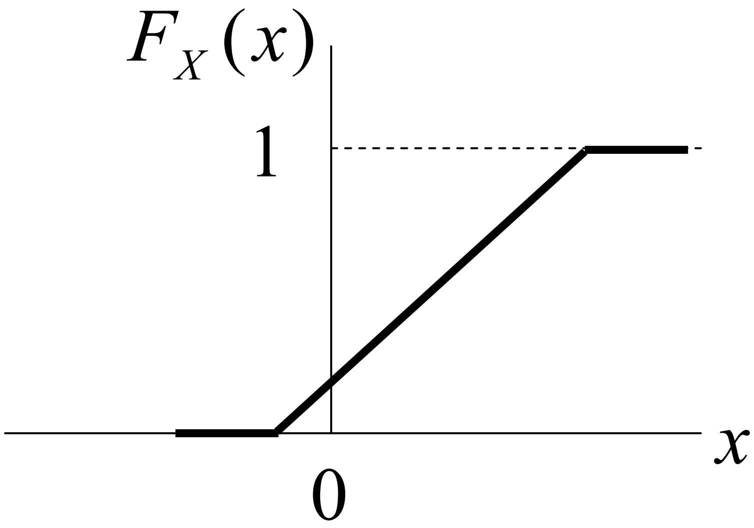
Random Variables

- Used to describe events in which the number of possible outcomes is infinite
- Values of the outcomes can not be predicted with certainty
- Distribution of outcome values is known

Probability Distribution Functions

- **The probability of the event that the random variable *X* is less than or equal to the allowed value *x*:**

$$F_X(x) = P(X \le x)$$



Probability Distribution Functions

- **Properties:**

$$0 \le F_X(x) \le 1 \qquad -\infty < x < \infty$$

$$F_X(-\infty) = 0 \quad \text{and} \quad F_X(\infty) = 1$$

$$F_X(x) \text{ is nondecreasing as } x \text{ increases}$$

$$P(x_1 < X \le x_2) = F_X(x_2) - F_X(x_1)$$

Probability Density Functions (PDF)

- **Derivative of probability distribution function,**

$$\boxed{f_X(x) = \frac{dF_X(x)}{dx}}$$

- **Interpretation:**

$$f_X(x) = P(x < X \leq x + dx)$$

Probability Density Functions (PDF)

- **Properties**

$$f_X(x) \geq 0 \qquad -\infty < x < \infty$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

$$\int_{x_1}^{x_2} f_X(x)dx = P(x_1 < X \leq x_2)$$

Mean and Variance

- <u>Expectation</u> or <u>Mean</u> of a random variable ***X***

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

- <u>Variance</u> of a random variable ***X***

$$\mu = E(X)$$

$$Var(X) = \sigma^2 = E\left[(X-\mu)^2\right]$$

$$Var(X) = \sigma^2 = E(X^2) - \left[E(X)\right]^2$$

Mean and Variance

- **Variance properties (if *X* and *Y* are independent)**

$$Var(X+Y) = Var(X) + Var(Y)$$

$$Var(aX) = a^2 Var(X)$$

$$Var(a_1 X_1 + \cdots + a_n X_n + b) = \\ a_1^2 Var(X_1) + \cdots + a_n^2 Var(X_n)$$

Covariance and Correlation

- **Covariance of *X* and *Y*:**

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$$

$$Cov(X,Y) = Cov(Y,X)$$

- **Correlation of *X* and *Y*:**

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \qquad -1 \le \rho_{XY} \le 1$$

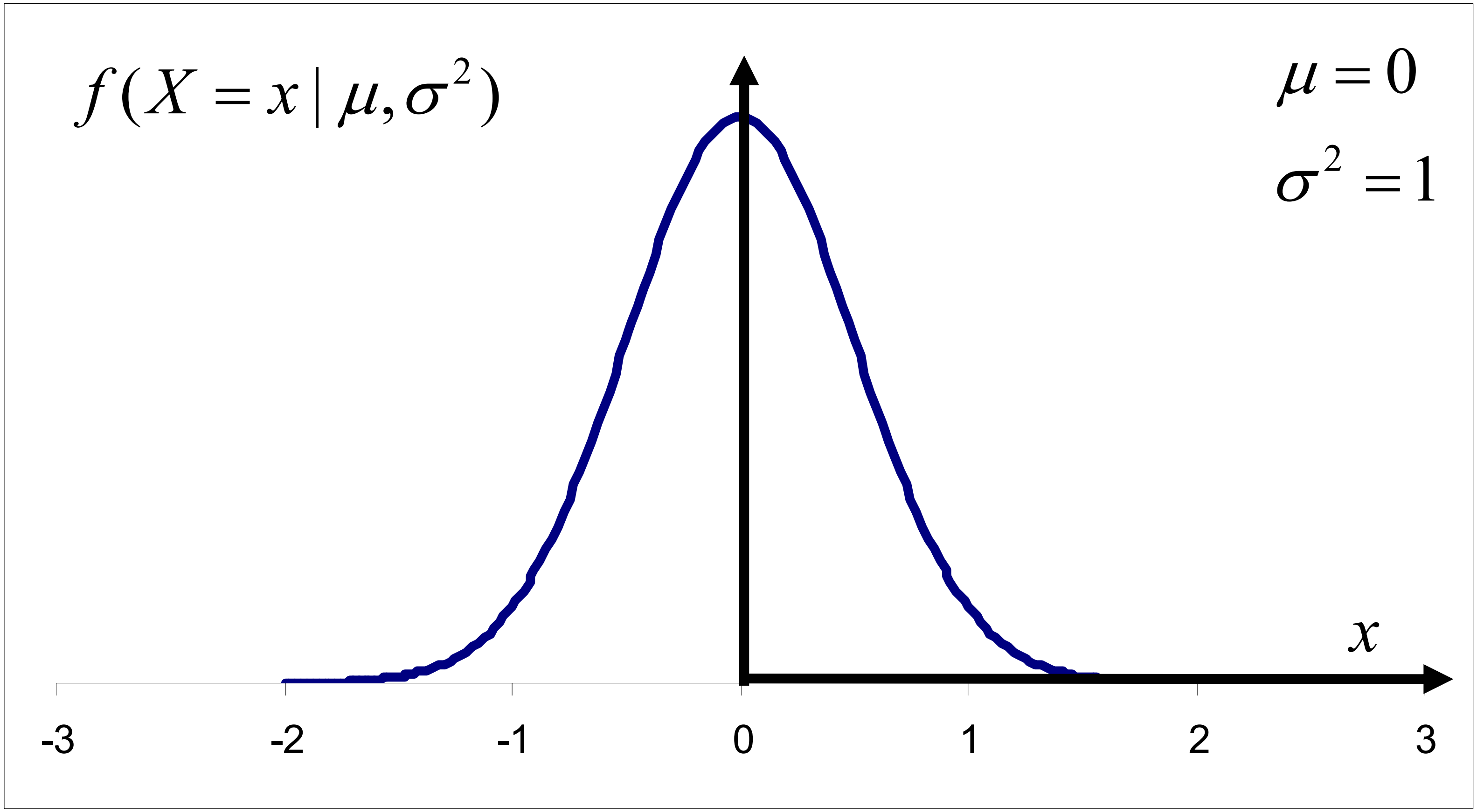
Gaussian (Normal) Distribution

$$f(X = x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$\mu = E[x] = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\sigma^2 = E(X^2) - [E(X)]^2 = \frac{1}{N}\sum_{n=1}^{N} x_n^2 - \left[\frac{1}{N}\sum_{n=1}^{N} x_n\right]^2$$

Gaussian (Normal) Distribution



Multivariate Distributions

- **Characterize more than one random variable at a time**
- **Example: Features for speech recognition**
  - In speech recognition we typically compute 39-dimensional feature vectors (more about that later)
  - 100 feature vectors per second of audio
- **Often want to compute the likelihood of the observed features given known (estimated) distribution which is being used to model some part of a phoneme (more about that later!).**

Multivariate Gaussian Distribution

$$f(X = x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} \underline{|\Sigma|^{1/2}}} \exp\left[-\frac{1}{2} (x - \mu)^{\mathrm{T}} \Sigma^{-1} (x - u) \right]$$

Determinant of covariance matrix

Distribution covariance matrix

Distribution mean vector

Observed vector of random variables (features)

Diagonal Covariance Assumption

- **Most speech recognition systems assume diagonal covariance matrices**

- **Data sparseness issue:**

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & 0 & 0 & 0 \\ 0 & \sigma_{22}^2 & 0 & 0 \\ 0 & 0 & \sigma_{33}^2 & 0 \\ 0 & 0 & 0 & \sigma_{44}^2 \end{bmatrix} \longrightarrow \boxed{|\Sigma| = \prod_{n=1}^{d} \sigma_{nn}^2}$$

Diagonal Covariance Assumption

- **Inverting a diagonal matrix involves simply inverting the elements along the diagonal:**

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_{11}^2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_{22}^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_{33}^2} & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_{44}^2} \end{bmatrix}$$

Multivariate Gaussians with Diagonal Covariance Matrices

- **Assuming a diagonal covariance matrix,**

$$f(X = x \mid \mu, \Sigma) = C * \exp\left[-\frac{1}{2} \sum_{d=1}^{n} \frac{(x[d] - \mu[d])^2}{\sigma^2[d]} \right]$$

$$C = \frac{1}{(2\pi)^{n/2} \left(\prod_{d=1}^{n} \sigma^2[d] \right)^{1/2}}$$

$$\sigma^2[d] = \sigma^2_{dd}$$

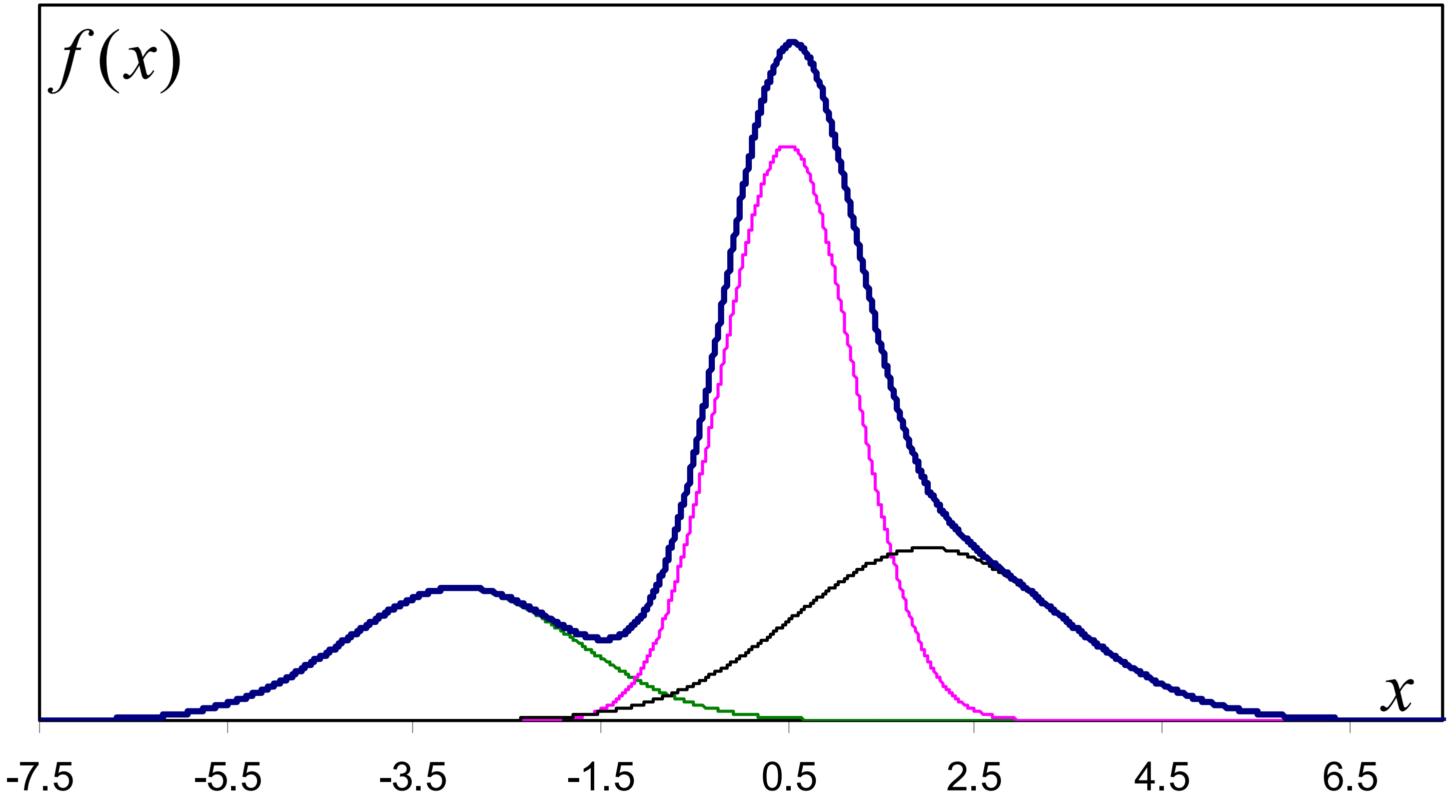
Multivariate *Mixture* Gaussians

- **Distribution is governed by several Gaussian density functions,**
- **Sum of Gaussians (w_m = mixture weight)**

$$f(x) = \sum_{m=1}^{M} w_m \mathrm{N}_m(x; \mu_m, \Sigma_m)$$

$$= \sum_{m=1}^{M} \frac{w_m}{(2\pi)^{n/2} \left|\Sigma_m\right|^{1/2}} \exp\left[-\frac{1}{2} (x - \mu_m)^{\mathrm{T}} \Sigma_m^{-1} (x - u_m) \right]$$

Multiple Mixtures (1-D case)



Example: 3 mixtures used to model underlying random process of 3 Gaussians

Speech Recognition
Problem Formulation

Problem Description

- **Given a sequence of observations (evidence) from an audio signal,**

$$\mathrm{O} = o_1 o_2 \cdots o_T$$

- **Determine the underlying word sequence,**

$$\mathrm{W} = w_1 w_2 \cdots w_m$$

- **Number of words (m) unknown, observation sequence is variable length (T)**

Problem Formulation

- **Goal: Minimize the classification error rate**

$$\Downarrow$$

- **Solution: Maximize the Posterior Probability**

$$\hat{W} = \arg\max_{W} P(W \mid O)$$

- **Solution requires optimization over all possible word strings!**

Problem Formulation

- **Using Bayes Rule,**

$$P(\mathrm{W} \mid \mathrm{O}) = \frac{P(\mathrm{O} \mid \mathrm{W})P(\mathrm{W})}{P(\mathrm{O})}$$

- **Since P(O) does not impact optimization,**

$$\begin{aligned}\hat{\mathrm{W}} &= \underset{\mathrm{W}}{\arg\max}\, P(\mathrm{W} \mid \mathrm{O}) \\ &= \underset{\mathrm{W}}{\arg\max}\, P(\mathrm{O} \mid \mathrm{W})P(\mathrm{W})\end{aligned}$$

Problem Formulation

- **Let's assume words can be represented by a sequence of states, S,**

$$\hat{W} = \arg\max_{W} P(O \mid W) P(W)$$

$$= \arg\max_{W} \sum_{S} P(O \mid S) P(S \mid W) P(W)$$

- **Words → Phonemes → States**
- **States represent smaller pieces of phonemes**

Problem Formulation

- **Optimize:** $$\hat{W} = \arg\max_{W} \sum_{S} P(O \mid S) P(S \mid W) P(W)$$

- **Practical Realization,**

O	Observation (feature) sequence
$P(O \mid S)$	Acoustic Model
$P(S \mid W)$	Lexicon / Pronunciation Model
$P(W)$	Language Model

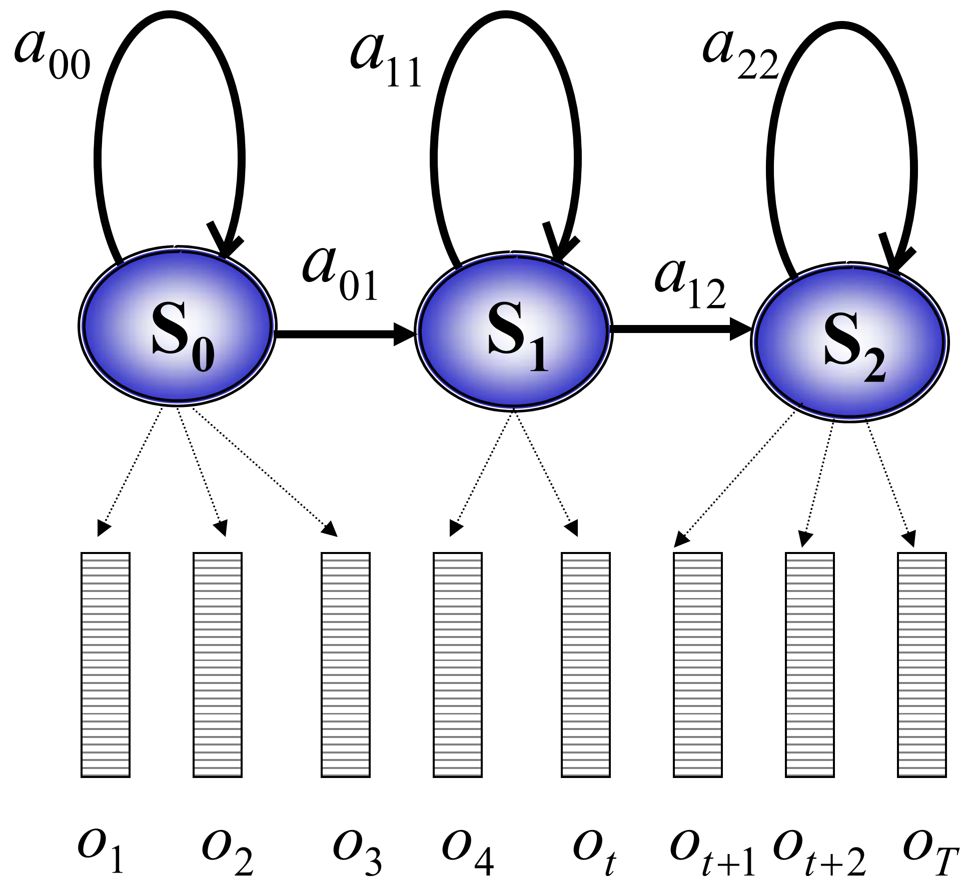
T-61.184

Problem Formulation

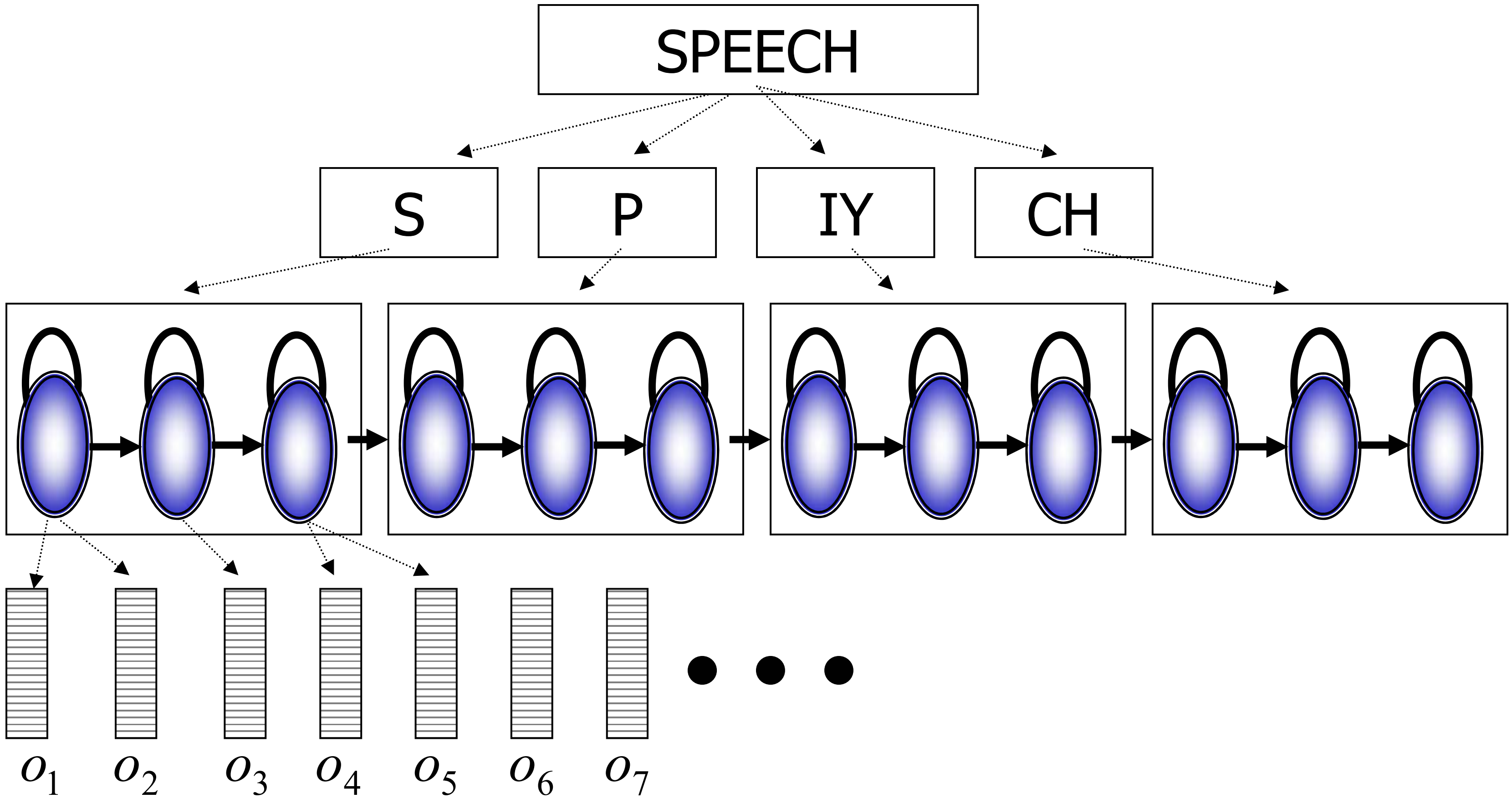
- **Optimization desires most likely word sequence given observations (evidence)**
- **Can not evaluate all possible word / state sequences (too many possibilities!)**
- **We need:**
  - To define a representation for modeling states (HMMs…)
  - A means for "approximately" searching for the best word / state sequence given the evidence (Viterbi Algorithm)
  - And a few other tricks up our sleeves to make it FASTER!

Hidden Markov Models (HMMs)

- **Observation vectors are assumed to be "generated" by a Markov Model**
- **HMM: A finite-state machine that at each time t that a state j is entered, an observation is emitted with probability density $b_j(o_t)$**
- **Transition from state i to state j modeled with probability a_{ij}**

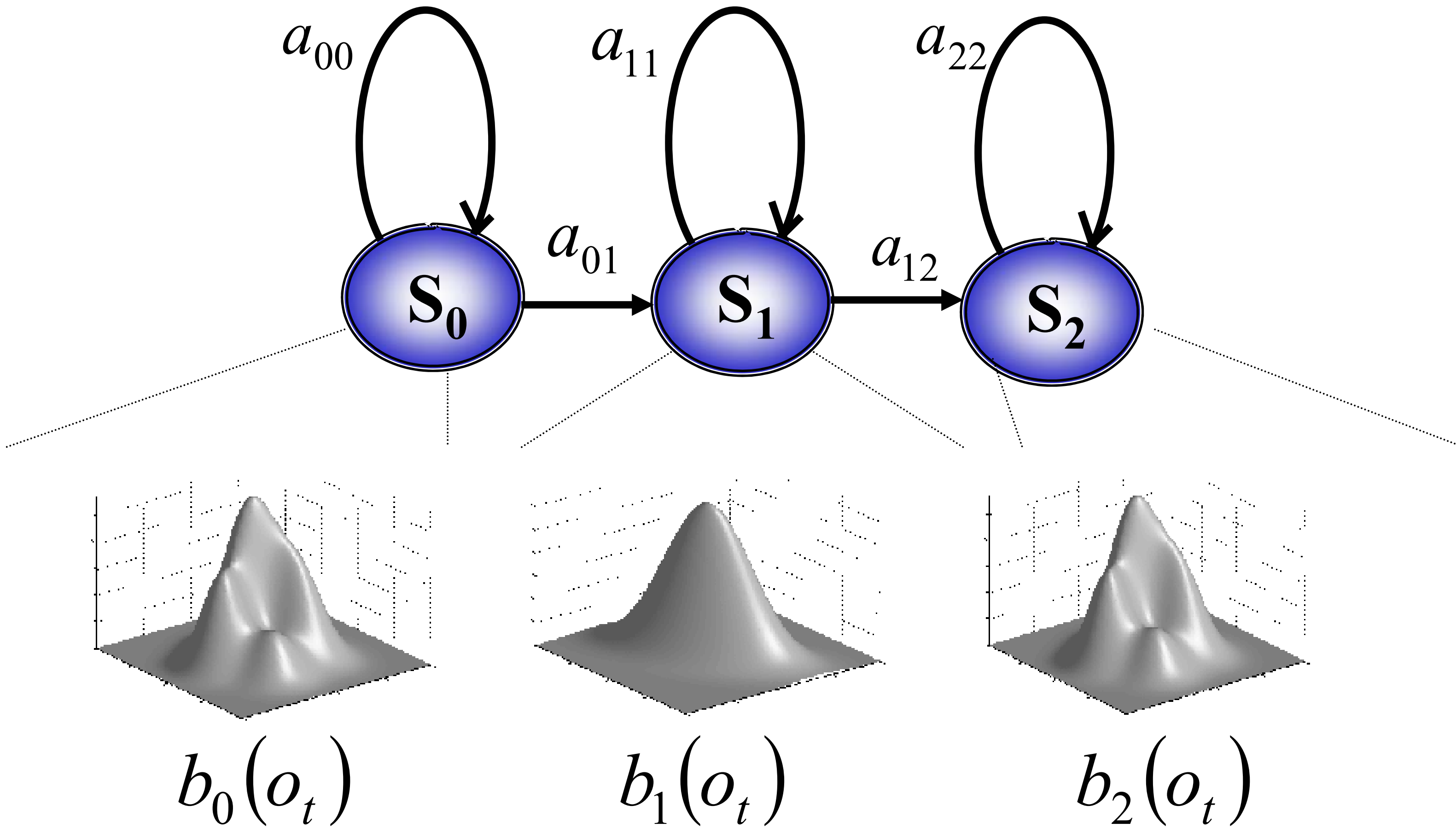


"Beads-on-a-String" HMM Representation



T-61.184

Modeled Probability Distribution



HMMs with Mixtures of Gaussians

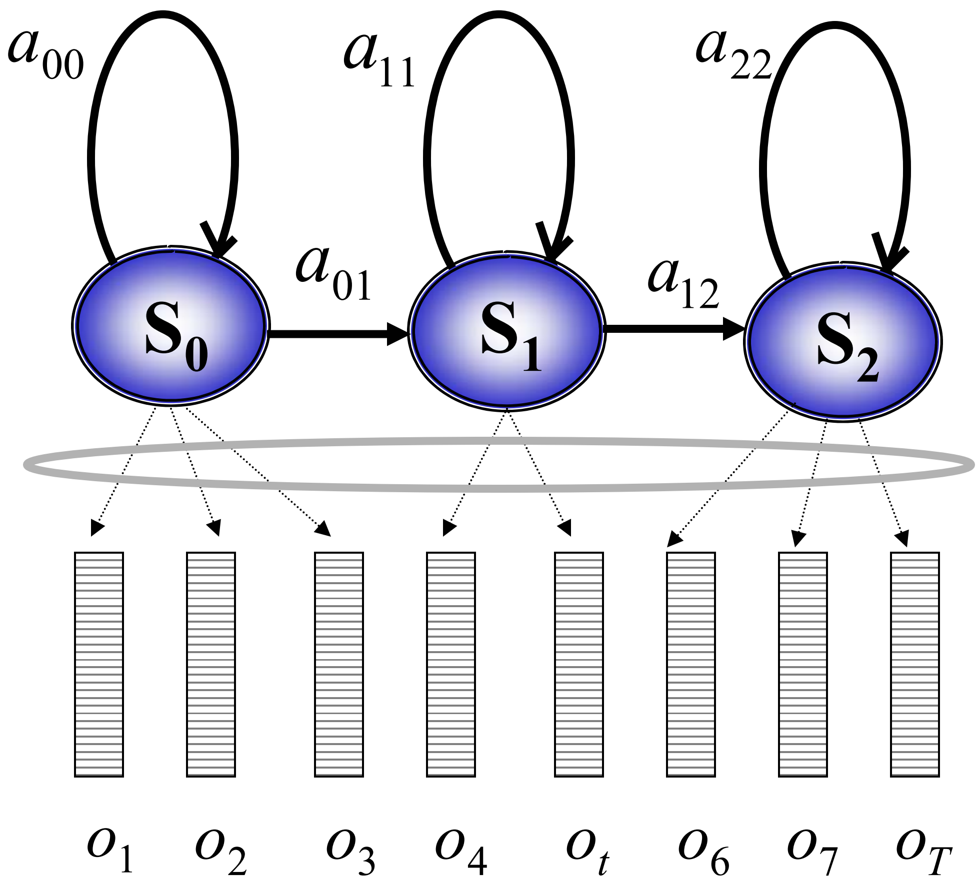
- **Multivariate Mixture-Gaussian distribution,**

$$b_j(o_t) = \sum_{m=1}^{M} \frac{w_m}{\sqrt{(2\pi)^n |\Sigma_j|}} e^{-\frac{1}{2}(o_t - \mu_j)^T \Sigma_j^{-1} (o_t - \mu_j)}$$

- **Model parameters are (1) means, (2) variances, and (3) mixture weights**
- **Sum of Gaussians can model complex distributions**

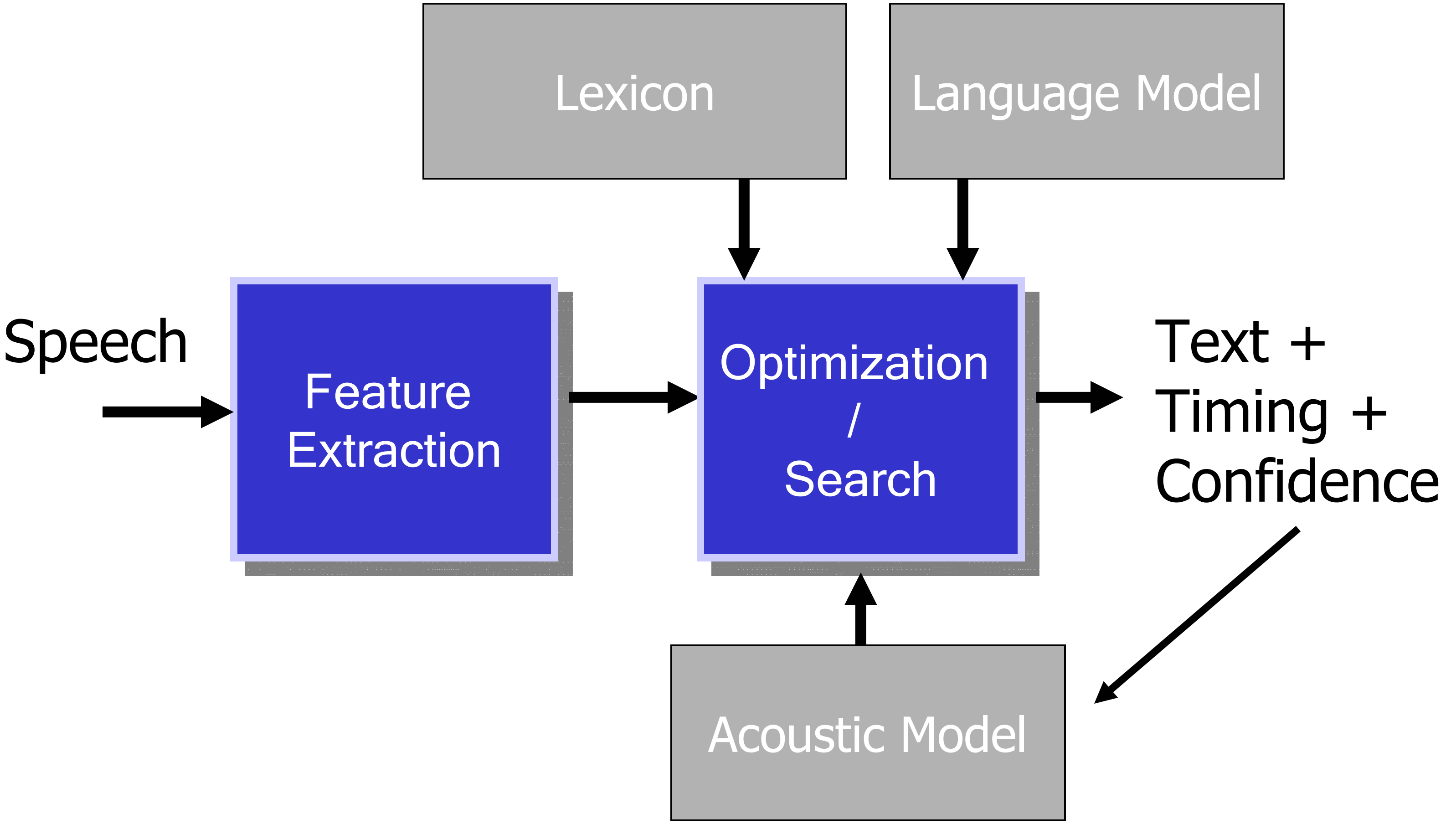
Hidden Markov Model Based ASR

- **Observation sequence assumed to be known**
- **Probability of a particular state sequence can be computed**
- **Underlying state sequence is unknown, "hidden"**



T-61.184

Components of a Speech Recognizer



Topics for Next Time

- An introduction to hearing
- Speech Detection
- Frame-based Speech Analysis
- Feature Extraction Methods for Speech Recognition