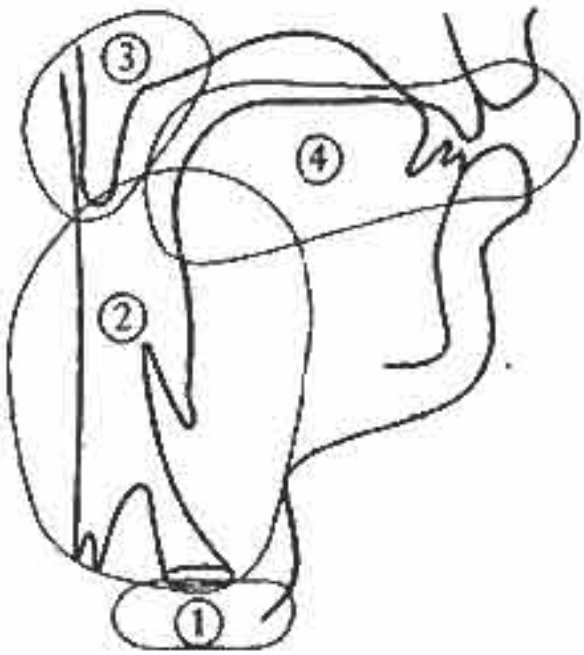
6.551J/HST714J ACOUSTICS OF SPEECH AND HEARING

Kenneth N Stevens

Lecture 20 *11/18/04*

Overview of the vocal tract, listing the various articulators involved in the production of speech sounds



Midsagittal section of the vocal tract showing four regions

1. The vocal folds. Their stiffness can be adjusted.
2. The pharyngeal region, including the glottis (space between the vocal folds)
3. The soft palate. Opening of velopharyngeal port can be adjusted (for nasal consonants and vowels)
4. Oral region. Tongue-body position, tongue blade constriction, and lip opening can be manipulated.

Not shown is the respiratory system (below item 1)

Classification of speech sounds

Table 5.1 List of articulator-free features and their values for different classes of segments

Feature	Vowels	Glides	Consonants	n,l	t,b	s,z	θ,ð
Vocalic	+	−	−	−	−	−	−
Consonantal	−	−	+	+	+	+	+
Continuant				−	−	+	+
Sonorant				+	−	−	−
Strident						+	−

Note: The segment labels at the top of the last four columns are examples of the segments with the given feature values.

From *Acoustic Phonetics* by Kenneth Stevens. (c) 1998, MIT Press. Used with permission.

Table 5.3 Feature values for some vowels and glides in English

Feature	i	ɪ	e	ɛ	æ	ɑ	ɔ	o	ʌ	ʊ	u	ɚ	w	j	h	l
Vocalic	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+
Consonantal	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Round	−	−	−	−	−	−	+	+	−	+	+	−	+	−		−
Anterior												−				+
Distributed												−				−
Lateral												−				+
High	+	+	−	−	−	−	−	−	−	+	+	−	+	+		−
Low	−	−	−	−	+	+	+	−	−	−	−	−	−	−		−
Back	−	−	−	−	−	+	+	+	+	+	+	+	+	−		+
Nasal																
Advanced tongue root	+	−	+	−	−	−	−	+	−	−	+	−	+	+		
Constricted tongue root	−	−	−	−	−	+	+	−	−	−	−	−	−	−		
Spread glottis															+	
Constricted glottis															−	
Stiff vocal folds																
Slack vocal folds																

Table 5.4 Feature values for some consonants in English

Feature	p	t	k	b	d	g	m	n	ŋ	f	θ	s	ʃ	v	ð	z	ʒ	ʤ	ʧ	j	l	r
Vocalic	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Consonantal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Continuant	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	−	+
Sonorant	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	+	+
Strident										+	−	+	+	+	−	+	+	+	+			
Lips	+			+			+			+				+								
Tongue blade		+			+			+			+	+	+		+	+	+	+	+	+	+	+
Tongue body			+			+			+													
Round	−			−			−			−				−								
Anterior		+			+			+			+	+	−		+	+	−	−	−	+	−	
Distributed		−			−			−			−	−	+		−	−	+	+	+	−	−	
Lateral							−	−	−												+	−
High			+			+			+												−	−
Low			−			−			−												−	−
Back			+			+			+												+	+
Nasal							+	+	+												−	−
Advanced tongue root																						
Constricted tongue root																						
Spread glottis	+	+	+	−	−	−				+	+	+	+	−	−	−	−	−	+	−		
Constricted glottis	−	−	−	−	−	−				−	−	−	−	−	−	−	−	−	−	−		
Stiff vocal folds	+	+	+	−	−	−				+	+	+	+	−	−	−	−	−	+	−		
Slack vocal folds	−	−	−	+	+	+				−	−	−	−	+	+	+	+	+	−	+		

From *Acoustic Phonetics* by Kenneth Stevens. (c) 1998, MIT Press. Used with permission.

Some minimal pairs of words, i.e., pairs in which one feature of one segment is changed.

pat	bat	[stiff vocal folds] or "voicing" of initial consonant
bat	mat	the feature [nasal]
pat	fat	the feature [continuant]
fat	vat	the feature [stiff vocal folds]
feed	fade	the feature [high] for the vowel
pat	pet	the feature [low] for the vowel
thin	sin	the feature [strident]
thin	thing	change from [tongue blade] to [tongue body]

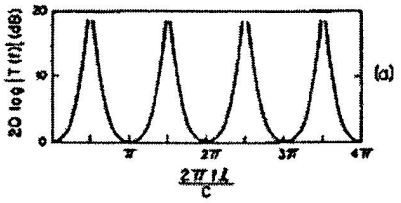
Distinction between vowels and consonants

The constriction in the vocal tract is narrower for consonants than for vowels. For vowels the constriction should not be so narrow that pressure builds up behind the constriction. During a vowel the acoustic amplitude at low frequencies reaches a peak. During a consonant the amplitude is lower.

Vowel features

Vowels in English are distinguished by four features: back, high, low, advanced tongue root (ATR). Vowels with +ATR are sometimes called tense vowels as opposed to lax vowels, e.g., /i/ is tense, /ɪ/ is lax.

lossless



Transfer function for vowel with uniform cross-sectional area

with loss

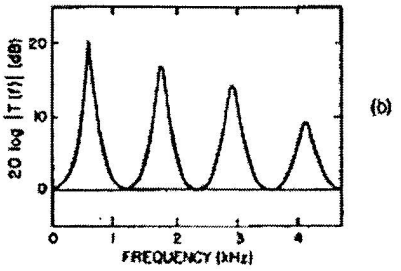


Figure 3.31 (a) Plot of magnitude of transfer function $T(f) = U_L/U_1$, expressed in decibels for an ideal uniform, lossless acoustic tube, shown in figure 3.8. (b) Magnitude of transfer function $T(f)$ for an ideal uniform tube of length 15 cm with losses similar to those occurring in the vocal tract

From *Acoustic Phonetics* by Kenneth Stevens. (c) 1998, MIT Press. Used with permission.

Some vowel spectra

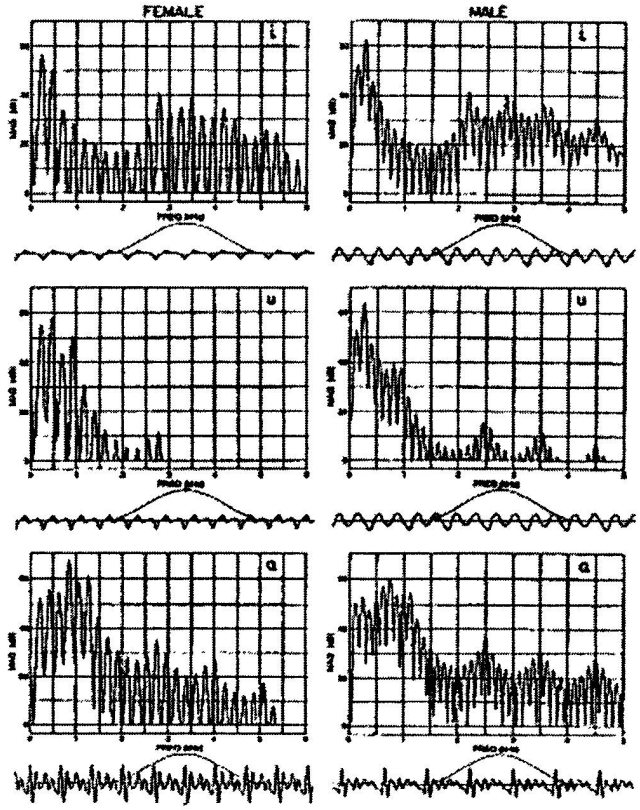


Figure 6.4 Spectra of the two high vowels /i/ and /u/ (top and middle panels) and the low vowel /a/ (bottom panel), with formant frequencies and fundamental frequencies appropriate for an adult female speaker (left column) and an adult male speaker (right column). The waveform is shown below each panel. The window used to calculate the spectrum is shown on each waveform. The vowels were synthesized using a Klatt synthesizer (Klatt and Klatt, 1990).

From *Acoustic Phonetics* by Kenneth Stevens. (c) 1998, MIT Press. Used with permission.

Vocal tract shapes for vowels

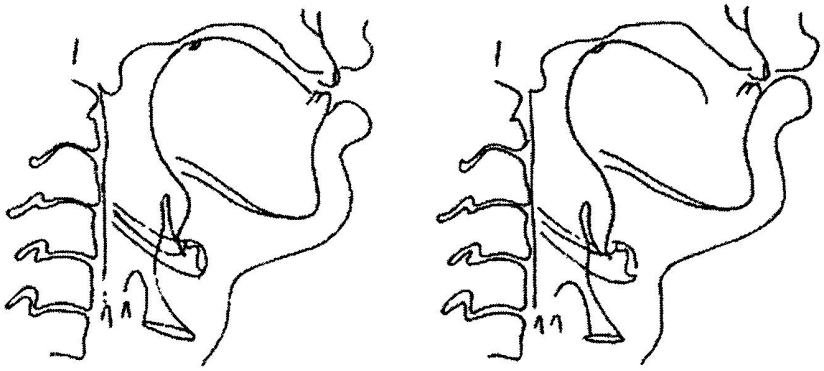


Figure 6.2 Midsagittal vocal tract configurations for the high vowels /i/ (left) and /u/ (right). Adult male speaker of English. (From Perkell, 1969.)

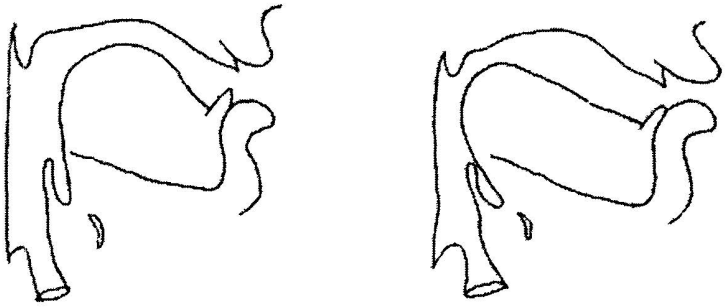


Figure 6.7 Midsagittal vocal tract configurations for the non-low, non-high vowels /e/ (left) and /o/ (right). Adult male speaker of French. (Adapted from Bothorel et al., 1986.)

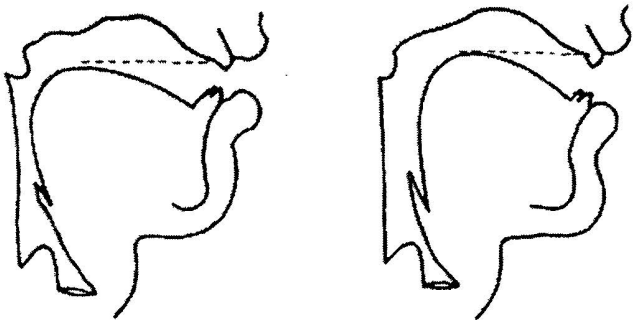


Figure 6.6 Midsagittal vocal tract configurations for the low vowels /ɑ/ (left) and /æ/ (right). The horizontal lines show the approximate location of the plane of the surfaces of the upper teeth. Adult male speaker of English. (From Perkell, 1969.)

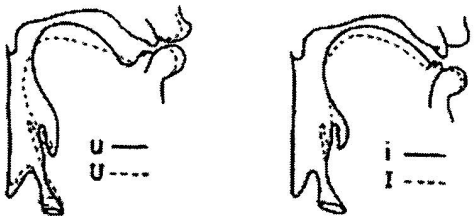
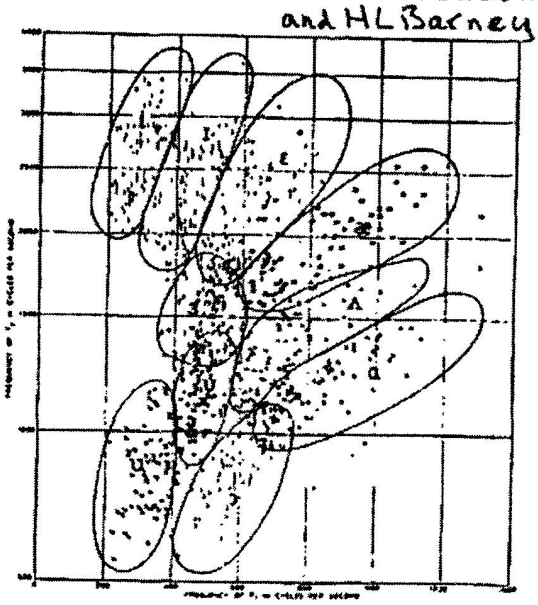


Figure 6.23 Comparing midsagittal vocal tract configurations for tense and lax high vowels for a speaker of American English. (Data from Perkell, 1969)

From *Acoustic Phonetics* by Kenneth Stevens. (c) 1998, MIT Press. Used with permission.

182 G. E. PETERSON and H L Barney

JASA 1952



Courtesy of the *Journal of the Acoustical Society of America.* Used with permission.

FIG. 8. Frequency of second formant *versus* frequency of first formant for ten vowels by 76 speakers.

TABLE II. Averages of fundamental and formant frequencies and formant amplitudes of vowels by 76 speakers.

		i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɝ
Fundamental frequencies (cps)	*M*	136	135	130	127	124	129	137	141	130	133
	W	235	232	223	210	212	216	232	231	221	218
	Ch	272	269	260	251	256	263	276	274	261	261
Formant frequencies (cps)											
F_1	*M*	270	390	530	660	730	570	440	300	640	490
	W	310	430	610	860	850	590	470	370	760	500
	Ch	370	530	690	1010	1030	680	560	430	850	560
F_2	*M*	2290	1990	1840	1720	1090	840	1020	870	1190	1350
	W	2790	2480	2330	2050	1220	920	1160	950	1400	1640
	Ch	3200	2730	2610	2320	1370	1060	1410	1170	1590	1820
F_3	*M*	3010	2550	2480	2410	2440	2410	2240	2240	2390	1690
	W	3310	3070	2990	2850	2810	2710	2680	2670	2780	1960
	Ch	3730	3600	3570	3320	3170	3180	3310	3260	3360	2160
Formant amplitudes (db)	L_1	−4	−3	−2	−1	−1	0	−1	−3	−1	−5
	L_2	−24	−23	−17	−12	−5	−7	−12	−19	−10	−15
	L_3	−28	−27	−24	−22	−28	−34	−34	−43	−27	−20

Courtesy of the *Journal of the Acoustical Society of America*. Used with permission.

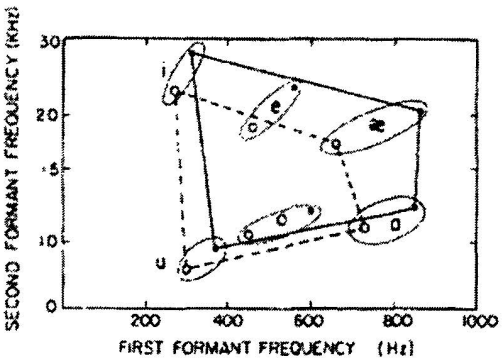


Figure 6.17 Plots of *F2* vs. *F1* for several vowels of American English. Open circles (joined by dashed lines) are data for adult male speakers and filled circles (solid lines) are for adult female speakers. The data for the vowels /i æ ɑ u/ are averages from Peterson and Barney (1952). Data for /ɛ ɔ/ are averages for two male and two female speakers.

From *Acoustic Phonetics* by Kenneth Stevens. (c) 1998, MIT Press. Used withpermission.

Source filter principle

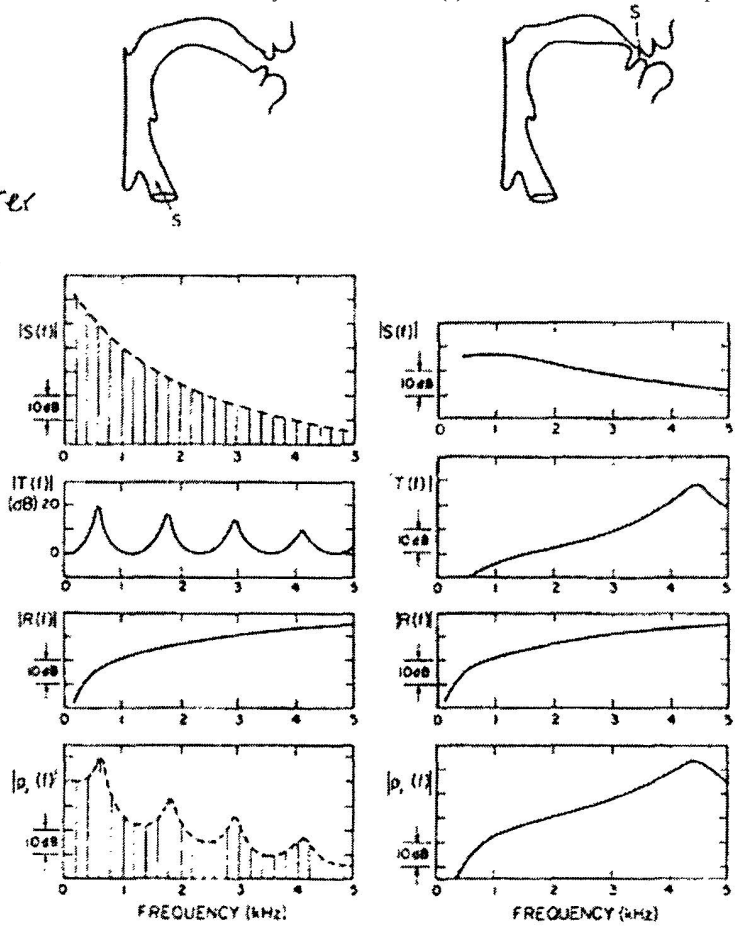
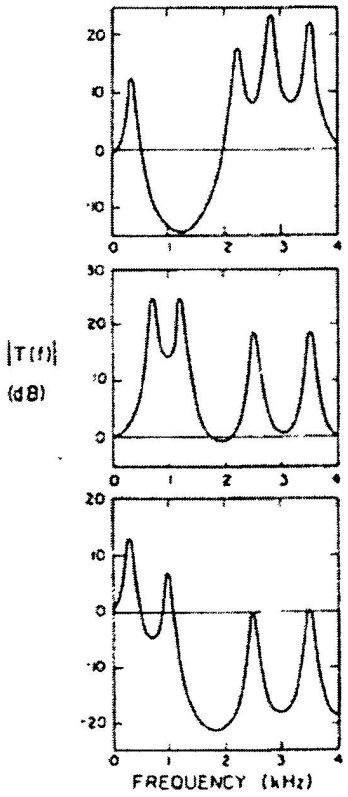


Figure 3.1 Sketches indicating components of the output spectrum $|p_l(f)|$, for a vowel and a fricative consonant. The output spectrum is the product of a source spectrum $S(f)$, a transfer function $T(f)$, and a radiation characteristic $R(f)$. The source spectra are similar to those derived in figures 2.10 and 2.33 in chapter 2. For the periodic source, $S(f)$ represents the amplitudes of spectral components; for the noise source, $S(f)$ is amplitude in a specified bandwidth. See text

From *Acoustic Phonetics* by Kenneth Stevens. (c) 1998, MIT Press. Used withpermission.

Explaining relative amplitudes of formant prominences



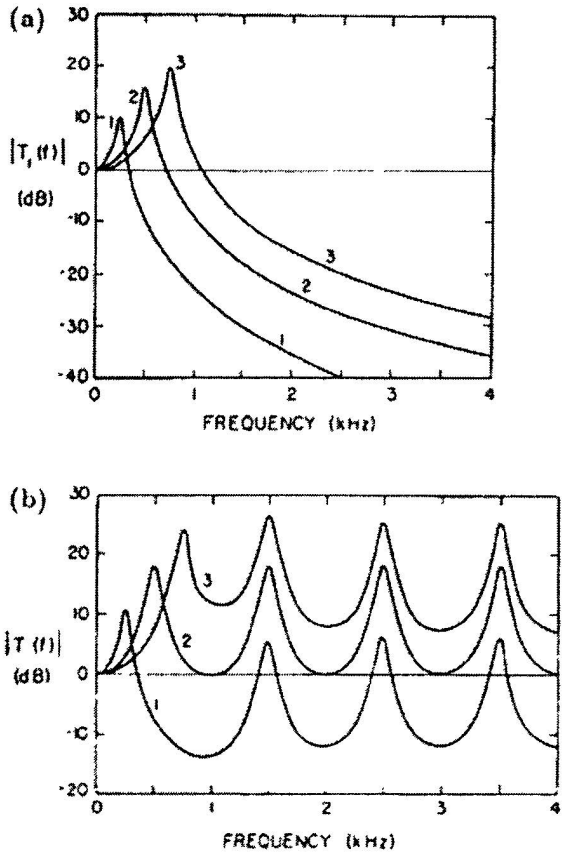


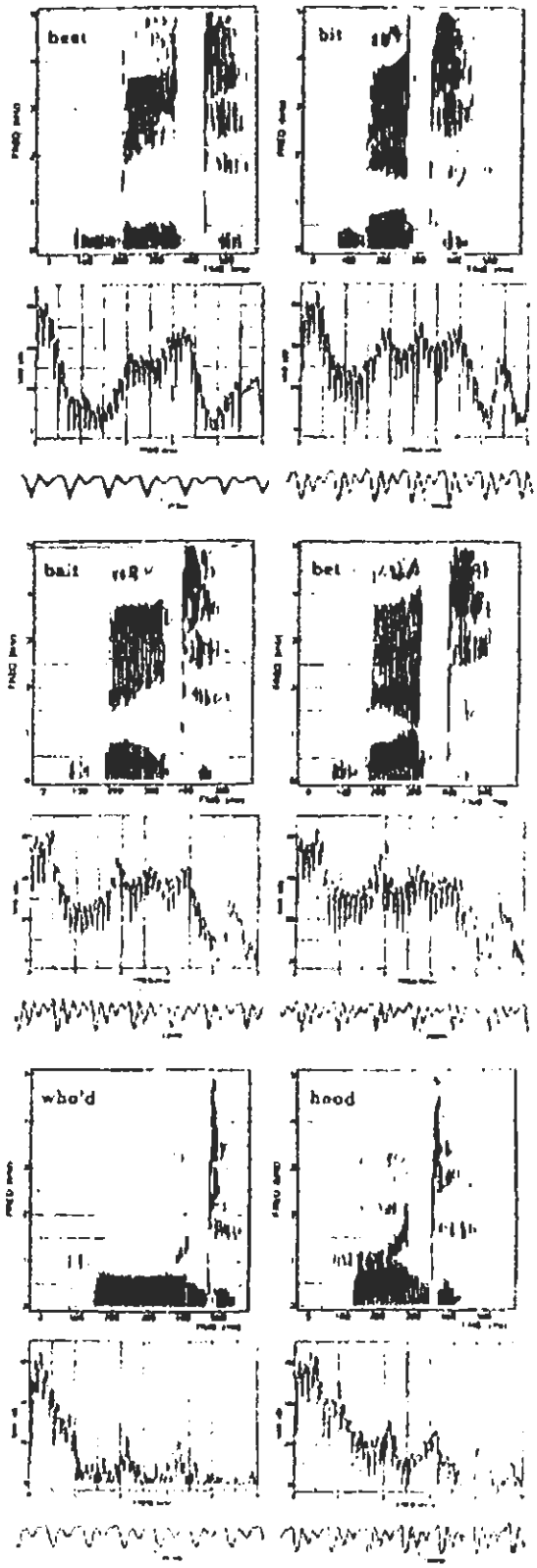
Figure 3.4 (a) The component of the vocal tract transfer function (in decibels) corresponding to the first formant for three different values of $F1$. Note the change in amplitude of the peak and the shift in level at higher frequencies. (b) The effect of a change in $F1$ on the overall transfer function, assuming formants above $F1$ remain fixed. The labels 1, 2, and 3 identify low, medium, and high values of $F1$.

Figure 3.5 Computed transfer functions for three different configurations of formant frequencies, illustrating changes in relative amplitudes of peaks and valleys in the transfer function. Bandwidths of all resonances are fixed at 80 Hz.

Spectrograms

Spectra

Waveforms



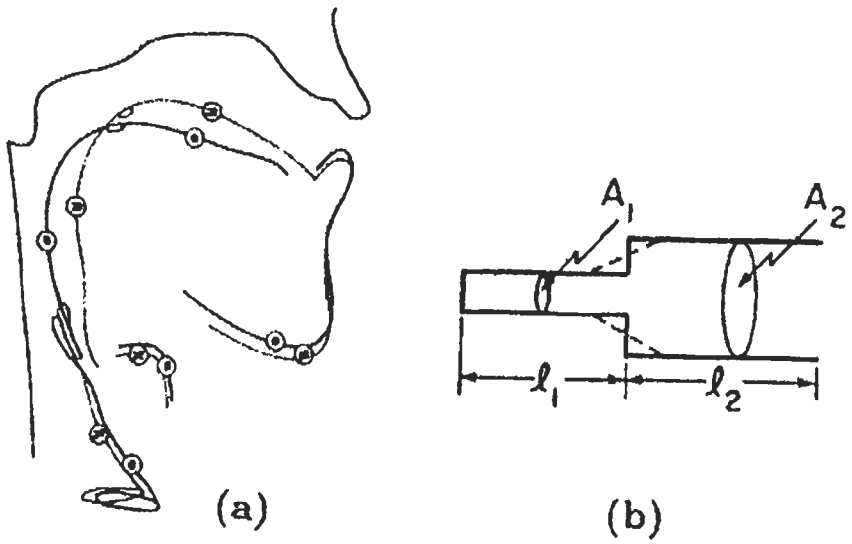


Figure 6.8 (a) Superimposed midsagittal configurations for the low vowels /æ/ and /ɑ/. (From Perkell, 1971.) (b) Model of low vowel vocal tract shape as a concatenation of two tubes. The dashed line indicates a tapered transition between the tubes.

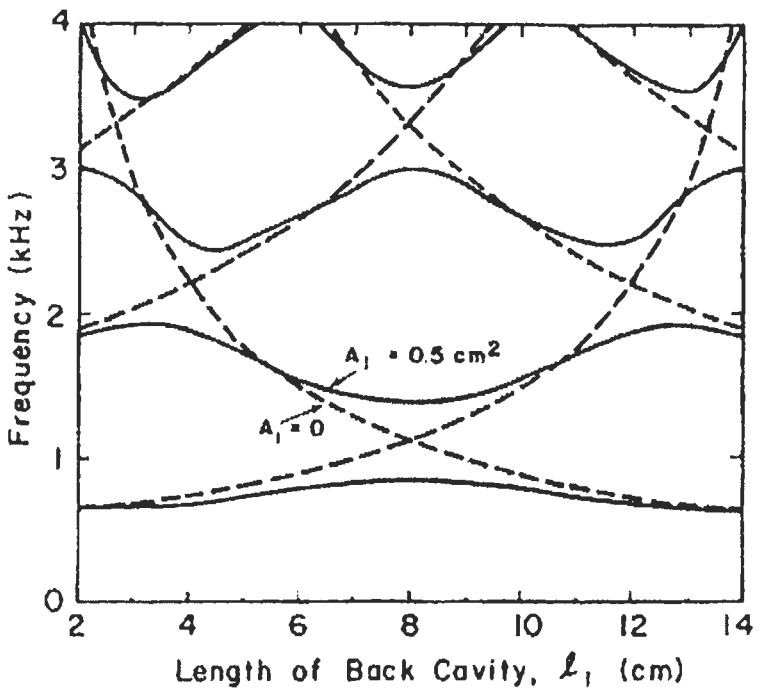


Figure 6.9 Frequencies of the first four natural frequencies for the nontapered configuration of figure 6.8, as the length ℓ_1 of the back cavity is manipulated. The total length $\ell_1 + \ell_2 = 16$ cm, and the cross-sectional area $A_2 = 3$ cm^2. The dashed line corresponds to the case where $A_1 = A_2$, and the solid line is for $A_1 = 0.5$ cm^2. The radiation impedance is assumed to be zero. (From K. N. Stevens, 1989.)