



# **Lecture 5: Understanding human speech**

**Zhizheng Wu**

# Outline

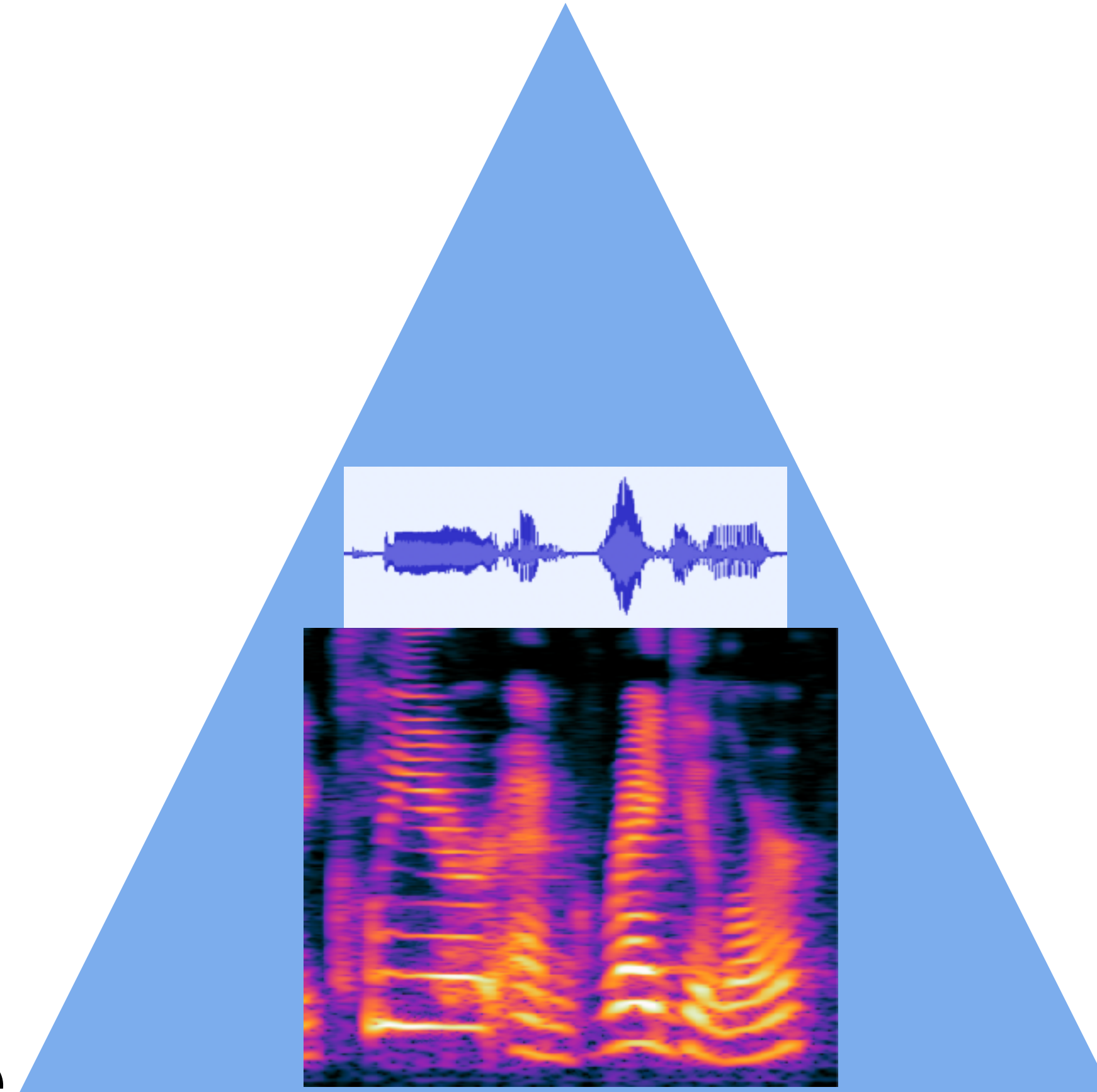
- ▶ Recap of speech production
- ▶ Phone, International Phonetic Alphabet, and Grapheme-to-phoneme conversion
- ▶ Articulatory phonetics
- ▶ Speech disorders

# Recap

- ▶ Speech representation in time and frequency domain
- ▶ Speech production and the source-filter model



Content

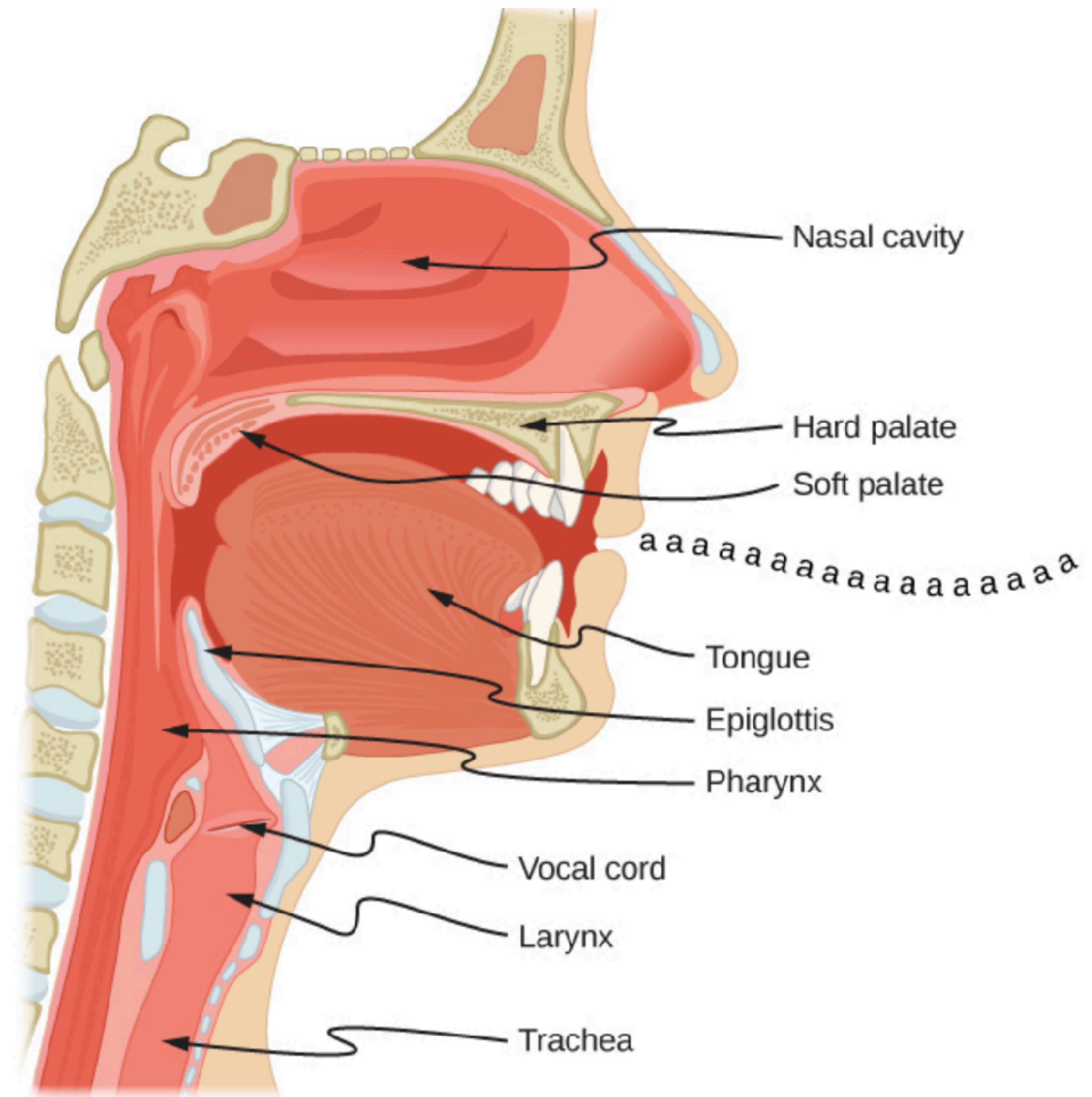


Timbre

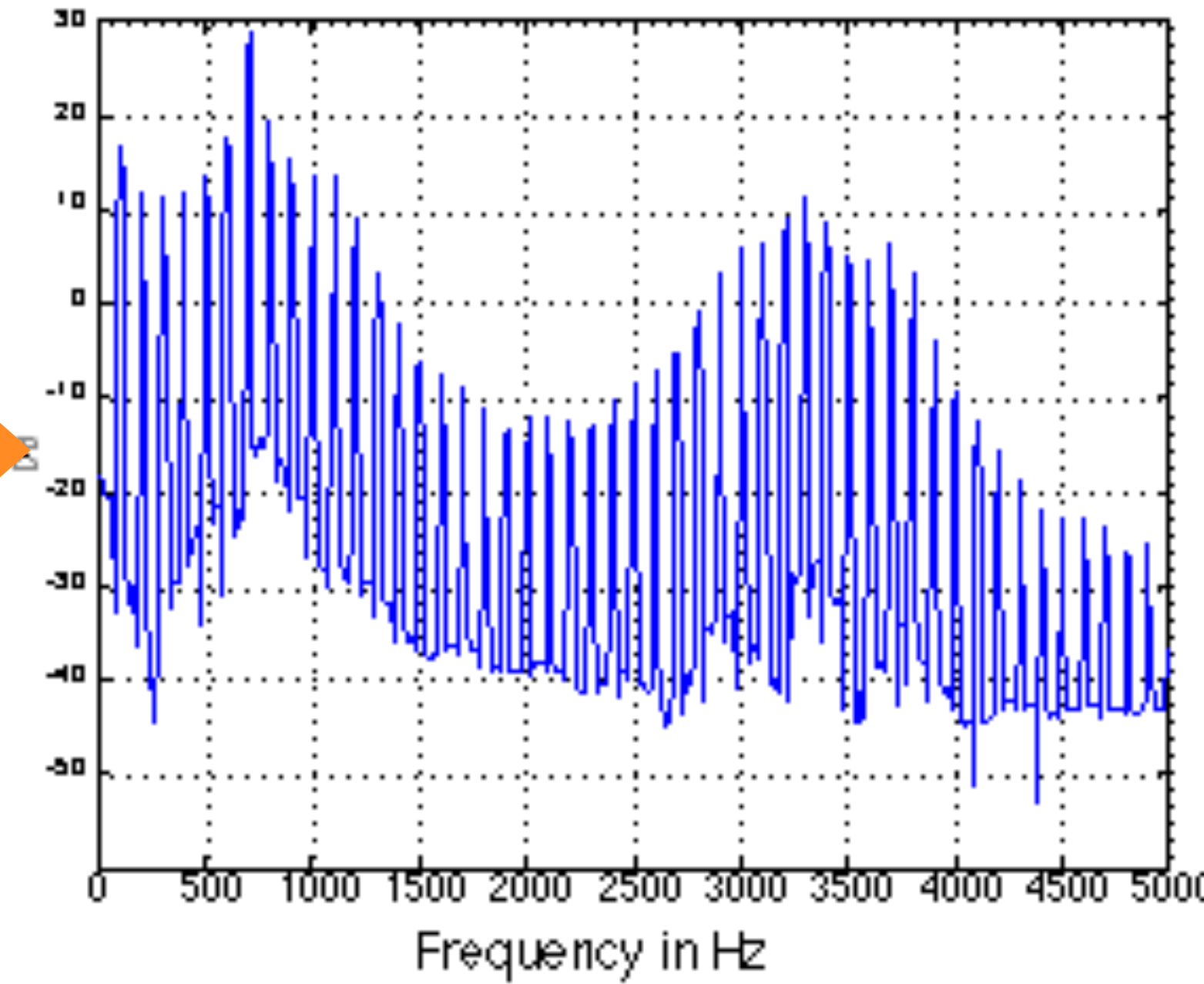
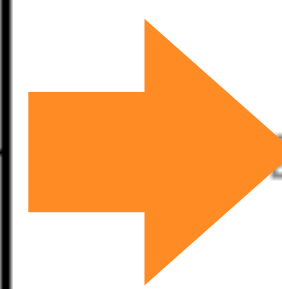
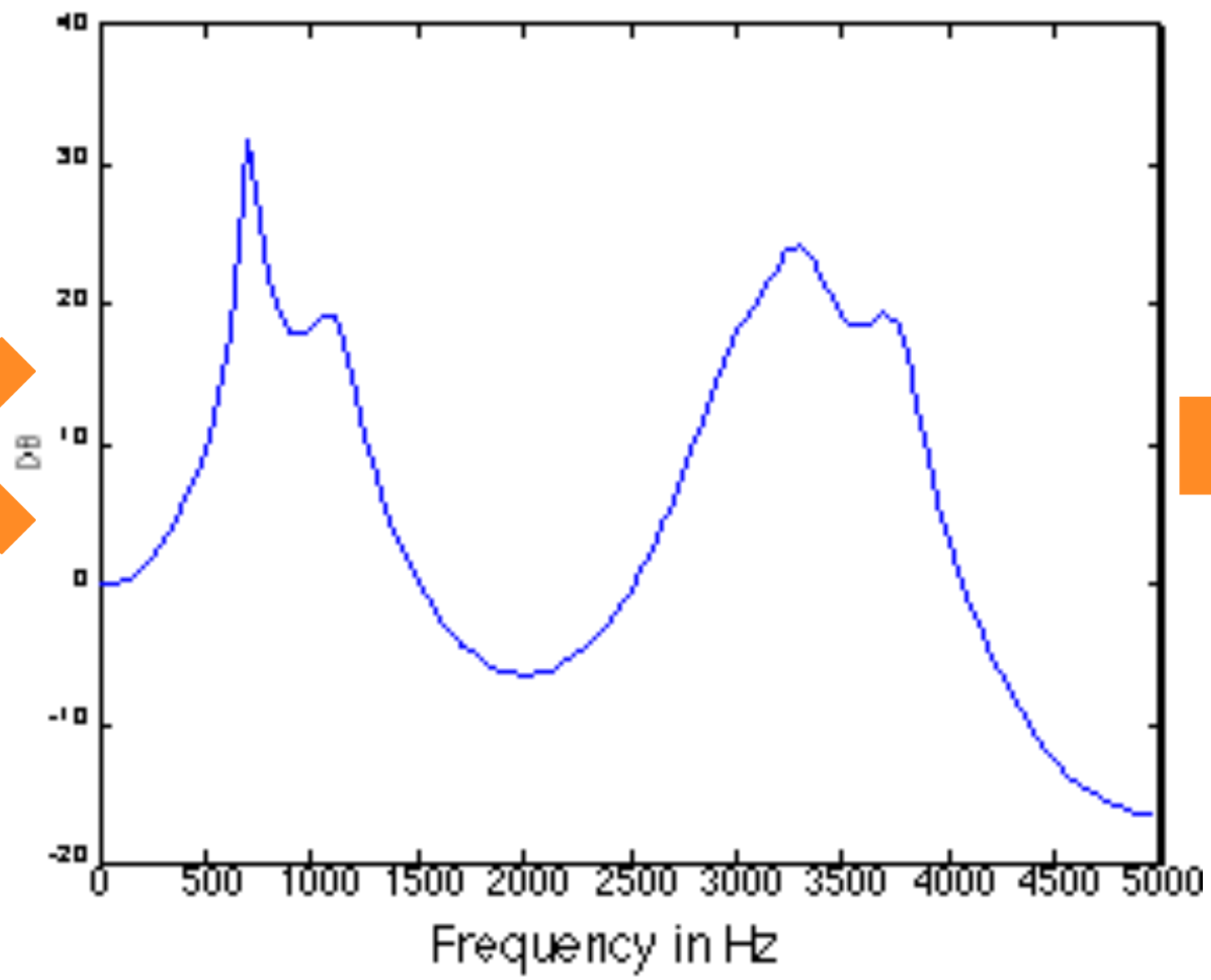
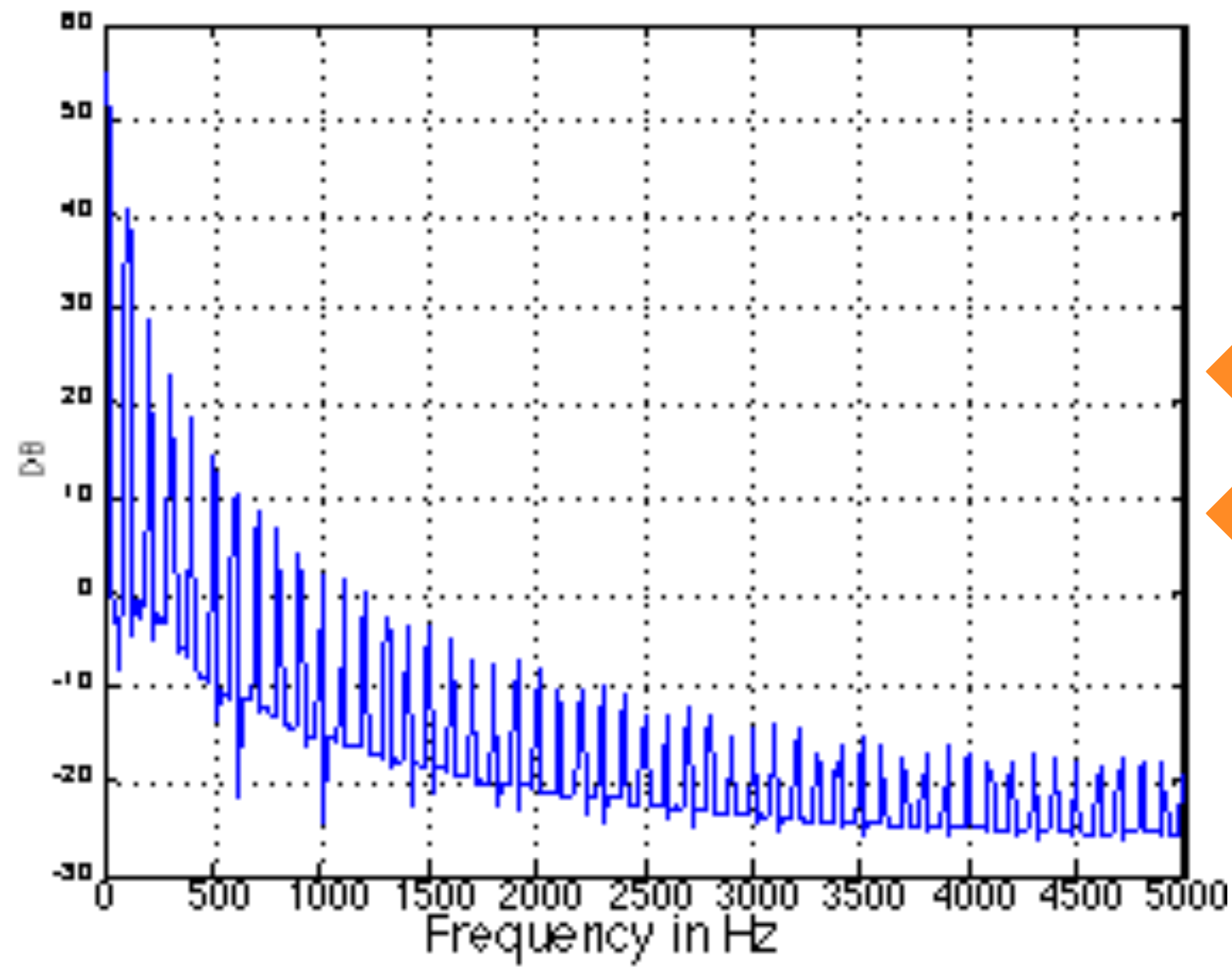
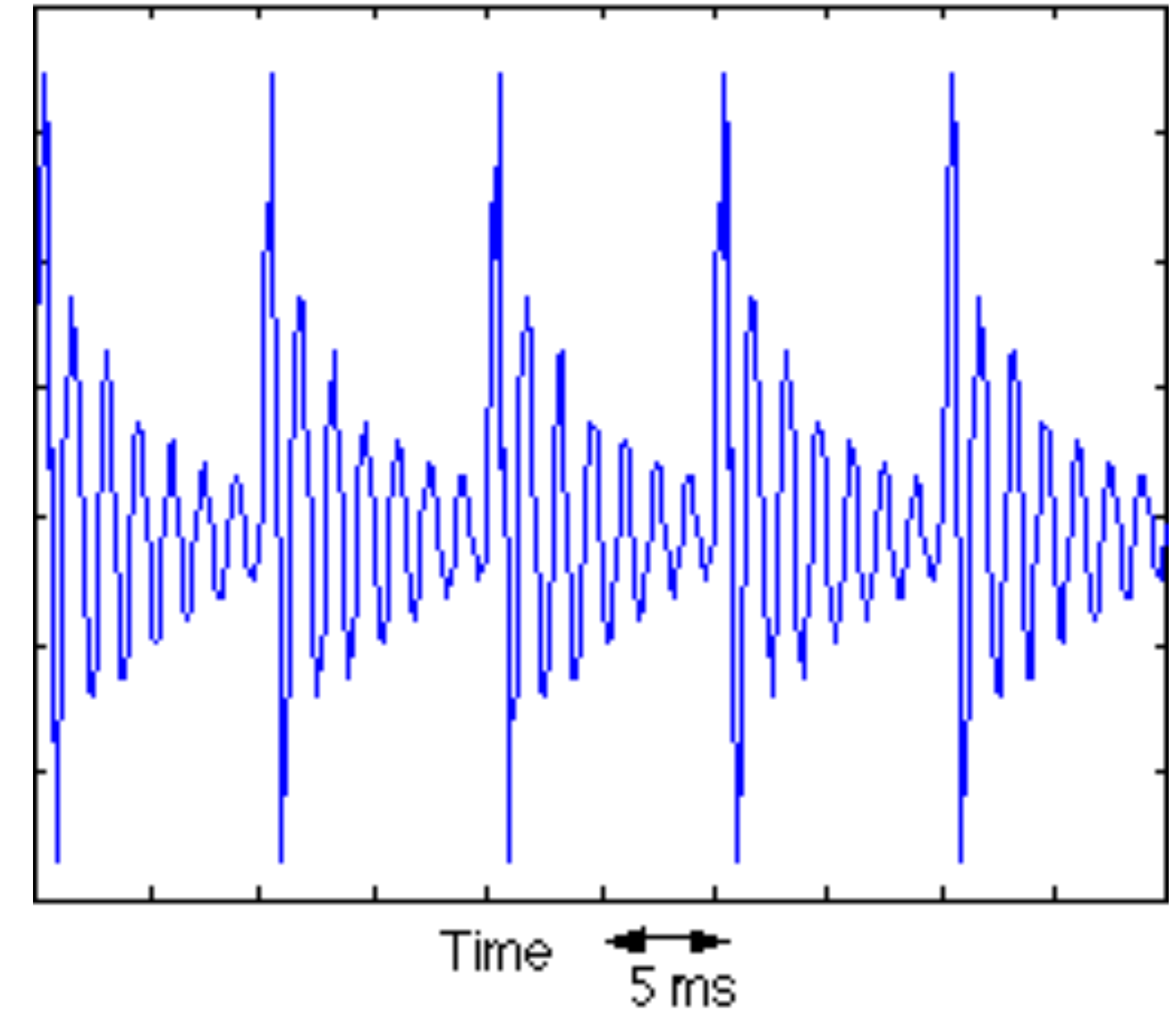
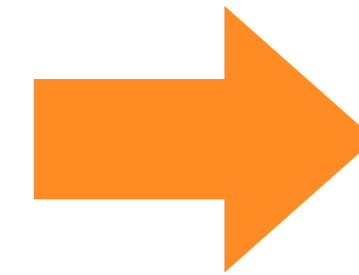
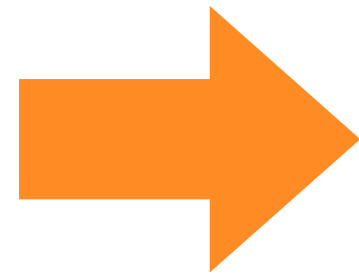
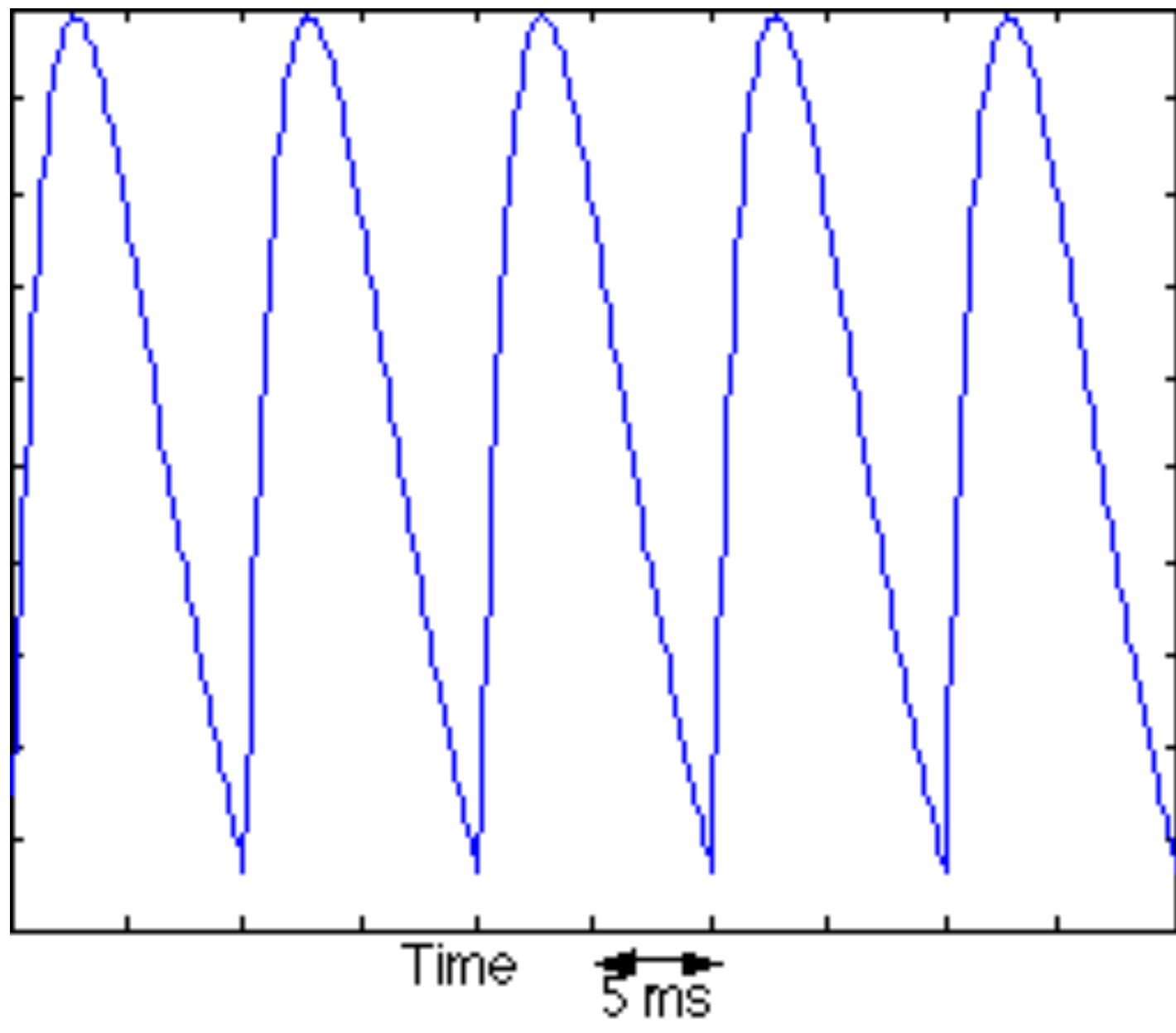
Prosody

# Speech production

- ▶ Source-filter model
  - Source produces an initial sound
  - Vocal tract filter modifies it
- ▶ Source
  - An input of acoustic energy into the speech production system
- ▶ Vocal tract filter
  - Articulators: tongue, teeth, lips, velum etc







Here are the words for "mom" in several different languages:

- English: Mom
- Spanish: Mamá
- French: Maman
- German: Mama
- Italian: Mamma
- Portuguese: Mãe
- Dutch: Moeder
- Russian: Mama (Mama)
- Chinese: 妈妈 (Māma)
- Japanese: 母 (Haha)
- Korean: 엄마 (Eomma)
- Arabic: أم (Umm)
- Hindi: माँ (Maan)
- Bengali: মা (Ma)

# Phone

- ▶ The pronunciation of a word can be represented as a sequence of phones
- ▶ The standard phonetic representation for transcribing the world's languages is the International Phonetic Alphabet (IPA)



tomato

/tə'meɪ.təʊ/

西红柿

xī hóng shì



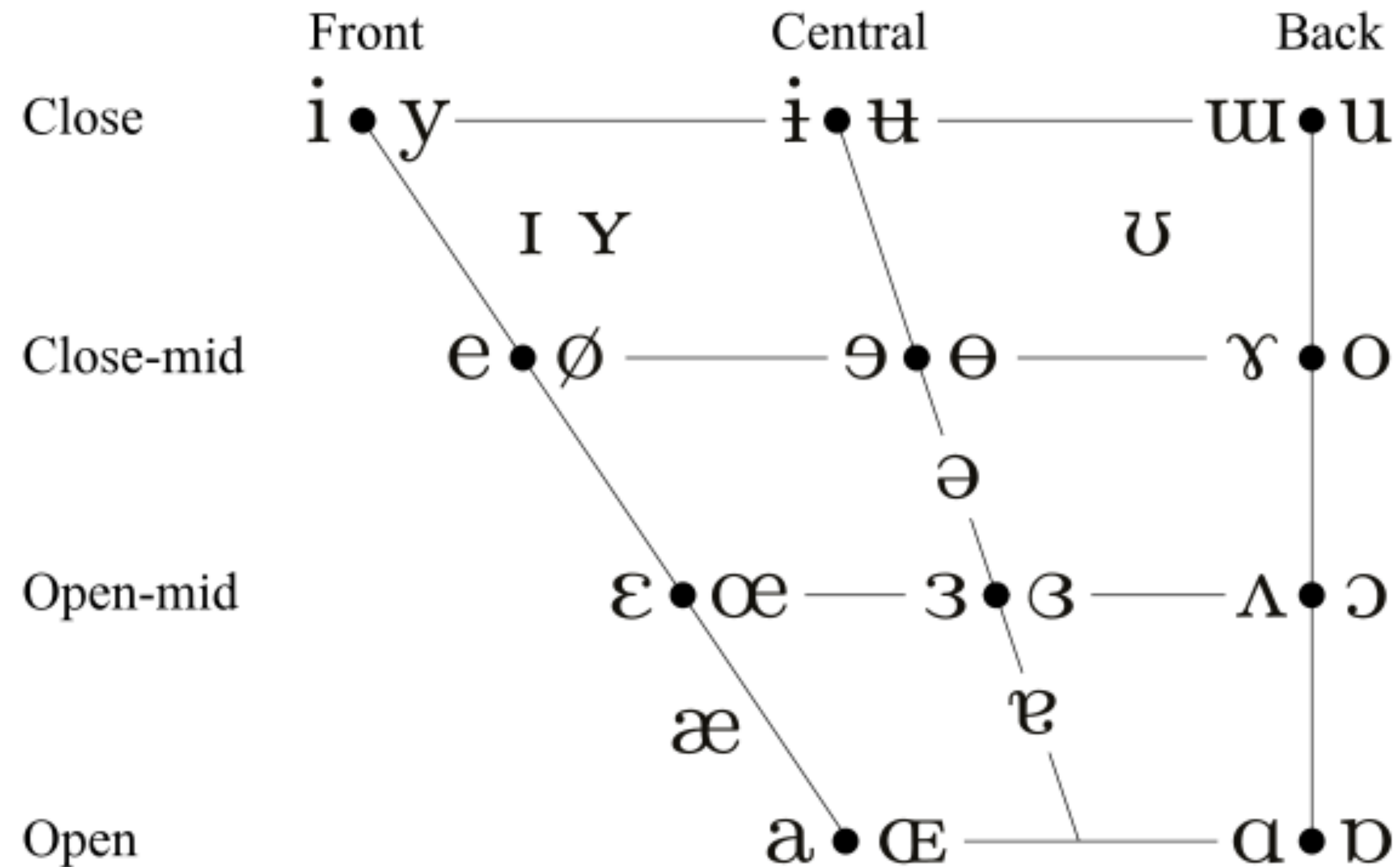
# International Phonetic Alphabet

## ► Consonants

	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			

# International Phonetic Alphabet

► Vowels



# Accent

- ▶ Same writing may have different pronunciation



tomato

/tə'meɪ.tou/

/tə'mɑː.tou/



# Grapheme to phoneme

- ▶ Grapheme: a letter or a group of letters that represent a single phoneme
- ▶ Phoneme: the smallest unit of sound that can distinguish one word from another in a particular language
- ▶ when a child says the sound /t/ this is a phoneme, but when they write the letter 't' this is a grapheme.

**Grapheme**

t o m a t o

**Phoneme**

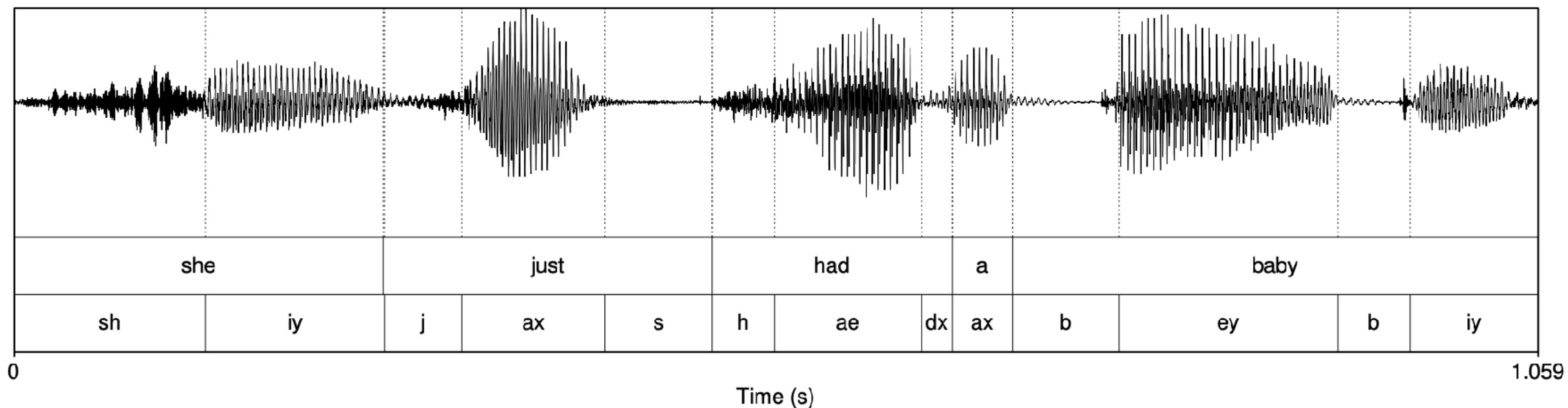
/t ə' m eɪ. t oʊ/

# Grapheme to phoneme conversion

- ▶ Build a set of rules or a statistical model to convert a sequence of graphemes to phonemes

<b>Grapheme</b>	<b>Phoneme</b>
ACCENT	AH <sub>0</sub> K S EH <sub>1</sub> N T
ACCENTS	AE <sub>1</sub> K S EH <sub>0</sub> N T S
ADDICT	AH <sub>0</sub> D IH <sub>1</sub> K T
ADDICTS	AH <sub>0</sub> D IH <sub>1</sub> K T S
ADVOCATE	AE <sub>1</sub> D V AH <sub>0</sub> K EY <sub>2</sub> T
ADVOCATES	AE <sub>1</sub> D V AH <sub>0</sub> K EY <sub>2</sub> T S
AFFECT	AH <sub>0</sub> F EH <sub>1</sub> K T
AFFECTS	AH <sub>0</sub> F EH <sub>1</sub> K T S

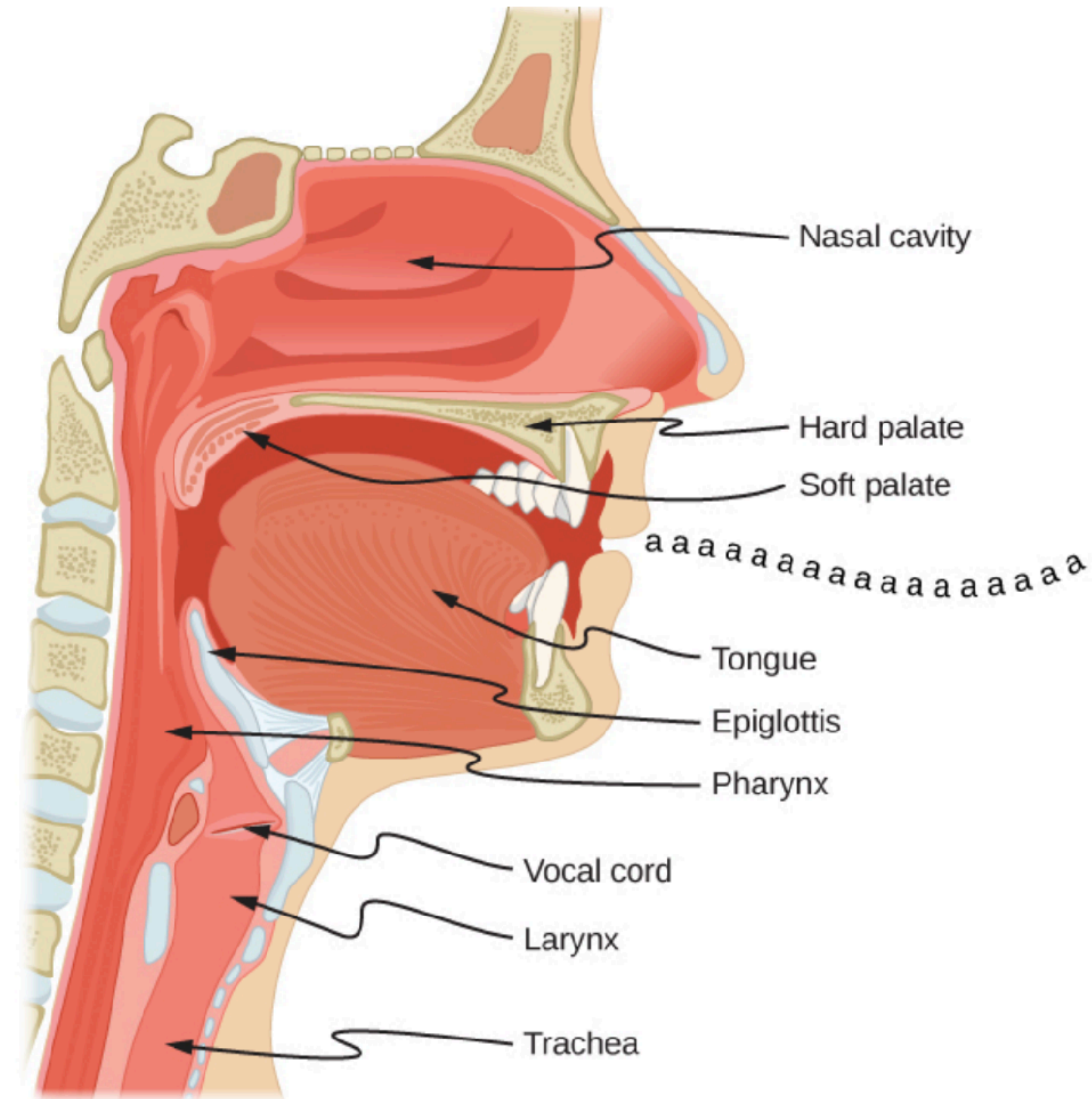
# Interpretation of Phones from a Waveform





# Articulatory Phonetics

- ▶ Articulatory phonetics studies how phones are produced as the various organs in the mouth, throat, and nose modify the airflow from the lungs.



# Vocal tract

- ▶ Vocal tract consists of **oral tract** and **nasal tract**
  - After the air leave the trachea, it can exit the body through the **mouth** or the **nose**
  - **Nasal sounds**: sounds made by air passing through the nose, they use both the oral and nasal tracts as resonating cavities
    - e.g. English [m], [n]
- ▶ Phones can be divided into two classes: vowels and consonants

# Vowels sounds

- ▶ Two types of vowel sounds

monophthongs

One vowel

Examples

Me, that, this, work

diphthongs

Two vowels

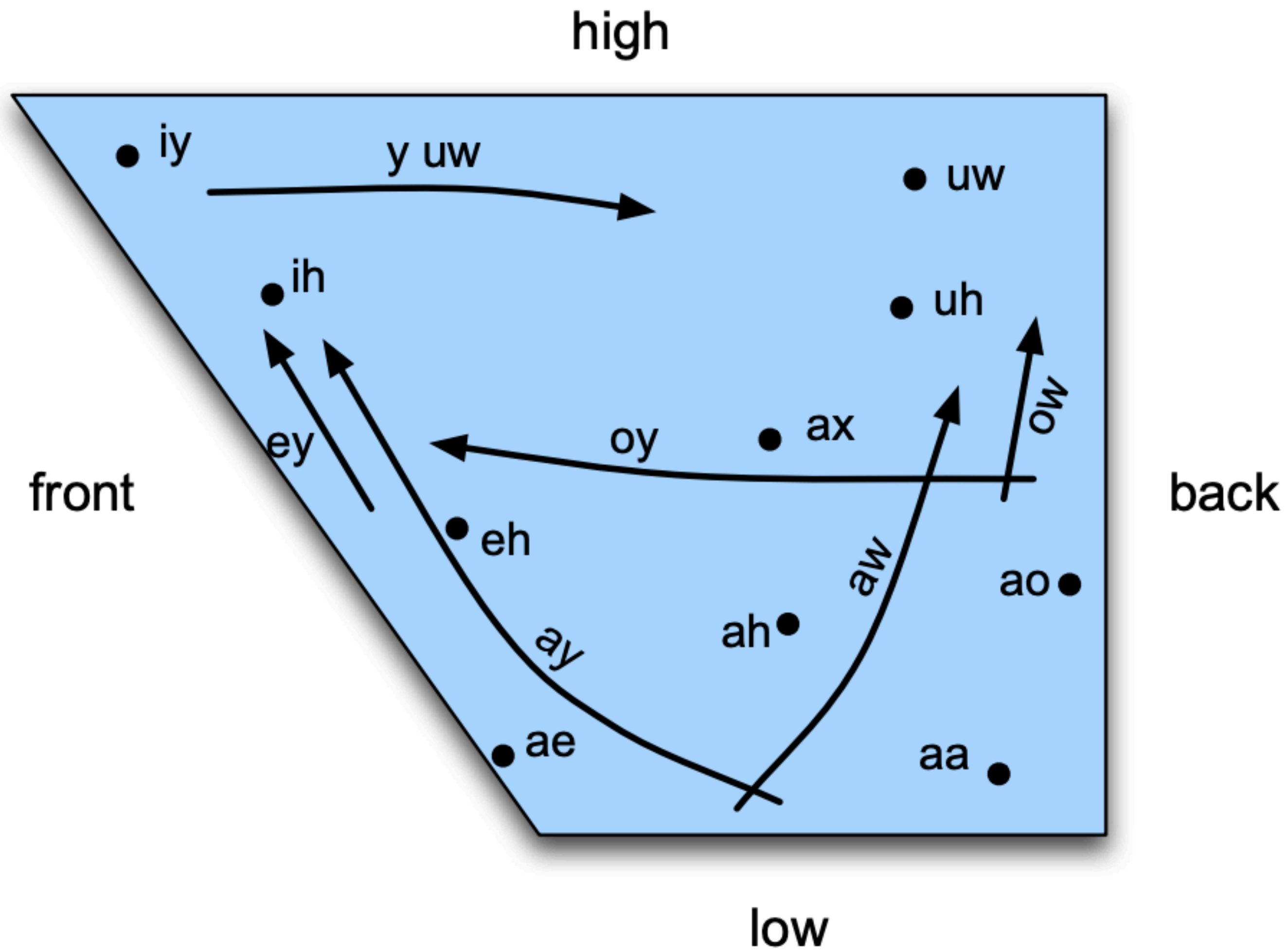
Play, town, slow, toy



# Vowels sounds

- ▶ Heights
  - the vertical position of the tongue relative to either the roof of the mouth
- ▶ Backness
  - the position of the tongue during the articulation of a vowel relative to the back of the mouth
- ▶ Roundedness
  - the amount of rounding in the lips during the articulation of a vowel

# Vowel space

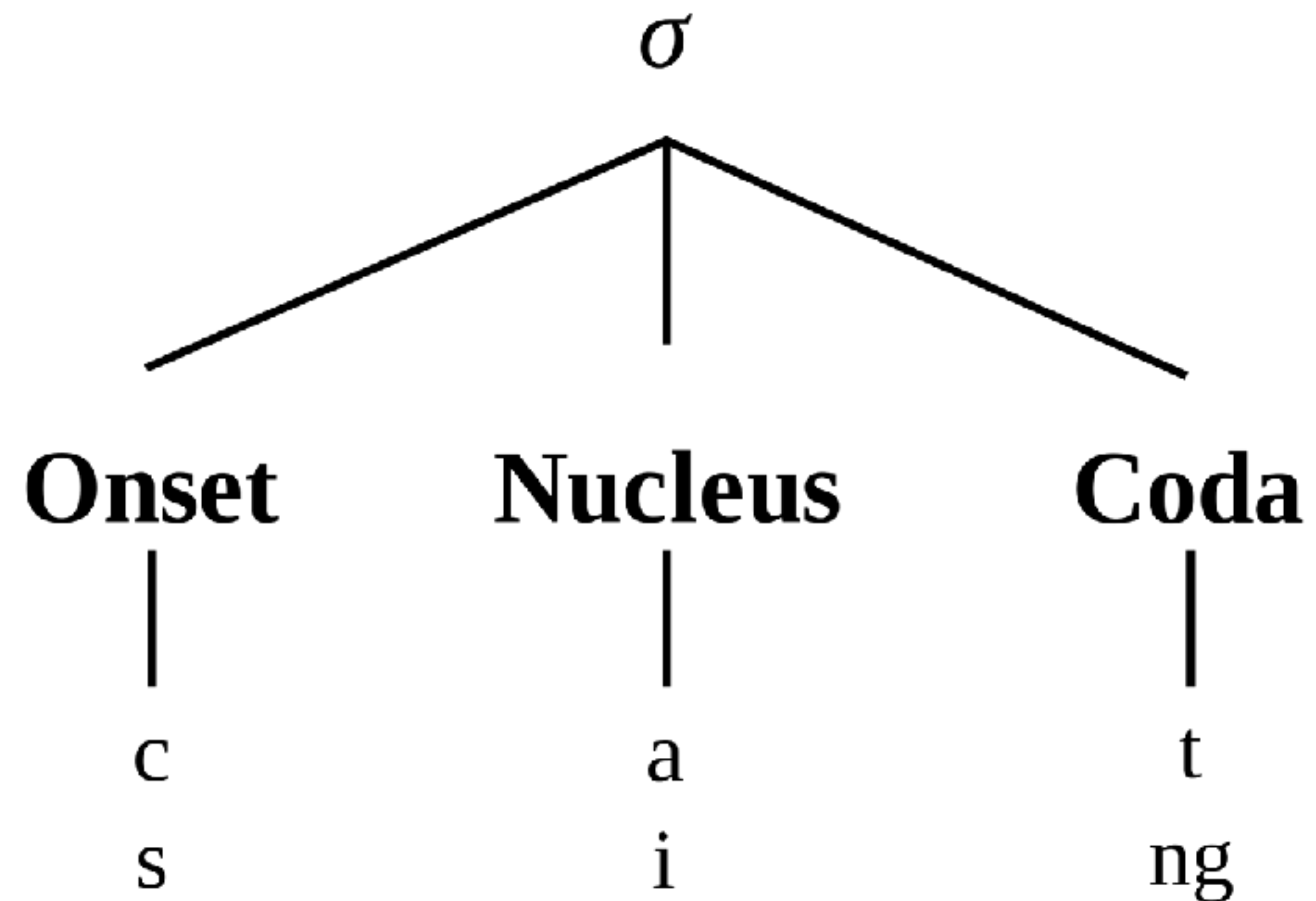


# Consonants

- ▶ A speech sound that is articulated with complete or partial closure of the vocal tract
- ▶ Place of articulations
  - where in the vocal tract the obstruction of the consonant occurs, and which speech organs are involved
- ▶ Manner of articulations
  - how air escapes from the vocal tract when the consonant sound is made
- ▶ Phonation
  - how the vocal cords vibrate during the articulation

# Syllable

- ▶ a unit of organization for a sequence of speech sounds
  - typically made up of a syllable nucleus (most often a vowel) with optional initial and final margins (typically, consonants).
- ▶ Syllables are often considered the phonological "building blocks" of words.





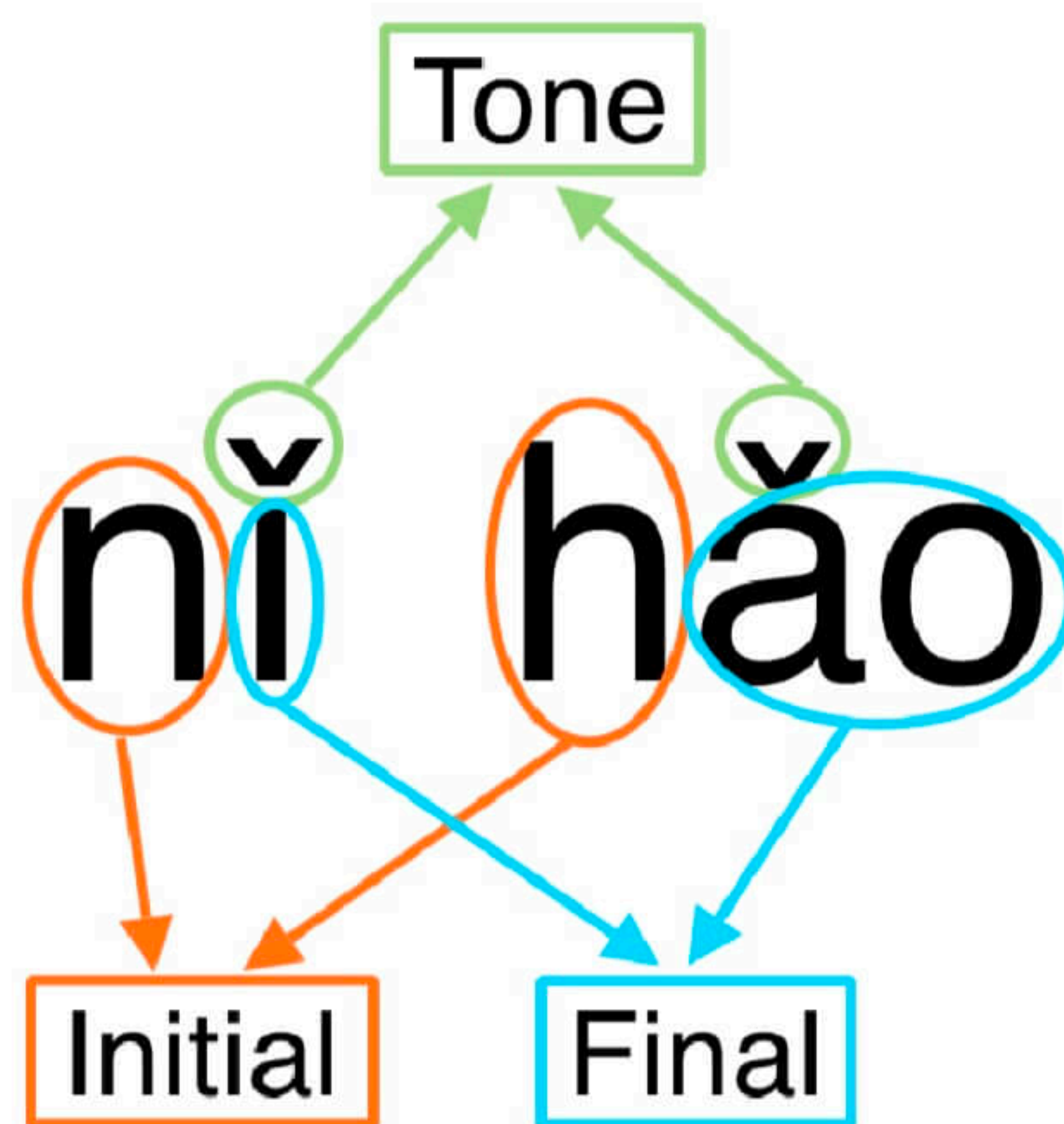
# Syllable

## ► Examples

- Congratulations      5 syllables: con-grat-u-la-tions
- International      5 syllables: in-ter-na-tio-nal
- Water
- Group
- Categorization      6 syllables: cat-e-go-ri-za-tion

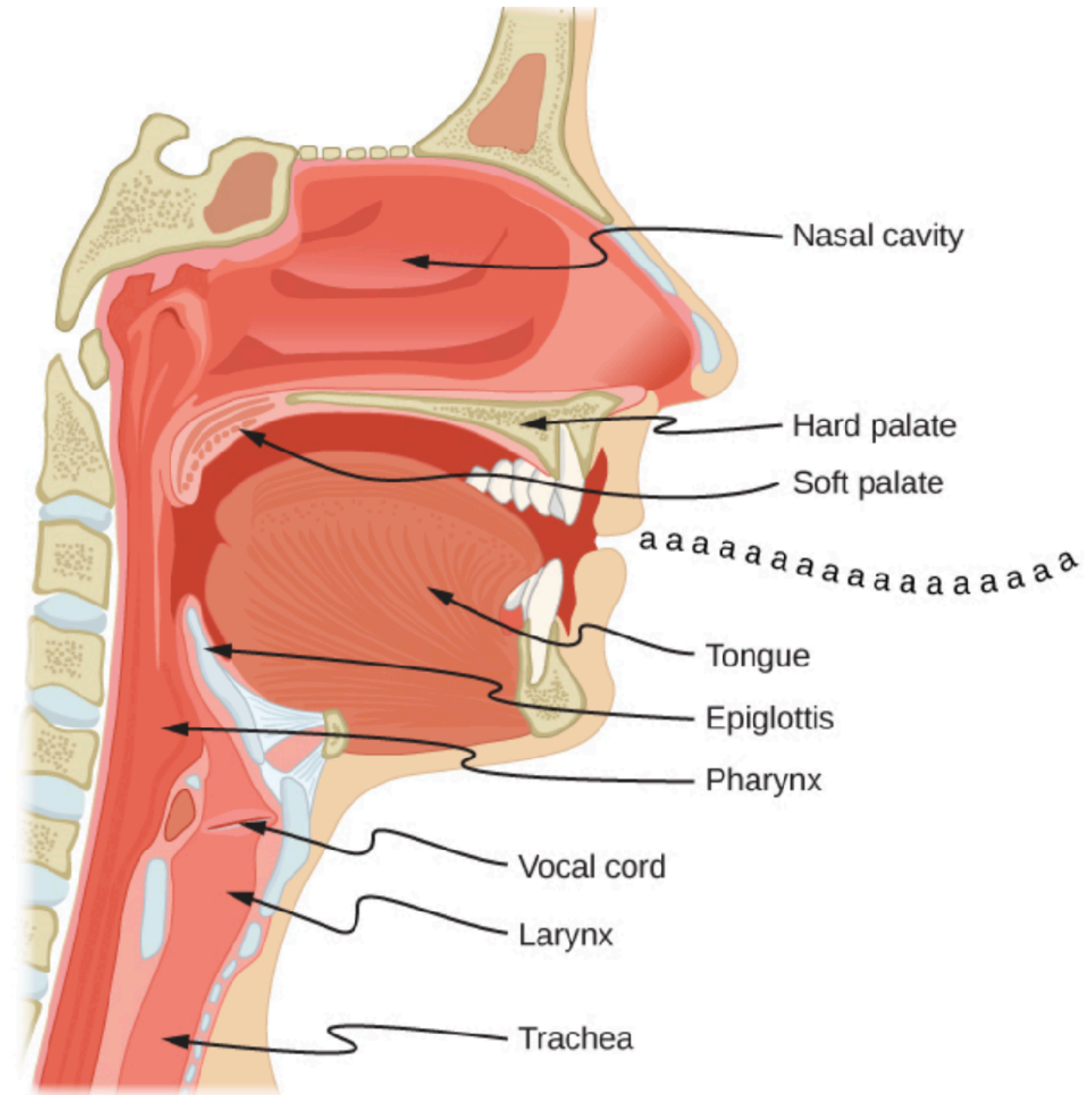
# Syllable: Mandarin Chinese

- ▶ There are about 1300 syllables
  - Each syllable consists of an initial, a final and a tone



# Speech production

- ▶ Larynx
- ▶ Vocal tract
- ▶ Brain
- ▶ Etc





# Speech disorders





# Laryngectomy: removal of voice box



# Summary

- ▶ International Phonetic Alphabet, and Grapheme-to-phoneme conversion
- ▶ Articulatory phonetics: vowels and consonants
- ▶ Speech disorders and the need of speech augmentation