CS460/626 : Natural Language Processing/Speech, NLP and the Web
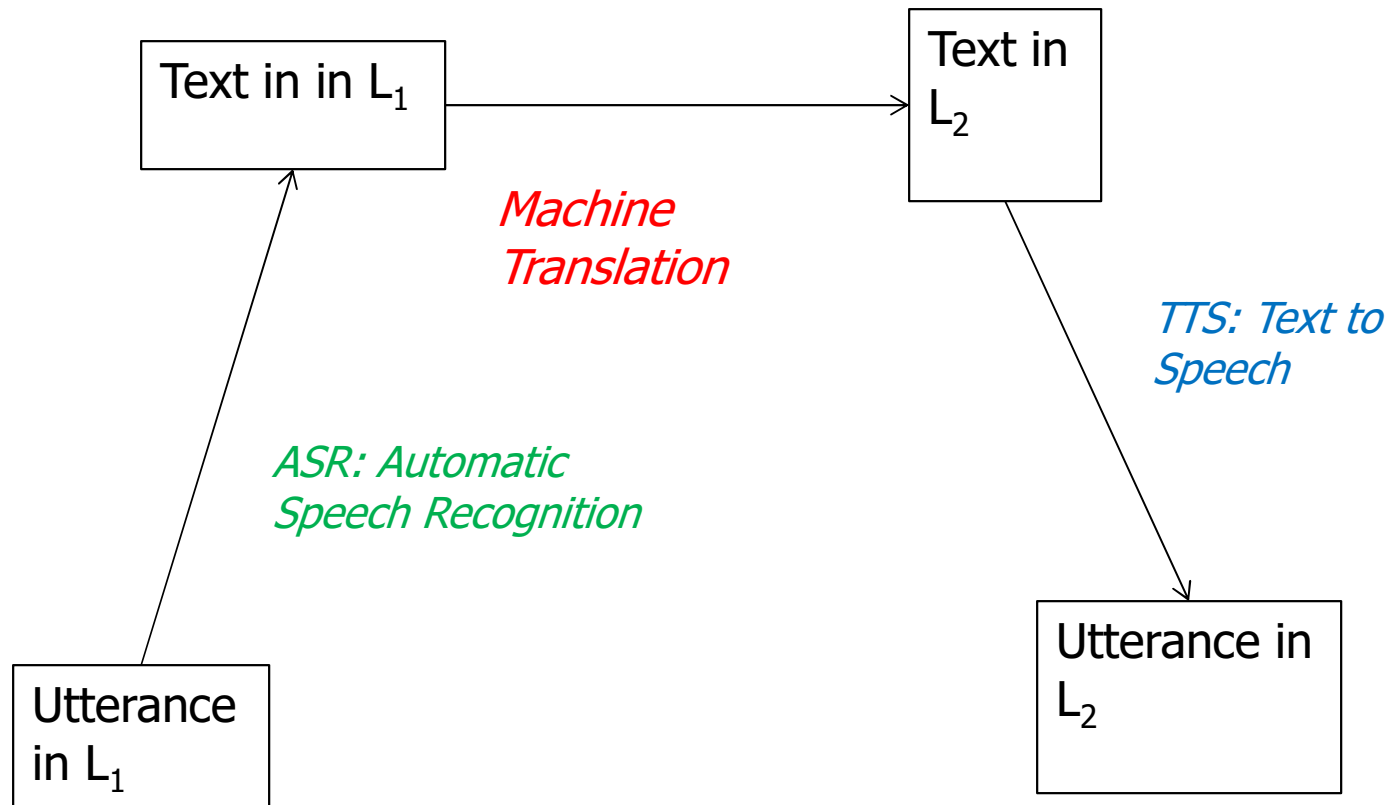
Lecture 28, 29:
Phonetics, Phonology and Speech; introduce transliteration

Pushpak Bhattacharyya
CSE Dept.,
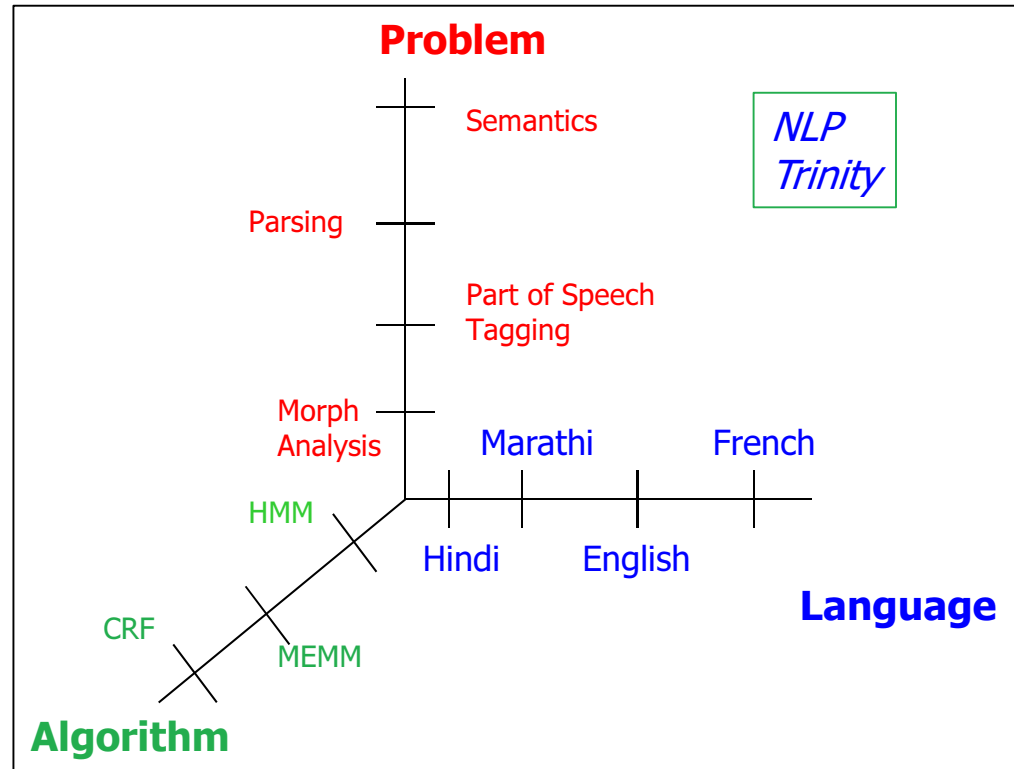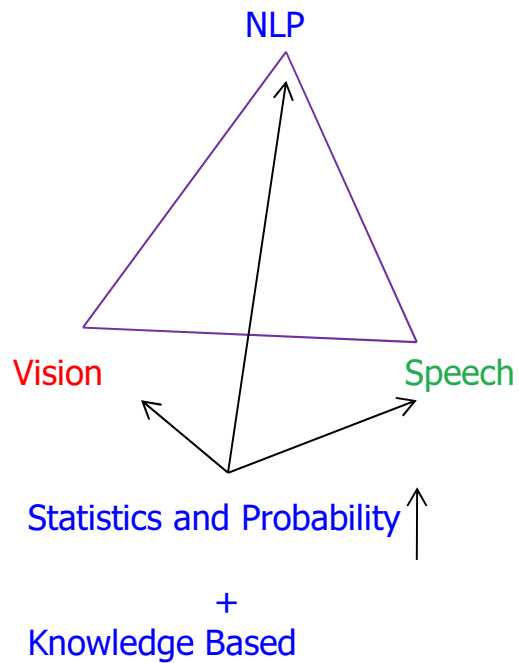IIT Bombay

28th and 29th Oct, 2012

Speech and NLP

- Speech is the "original" language data

- Writing system came much later!

- Word boundary and pause can completely alter the meaning of utterances

    - *aa jaayenge/aaj aayenge*
    - *I got a plate/I got up late*
    - *When it rains cats and dogs, run for cover/When it rains, cats and dogs run for cover*

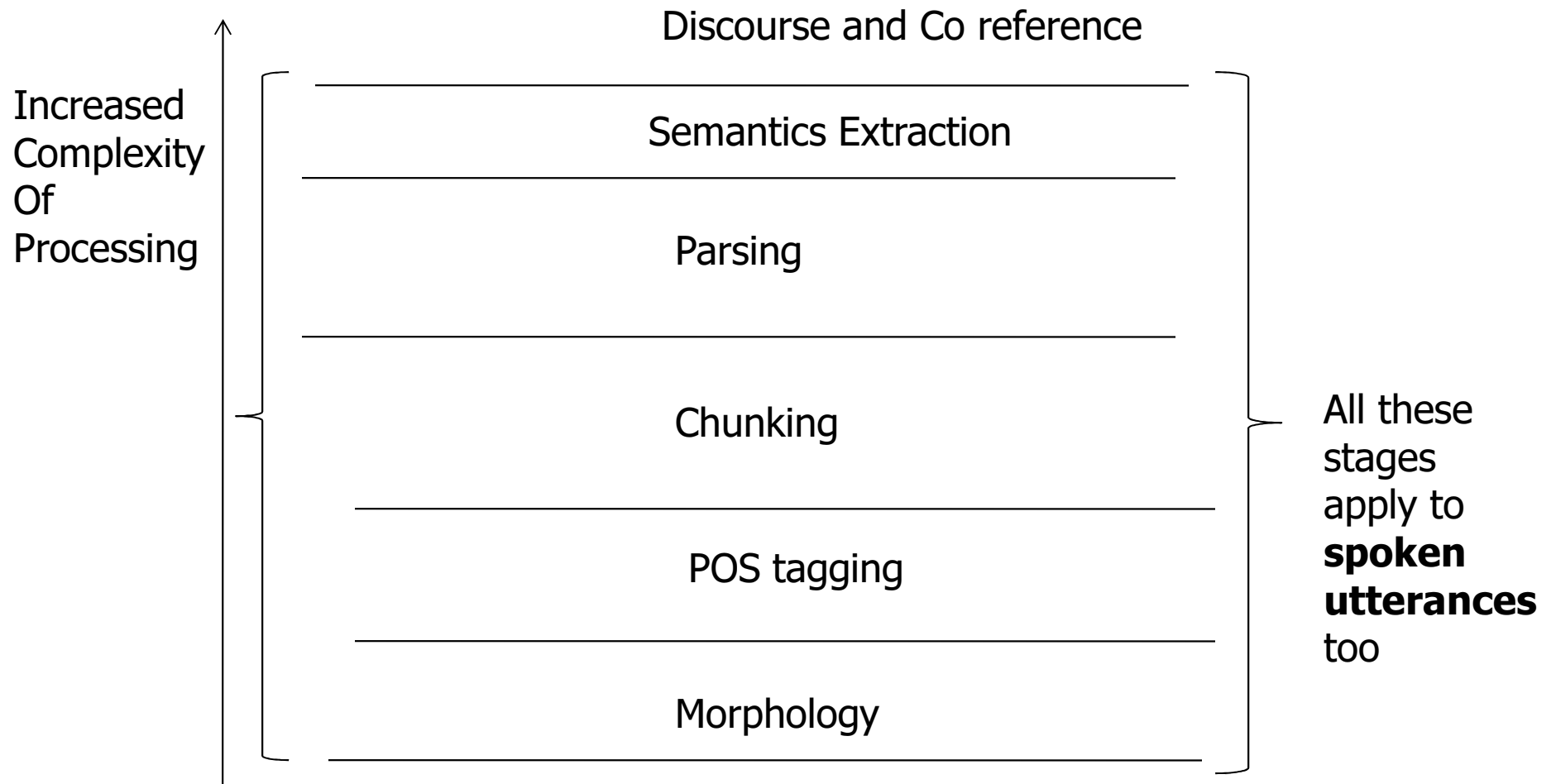- Speech to Speech Machine Translation: *killer* application

A vision

Text in in L_1 → Text in L_2

Machine Translation

TTS: Text to Speech

ASR: Automatic Speech Recognition

Utterance in L_1

Utterance in L_2

The trinity

NLP

Vision   Speech

Statistics and Probability

\+

Knowledge Based

Problem

Semantics

NLP
Trinity

Parsing

Part of Speech
Tagging

Morph
Analysis   Marathi   French

HMM

Hindi   English

Language

CRF

MEMM

Algorithm

NLP Layer and speech

Increased
Complexity
Of
Processing

Discourse and Co reference

Semantics Extraction

Parsing

Chunking

POS tagging

Morphology

All these stages apply to **spoken utterances** too

Probabilistic Speech Recognition

- **Problem Definition : Given a sequence of speech signals, identify the words.**

- **2 steps :**

  - **Segmentation (Word Boundary Detection)**

  - **Identify the word**

- **Isolated Word Recognition :**

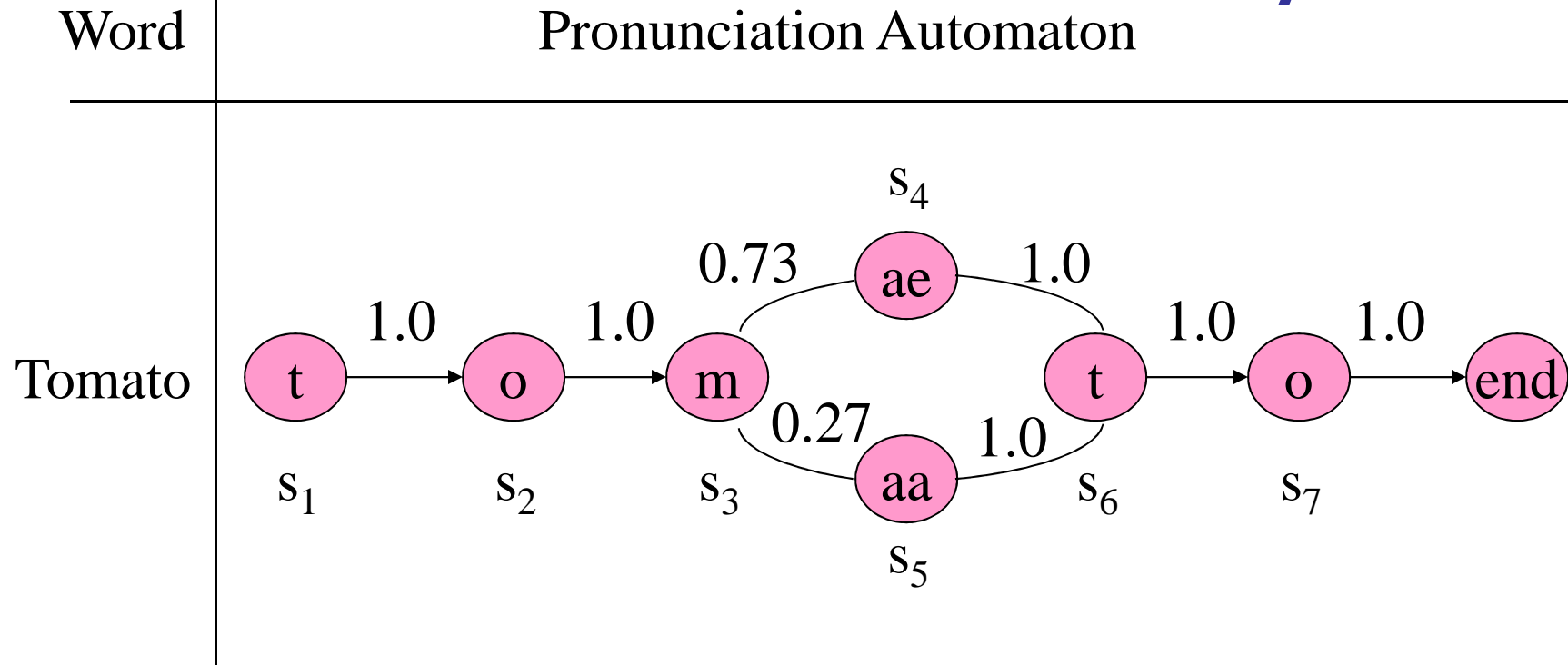 - **Identify W given SS (speech signal)**

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

Speech recognition: Identifying the word

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

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

- *P(SS/W)* = likelihood called "phonological model" → intuitively more tractable!
- *P(W)* = prior probability called "language model"

$$P(W) = \frac{\#\ W\ \text{appears in the corpus}}{\#\ \text{words in the corpus}}$$

Pronunciation Dictionary

Word	Pronunciation Automaton

- *P(SS/W)* is maintained in this way.
- *P(t o m ae t o |Word is "tomato")* = Product of arc probabilities

Grapheme to phoneme mapping is not unique

- The plural morpheme:
  - -s:
    - /s/ (cats)
    - /z/ (dogs)
    - /iz/ (bushes)

- Different sounds

Representing sound can be challenging (as its meaning)

- Afrikaans: **bromponie** a motor scooter (literally, a growling or muttering pony)
- IsiNdebele: **U-Linda** mind the village until the father's return
- Setswana: **bitlisisa** a sore eye that has been rubbed
- Tshivenda: **mmbwe** a round pebble taken from a crocodile's stomach and swallowed by a chief

  mvula-tshikole rain with sunshine
- Xitsonga: **byatabyata** to try to say something but fail for lack of words

  kentenga to find oneself suddenly without some vital item (of a man whose only wife has run away, or when the roof of a hut has blown off)

    (The above are African languages)

CMU Pronunciation dictionary

- machine-readable pronunciation dictionary for North American English that contains over **125,000 words** and their transcriptions.

- The current phoneme set contains 39 phonemes

"Parallel" Corpus

```
Phoneme Example Translation
-------- -------- ------------

AA     odd     AA D
AE     at      AE T
AH     hut     HH AH T
AO     ought   AO T
AW     cow     K AW
AY     hide    HH AY D
B      be      B IY
```

"Parallel" Corpus cntd

Phoneme Example Translation

------- ------- ------------

CH   cheese CH IY Z

D   dee D IY

DH   thee DH IY EH Ed EH D

ER   hurt HH ER T

EY   ate EY T

F   fee F IY

G   green G R IY N

HH   he HH IY

IH   it IH T

IY   eat IY T

JH   gee JH IY

A Statistical Machine Translation like task

- First obtain the Carnegie Mellon University's Pronouncing Dictionary
- Train and Test the following Statistical Machine Learning Algorithms
- HMM - For HMM we can use either Natural Language Toolkit or you can use GIZA++ with MOSES

Phonetics and Phonology

- **Phonetics: The study of speech sounds**
  - Articulatory
  - Acoustic
  - Auditory

- **Phonology: the structure and patterning of sounds**

- **Phonetic Transcription:**
  - A writing system for representing speech sounds

The need for phonetic transcription

- **Eccentricity of English Spelling**
  - Put/Putt
  - Car/Kite
  - Rough/Puff
  - 'Fish' can be spelt 'ghoti'; (Bernard Shaw: 'laugh', 'women', 'nation')

- **A standardized system for representing sounds in languages**
  - IPA (International)
  - ARPABET (mainly US)

IPA and ARPAbet vowels

ARPAbet Symbol	IPA Symbol	Word	ARPAbet Transcription
[iy]	[i]	lily	[l ih l iy]
[ih]	[ɪ]	lily	[l ih l iy]
[ey]	[eɪ]	daisy	[d ey z iy]
[eh]	[ɛ]	pen	[p eh n]
[ae]	[æ]	aster	[ae s t axr]
[aa]	[ɑ]	poppy	[p aa p iy]
[ao]	[ɔ]	orchid	[ao r k ix d]
[uh]	[ʊ]	wood	[w uh d]
[ow]	[oʊ]	lotus	[l ow dx ax s]
[uw]	[u]	tulip	[t uw l ix p]
[ah]	[ʌ]	buttercup	[b ah dx axr k ah p]
[er]	[ɝ]	bird	[b er d]
[ay]	[aɪ]	iris	[ay r ix s]
[aw]	[aʊ]	sunflower	[s ah n f l aw axr]
[oy]	[ɔɪ]	soil	[s oy l]

Reduced and uncommon phones

ARPAbet Symbol	IPA Symbol	Word	ARPAbet Transcription
[ax]	[ə]	lotus	[l ow dx ax s]
[axr]	[ɚ]	heather	[h eh dh axr]
[ix]	[ɨ]	tulip	[t uw l ix p]
[ux]	[ʉ]	dude[1]	[d ux d]

IPA and ARPAbet consonents

ARPAbet Symbol	IPA Symbol	Word	ARPAbet Transcription
[p]	[p]	parsley	[p aa r s l iy]
[t]	[t]	tea	[t iy]
[k]	[k]	cook	[k uh k]
[b]	[b]	bay	[b ey]
[d]	[d]	dill	[d ih l]
[g]	[g]	garlic	[g aa r l ix k]
[m]	[m]	mint	[m ih n t]
[n]	[n]	nutmeg	[n ah t m eh g]
[ng]	[ŋ]	baking	[b ey k ix ng]
[f]	[f]	flour	[f l aw axr]
[v]	[v]	clove	[k l ow v]
[th]	[θ]	thick	[th ih k]
[dh]	[ð]	those	[dh ow z]
[s]	[s]	soup	[s uw p]
[z]	[z]	eggs	[eh g z]
[sh]	[ʃ]	squash	[s k w aa sh]
[zh]	[ʒ]	ambrosia	[ae m b r ow zh ax]
[ch]	[tʃ]	cherry	[ch eh r iy]
[jh]	[dʒ]	jar	[jh aa r]
[l]	[l]	licorice	[l ih k axr ix sh]
[w]	[w]	kiwi	[k iy w iy]
[r]	[r]	rice	[r ay s]
[y]	[j]	yellow	[y eh l ow]
[h]	[h]	honey	[h ah n iy]
Less commonly used phones and allophones			
[q]	[ʔ]	uh-oh	[q ah q ow]
[dx]	[ɾ]	butter	[b ah dx axr]
[nx]	[ɾ̃]	winner	[w ih nx axr]
[el]	[l̩]	table	[t ey b el]

Text Input Methods: Keyboard

English QWERTY

a alveolar ridge
b hard palate
c velum
d vocal folds

nasal cavity

upper lip

teeth

lower lip

tongue

tongue root
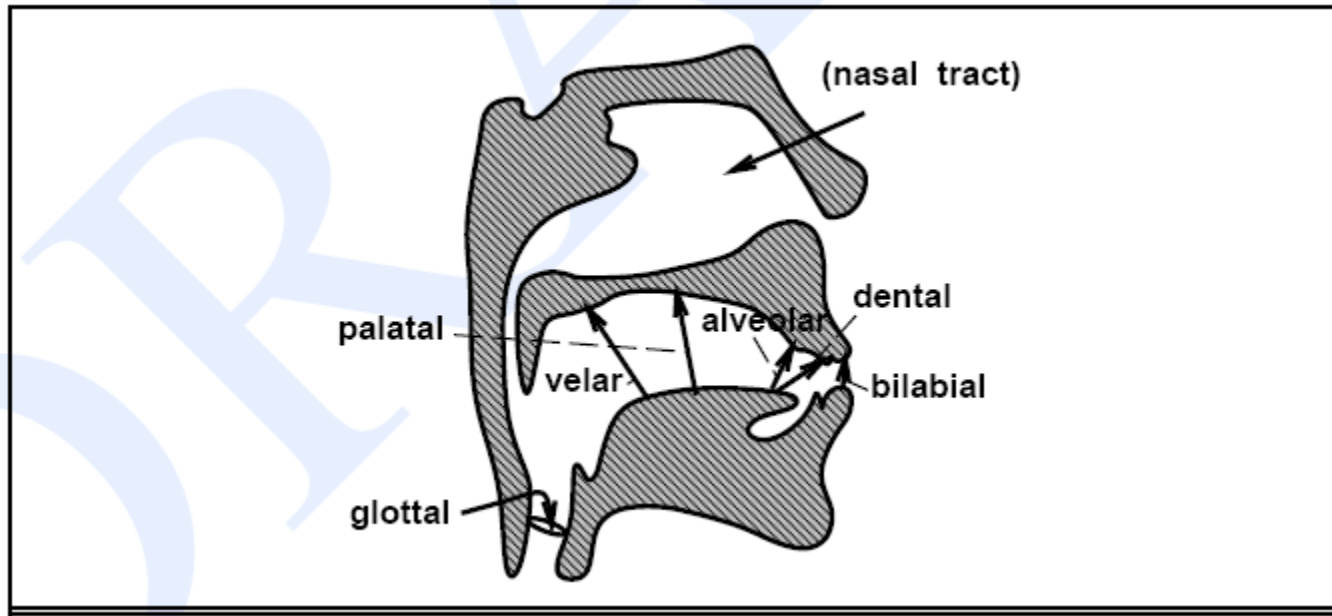epiglottis

pharynx

larynx

trachea

esophagus
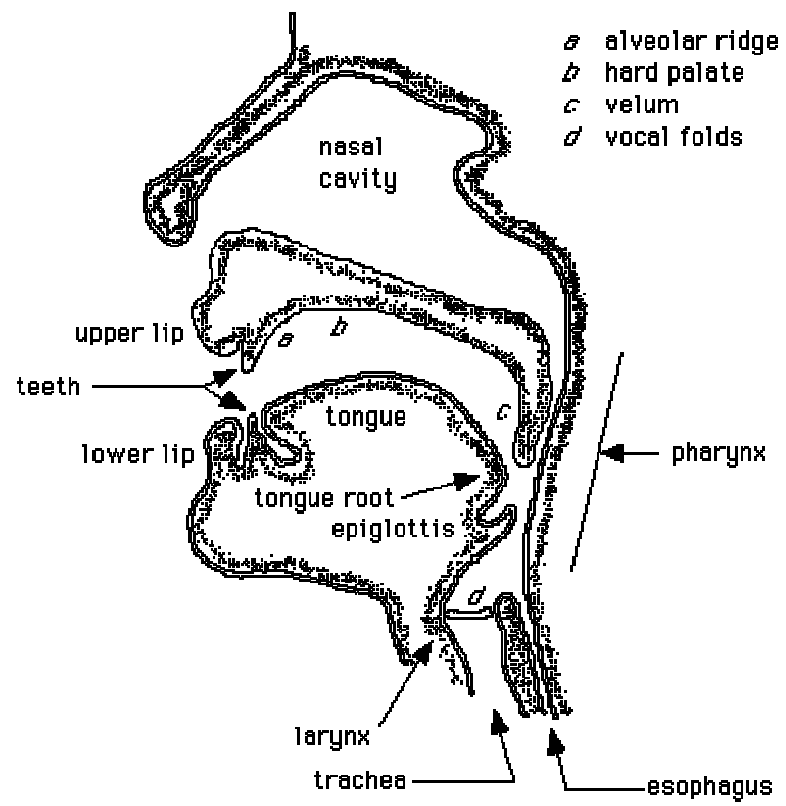
Classification

- Manner of articulation
- Place of articulation
- Voicedness

Ancient 5 x 5 Indian Classification of Consonants

Group						
क वर्ग	क	ख	ग	घ	ङ	**Velar**
च वर्ग	च	छ	ज	झ	ञ	**Palatal**
ट वर्ग	ट	ठ	ड	ढ	ण	**Alveolar**
त वर्ग	त	थ	द	ध	न	**Dental**
प वर्ग	प	फ	ब	भ	म	**Labial**

Stops

- /p/ - voiceless bilabial
- /b/ - voiced bilabial
- /t/ - voiceless alveolar
- /d/ - voiced alveolar
- /k/ - voiceless velar
- /g/ - voiced velar

Fricatives

- /f/
- /v/
- /th/
- /dh/
- /s/
- /sh/
- /zh/
- /h/

Affricates

- /ch/
- /jh/

Nasals

- /m/
- /n/
- /ng/

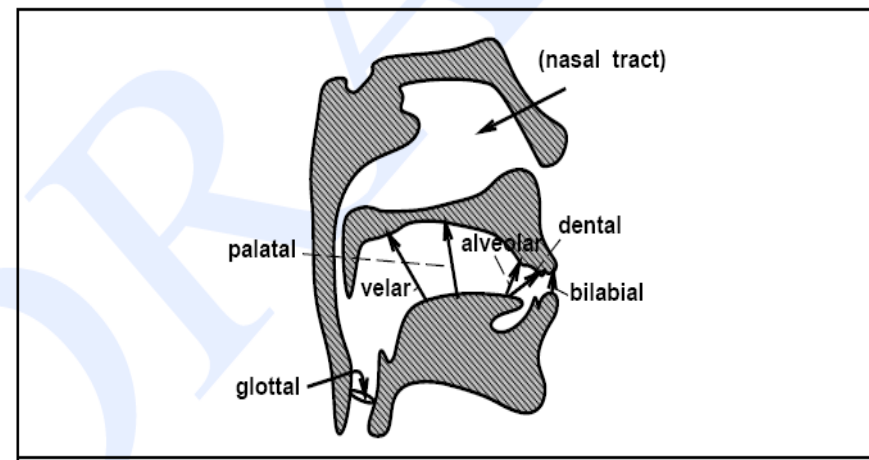| Manner of Articulation		Place of Articulation						
		Bilabial	Labio-dental	Dental	Alveolar	Palatal	Velar	Glottal
Stops	Voiceless	p (pat)			t (tack)		k (cat)	
	Voiced	b (bat)			d (dig)		g (get)	
Fricatives	Voiceless		f (fat)	θ (thin)	s (sat)	š (fish)		h (hat)
	Voiced		v (vat)	ð (then)	z (zap)	ž (azure)		
Affricates	Voiceless					č (church)		
	Voiced					ǰ (judge)		
Nasals		m (mat)			n (nat)		ŋ (sing)	
Liquids					l (late)	r (rate)		
Glides		w (win)				y (yet)		

The plural sound

- Cats, racks … /s/
- dogs, rags … /z/
- Bushes, classes … /iz/

Hypotheses?

Place of Articulation

- **Labial:** Two lips coming together
  - [p] as in *possum,* [b] as in *bear*
- **Dental: Tongue against the teeth**
  - [th] of *thing or the [dh] of though*
- **Alveolar: Alveolar ridge is the portion of the roof of the mouth just behind the** upper teeth; tip of the tongue against the alveolar ridge.
  - Phones [s], [z], [t], and [d]
- P**alatal:** Roof of the mouth; blade of the tongue against this rising back of the alveolar ridge
  - sounds [sh] (*shrimp*), [ch] (*china*), [zh] (*Asian*), and [jh] (*jar*)
- **Velar:** Movable muscular flap at the back of the roof of the mouth; back of the tongue up against the velum
  - sounds [k] (*cuckoo*), [g] (*goose*), and [N] (*kingfisher*)
- G**lottal: closing the glottis (by bringing the** vocal folds together)
 - glottal stop [q] (IPA [P]) is made by closing the glotis (Urdu: gam: sadness)

Manner of Articulation: Stops and Nasals

- All consonants are produced by restriction of airflow
- **Manner of Articulation;** how the restriction is produced**:**
  - complete or partial stoppage
- A **<u>stop</u> is a consonant in which airflow is completely blocked for a short time**
- **English has** voiced stops like [b], [d], and [g] as well as unvoiced stops like [p], [t], and [k].
- Stops are also called **plosives**
- **Nasal sounds [n], [m], and [ng] are made by lowering the velum and allowing** air to pass into the nasal cavity
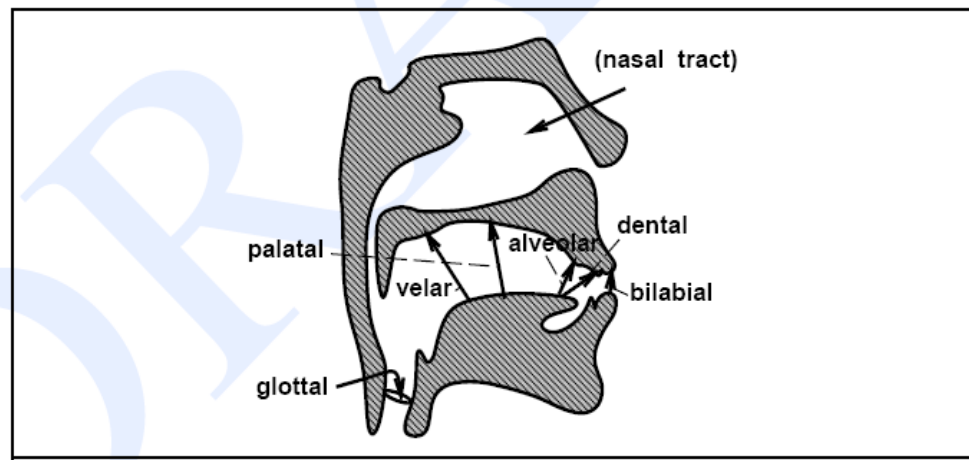
Fricatives

- **Fricatives**, airflow is constricted but not cut off completely. The turbulent airflow that results from the constriction produces a characteristic "hissing" sound.
- The English **labiodental fricatives** [f] and [v] are produced by pressing the lower lip against the upper teeth, allowing a restricted airflow between the upper teeth.
- The **dental fricatives** [th] and [dh] allow air to flow around the tongue between the teeth.
- The **alveolar fricatives** [s] and [z] are produced with the tongue against the alveolar ridge, forcing air over the edge of the teeth.
- In the **palato-alveolar fricatives** [sh] and [zh] the tongue is at the back of the alveolar ridge forcing air through a groove formed in the tongue.
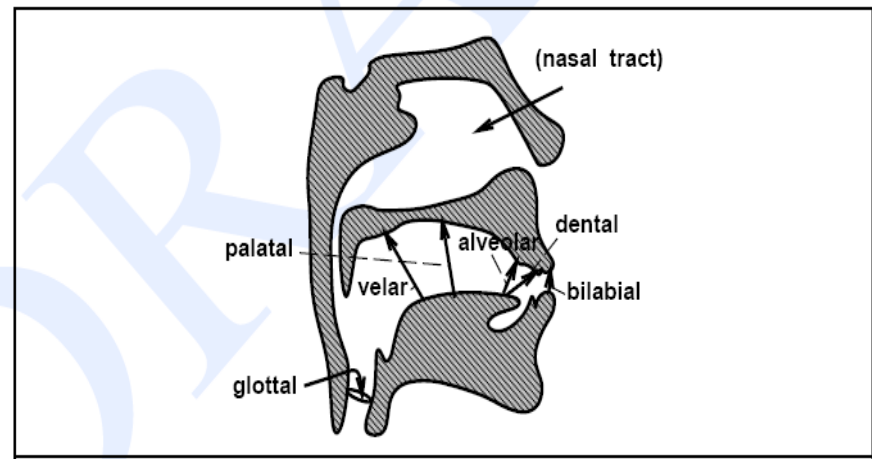
Affricates, Laterals/Liquids and Taps/Flaps

- **Affricates** are stops followed immediately by fricatives
  - **English** [ch] (*chicken); Marathi *chaa (e.g., gharaachaa; of the house)*
- **Lateral** or **Liquids:** tip of the tongue up against the alveolar ridge or the teeth, with one or both sides of the tongue lowered to allow air to flow over it
  - [l] (*learn)*
- **Tap or flap:** quick motion of the tongue against the alveolar ridge
  - [dx] (IPA [R])
  - The consonant in the middle of the word *lotus ([l ow dx ax s]) is a tap in most* dialects of American English
  - speakers of many UK dialects would use a [t] instead of a tap in this word.
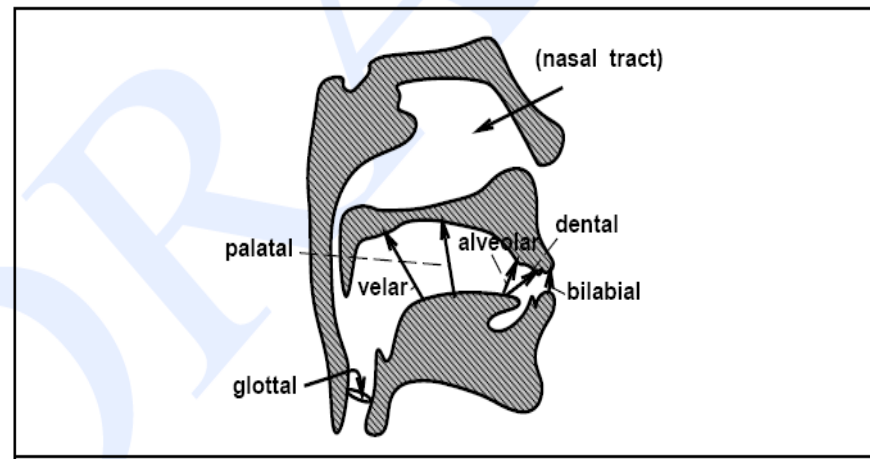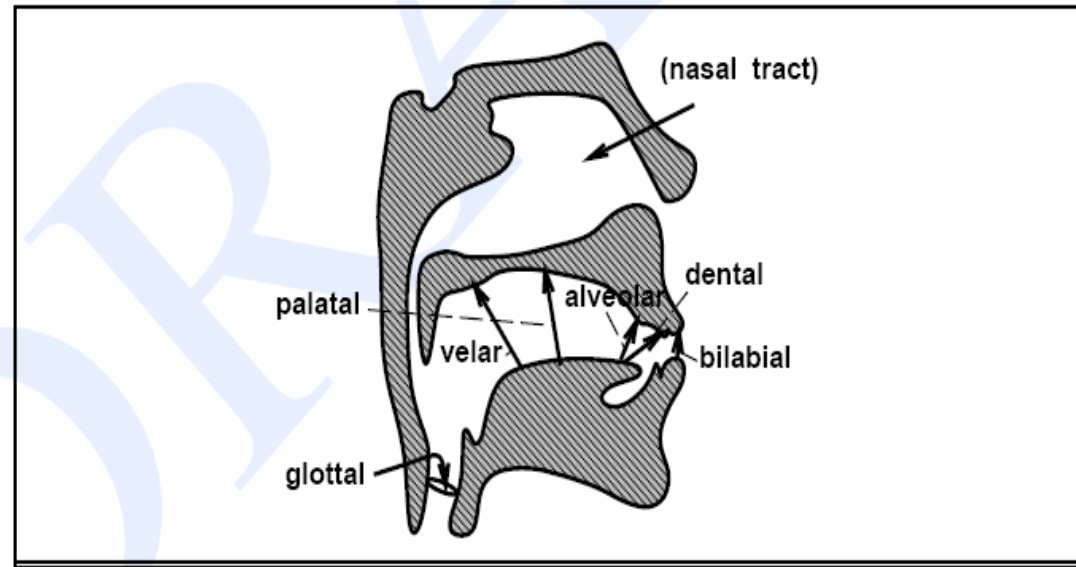
Articulation of consonants: Larynx action/glottis state (1/2)

- Vocal cords are pulled apart. The air passes freely through the glottis. This is called the voicelessness state and sounds produced with this configuration of the vocal cords are called **voiceless: p t k f θ s ʃ tʃ**
- Vocal cords are pulled close together. The air passing through the glottis causes the vocal cords to vibrate. This is called the voicing state and sounds produced with this configuration of the vocal cords are called **voiced: b d g v ð z ʒ dʒ**
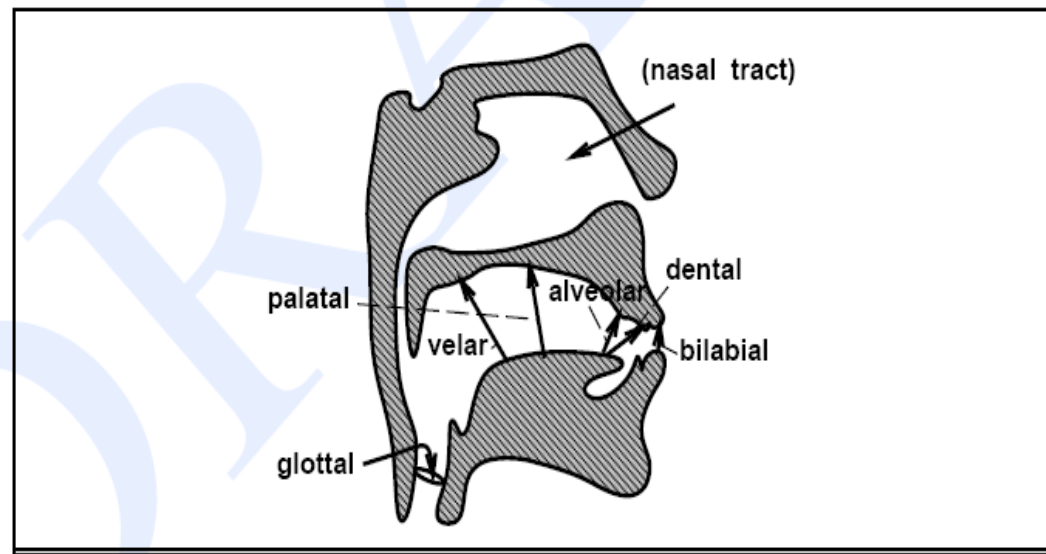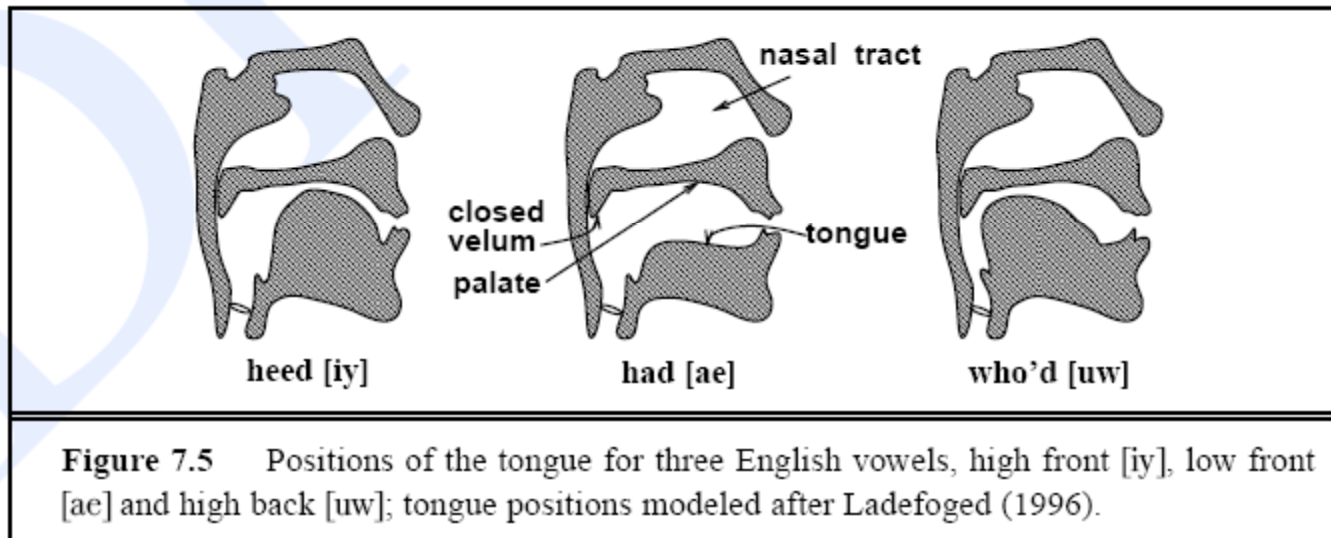
Articulation of consonants: Larynx action/glottis state (2/2)

- Vocal cords are apart at the back and pulled together at the front. This is called the **whisper** state.
- Vocal cords assume the voicing state but are relaxed. This is called the **murmur** state.

Vowels (1/2)

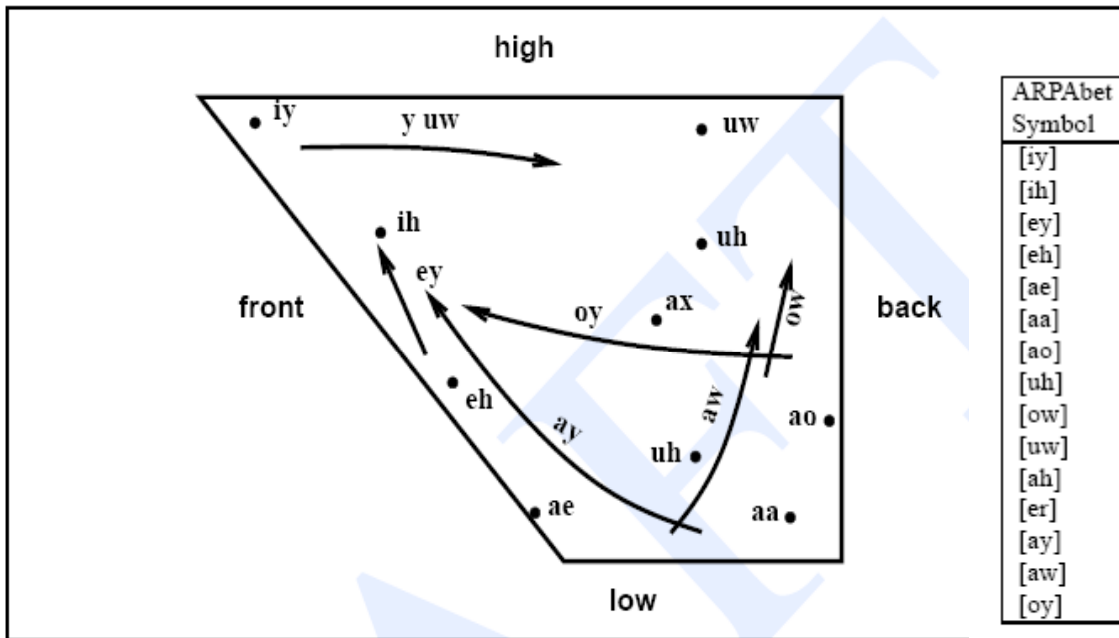**Figure 7.5** Positions of the tongue for three English vowels, high front [iy], low front [ae] and high back [uw]; tongue positions modeled after Ladefoged (1996).

Vowels (2/2)

Figure 7.6 Qualities of English vowels (after Ladefoged (1993)).

ARPAbet Symbol	IPA Symbol	Word	ARPAbet Transcription
[iy]	[i]	lily	[l ih l iy]
[ih]	[ɪ]	lily	[l ih l iy]
[ey]	[eɪ]	daisy	[d ey z iy]
[eh]	[ɛ]	pen	[p eh n]
[ae]	[æ]	aster	[ae s t axr]
[aa]	[ɑ]	poppy	[p aa p iy]
[ao]	[ɔ]	orchid	[ao r k ix d]
[uh]	[ʊ]	wood	[w uh d]
[ow]	[oʊ]	lotus	[l ow dx ax s]
[uw]	[u]	tulip	[t uw l ix p]
[ah]	[ʌ]	buttercup	[b ah dx axr k ah p]
[er]	[ɝ]	bird	[b er d]
[ay]	[aɪ]	iris	[ay r ix s]
[aw]	[aʊ]	sunflower	[s ah n f l aw axr]
[oy]	[ɔɪ]	soil	[s oy l]
Reduced and uncommon phones			
[ax]	[ə]	lotus	[l ow dx ax s]
[axr]	[ɚ]	heather	[h eh dh axr]
[ix]	[ɨ]	tulip	[t uw l ix p]
[ux]	[ʉ]	dude[1]	[d ux d]

Phonology: Syllables
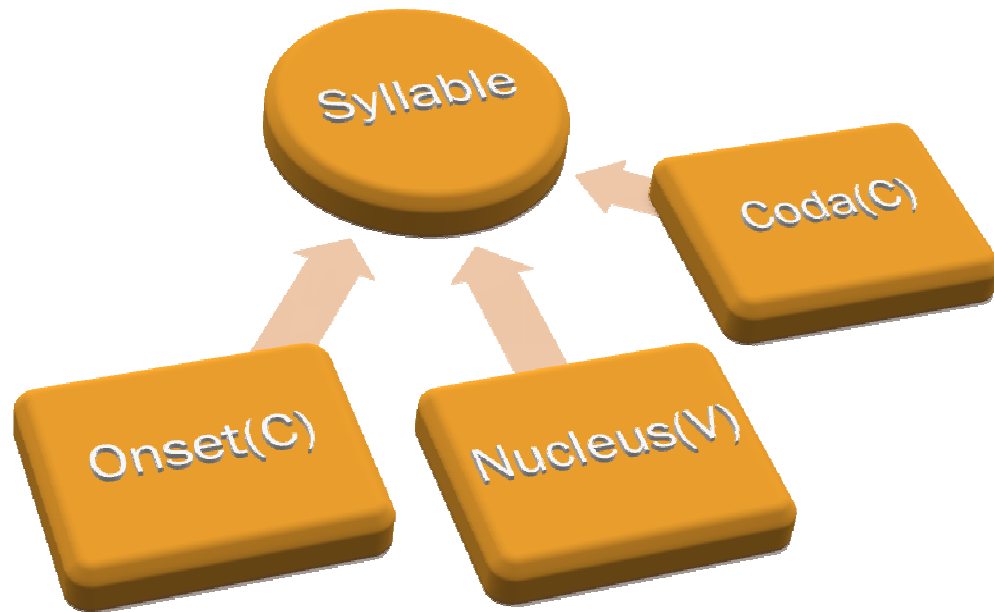
Basic of syllables

"Syllable is a unit of spoken language consisting of a single uninterrupted sound formed generally by a Vowel and preceded or followed by one or more consonants."

- Vowels are the heart of a syllable (Most Sonorous Element) (*svayam raajate iti svaraH)*
- Consonants act as sounds attached to vowels.

Syllable structure

- A syllable consists of 3 major parts:-
  - Onset (C)
  - Nucleus (V)
  - Coda (C)
- Vowels sit in the Nucleus of a syllable
- Consonants may get attached as Onset or Coda.
- Basic structure - CV

Possible syllable structures

- The Nucleus is always present
- Onset and Coda may be absent
- Possible structures
  - V
  - CV
  - VC
  - CVC

syllable theories

- ## Prominence Theory
  - E.g. *entertaining* /entəteɪnɪŋ/
  - The peaks of prominence: vowels /e ə eɪ ɪ/
  - Number of syllables: 4
- ## Chest Pulse Theory
  - Based on muscular activities
- ## Sonority Theory
  - Based on relative soundness of segment within words