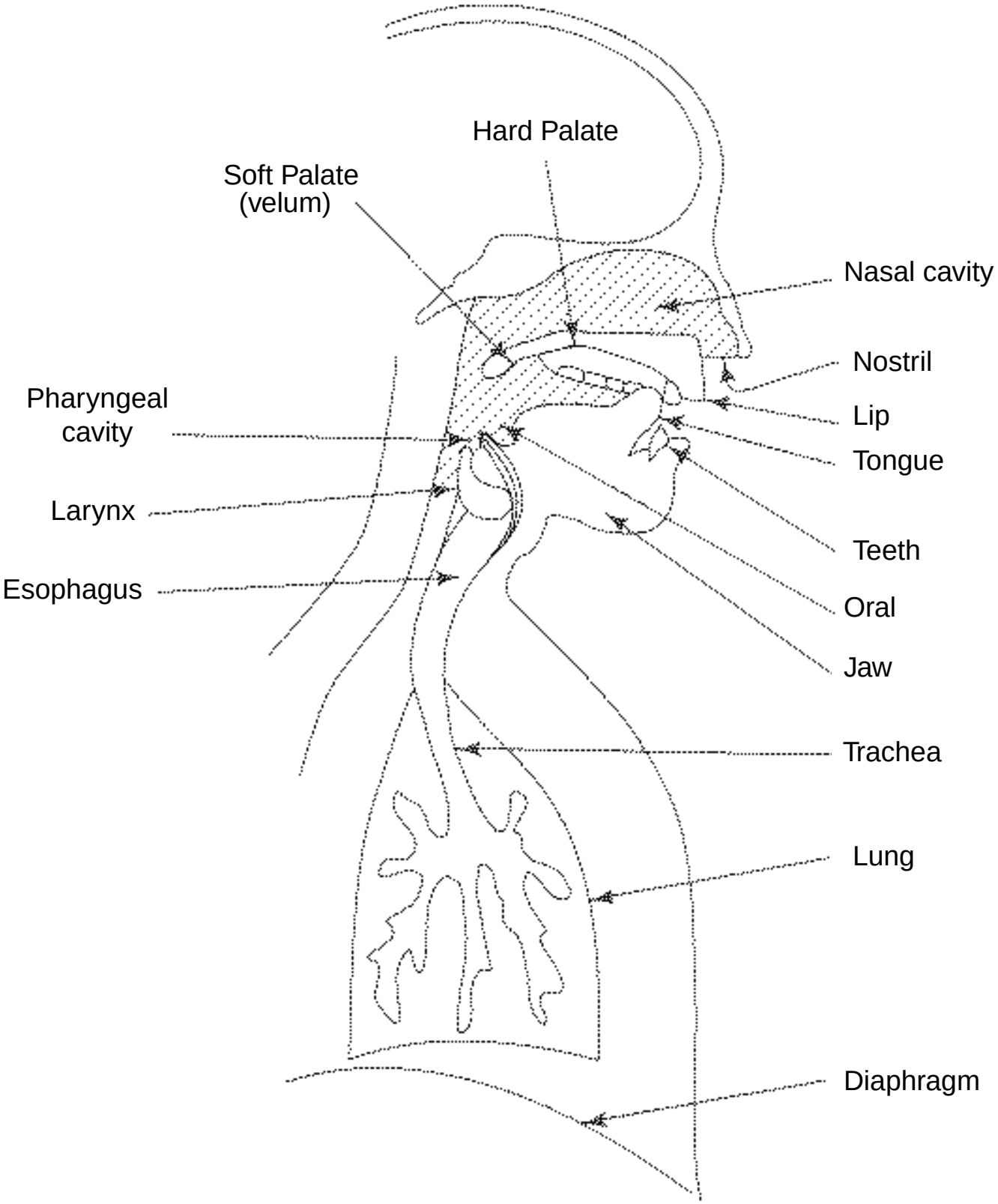
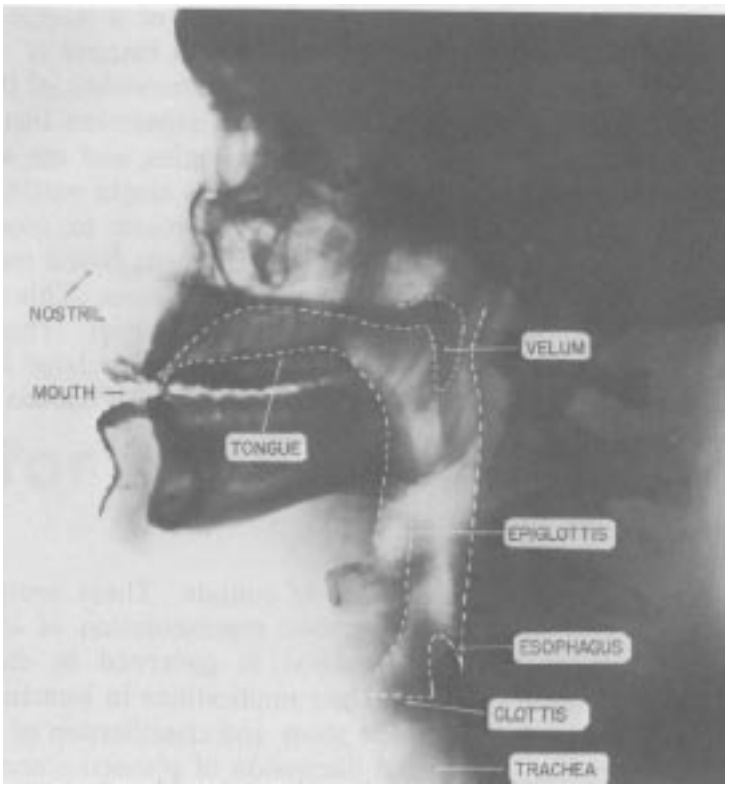
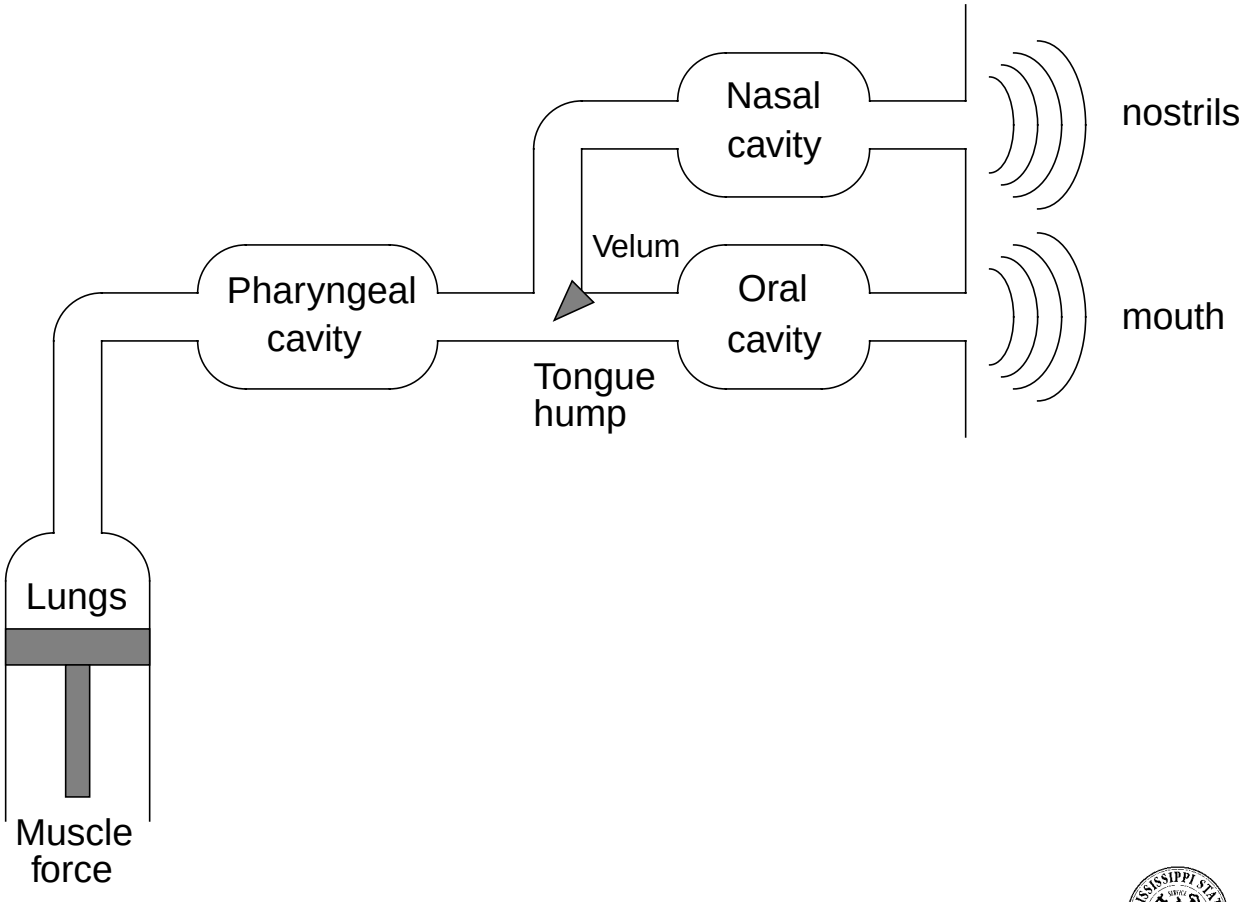
Speech Production Physiology

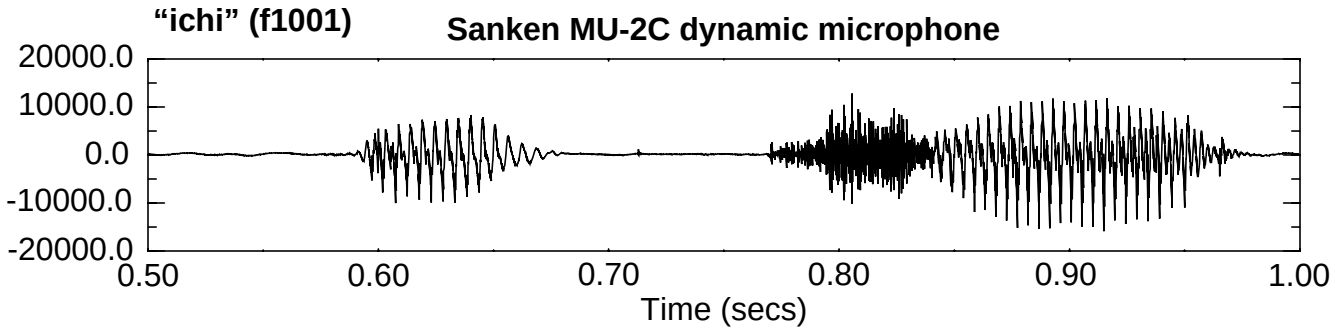


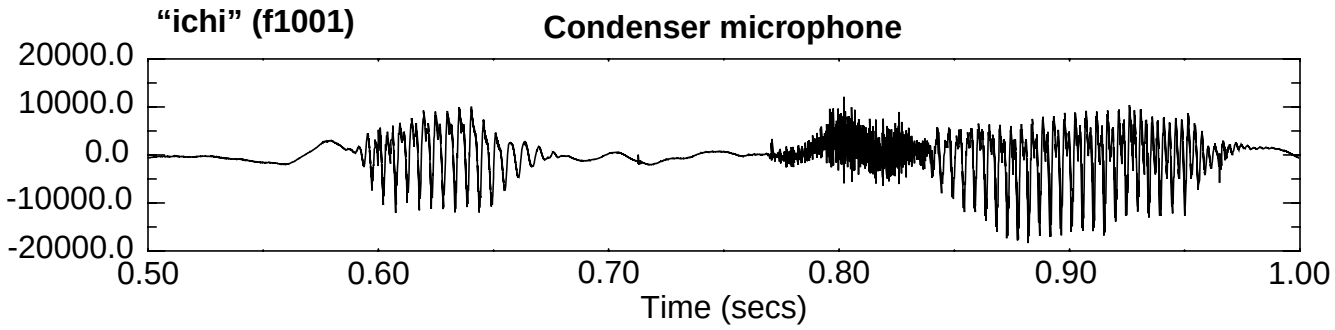
A Block Diagram of Human Speech Production

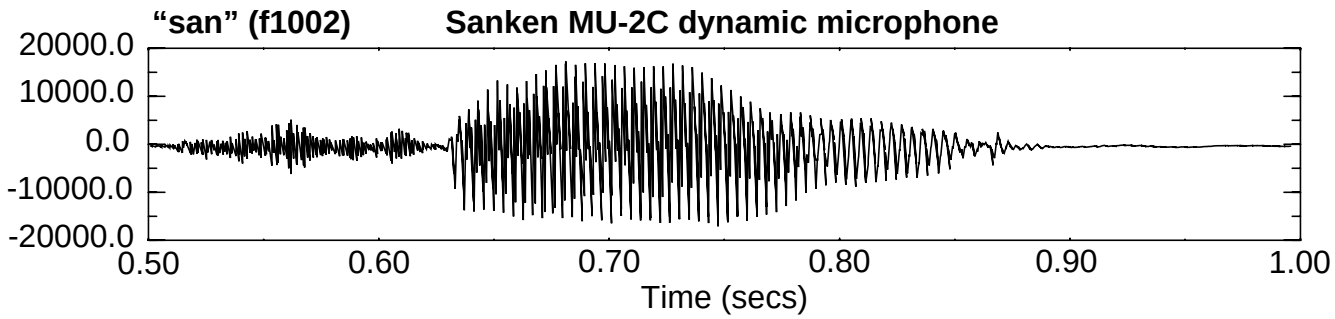


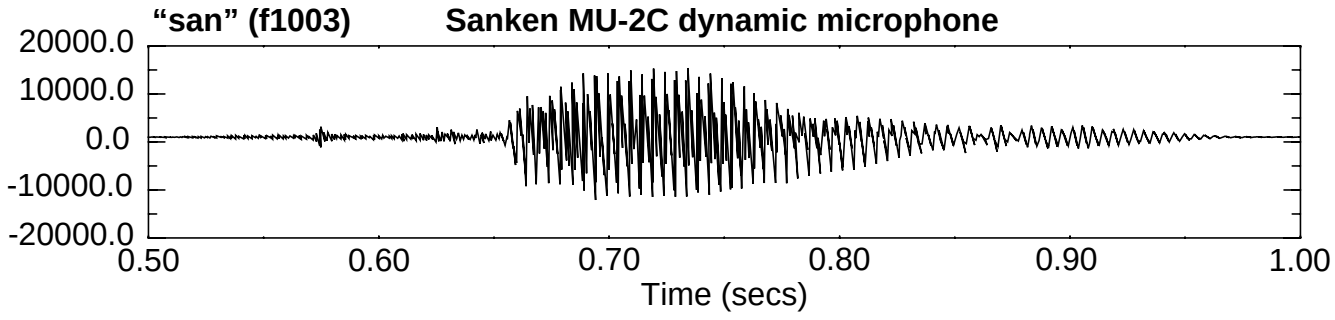


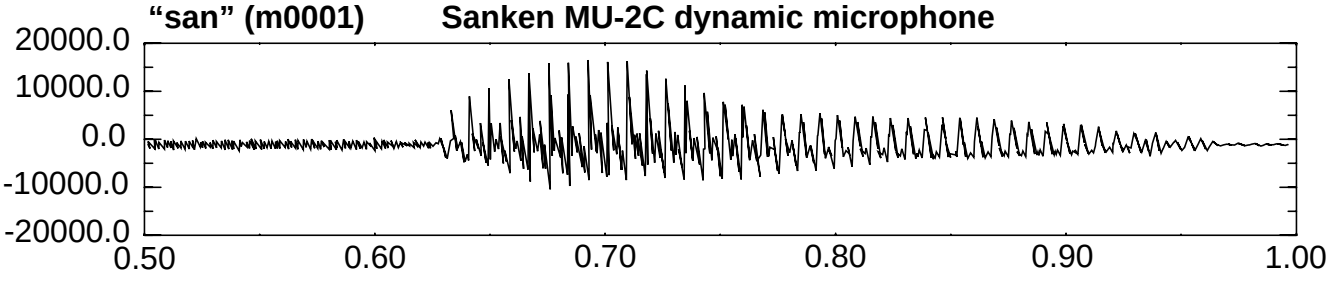
What Does A Speech Signal Look Like?





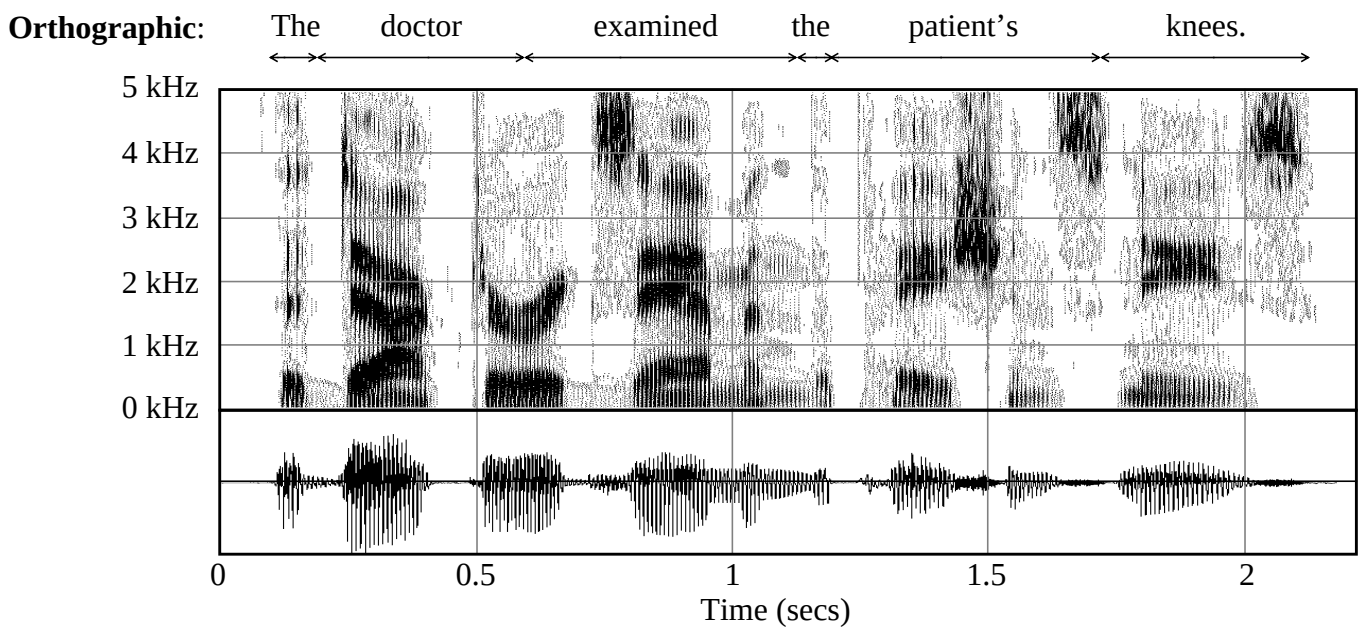




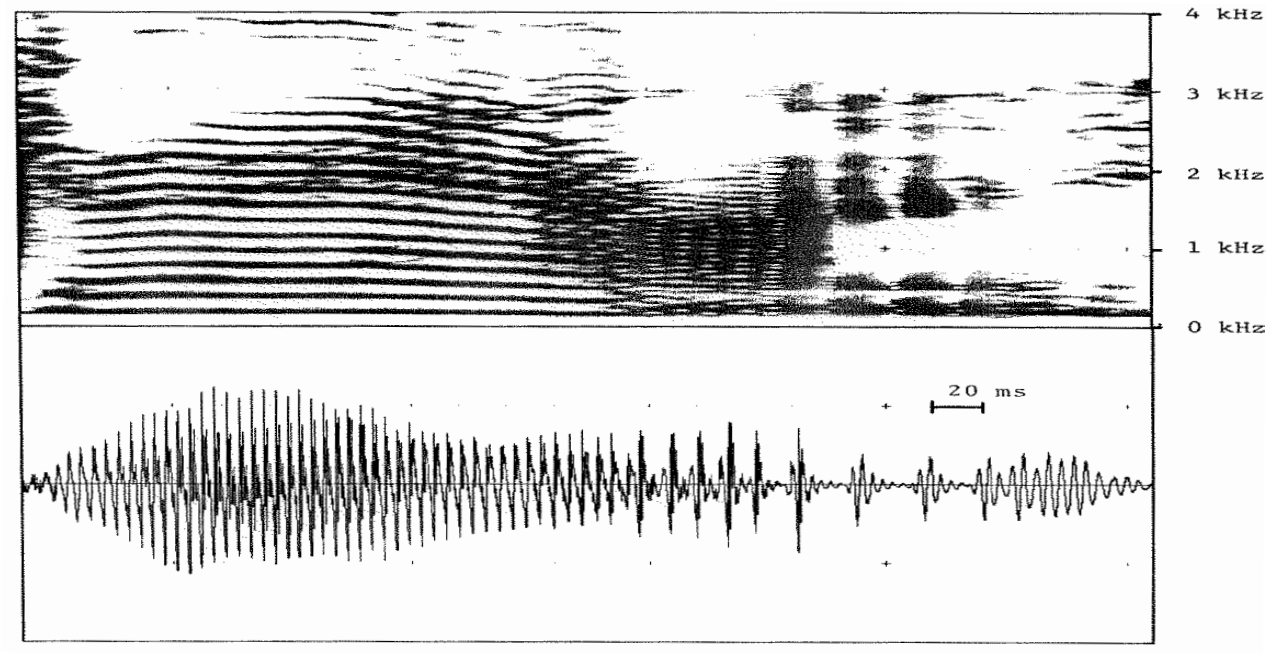


What Does A Speech Signal Spectrogram Look Like?

Standard wideband spectrogram ($f_s = 10\ kHz$, $T_w = 6\ ms$):



Narrowband Spectrogram ($f_s = 8\ kHz$, $T_w = 30\ ms$):



"Drown" (female)

Phonemics and Phonetics

Some simple definitions:

Phoneme:
- an ideal sound unit with a complete set of articulatory gestures.
- the basic theoretical unit for describing how speech conveys linguistic meaning.
- (For English, there are about 42 phonemes.)
- Types of phonemes: vowels, semivowels, dipthongs, and consonants.

Phonemics:
- the study of abstract units and their relationships in a language

Phone:
- the actual sounds that are produced in speaking (for example, “d” in letter pronounced “l e d er”).

Phonetics:
- the study of the actual sounds of the language

Allophones:
- the collection of all minor variants of a given sound (“t” in eight versus “t” in “top”)
- Monophones, Biphones, Triphones — sequences of one, two, and three phones. Most often used to describe acoustic models.

Three branches of phonetics:

- **Articulatory phonetics**: manner in which the speech sounds are produced by the articulators of the vocal system.
- **Acoustic phonetics**: sounds of speech through the analysis of the speech waveform and spectrum
- **Auditory phonetics**: studies the perceptual response to speech sounds as reflected in listener trials.

Issues:

- Broad phonemic transcriptions vs. narrow phonetic transcriptions

Phonemic and Phonetic Transcription - Standards

Major governing bodies for phonetic alphabets:

- International Phonetic Alphabet (**IPA**) — over 100 years of history
- **ARPAbet** — developed in the late 1970's to support ARPA research
- **TIMIT** — TI/MIT variant of ARPAbet used for the TIMIT corpus
- **Worldbet** — developed recently by Jim Hieronymous (AT&T) to deal with multiple languages within a single ASCII system

Example:

CONSONANTS

The Worldbet representation of each IPA symbol is written below it. IPA symbols in parentheses are rare phonemes, for which no machine-readable coding has yet been proposed. (In these cases a coding employing diacritics is proposed.)

	Bi-labial	Labio-dental	Dental	Alveolar	Post-alveolar	Retro-flex	Palatal	Velar	Uvular	Pharyn-geal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
	p b			t d		tr dr	c J	k g	q Q		?
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	(ɴ)		
	m	M		n		nr	n~	N	Nq		
Trill	ʙ			r					ʀ		
	B			r					R		
Tap or Flap				ɾ ɾ	ɾ	ɽ					
				t(d(	r(	rr					
Fricative	ɸ ʒ	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç (j)	x ɣ	χ ʁ	ħ ʕ	h (ɦ)
	F V	f v	T D	s z	S Z	sr zr	C jʌ	x G	X K	H !	h hv
Lateral fricative				(ɬ) (ɮ)							
				hl Zl							
Approxi-mant		(ʋ)		ɹ		ɻ	j	(ɱ)			
		V[		9		9r	j	4)			
Lateral approx.				l		ɭ	ʎ	(ʟ)			
				l		lr	L	Lg			
Ejective stop	p'			t'		ʈ'	c'	k'	q'		
	p>			t>		t> r	c>	k>	q>		
Implosive	ɓ			ɗ				ɠ ɠ			
	p<b<			t< d<			c< J<	k<g<	q<Q<		

VOWELS

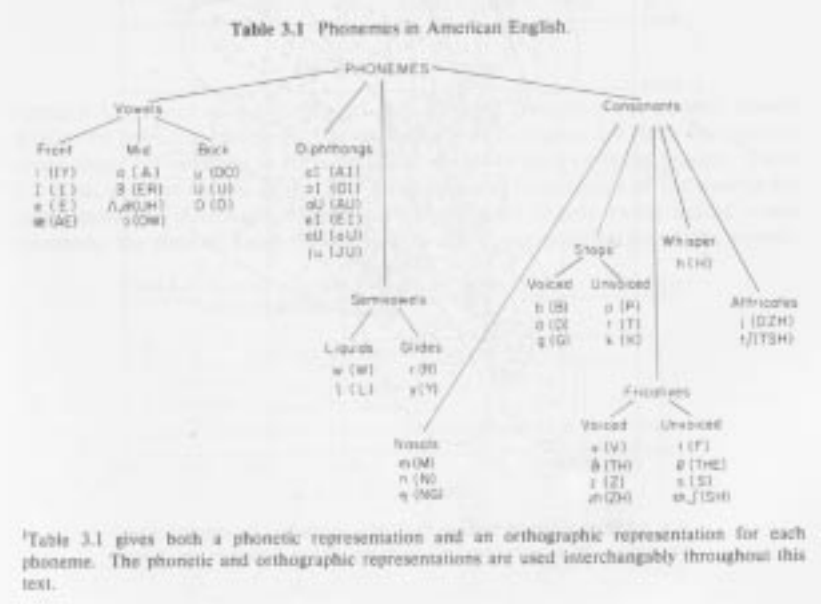
	Front		Central		Back	
Close	i y		i u		ɯ u	
	i y		ix ux		4 u	
		ɪ ʏ	ɪ†		ʊ	
		I Y	Ix		U	
Close-mid	e ø				ɤ o	
	e 7				2 o	
			ɔ o			
			ʎ ox			
Open-mid	ɛ œ				ʌ ɔ	
	E 8				Λ >	
	æ		ɐ			
	@		ax			
Open	a œ				ɑ ɒ	
	a 6				A 5	

†= iis not an approved IPA symbol, but it is in such common use that we have propose Ix as the most natural ASCII representation for a "centralized r".

TABLE 1: Worldbet Consonant and Vowel Symbols

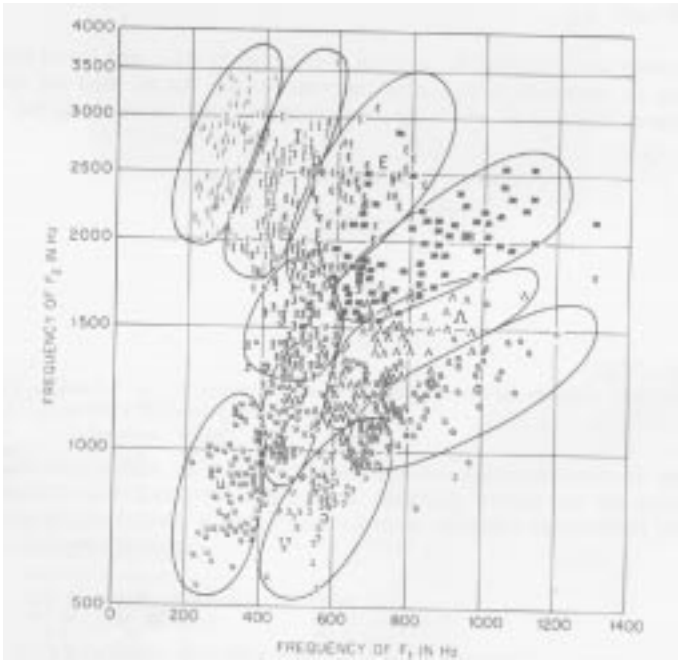
The Vowel Space

Each fundamental speech sound can be categorized according to the position of the articulators. This is known as the study of Acoustic Phonetics.



[1]Table 3.1 gives both a phonetic representation and an orthographic representation for each phoneme. The phonetic and orthographic representations are used interchangably throughout this text.

We can characterize a vowel sound by the locations of the first and second spectral resonances, known as a formant frequencies:



Some voiced sounds, such as diphthongs, are transitional sounds that move from one vowel location to another.

Average Formant Frequency Locations

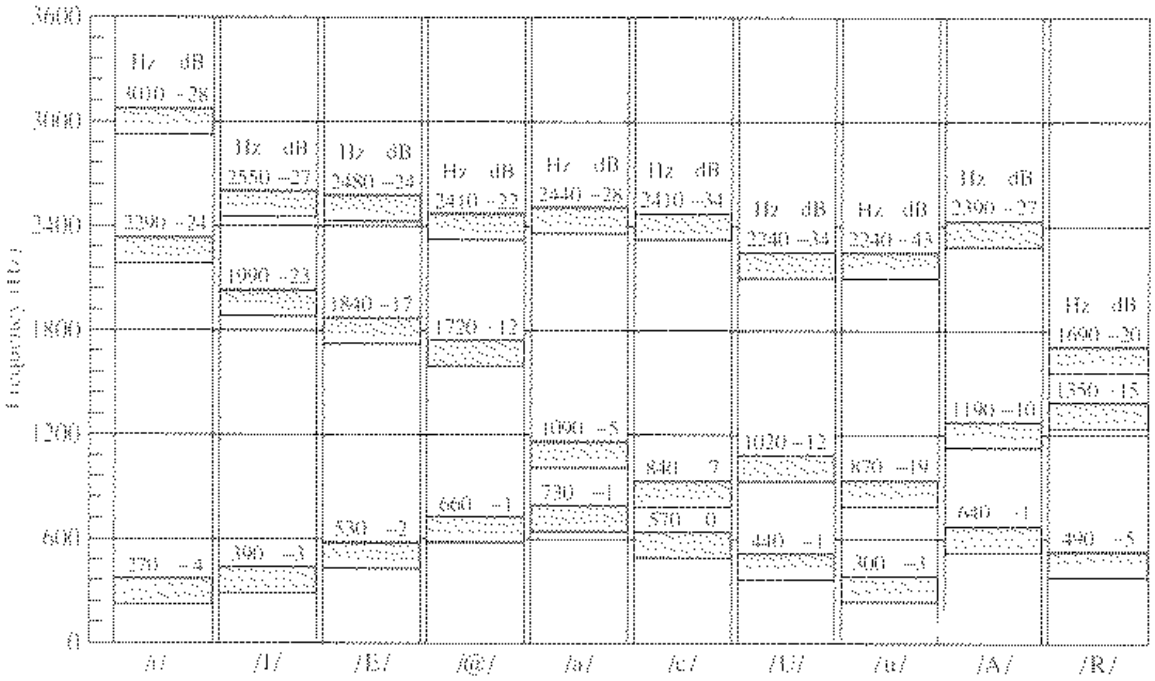
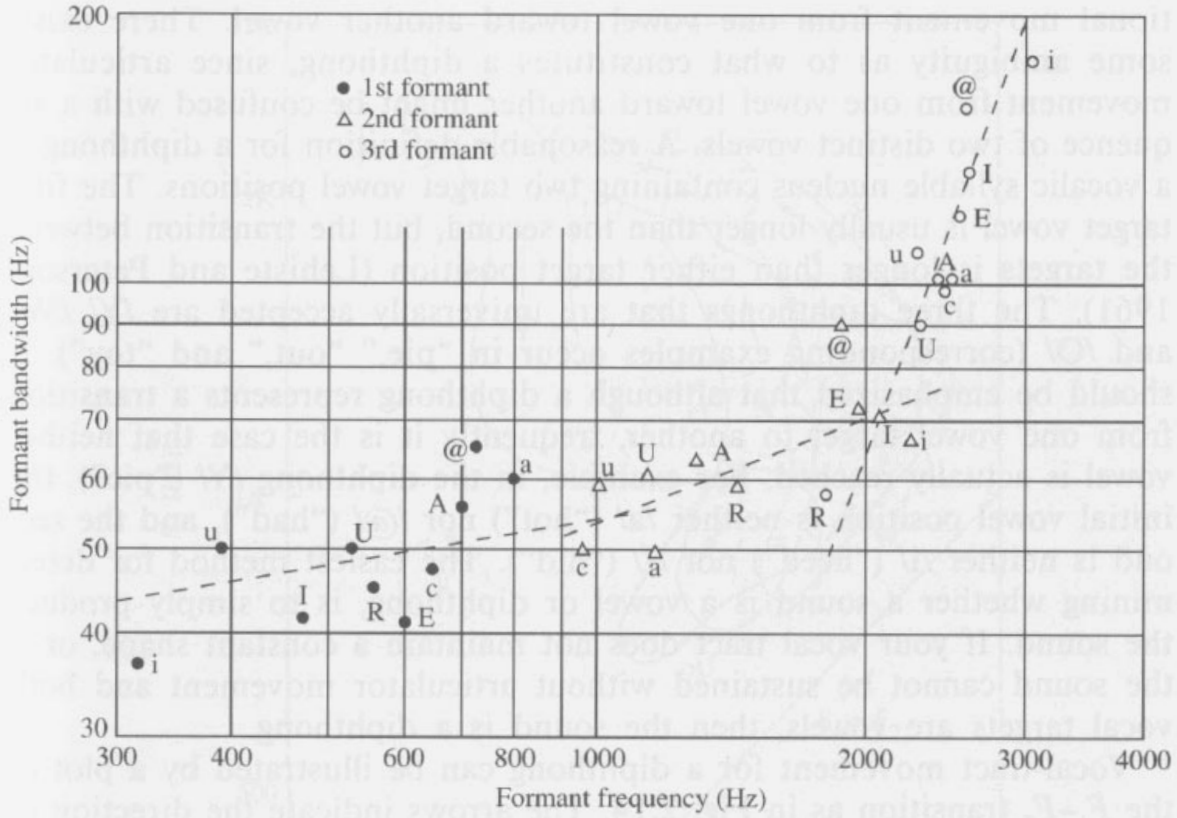


FIGURE 2.11. Average formant locations for vowels in American English (Peterson and Barney, 1952).

Rules Relating Formant Frequencies and Vocal-Tract Characteristics:

- Length Rule: The average frequencies of the vowel formants are inversely proportional to the length of the pharyngeal-oral tract
- F_1 Rule — Oral Constriction: The frequency of F_1 is lowered by any constriction in the front half of the oral section of the vocal tract.
- F_1 Rule — Pharyngeal Constriction: The frequency of F_1 is raised by constriction of the pharynx
- F_2 Rule — Back Tongue Constriction: The frequency of F_2 tends to be lowered by a back tongue constriction.
- F_2 Rule — Front Tongue Constriction: The frequency of F_2 is raised by a front tongue constriction.
- Lip-Rounding Rule: The frequencies of all formants are lowered by lip-rounding.

What About Bandwidth?



Vowels	F_1 Avg.	F_1 Extremes		F_2 Avg.	F_2 Extremes		F_3 Avg.	F_3 Extremes	
i	38	30	80	66	30	120	171	60	300
I	42	30	100	71	40	120	142	60	300
E	42	30	120	72	30	140	126	50	300
@	65	30	140	90	40	200	156	50	300
a	60	30	160	50	30	80	102	40	300
c	47	30	120	50	30	200	98	40	240
u	50	30	120	58	30	200	107	50	200
U	51	30	100	61	30	140	90	40	200
A	56	30	140	63	30	140	102	50	300
R	46	30	80	59	30	120	58	40	120
Avg.	49.7			64.0			115.2		

FIGURE 2.13. Average formant bandwidths for vowels in American English (Dunn, 1961).

When We Put This All Together:

We Have An Acoustic Theory of Speech Production

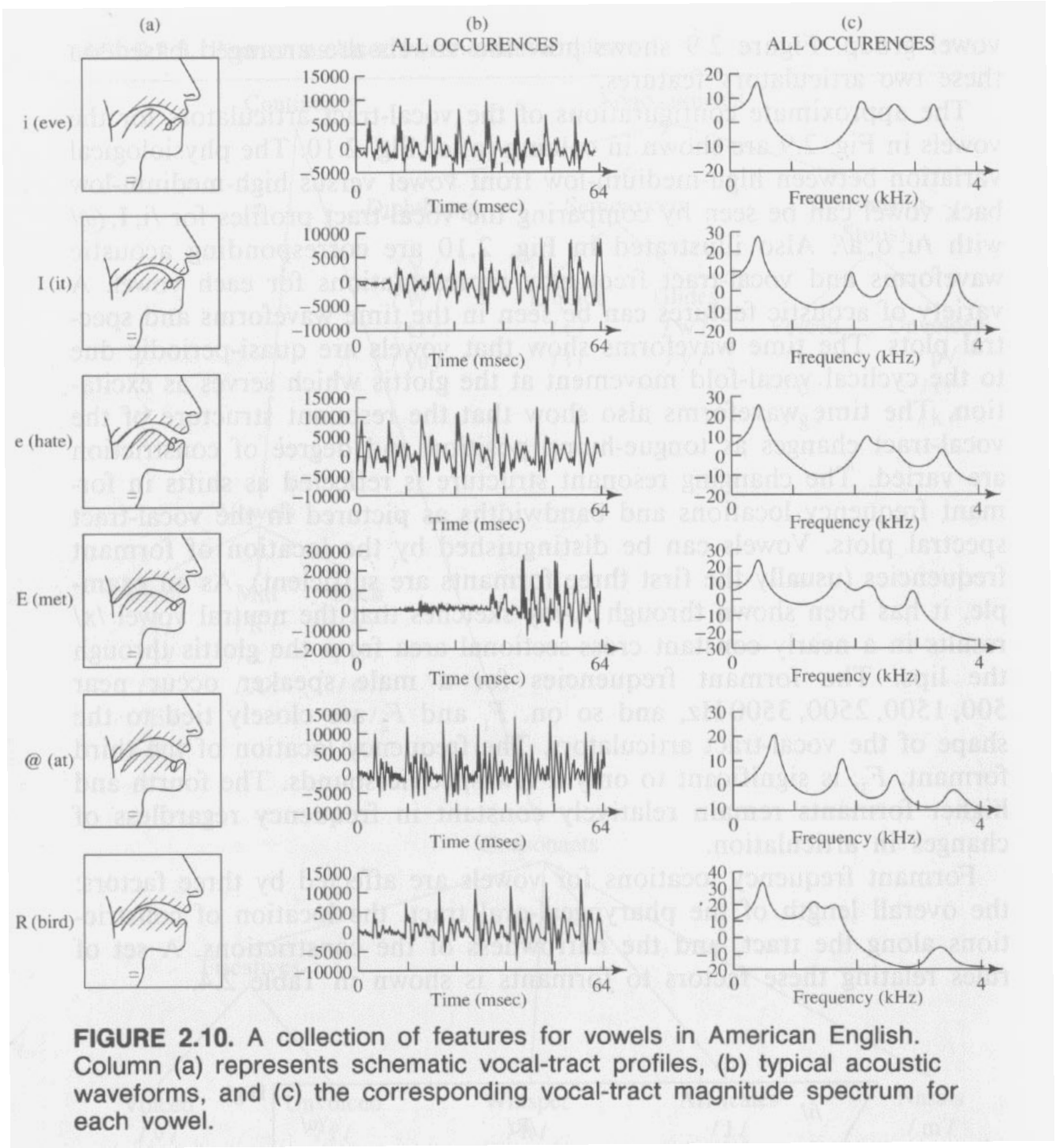


FIGURE 2.10. A collection of features for vowels in American English. Column (a) represents schematic vocal-tract profiles, (b) typical acoustic waveforms, and (c) the corresponding vocal-tract magnitude spectrum for each vowel.

Consonants Can Be Similarly Classified

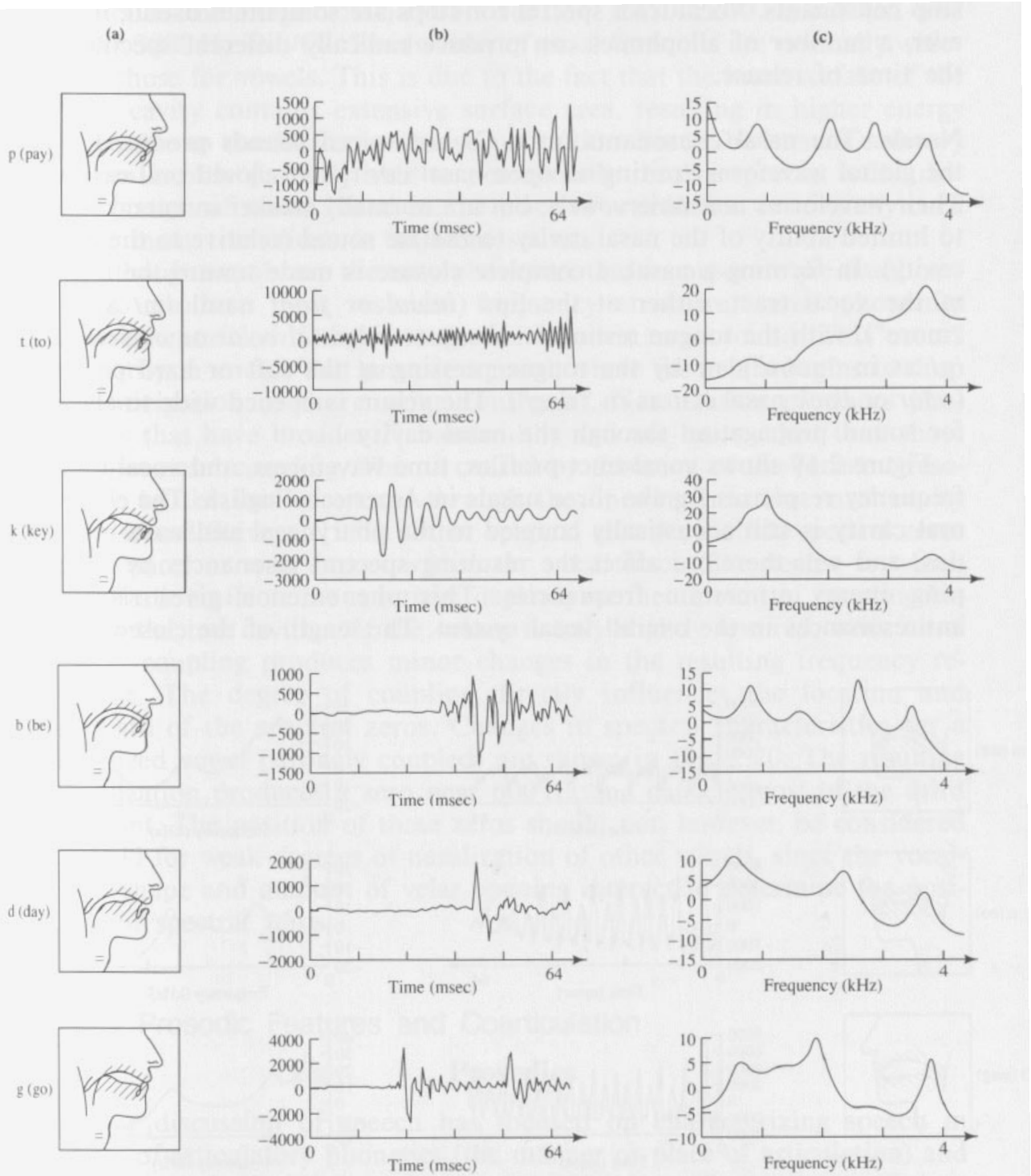


FIGURE 2.18. A collection of features for voiced and unvoiced stops in American English. Column (a) represents schematic vocal-tract profiles just prior to release, (b) typical acoustic waveforms, and (c) the corresponding vocal-tract magnitude spectra.