Sounds

Syllabus & Intros

Name

Where you're from

Year in school

Have you programmed?

What do you want to get out of this class?

Fun fact

International Phonetic Alphabet

Linguistics

What is IPA?

A set of symbols designed by the International Phonetic Association, used for the transcription of speech sounds in language-related fields.

Why do we need IPA?

It is the closest thing to a universal phonetic writing system.

It is orthographically transparent (one symbol for each sound)

/ˈsɪm.bəlz/ [ˈsɪ̃m.bl̩z]

Combinations of letters, diacritics, brackets

Broad transcription - conventional sounds of the language (or phonemes)

Narrow transcription - actual sounds (or phones)

Four Types of Symbols

1. Same as English
2. Same as English, but more specific
3. Looks like English, but different
4. Not in Modern English

Same as English

14 letters:

/b, d, f, h, k, l, m, n, p, s, t, v, w, z/

Same, but more specific

/i/ is the *i* in *machine*, not in *pit* or *idol*

/ɪ/ is the *i* in *pit*

/u/ is the *u* in *hula*, not in *union* or *up*

/g/ is the *g* in *gift*, not in *gin*

Looks like English, but not

/j/ (“yod”) is the *y* in *yes*

Symbols not in English (I)

/æ/ (“ash”) is the vowel in *hat*

/ɑ/ (“script a”) is the first vowel in *father*

/ɛ/ (“epsilon”) is the vowel in *get*

/ɔ/ (“open o”) is the vowel in *law*

Symbols not in English (II)

/ʊ/ (“upsilon”) is the vowel in *book*

/ʌ/ (“caret”) is the vowel in *up*

/ə/ (“schwa”) is the first vowel in *above*

Don’t worry abou the distinction between /ʌ/ and /ə/ for now

Symbols not in English (III)

/ŋ/ (“engma”) is the last sound in *song*

/θ/ (“theta”) is the first sound in *thin*

/ð/ (“eth”) is the first sound in *then*

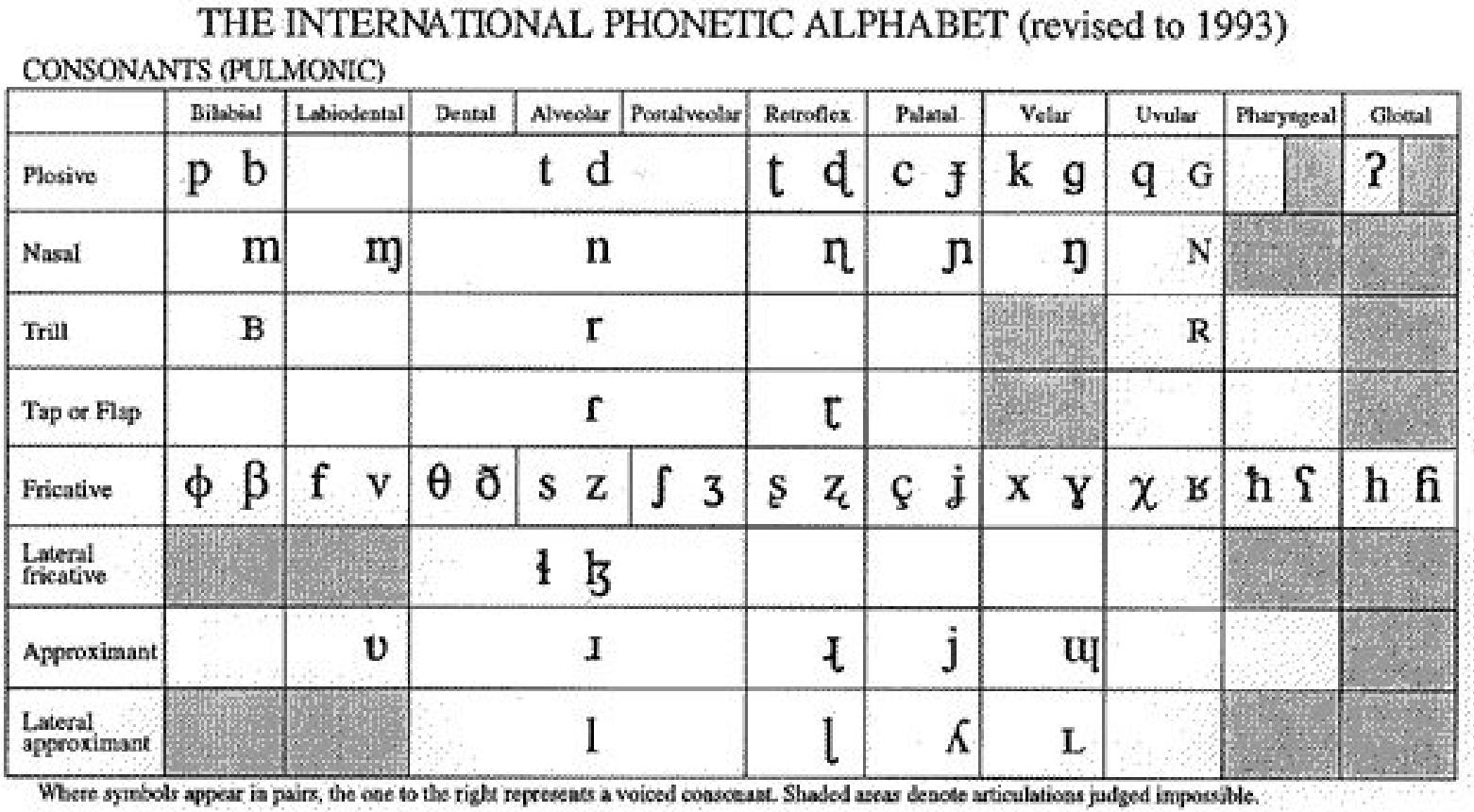
/ʃ/ (“esh”) is the first sound in *she*

Symbols not in English (IV)

/ʒ/ (“ezh”) is the second consonant in *vision*

/ɹ/ is the American *r* sound

/ʔ/ (“glottal stop”) is the sound in the middle of *uh-oh*



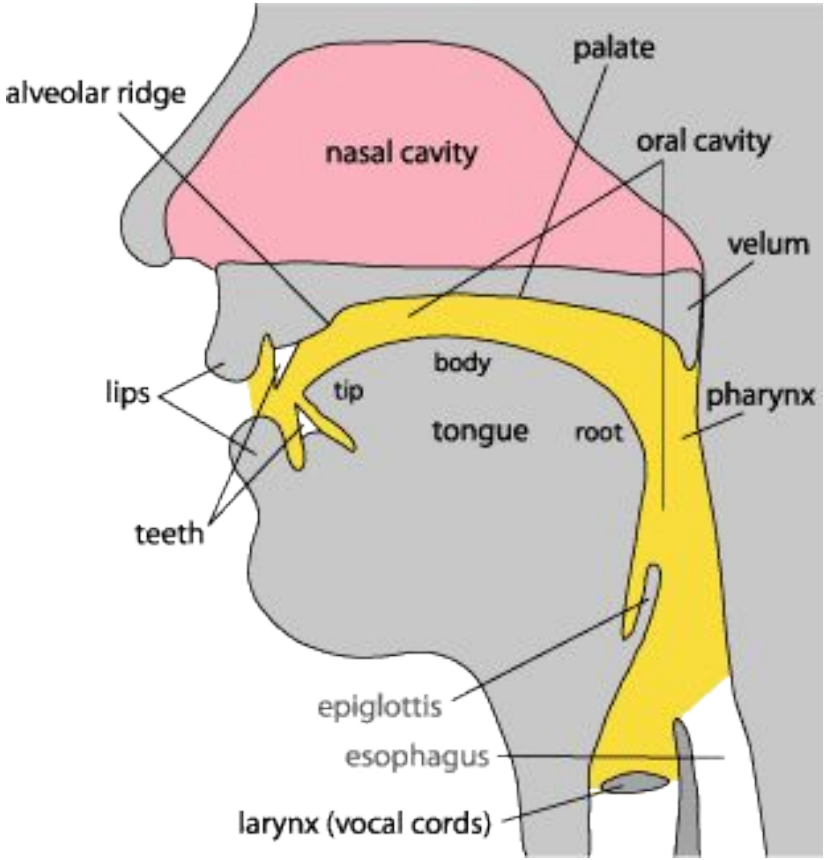
THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

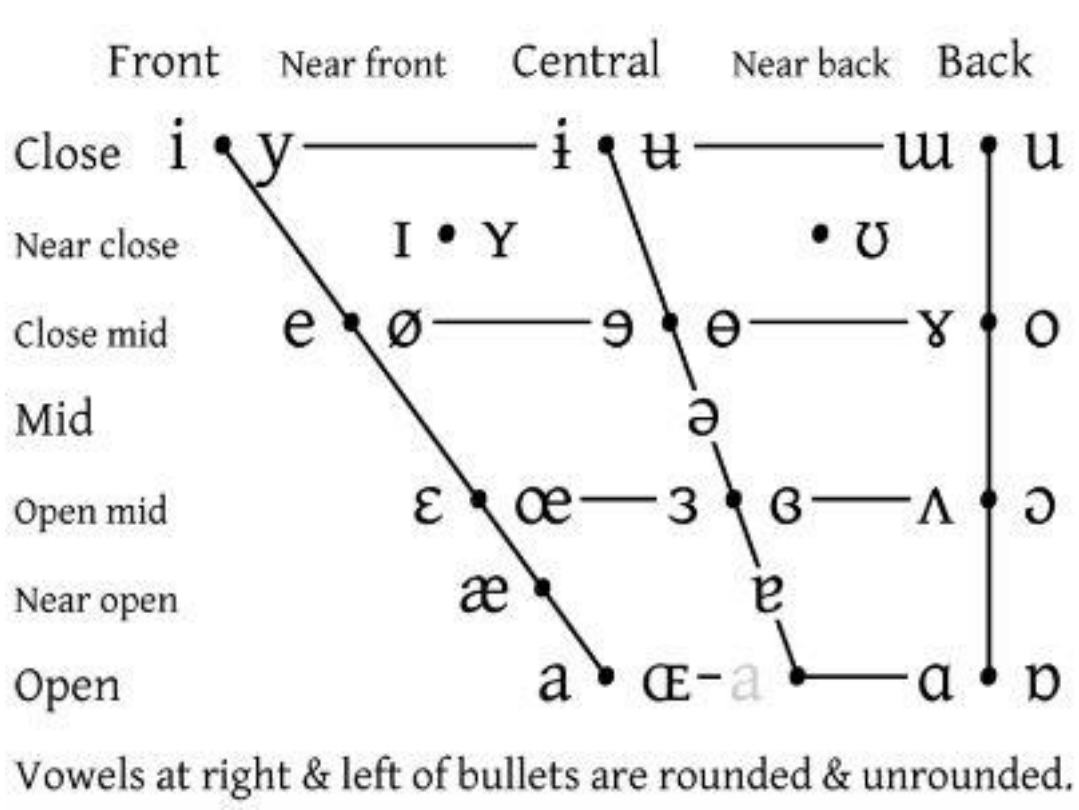
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant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

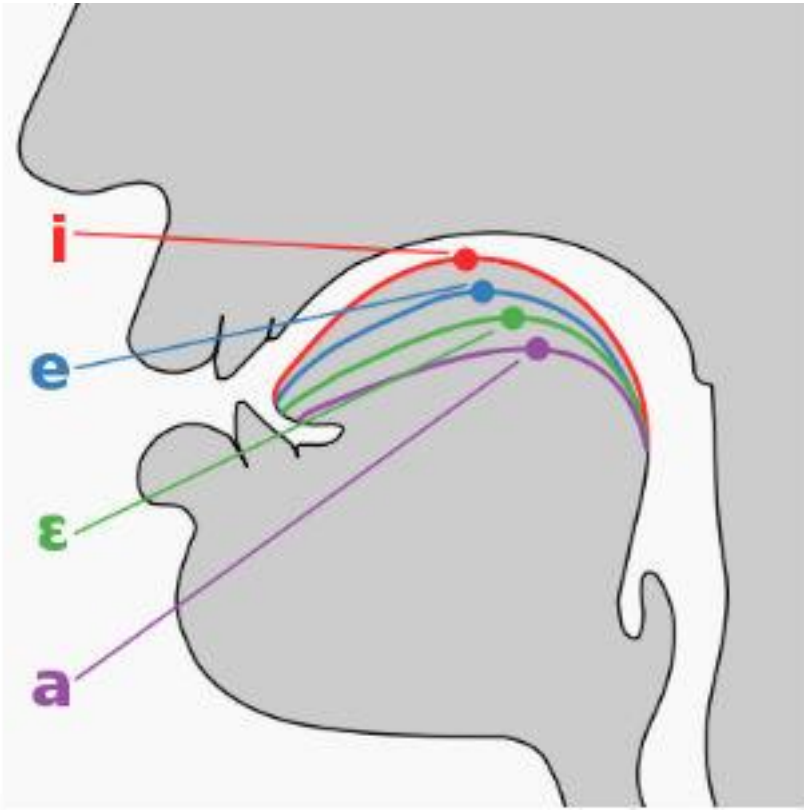
Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Consonants



The Speech Mechanism


Front
Near front
Central
Near back
Back
Close
Near close
Close mid
Mid
Open mid
Near open
Open
i y ɨ ʉ ɯ u
ɪ ʏ ʊ
e ø ɘ ɵ ɤ o
ə
ɛ œ ɜ ɞ ʌ ɔ
æ ɐ
a ɶ ɑ ɒ
Vowels at right & left of bullets are rounded & unrounded.


i
e
ɛ
a

Vowels

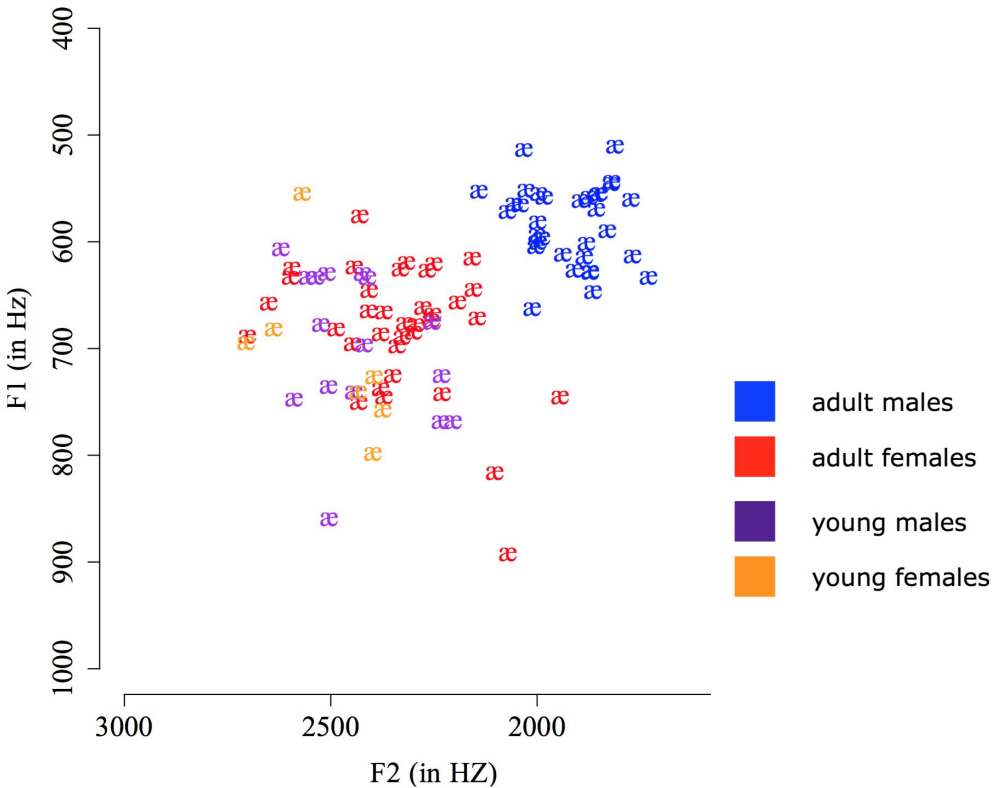
Transcription Practice

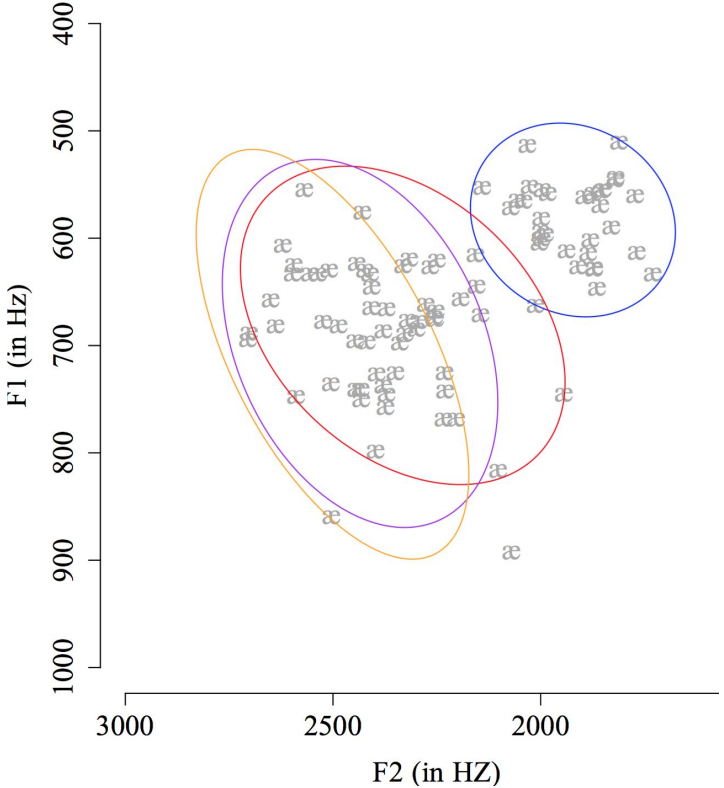
cat
sing
build
thin
bees
think
this
tacks
strengths
enough

Transcription Practice

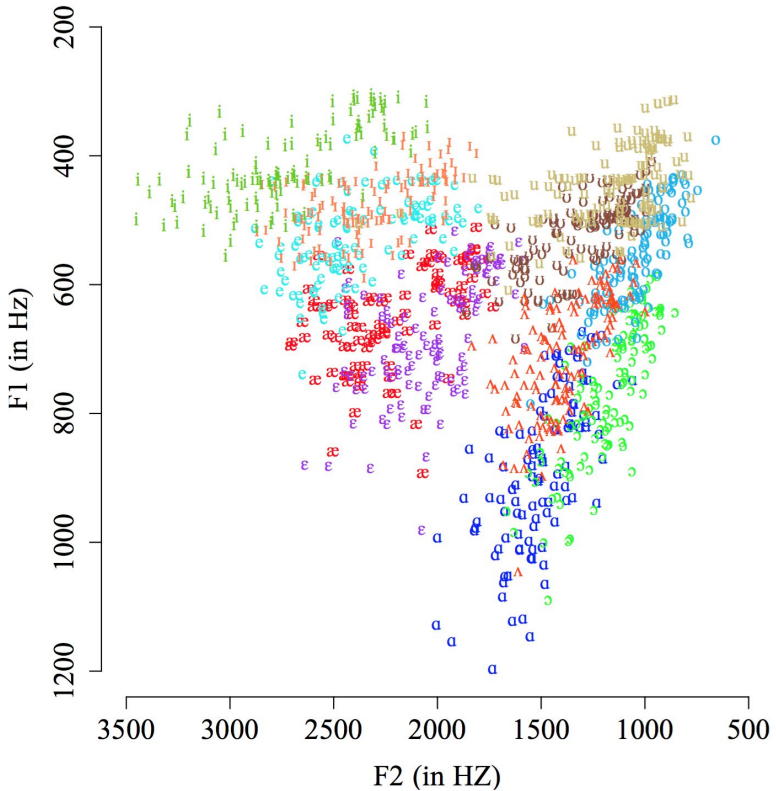
cat	/kæt/
sing	/sɪŋ/
build	/bɪld/
thin	/θɪn/
bees	/biz/
think	/θɪŋk/
this	/ðɪs/
tacks	/tæks/
strengths	/stɹɛŋθs/
enough	/inʌf/

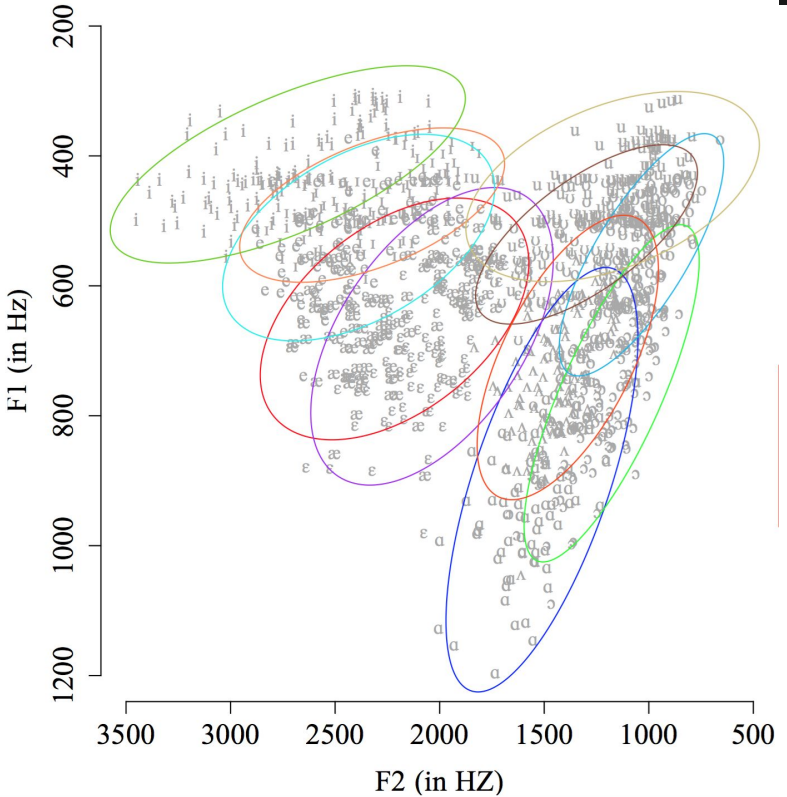
Lack of Invariance





Lack of Invariance





Lack of Invariance

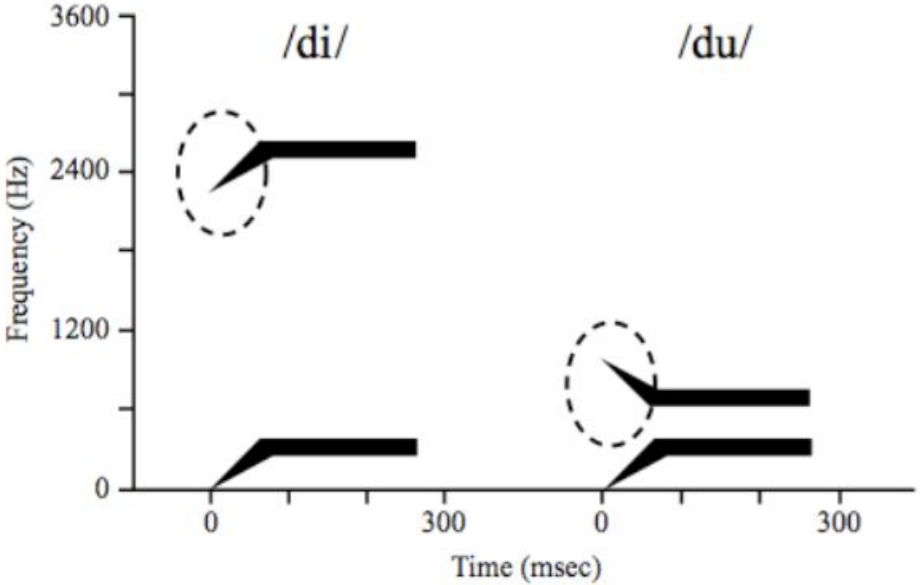


Figure 1. Spectrographic patterns for the two two-formant synthetic syllables /di/ and /du/. Note the difference in formant transitions, marked by the dotted ovals.

Lack of Invariance

There is no 1:1 correlation between the physical properties of speech sounds and the perception of speech sounds

Adaptation

Given that the speech signal is noisy and there is a lack of invariance, we have to adapt to the signal.

$$P(\text{Message}|\ \text{Signal}) \propto P(\text{Signal}|\ \text{Message})\ P(\text{Message})$$

New beliefs — Likelihood — Prior Beliefs

PROLOG BREAK

Speech Recognition

- small vocab/many users vs. large vocab/few users
- discrete vs. continuous speech

Speech Recognition

How it works: Set-up (enrollment)

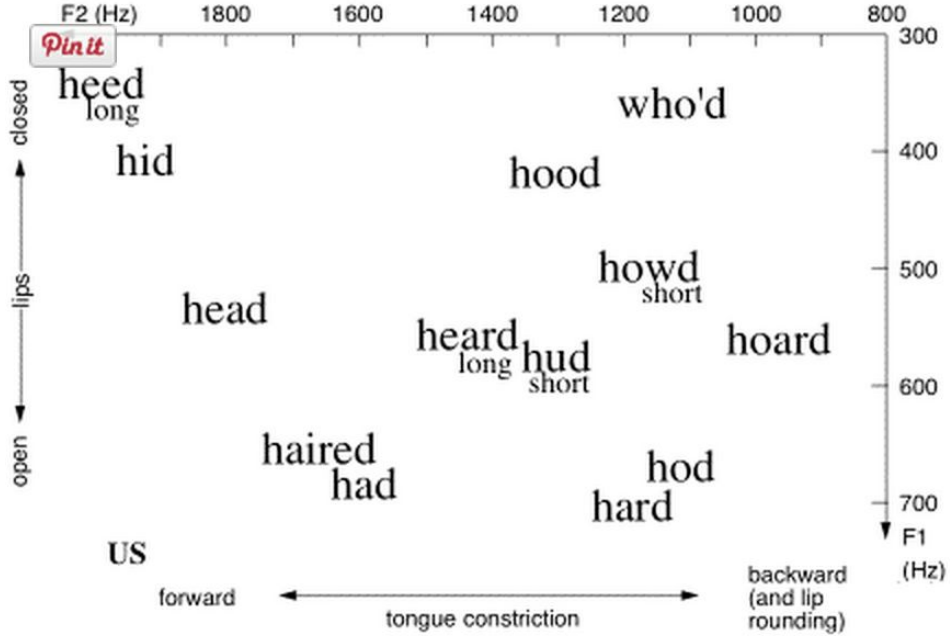
- set of words/phrases new users say before using the system
- how are these selected?

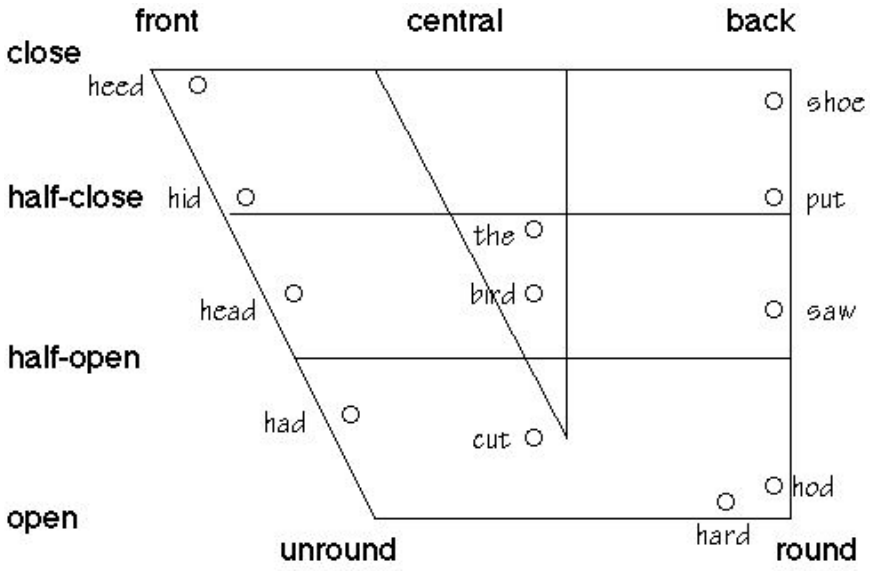
Speech Recognition

What information can a program use?

- Waveform of speech
- Decode sounds using formant information

Speech Recognition





Ghonim et al., 2007, 2013

https://www.llas.ac.uk/materialsbank/mb081/page_10.htm

Speech Recognition

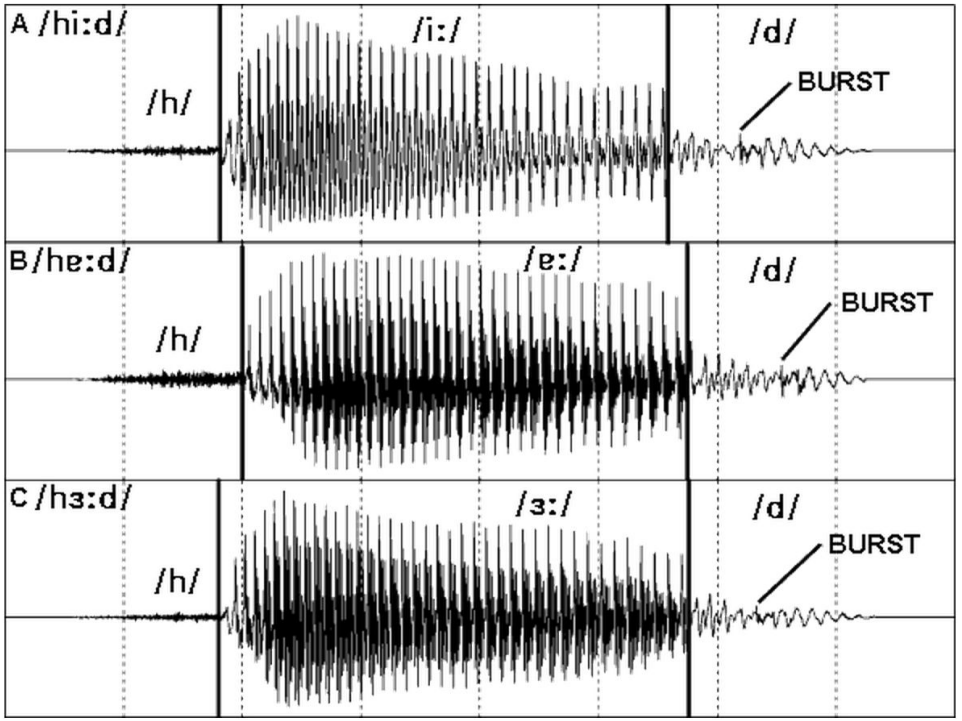
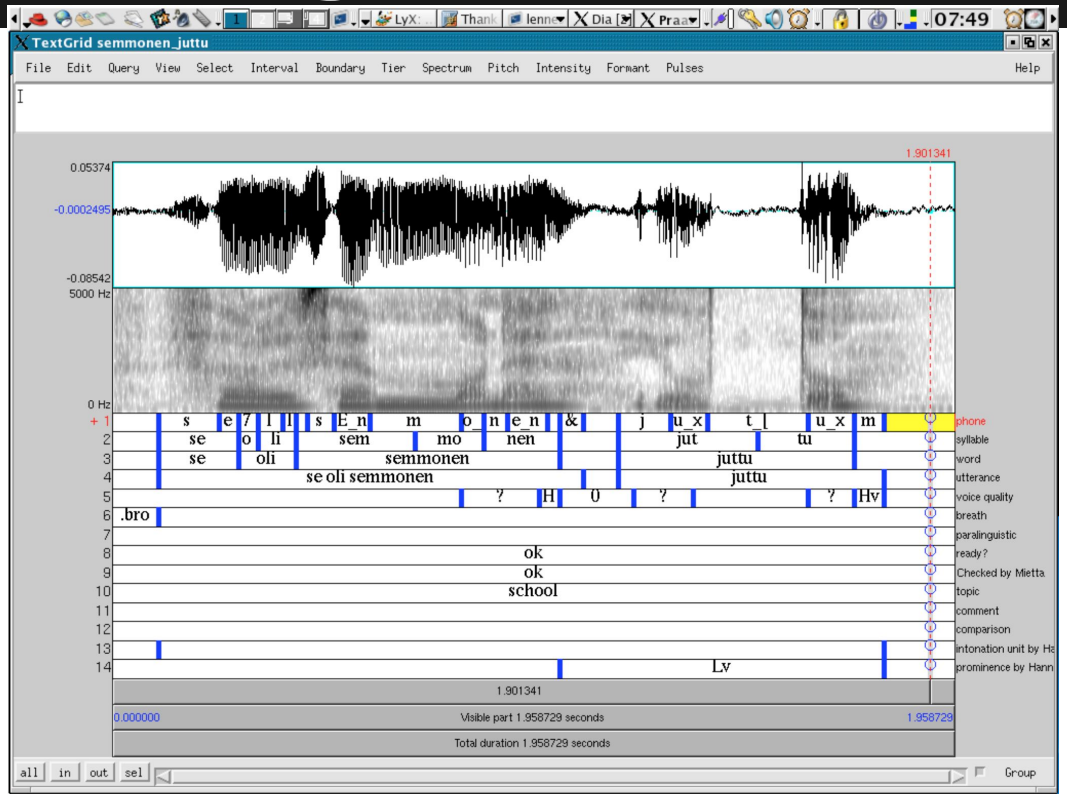


Figure 5: Three long vowels in an /h_d/ context.

http://clas.mq.edu.au/speech/acoustics/waveforms/speech_waveforms.html

Speech Recognition



Lennes, M., 2005

Try Out Praat on Your Own

- Partner up
- Say vowels into Praat
- Compare yours to your partner’s
  - do they seem similar?
  - how are they different?
  - what differences do you see between different vowels for the same person?

Clustering

Machine Learning

An area at the intersection of mathematics, computer science, engineering and data science that focuses predominantly on one question:

Given some data, how I can I learn a model to predict X.

k-Means Classifier

Initialize:

1. Pick a number *k* and say that there are that many classes of thing.
2. Randomly assign all the data points to a class.

k-Means Classifier

Search:

1. Calculate the mean of each class.
2. Calculate the probability of each data point is from that class.
3. Reassign the data points according to which class is most likely for that data point.
4. Rinse and repeat until convergence.

Quick and Dirty k-Means

https://mollicaf.shinyapps.io/k-Means/