Articulatory phonetics

UNIVERSITY OF TORONTO

# Sounds and transcriptions

- We are often interested in the meaning of an utterance
- In English, we often transcribe utterances as word tokens
  - We write: <How to recognize speech>
- Is this "what was said?"
  - We might write instead: <How to wreck a nice beach>
  - We can transcribe (or even adopt) foreign words
    - 沙发 = <sofa>, not <sandy hair>
  - We can even transcribe brand new words
    - <skibidi toilet>
- We can instead transcribe "speech sounds"

Phones and phonetics

- **Phonetics** is the study of speech sounds
- A **phone** is a unit of speech
  - Denoted with square braces: [t], [tʰ], [u]
  - Language-independent
- Phones which are perceived "similarly" are grouped into **phonemes**
  - Denoted with slashes: /t/, /u/
  - [t],[tʰ] ↦ /t/
  - Language-dependent
- Transcriptions are often in-between:
  - [ˈtʰuː] ↦ [tʰu] ↦ [tu] ↦ /tu/
- We will be *very* loose with the distinction

Phonetic transcription

- Often, we assume that a **spoken utterance** can be **partitioned** into a **sequence** of **non-overlapping** phones.
  - Demarking the periods during which certain phones are being uttered is called **phonetic transcription**
  - This approach has problems (e.g., when *exactly* does one phoneme end and another begin?), but it's useful for **classification**.

Waveform

This looks periodic

[t]　　　[u]
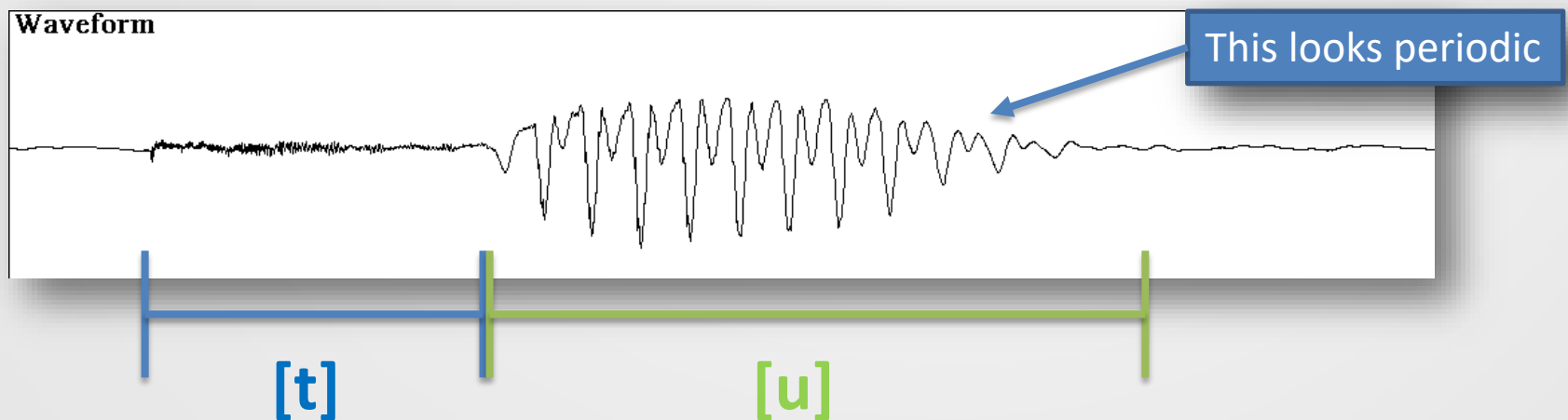
Phonetic alphabets

- There are several alphabets that categorize the sounds of speech.
  - The **International Phonetic Alphabet (IPA)** is popular, but it uses non-ASCII symbols.
  - The **TIMIT** phonetic alphabet will be used by **default** in this course.

  - Other popular alphabets include **ARPAbet**, **Worldbet**, and **OGIbet**, usually adding special cases.
    - E.g., [pcl] is the period of silence immediately before a [p].

TIMIT	IPA	e.g.
[iy]	[iʸ]	b*ea*t
[ih]	[ɪ]	b*i*t
[eh]	[ɛ]	b*e*t
[ae]	[æ]	b*a*t
[aa]	[ɑ]	B*o*b
[ah]	[ʌ]	b*u*t
[ao]	[ɔ]	b*ou*ght
[uh]	[ʊ]	b*oo*k
[uw]	[u]	b*oo*t
[ux]	[ʉ]	s*ui*t
[ax]	[ə]	*a*bout

UNIVERSITY OF
TORONTO

TIMITbet (incomplete)

Vowel	e.g.
[iy]	b*ea*t
[ih]	b*i*t
[eh]	b*e*t
[ae]	B*a*t
[aa]	B*o*b
[ah]	B*u*t
[ao]	b*ou*ght
[uh]	b*oo*k
[uw]	b*oo*t
[ux]	s*ui*t
[ax]	*a*bout

stop	e.g.
[b]	*B*il*b*o
[d]	*d*a*d*a
[g]	*G*a*g*a
[p]	*P*i*pp*in
[t]	*T*oo*t*s
[k]	*k*i*ck*

nasal	e.g.
[m]	*M*a*m*a
[n]	*n*oo*n*
[ng]	thi*ng*

fricative	e.g.
[s]	*S*ea
[f]	*F*rank
[z]	*Z*appa
[th]	*th*is
[sh]	*Sh*ip
[zh]	a*z*ure
[v]	*V*ogon
[dh]	*th*en

• • •

(Incomplete)

UNIVERSITY OF TORONTO

The vocal tract

Nasal cavity

Velum

Lips

Tongue

Jaw

Lungs
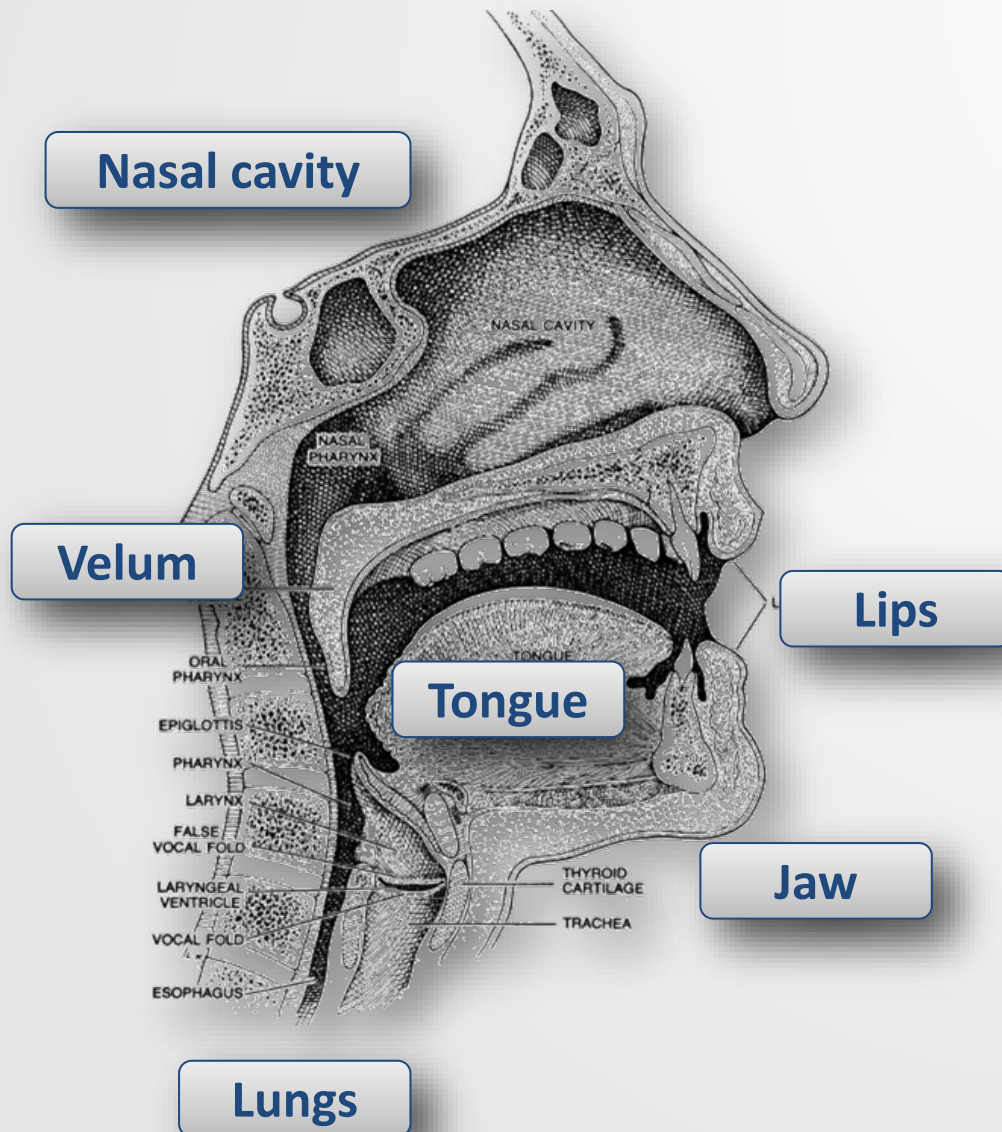
- Many physical structures are co-ordinated in the production of speech.
- Generally, sound is **generated** by passing air through the vocal tract.
- Sound is **modified** by constricting airflow in particular ways.
- We can classify phones by how they are **produced**

UNIVERSITY OF TORONTO

A taxonomy of phones

- Phones fall into two broad categories
- **Vowels** are
  - Always periodic
  - Produced with relatively unobstructed airflow
  - Use tongue, lips, and jaw to produce **resonances** in vocal tract, in turn generating **formants**
- **Consonants** are
  - Mostly noisy (not nasals, semivowels)
  - Produced by obstructing airflow
  - Classified by the **place** and **manner** of primary obstruction, as well as **voicing**

UNIVERSITY OF
TORONTO

Voicing and fundamental frequency

- **Voiced** phones are produced with vibrating **vocal folds**
  - The space between the folds is the **glottis**
- All vowels are voiced; consonants can be **unvoiced**
- F_0: *n.* (**fundamental frequency**), the rate of vibration (Hz)
  - Very indicative of speaker

	Avg F_0 (Hz)	Min F_0 (Hz)	Max F_0 (Hz)
Male	125	80	200
Female	225	150	350
Children	300	200	500

UNIVERSITY OF TORONTO

Vowels

- There are approximately **19** vowels in Canadian English, including **diphthongs** in which the articulators **move** over time.

- Vowels are distinguished primarily by their **formants**. (?)

other	e.g.
[er]	*B**er**t*
[axr]	*b**u**tter*

diphthong	e.g.
[ey]	*b**ai**t*
[ow]	*b**oa**t*
[ay]	*b**i**te*
[oy]	*b**oy***
[aw]	*b**ou**t*
[ux]	*s**ui**t*

| Mono-phthong	e.g.
[iy]	*b**ea**t*
[ih]	*b**i**t*
[eh]	*b**e**t*
[ae]	*b**a**t*
[aa]	*B**o**b*
[ao]	*b**ou**ght*
[ah]	*b**u**t*
[uh]	*b**oo**k*
[uw]	*b**oo**t*
[ax]	***a**bout*
[ix]	*ros**e**s*

UNIVERSITY OF TORONTO

Uniform tubes

- Formants and resonances can be approximated with tubes
- Many **musical instruments** are based on the idea of uniform (or, in many cases, bent) tubes.
- **Longer** tubes produce '**deeper**' sounds (lower frequencies).
  - A tube ½ the length of another will be 1 octave higher.

UNIVERSITY OF
TORONTO

# The uniform tube

Closed, vibrating end

glottis

Open, radiating end

lips

17 cm
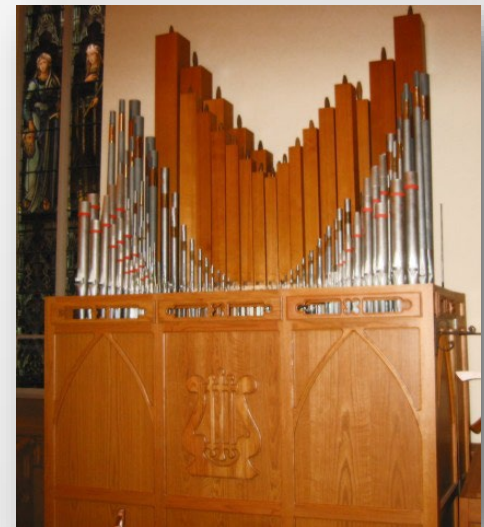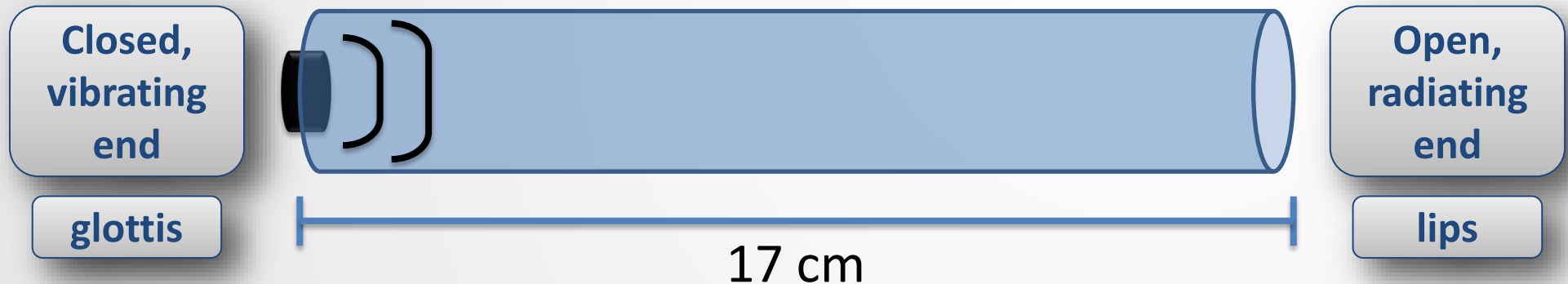
- The positions of the tongue, jaw, and lips change the **shape** and **cross-sectional** area of the vocal tract.

Vocal–tract area function

Area (cm²)
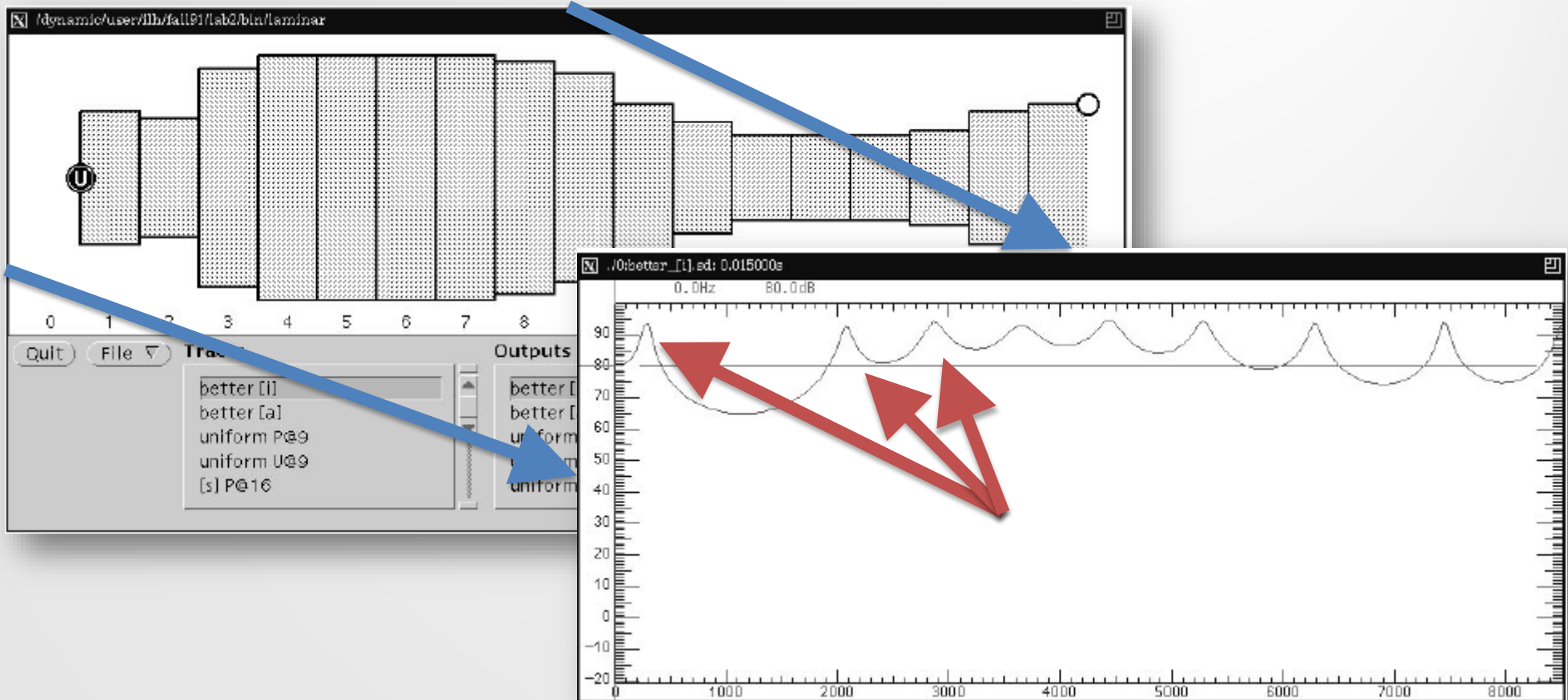
Distance from larynx (cm)

UNIVERSITY OF TORONTO
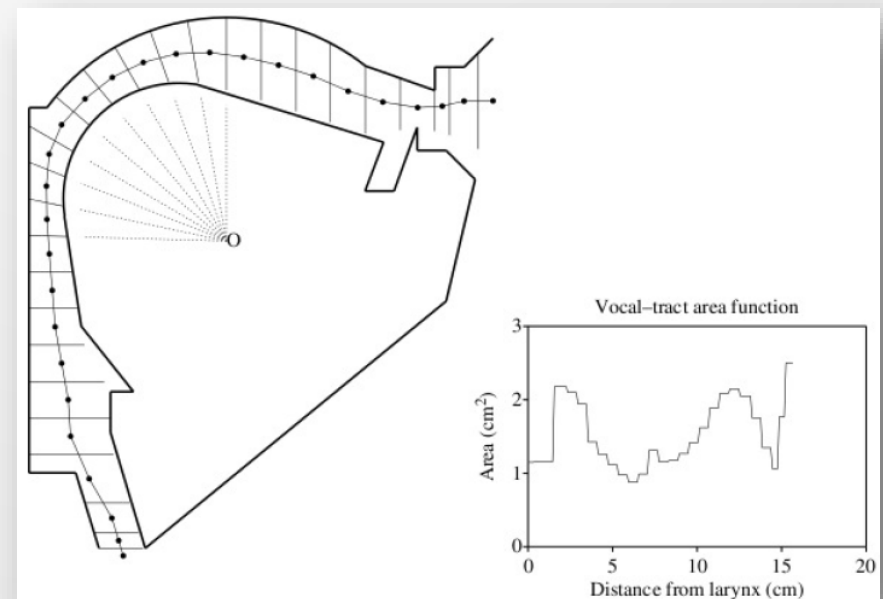
Vowels as concatenated tubes

- The vocal tract can be modelled as the concatenation of dozens, hundreds, or thousands of tubes.
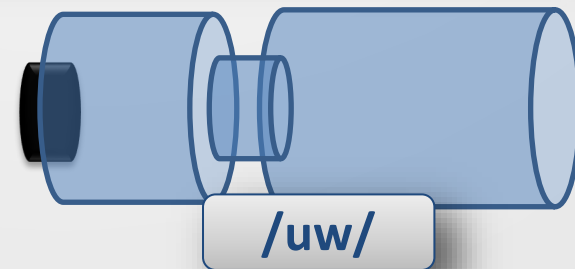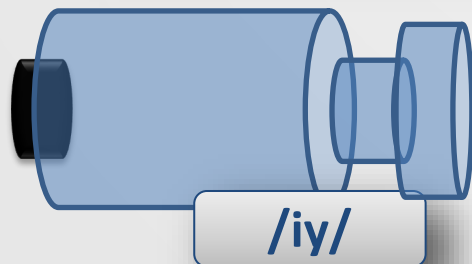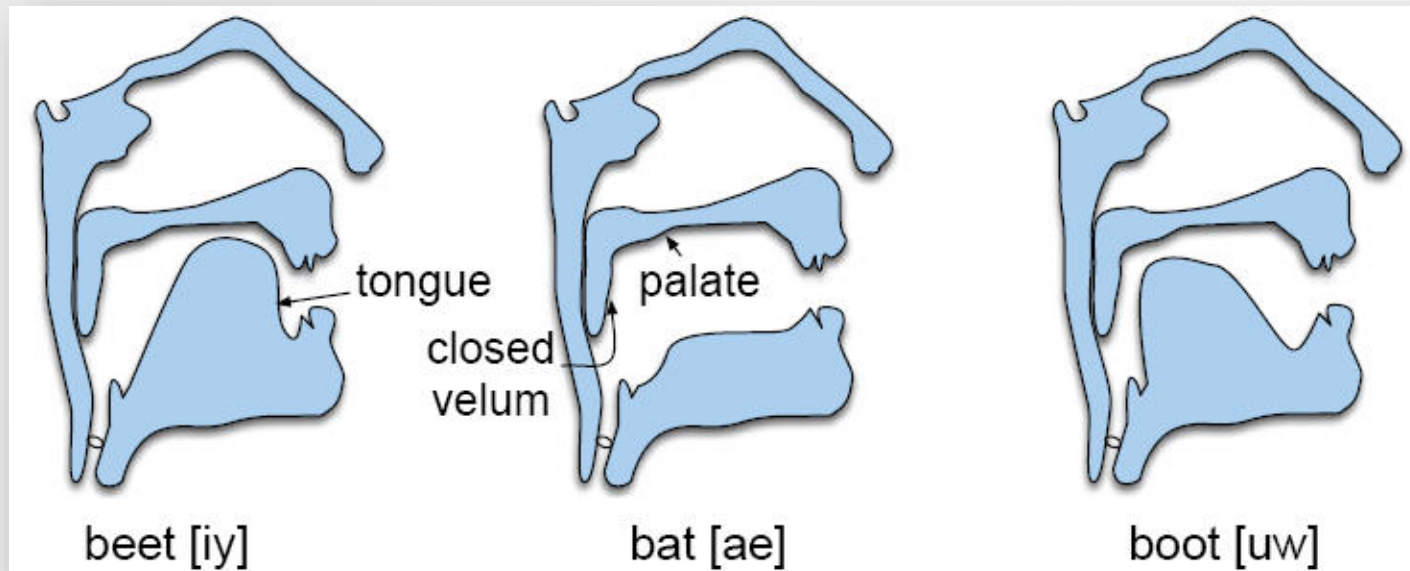
Waves in concatenated tubes

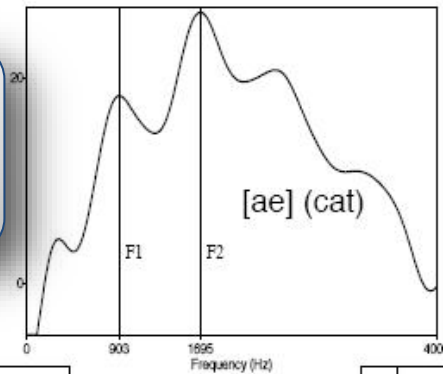- Reflections at tube boundaries produce resonances which amplify certain frequencies

Formants and vowels

- **Formant**: *n.* A concentration of energy within a frequency band. Ordered from low to high bands (e.g., F_1, F_2, F_3).
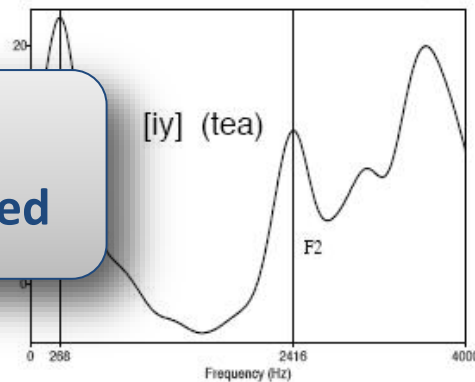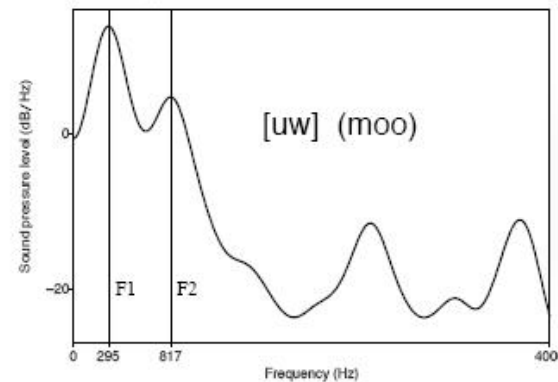

beet
/biʸt/

bat
/bæt/

bott
/bɑt/

boot
/but/
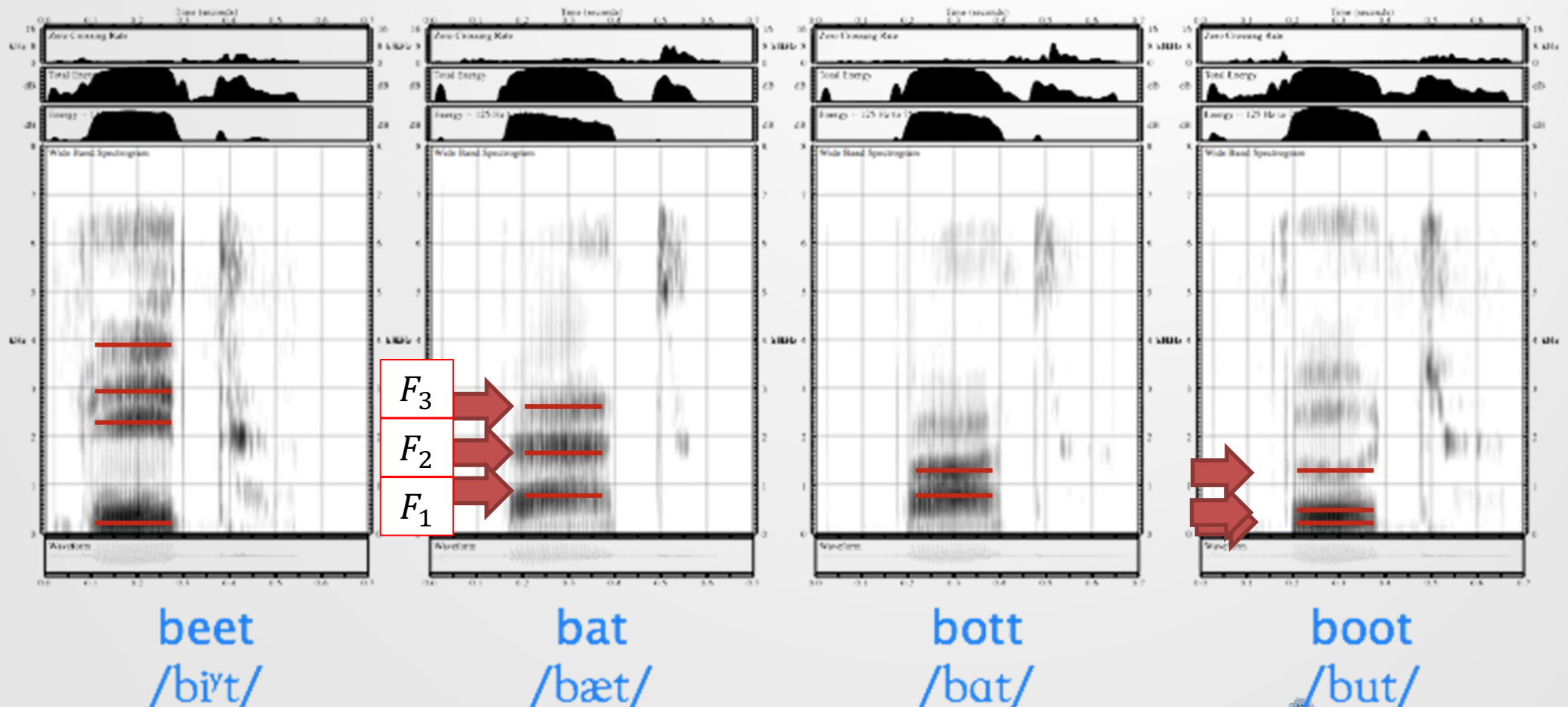
Tongues, lips, and formants

Front/ low/unrounded
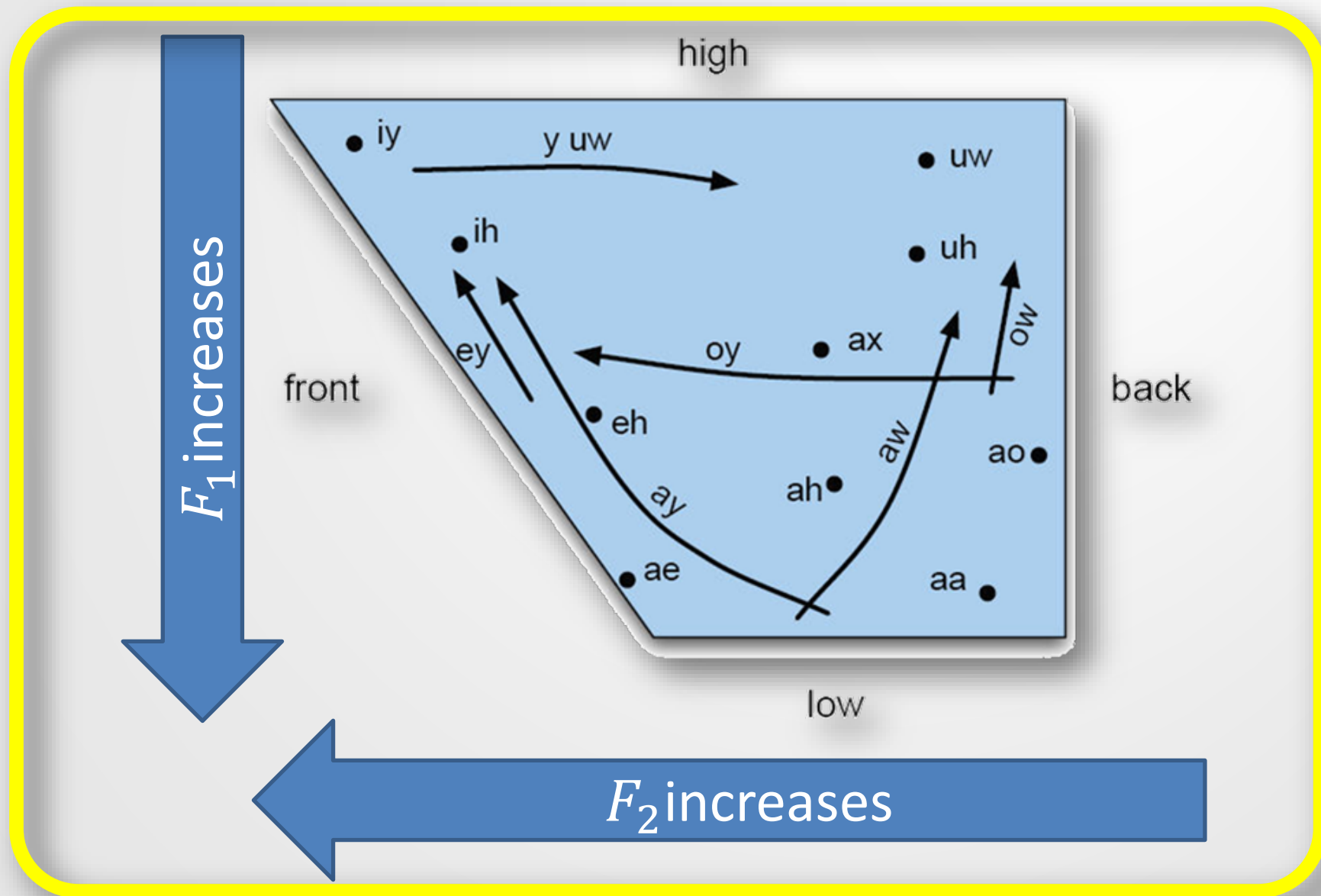
Front/ high/unrounded

Back/ high/rounded

UNIVERSITY OF TORONTO

The vowel trapezoid

Manner of articulation

- Consonants are classified by **place** and **manner** of obstruction
- For manner:
  - **Fricatives:**   **noisy**, with air passing through a tight constriction (e.g., '*shift*').
  - **Stops/plosives:**   **complete** vocal tract constriction and burst of energy (e.g., '*papa*').
  - **Nasals:**   air passes through the **nasal** cavity (e.g., '*mama*').
  - **Semivowels:**   similar to vowels, but typically with more constriction (e.g., '*wall*').
  - **Affricates:**   Alveolar stop followed by fricative.
 - **Taps:**   Quick collision of articulators ('bu<u>tt</u>er')

Place of articulation

- The **location** of the *primary constriction* can be:
  - **Alveolar**: constriction near the alveolar ridge (e.g., [*t*])
  - **Bilabial**: touching of the lips together (e.g., [*m*], [*p*])
  - **Dental**: constriction of/at the teeth (e.g., [*th*])
  - **Palatal**: constriction at the (hard) palate (e.g., [sh])
  - **Labiodental**: constriction between lip and teeth (e.g., [*f*])
  - **Velar**: constriction at or near the velum (e.g., [*k*]).
  - **Glottal:** constriction of the glottis ([q])

UNIVERSITY OF TORONTO
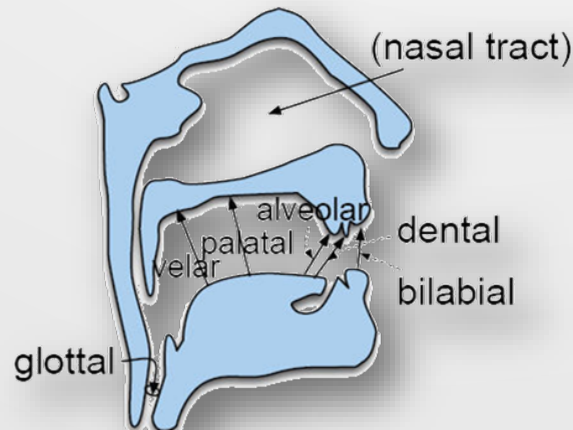
Fricatives

- **Fricatives** are caused by acoustic turbulence at a **narrow constriction** whose position determines the sound.

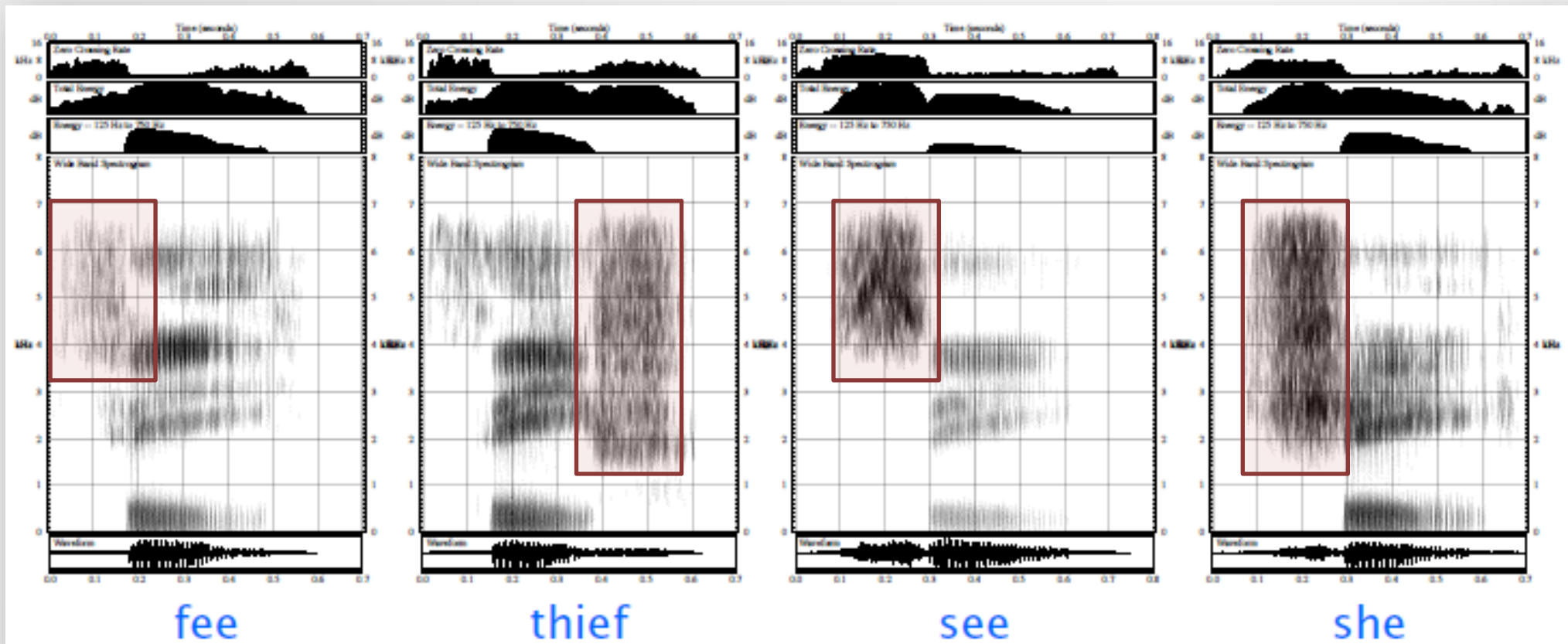| Labio-dental	dental	alveolar	palatal
[f]	[th]	[s]	[sh]

UNIVERSITY OF TORONTO
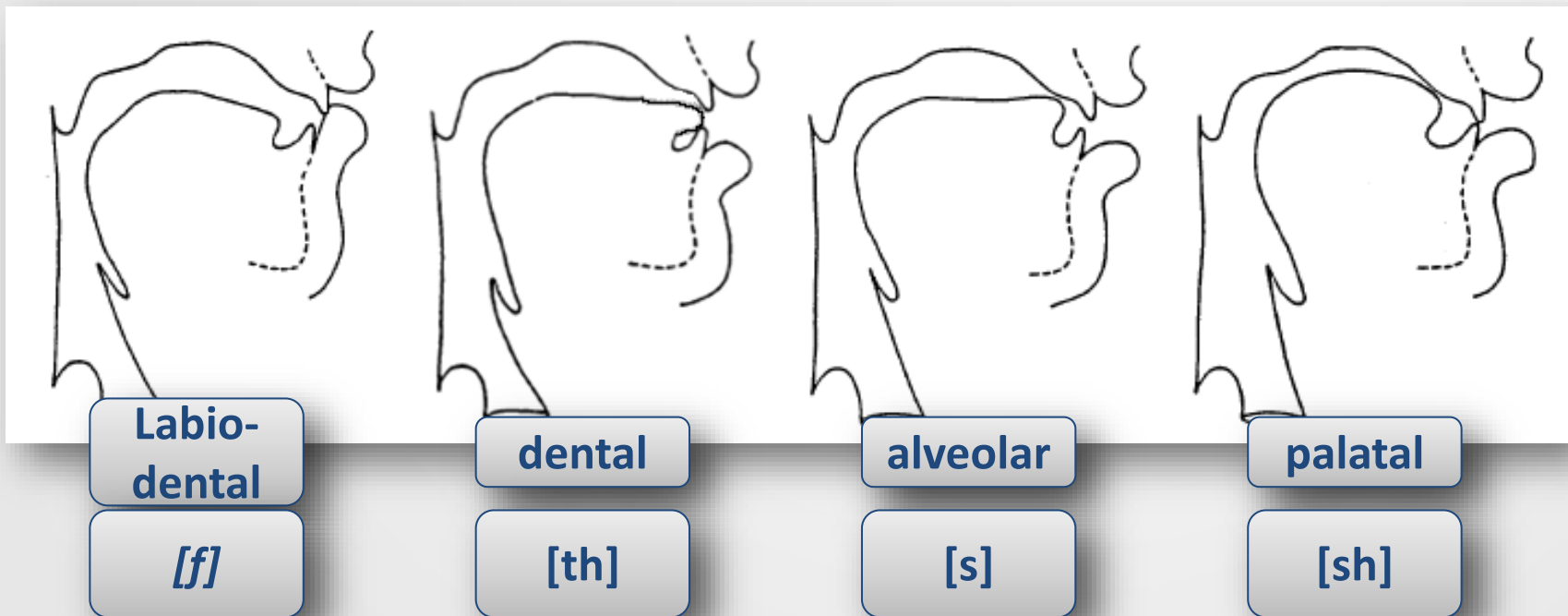
Fricatives

- **Fricatives** have four places of articulation.
- Each place of articulation has a **voiced** fricative (i.e., the folds can be vibrating), and an **unvoiced** fricative.
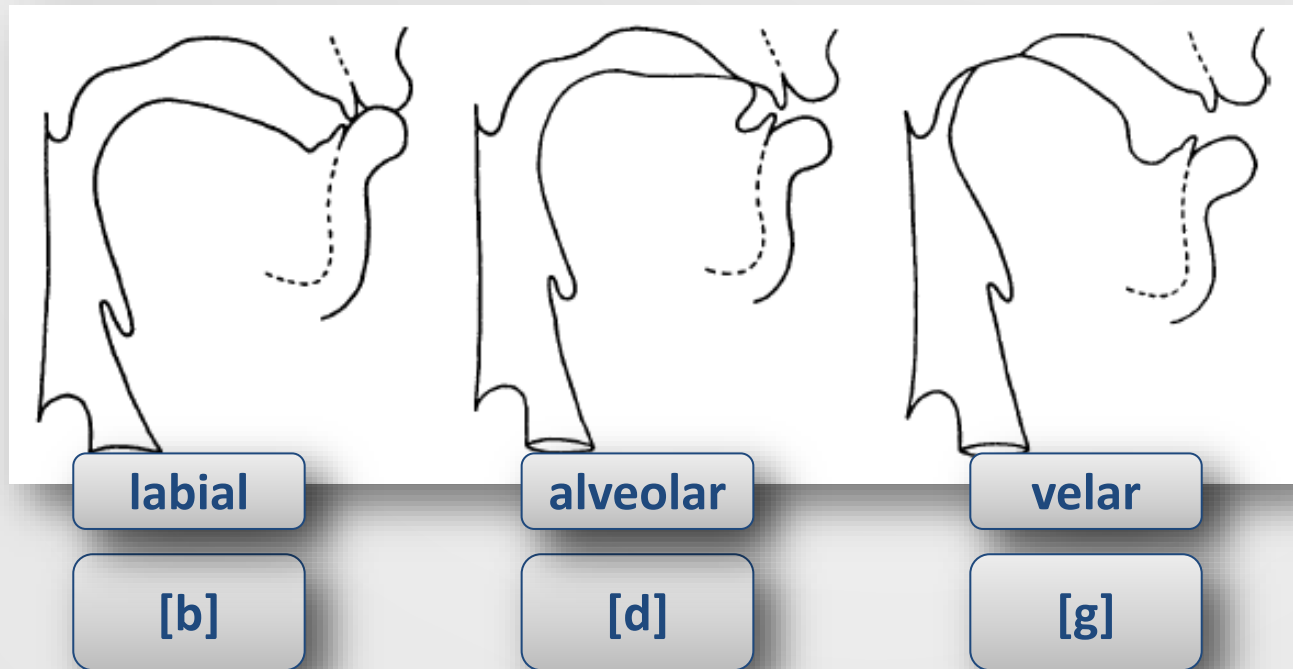
	Unvoiced		Voiced	
Labio-dental	*[f]*	*fee*	*[v]*	*Vendetta*
Dental	*[th]*	*thief*	*[dh]*	*Thee*
Alveolar	*[s]*	*see*	*[z]*	*Zardoz*
Palatal	*[sh]*	*she*	*[zh]*	*Zha-zha*

UNIVERSITY OF
TORONTO

Unvoiced fricatives

fee   thief   see   she

UNIVERSITY OF
TORONTO

Plosives (3/6)

- **Plosives** build pressure behind a **complete closure** in the vocal tract.
- A **sudden release** of this constriction results in **brief noise**.

labial — [b]

alveolar — [d]

velar — [g]

UNIVERSITY OF TORONTO

Plosives

- **Plosives** have three places of articulation:

|	Unvoiced		Voiced	
Labial	[p]	*__p__or__p__oise*	[b]	*__b__a__b__oon*
Alveolar	[t]	*__t__or__t__*	[d]	*__d__o__d__o*
Velar	[k]	*__k__i__ck__*	[g]	*__G__oo__g__le*

- **Voiced** stops are usually characterized by a "**voice bar**" during closure, indicating the vibrating vocal folds.
- Formant **transitions** are very **informative** in classification.

UNIVERSITY OF TORONTO

Formant transitions in plosives

poop   toot   kook
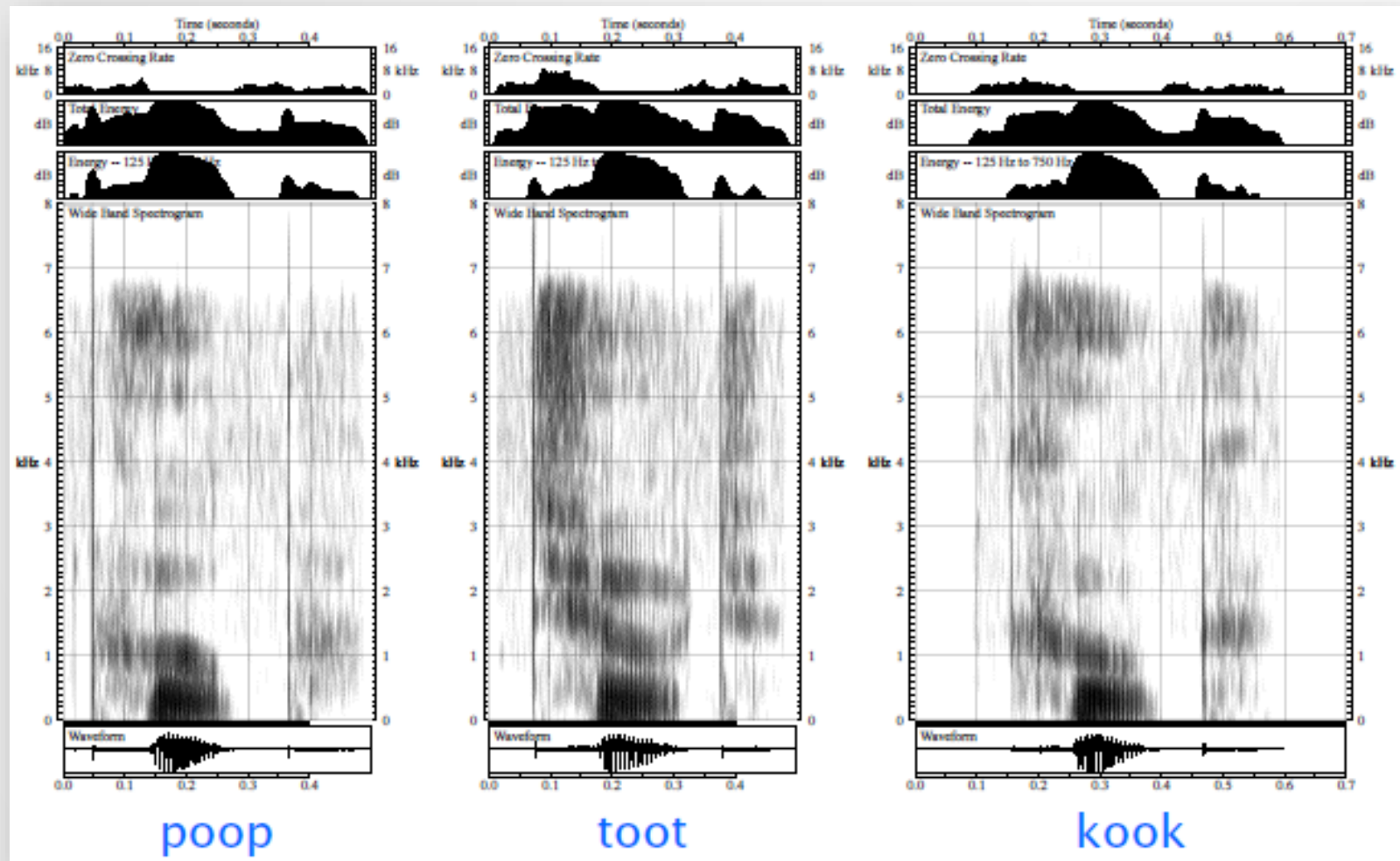
- Despite a **common** vowel, the **motion** of F_2 and F_3 into (and out of) the vowel helps identify the plosive.

UNIVERSITY OF TORONTO

Voicing in plosives
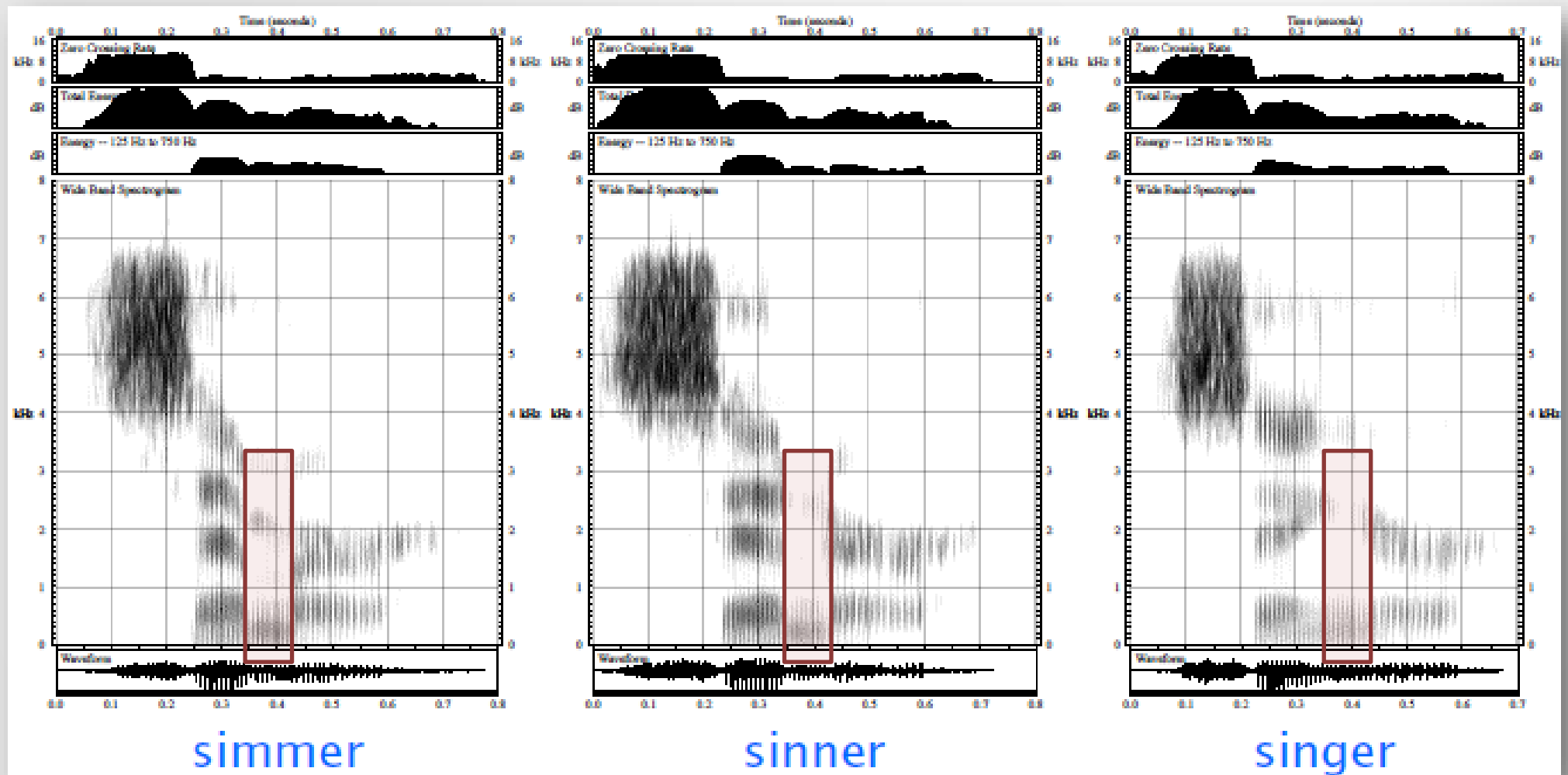

The "voice bar"

pop

bob
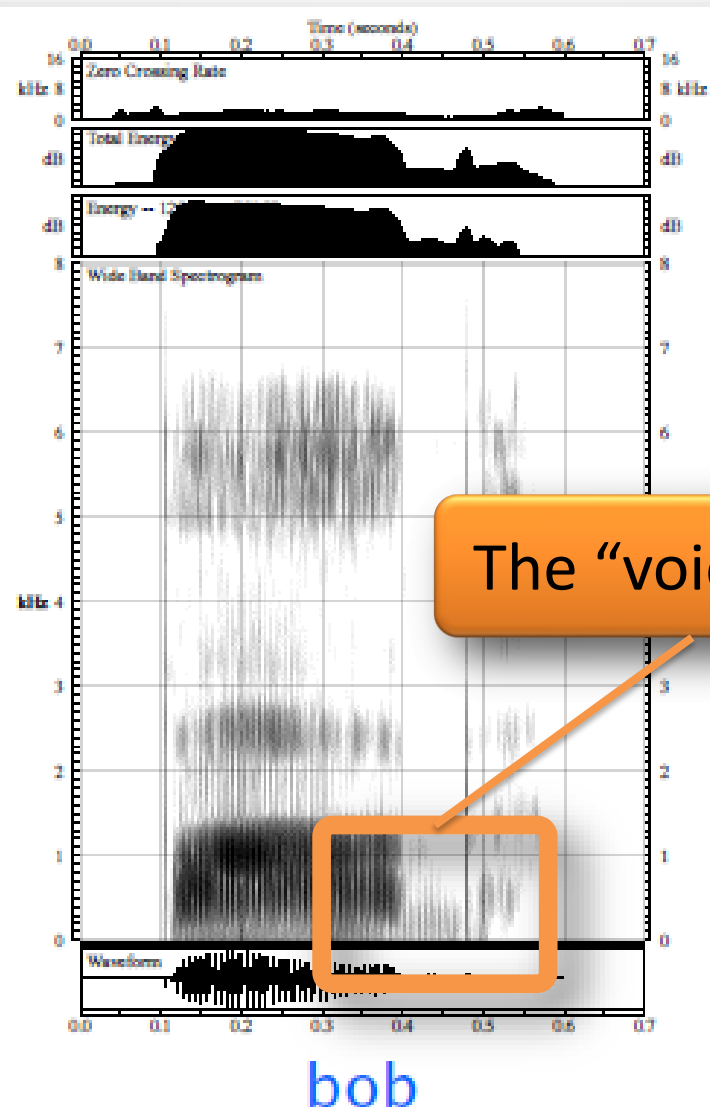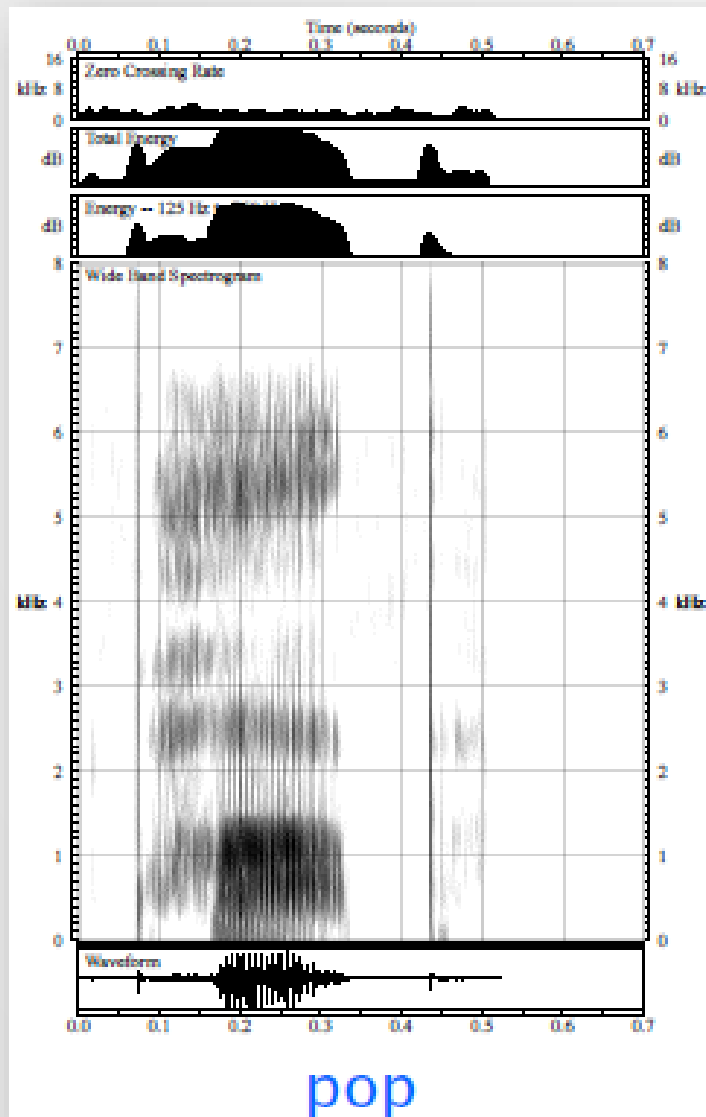
Nasals

- **Nasals** involve lowering the velum so that air passes through the **nasal cavity**.
- **Closures** in the oral cavity (at same positions as plosives) change the resonant characteristics of the nasal sonorant.


labial   alveolar   velar

[m]   [n]   [ng]
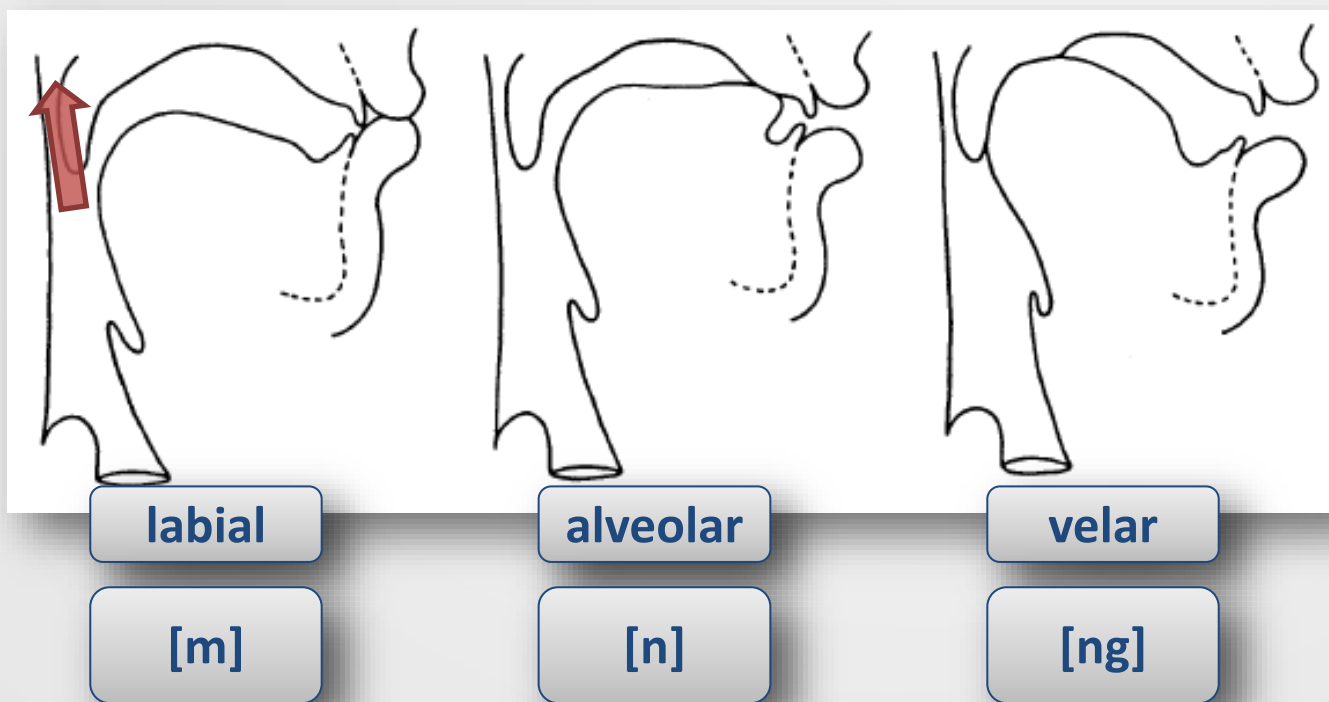
Formant transitions among nasals

simmer   sinner   singer

- Nasals often appear as two formants

UNIVERSITY OF TORONTO

Semivowels

- **Semivowels** act as consonants in syllables and involve constriction in the vocal tract, but there is **less turbulence**.
  - They also involve slower articulatory motion.
- **Laterals** involve airflow around the **sides** of the tongue.

/w/   /y/   /r/   /l/

UNIVERSITY OF TORONTO

Semivowels

we   ye   reed   lee
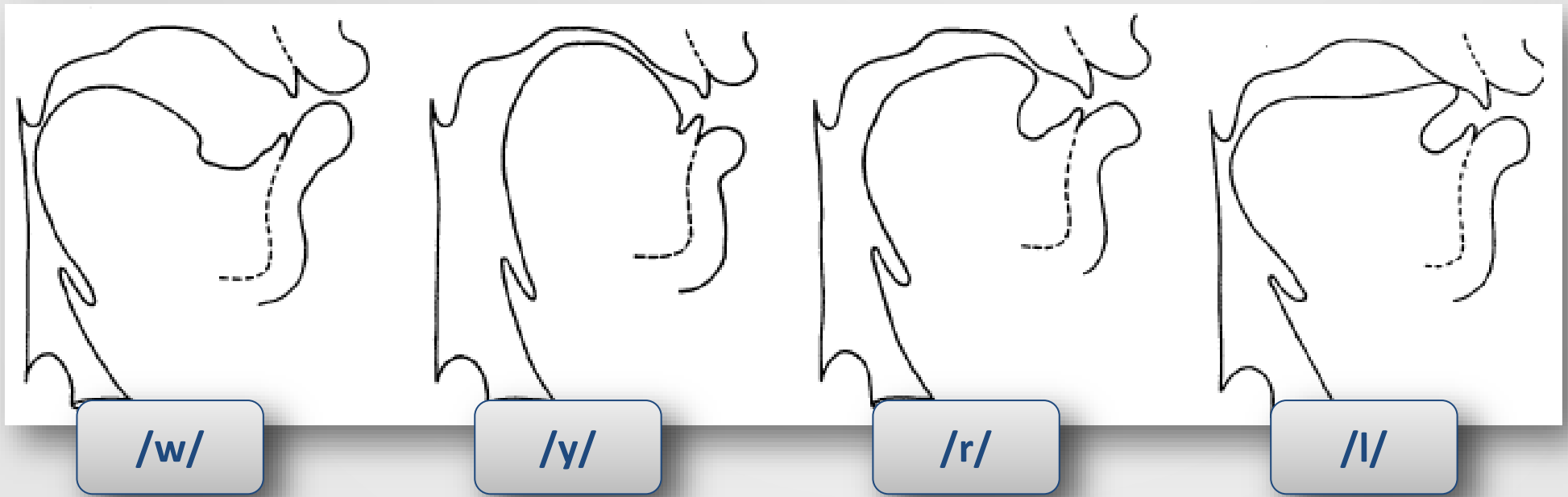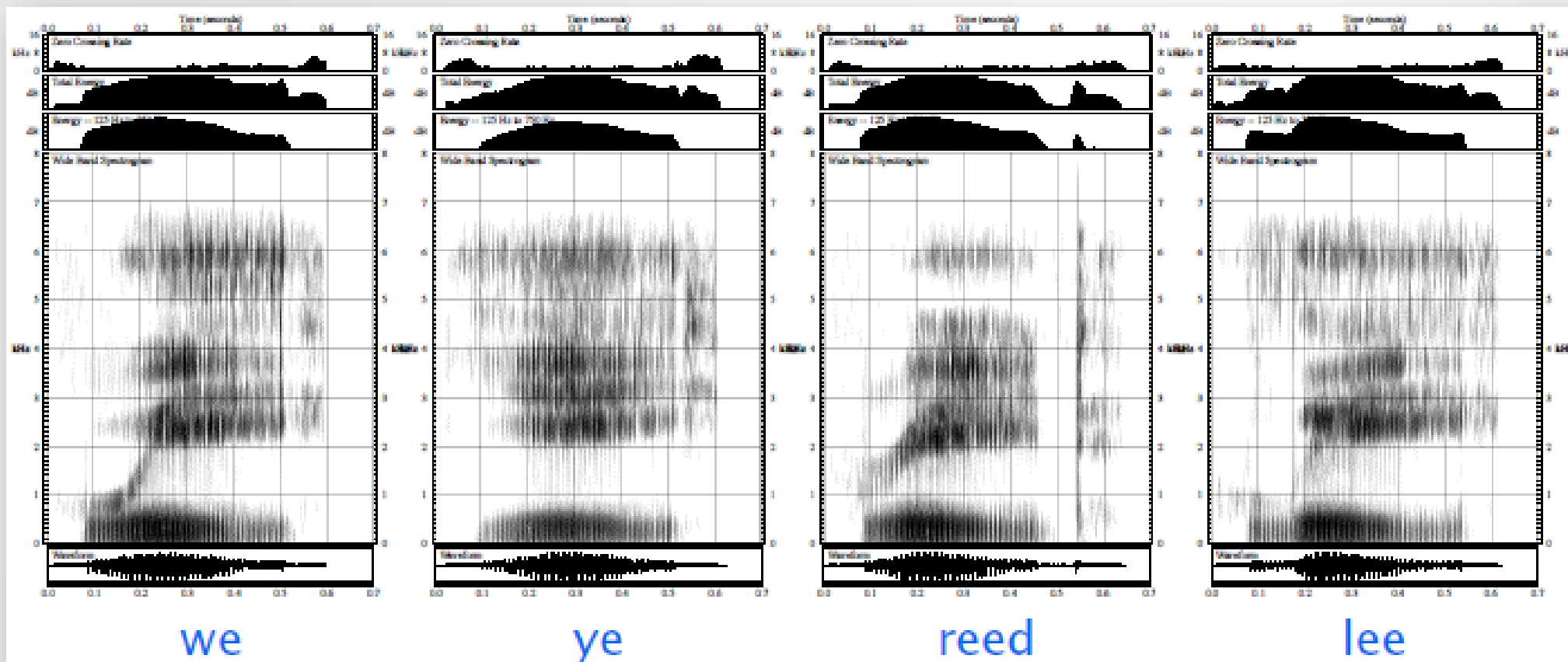
- Note the drastic formant transitions which are more typical of semivowels.

Affricates and aspirants

- There are two **affricates**: [jh] (voiced; e.g., *ju**dge***) and [ch] (unvoiced; e.g., *ch**ur**ch*).
  - These involve an **alveolar stop** followed by a **fricative**.
  - Voicing in [jh] is normally indicated by voice bars, as with plosives.

- There's only one **aspirant** in Canadian English: [h] (e.g., *h**at*)
  - This involves turbulence generated at the **glottis**,
  - In Canadian English, there is **no** constriction in the vocal tract.

UNIVERSITY OF TORONTO

Affricates and aspirants

each

huge

UNIVERSITY OF
TORONTO
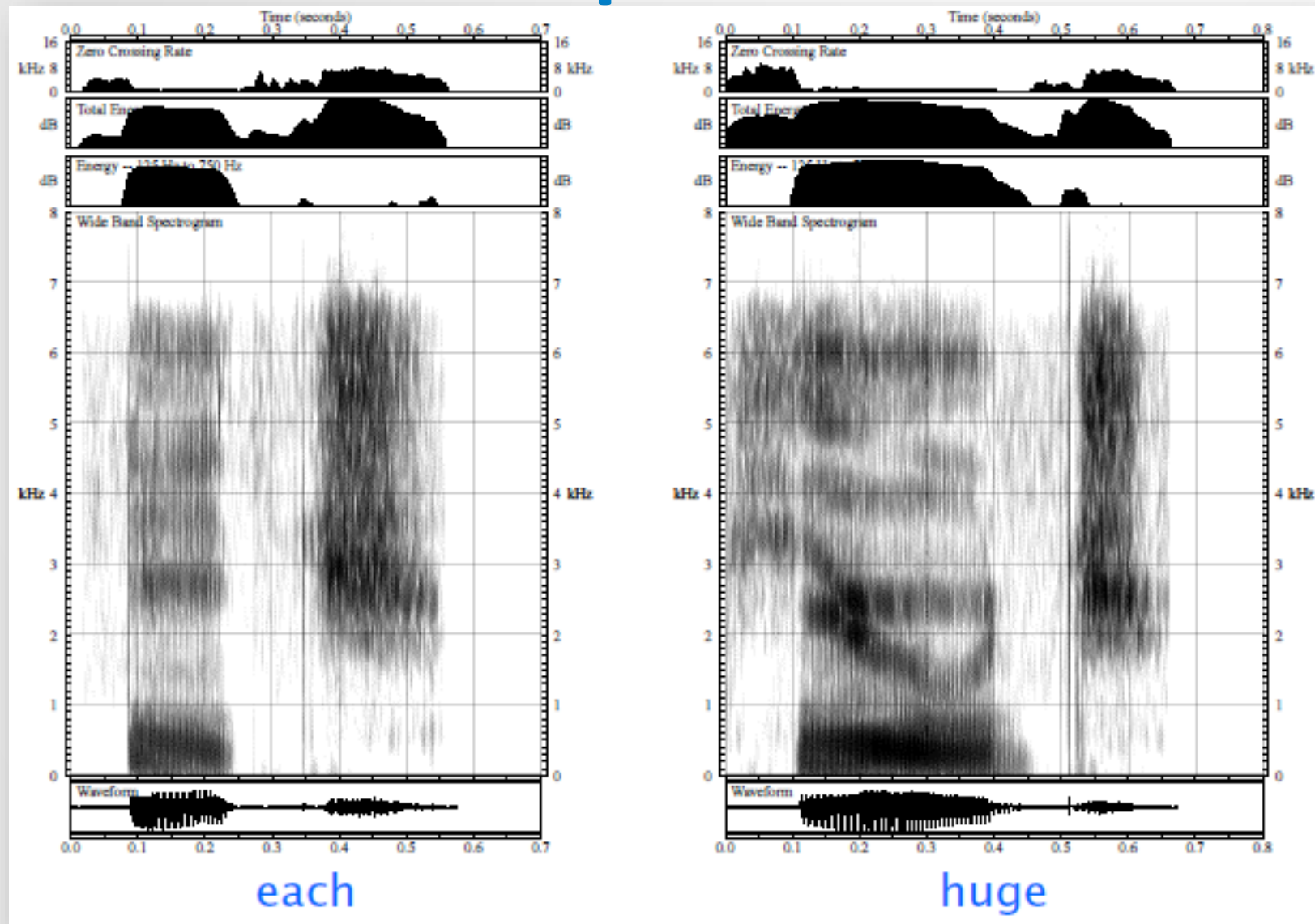
Other topics in phonetics

- The grouping of phones into **syllables**
  - Consisting of a vowel (**nucleus**), and optionally preceding (**onset**) and succeeding (**coda**) consonants
  - Only certain sequences are permissible in English
  - Syllables may be made more **prominent** via pitch, duration, or loudness
- The **prosody**, or intonation and rhythm, of an utterance
  - Prominence can also indicate phrase boundaries
  - Gradual F0 movement (**tune**) can indicate a question or statement
- These are especially important to **text-to-speech**

UNIVERSITY OF TORONTO