Roadmap

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

Roadmap

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

- Week 1
  - Module 1: introduction, brief look at speech production and perception
- Week 2
  - Foundations: signals
  - Module 2: speech signals and the source-filter model
- Week 3
  - Foundations: phonetics

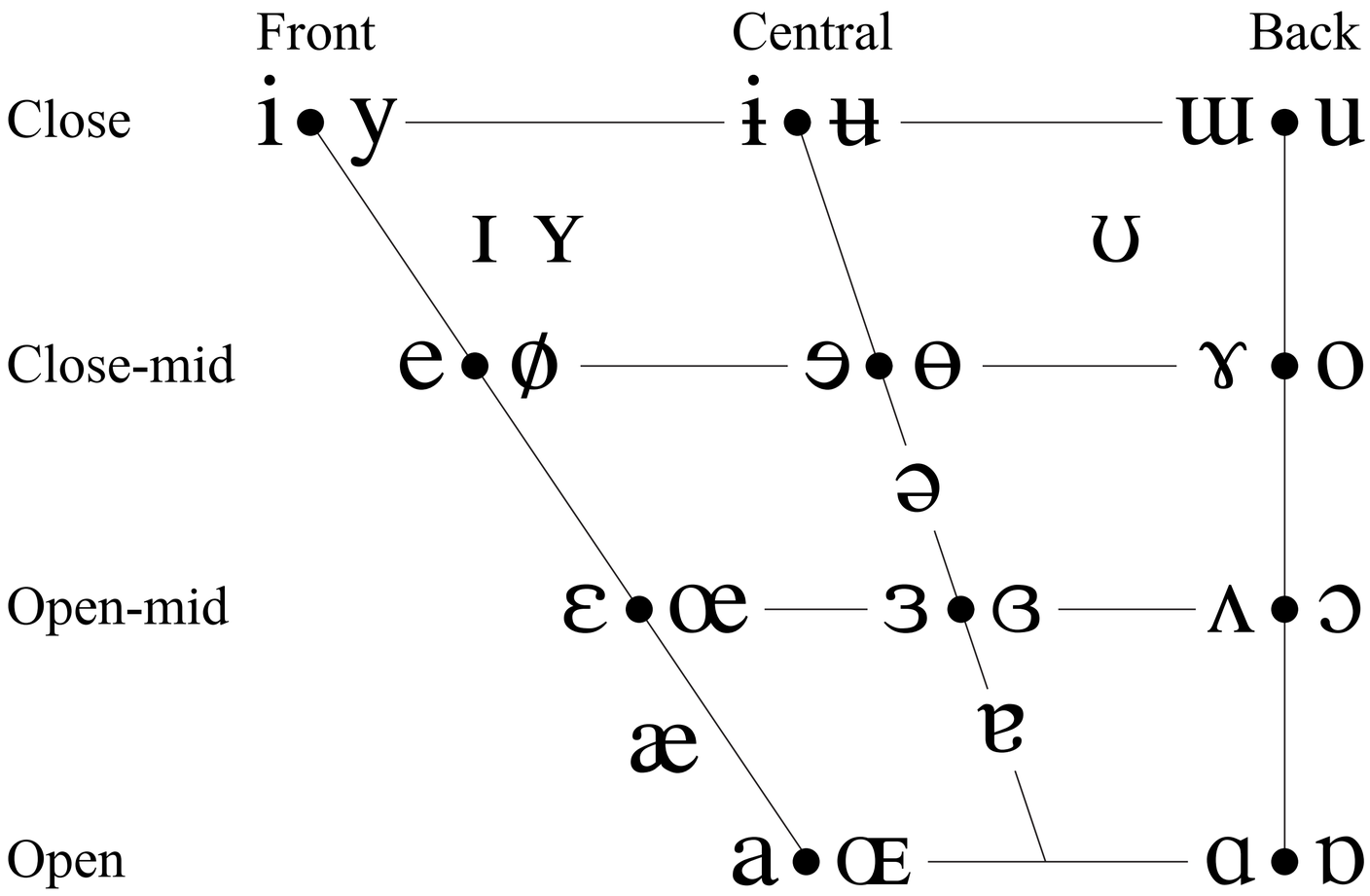
See slides from: Foundations for Speech Processing > Phonetics

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)

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		ⱱ		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			

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.


end of first half of lecture 03

Roadmap

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

Roadmap

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

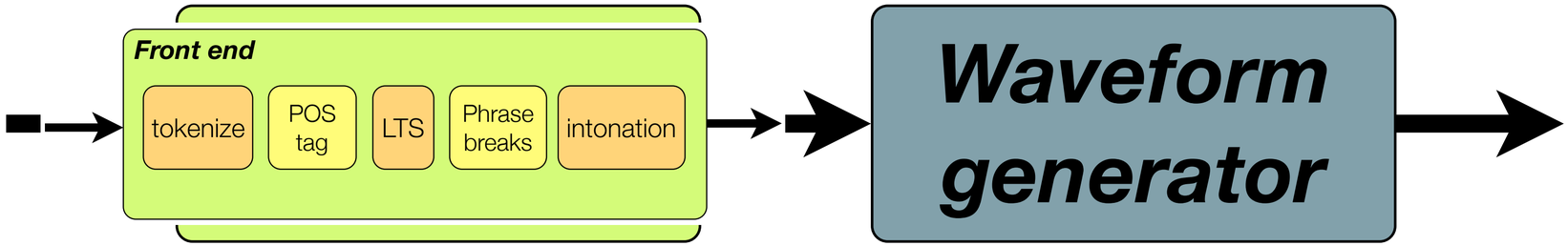
- Week 3
  - Module 3: text processing
- Week 4
  - Class trip
  - Module 4: pronunciation & prosody
- Week 5
  - Assignment Q&A
  - Module 5: waveform generation
- Week 6
  - Submission of first assignment

What you should already know

- From the videos & readings
  - text-to-speech pipeline
  - dealing with Non-Standard Words (NSWs)

What you should already know

- From the videos & readings
  - text-to-speech pipeline
  - dealing with Non-Standard Words (NSWs)



Speech synthesis - text processing

- Representing linguistic information using data structures
- Designing features for classifying Non-Standard Words (NSWs) into categories
- Writing algorithms to expand NSWs

How to represent linguistic information?

- The Heterogeneous Relation Graph (HRG) formalism (as used in Festival)
- Basic data structure to represent a linguistic item: **feature structure**
  - an unordered list of key-value pairs (*like a Python dictionary*)

$$word: \begin{bmatrix} \text{NAME} & abuse_1 \\ \text{POS} & noun \\ \text{TEXT} & abuse \\ \text{PRON} & /@buws/ \end{bmatrix}$$

Example taken from Taylor - Section 4.5

Nesting: values can themselves be feature structures

$$
\textit{phone}: \begin{bmatrix}
\text{NAME} & p \\
\text{STRESS} & 1 \\
\text{DISTINCTIVE FEATURES} & \begin{bmatrix}
\text{VOICED} & \textit{false} \\
\text{MANNER} & \textit{stop} \\
\text{PLACE} & \textit{bilabial}
\end{bmatrix}
\end{bmatrix}
$$

Example taken from Taylor - Section 4.5

How to represent linguistic information?

Data structures

- The Heterogeneous Relation Graph (HRG) formalism (as used in Festival)
- Basic data structure to represent a linguistic item: **feature structure**
  - an unordered list of key-value pairs (like a Python dictionary)
- **Relations** between linguistic items

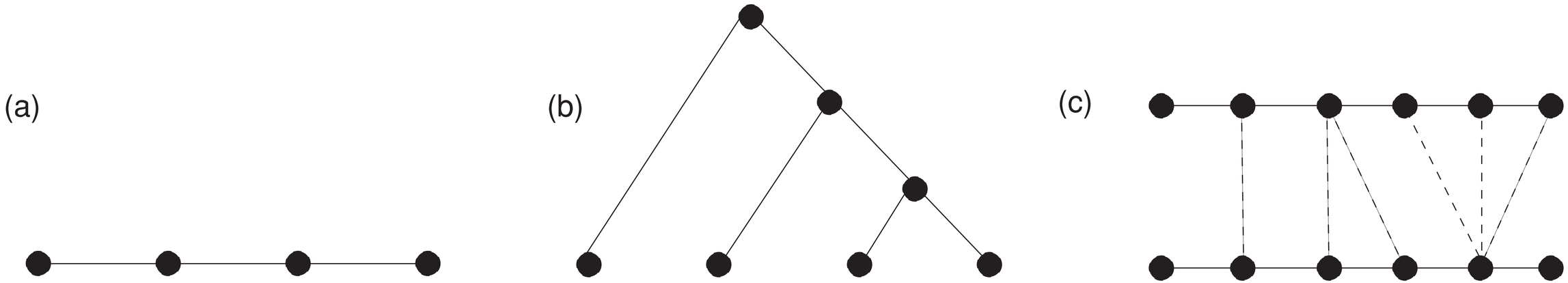
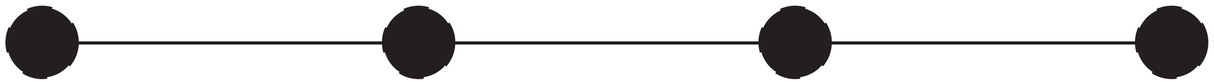


Figure 4.1 The three types of relation: (a) list relation, (b) tree relation and (c) ladder relation.

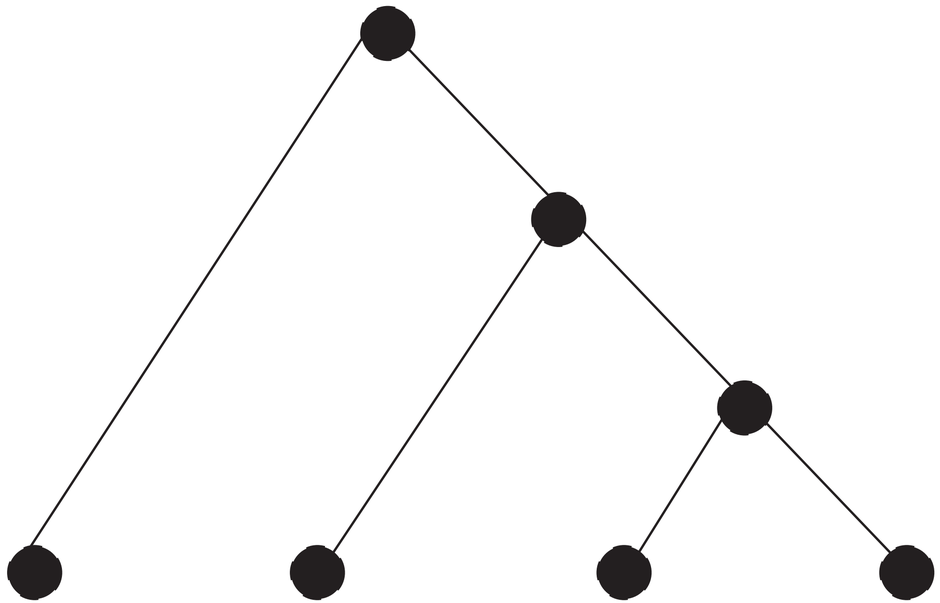
Taylor - Figure 4.1

List - for example, relation between the words in a sentence



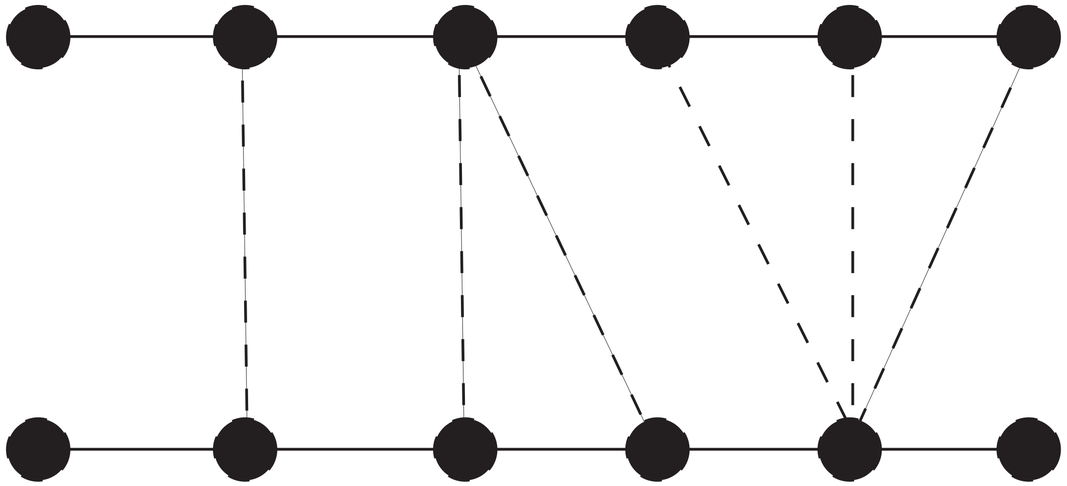
Taylor - Figure 4.1

Tree - for example, relation between words, syllables and phones



Taylor - Figure 4.1

Ladder - for example, relation between syllables and pitch accents



Taylor - Figure 4.1

Speech synthesis - text processing

- Representing linguistic information using data structures
- Designing features for classifying Non-Standard Words (NSWs) into categories
- Writing algorithms to expand NSWs

Design some features that might be useful for classifying NSWs

TABLE I. Taxonomy of non-standard words used in hand-tagging and in the text normalization models

	EXPN	abbreviation	*adv, N.Y, mph, gov't*
alpha	LSEQ	letter sequence	*CIA, D.C, CDs*
	ASWD	read as word	*CAT*, proper names
	MSPL	misspelling	*geogaphy*
	NUM	number (cardinal)	*12, 45, 1/2, 0·6*
	NORD	number (ordinal)	*May 7, 3rd, Bill Gates III*
	NTEL	telephone (or part of)	*212 555-4523*
	NDIG	number as digits	*Room 101*
N	NIDE	identifier	*747, 386, I5, pc110, 3A*
U	NADDR	number as street address	*5000 Pennsylvania, 4523 Forbes*
M	NZIP	zip code or PO Box	*91020*
B	NTIME	a (compound) time	*3·20, 11:45*
E	NDATE	a (compound) date	*2/2/99, 14/03/87* (or US) *03/14/87*
R	NYER	year(s)	*1998, 80s, 1900s, 2003*
S	MONEY	money (US or other)	*$3·45, HK$300, Y20,000, $200K*
	BMONEY	money tr/m/billions	*$3·45 billion*
	PRCT	percentage	*75%, 3·4%*
	SPLT	mixed or "split"	*WS99, x220, 2-car* (see also SLNT and PUNC examples)
	SLNT	not spoken,	word boundary or emphasis character:
M		word boundary	*M.bath, KENT*RLTY, _really_*
I	PUNC	not spoken,	non-standard punctuation: "***" in
S		phrase boundary	*$99,9K***Whites*, "…" in *DECIDE…Year*
C	FNSP	funny spelling	*sllooooooww, sh*t*
	URL	url, pathname or email	*http://apj.co.uk, /usr/local, phj@tpt.com*
	NONE	should be ignored	ascii art, formatting junk

Table I from Sproat et al,
"Normalization of non-standard words"
Computer Speech and Language
(2001) 15, 287–333
doi:10.1006/csla.2001.0169

Speech synthesis - text processing

- Representing linguistic information using data structures
- Designing features for classifying Non-Standard Words (NSWs) into categories
- Writing algorithms to expand NSWs

Write an algorithm to expand **`LSEQ`** (letter sequence) to words

- Your algorithm must handle these examples
  - IBM
  - DVD
  - UN
  - ABC

Write an algorithm to expand **`NUM`** (cardinal number) to words

- Your algorithm must handle these examples
  - 7
  - 21
  - -9
  - 3.1
  - 99.9

Write an algorithm to expand **`PRCT`** (percentage) to words

- Your algorithm must handle these examples
  - 50%
  - -30%
  - 4.5%