ECE 797:

Speech and Audio Processing

Hand-out for Lecture #1

Tuesday, January 13, 2004

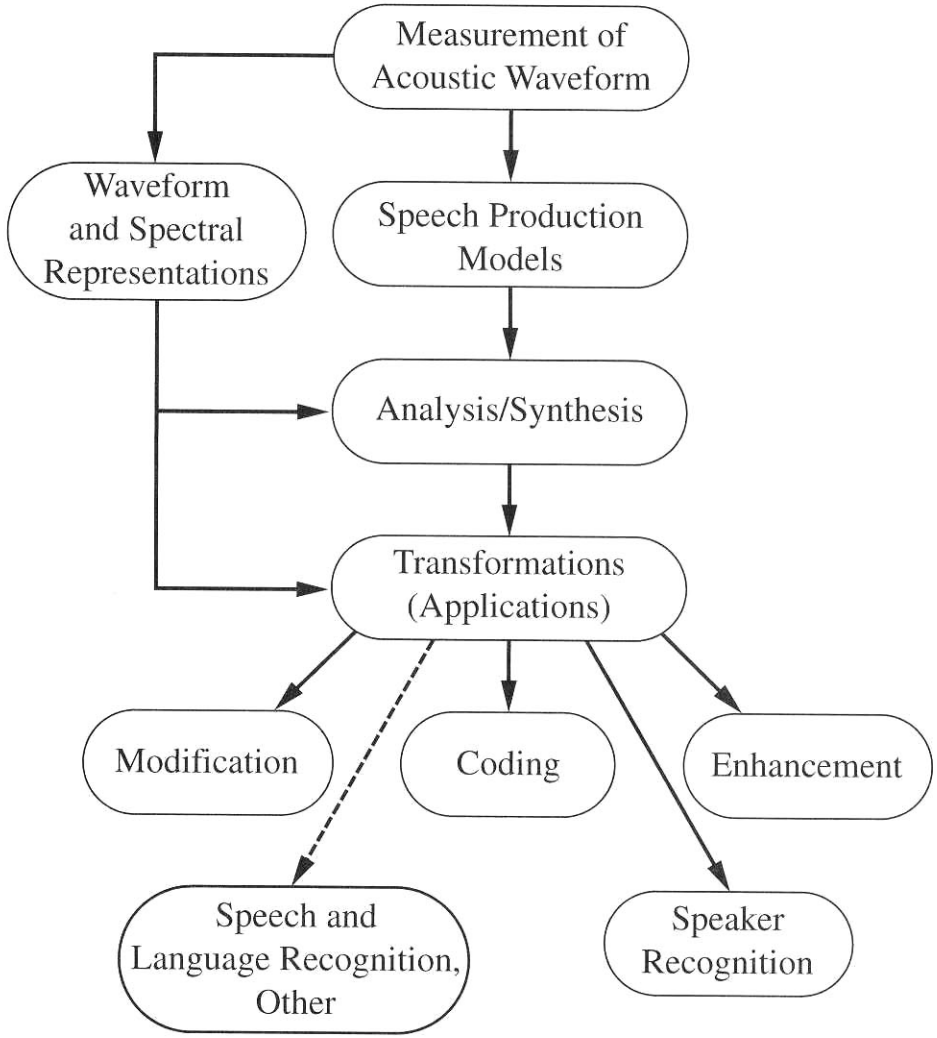

Figure 1.3 Discrete-time speech signal processing overview. Applications within the text include speech modification, coding, enhancement, and speaker recognition.

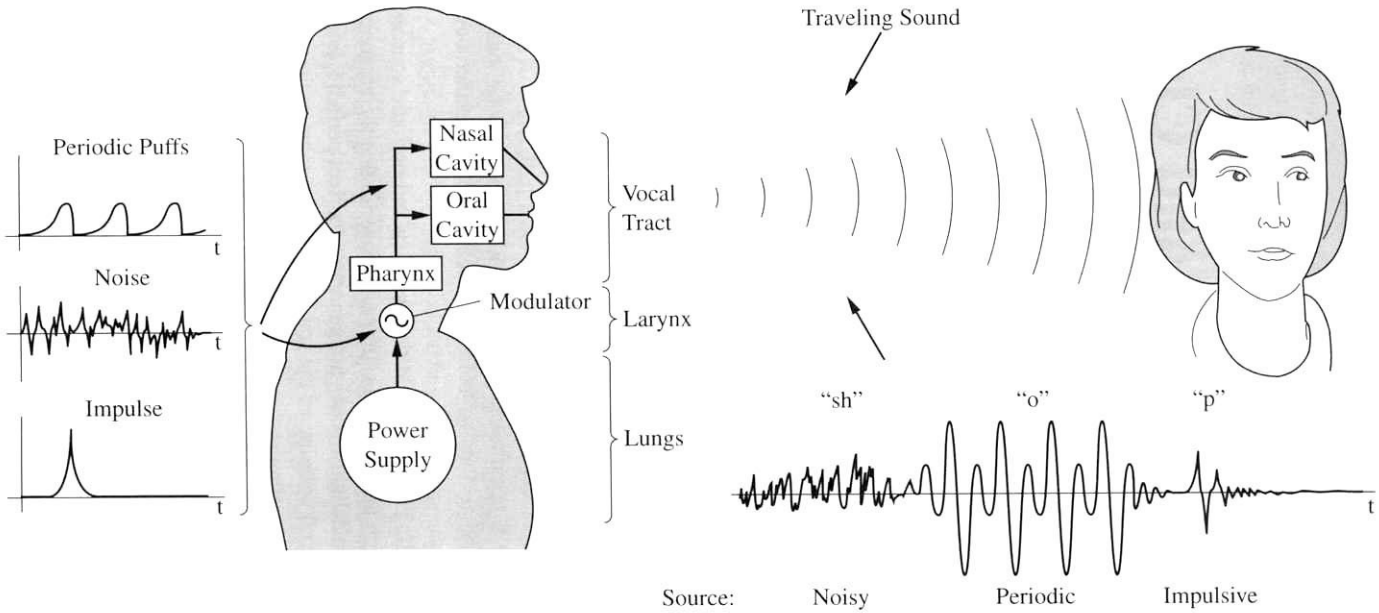

Figure 3.1 Simple view of speech production. The sound sources are idealized as periodic, impulsive, or (white) noise and can occur in the larynx or vocal tract.

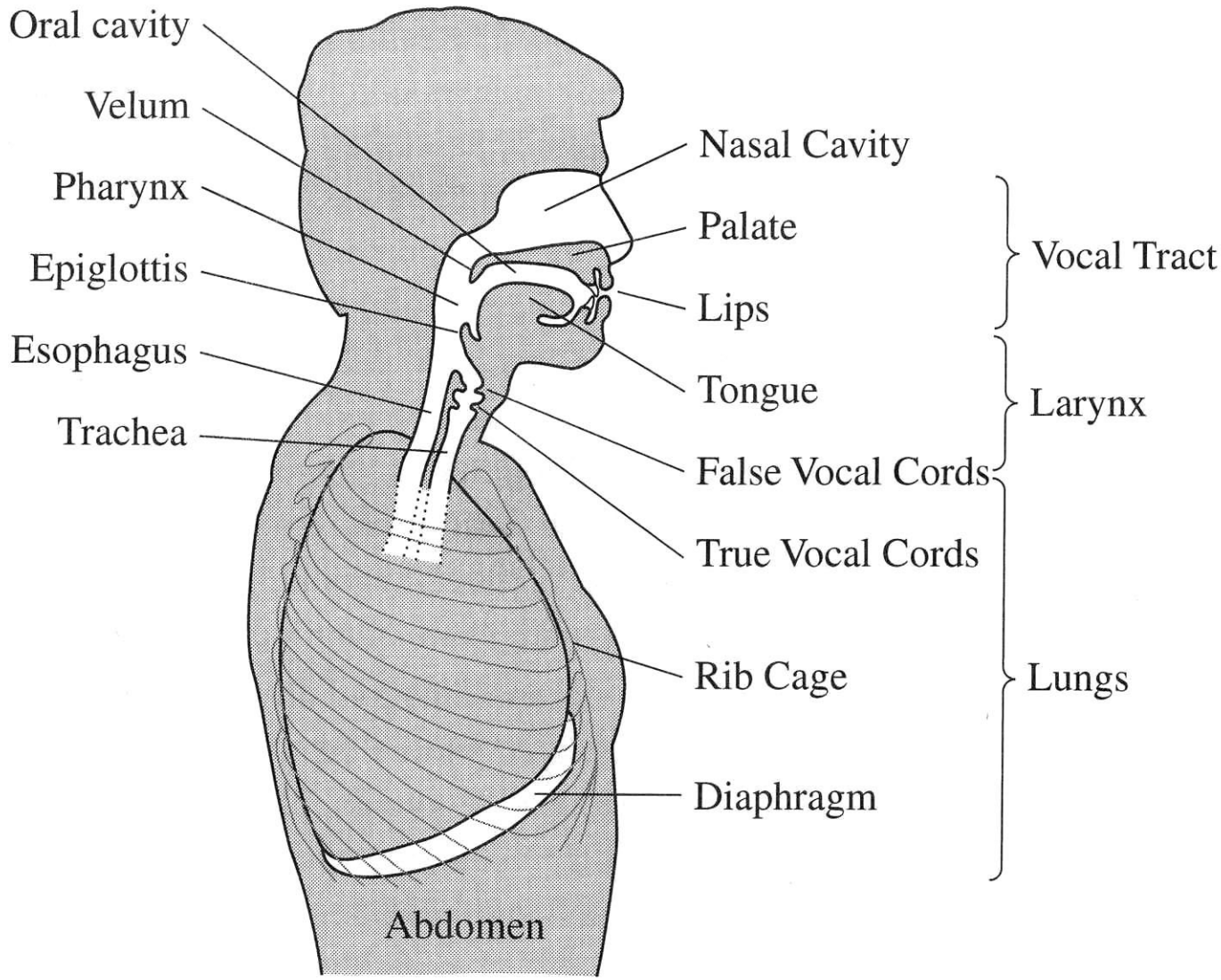

Figure 3.2 Cross-sectional view of the anatomy of speech production.

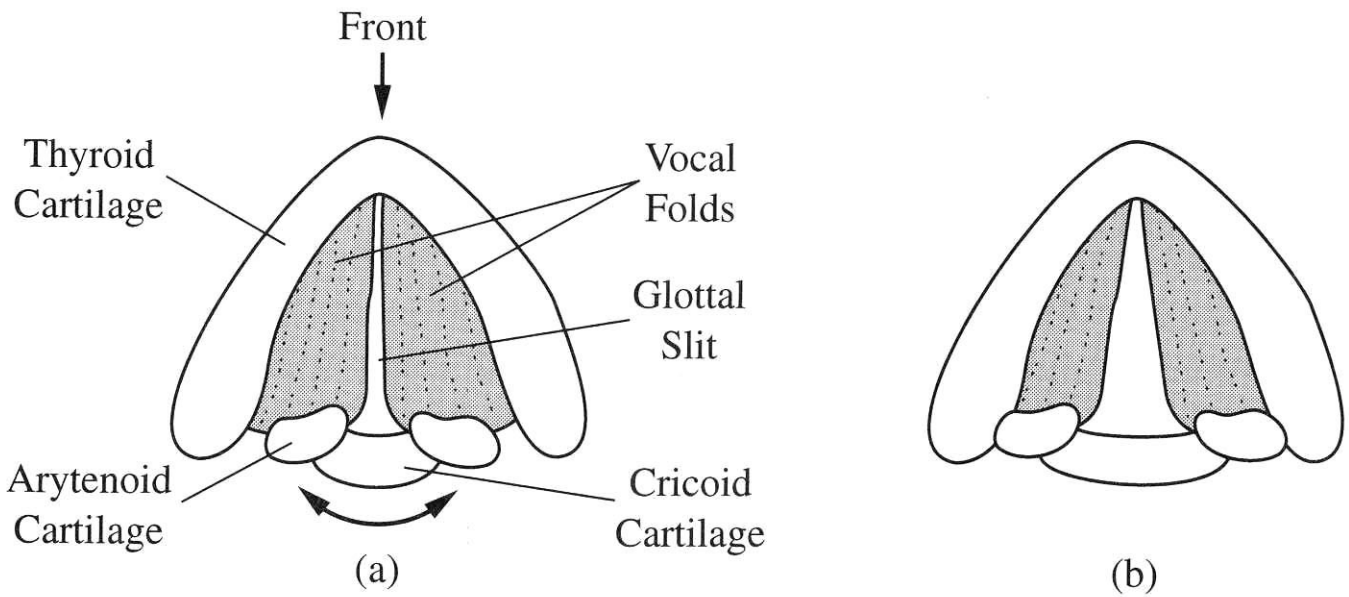

Figure 3.3 Sketches of downward-looking view of the human larynx: (a) voicing; (b) breathing.

SOURCE: K.N. Stevens, *Acoustic Phonetics*, The MIT Press [33].

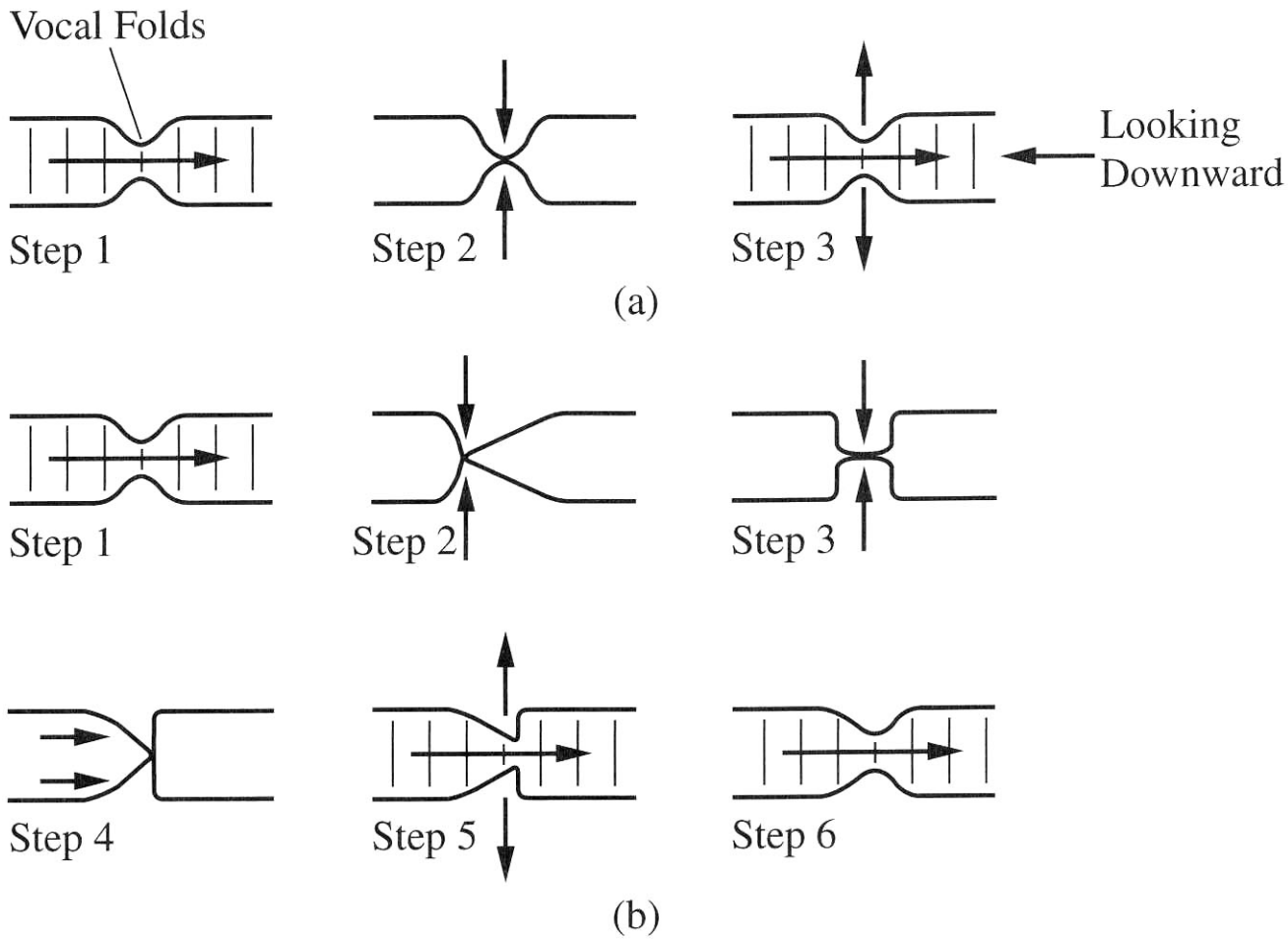

Figure 3.4 Bernoulli's Principle in the glottis: (a) basic horizontal open/close voicing cycle; (b) refinement of (a) with vertical vocal fold motion. Vertical lines represent airflow in the direction of the arrows.

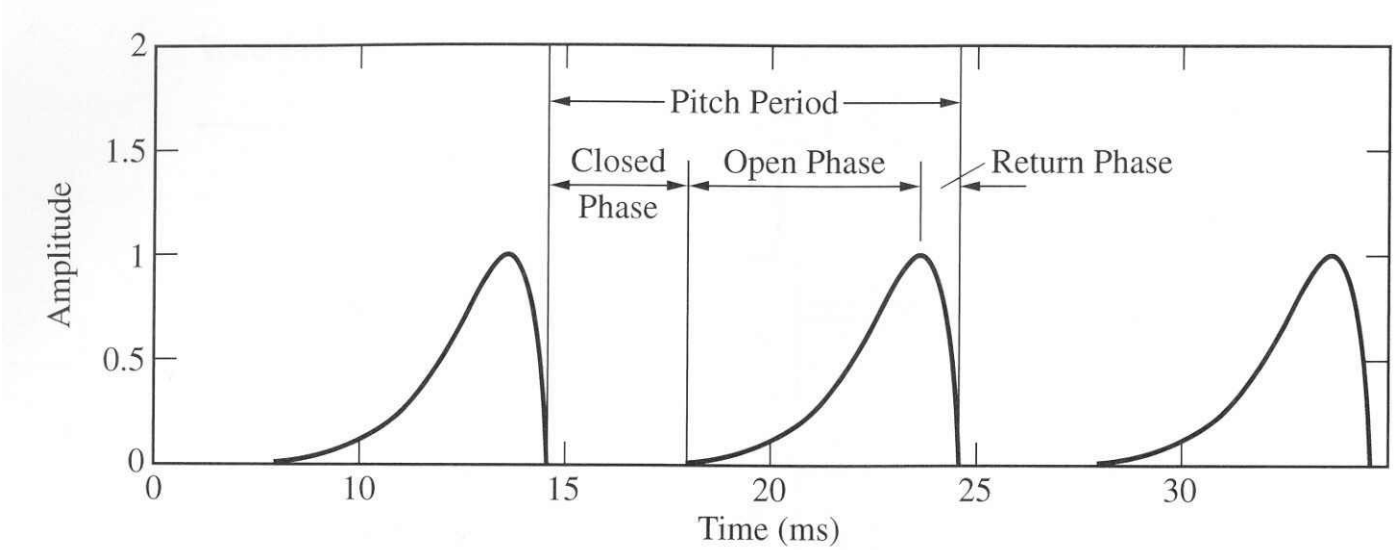

Figure 3.6 Illustration of periodic glottal airflow velocity.

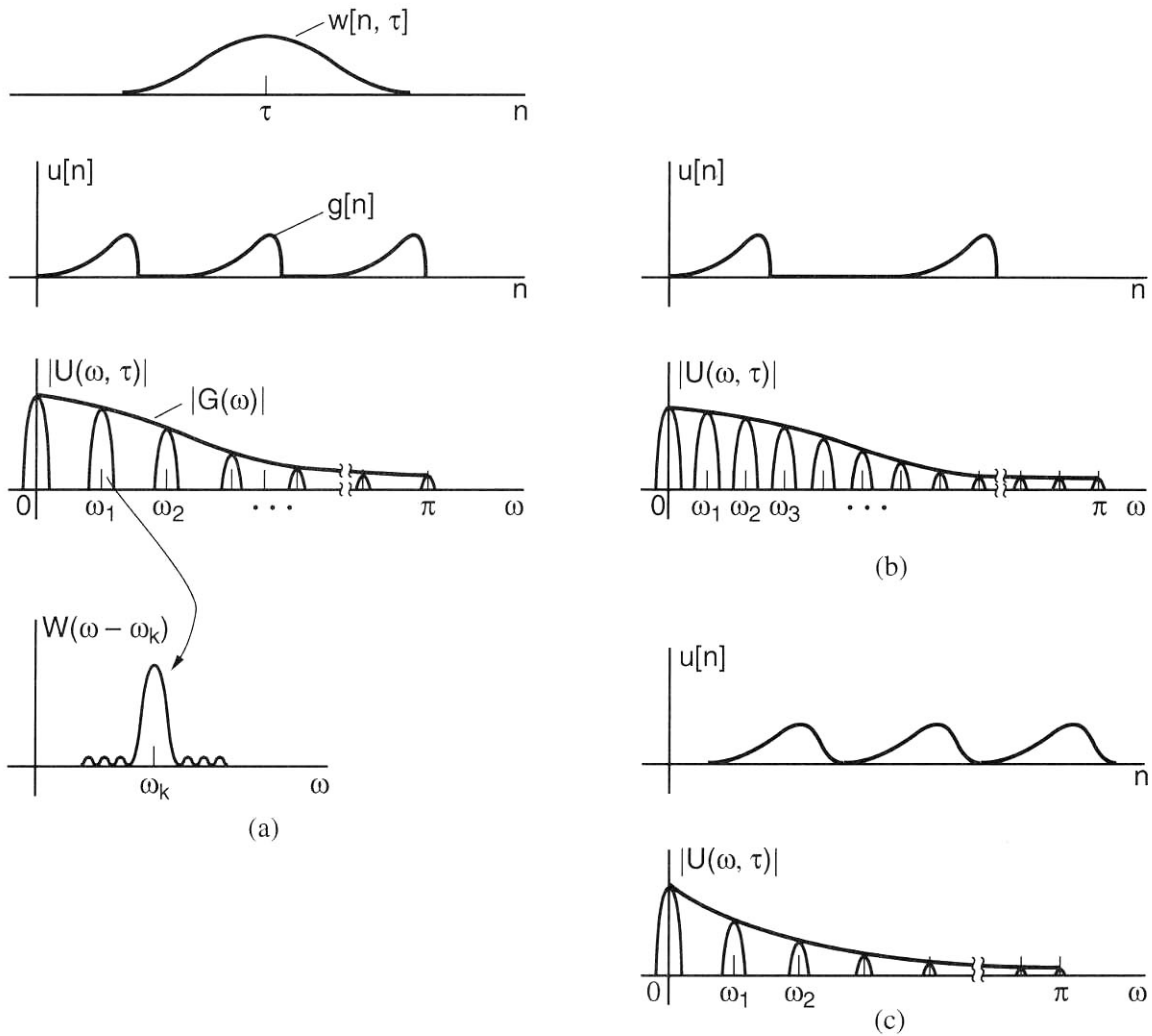

Figure 3.7 Illustration of periodic glottal flow in Example 3.1: (a) typical glottal flow and its spectrum; (b) same as (a) with lower pitch; and (c) same as (a) with "softer" or more "relaxed" glottal flow.

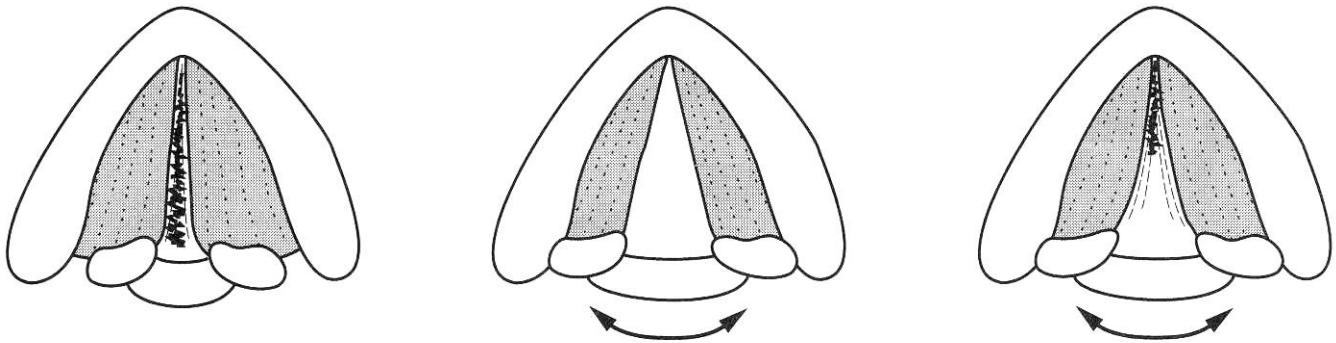

Figure 3.8 Sketches of various vocal fold configurations: (a) aspiration (whispering), (b) voicing, and (c) aspirated voicing. Arrows indicate vocal fold vibration, while ragged lines indicate turbulence.

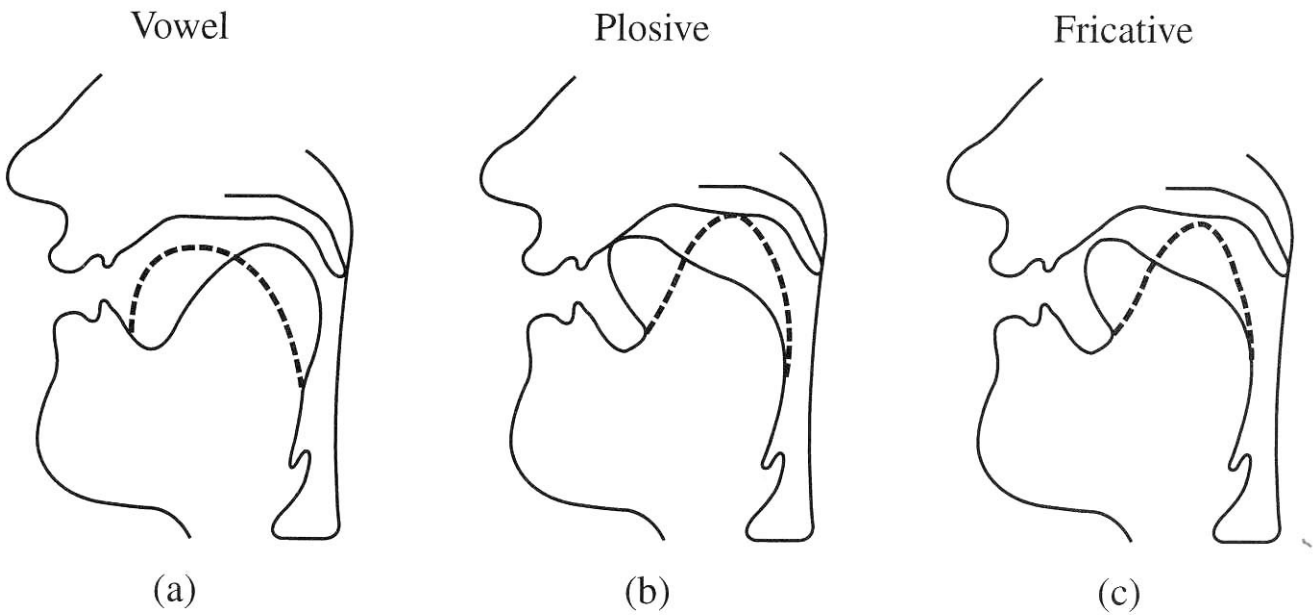

Figure 3.10 Illustration of changing vocal tract shapes for (a) vowels (having a periodic source), (b) plosives (having an impulsive source), and (c) fricatives (having a noise source).

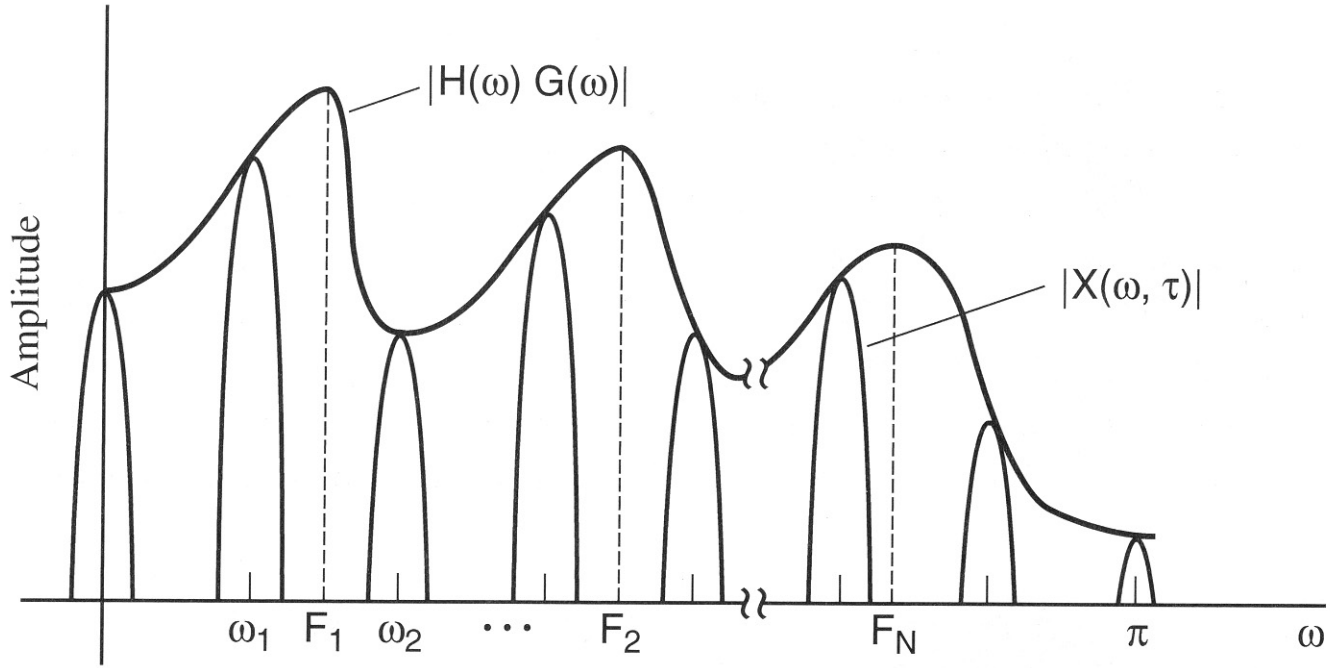

Figure 3.11 Illustration of relation of glottal source harmonics $\omega_1, \omega_2, \ldots \omega_N$, vocal tract formants $F_1, F_2, \ldots F_M$, and the spectral envelope $|H(\omega)G(\omega)|$.

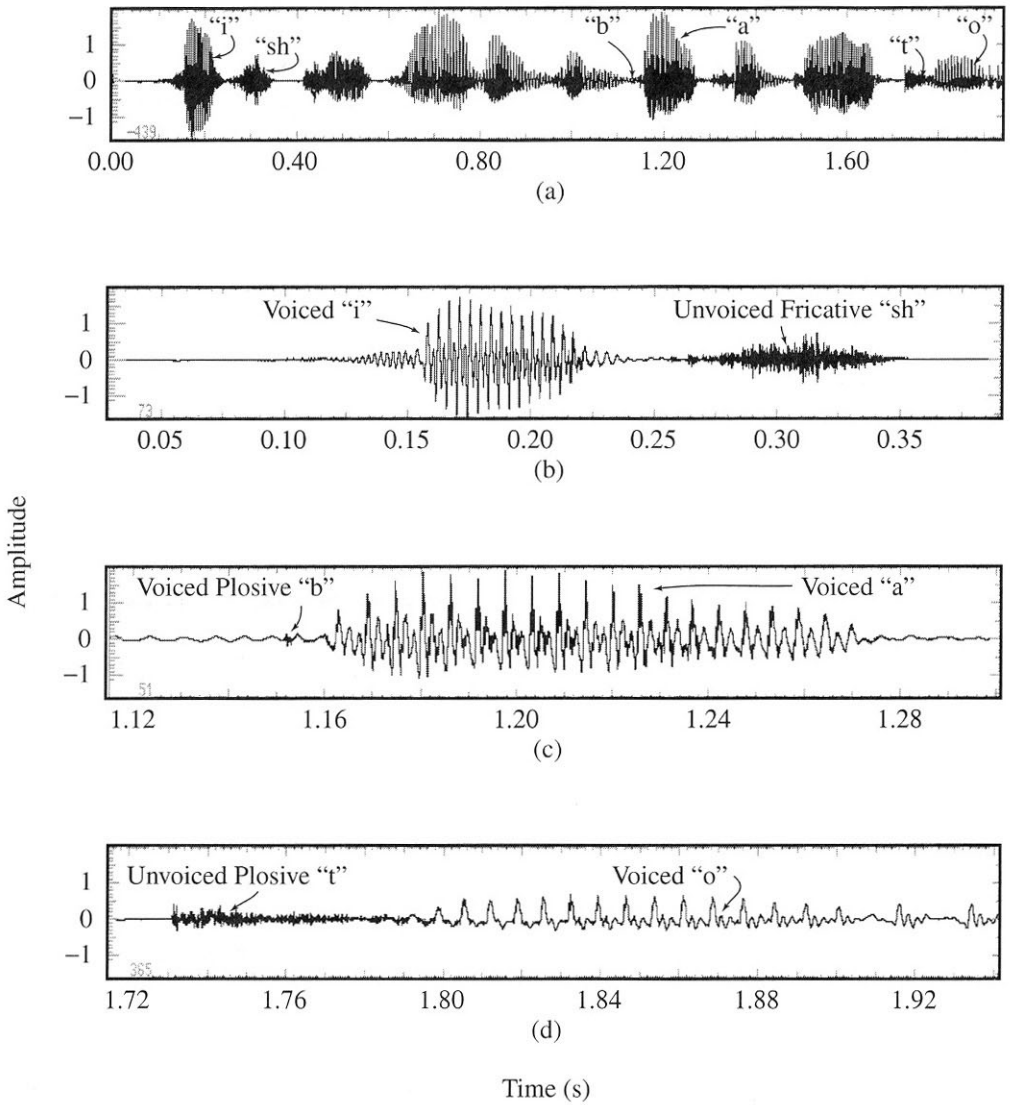

Figure 3.13 Examples of voiced, fricative, and plosive sounds in the sentence, "Which tea party did Baker go to?": (a) speech waveform; (b)–(d) magnified voiced, fricative, and plosive sounds from (a). (Note the "sh" is a component of an affricate to be studied in Section 3.4.6.)

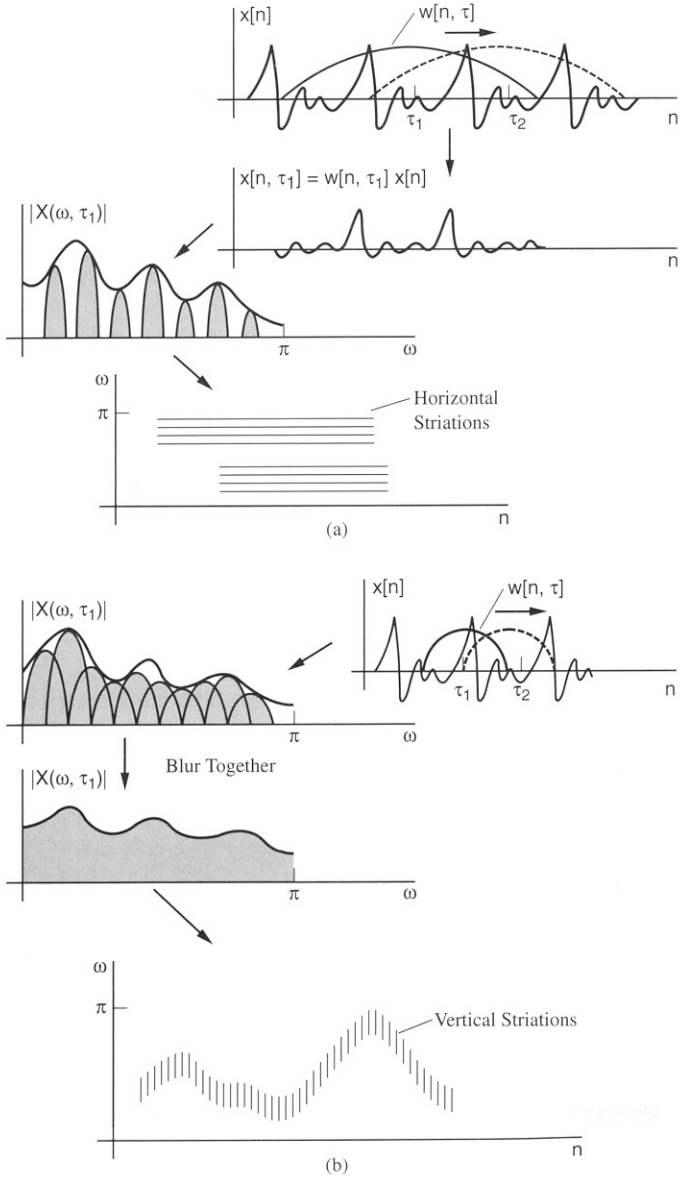

Figure 3.14 Formation of (a) the narrowband and (b) the wideband spectrograms.

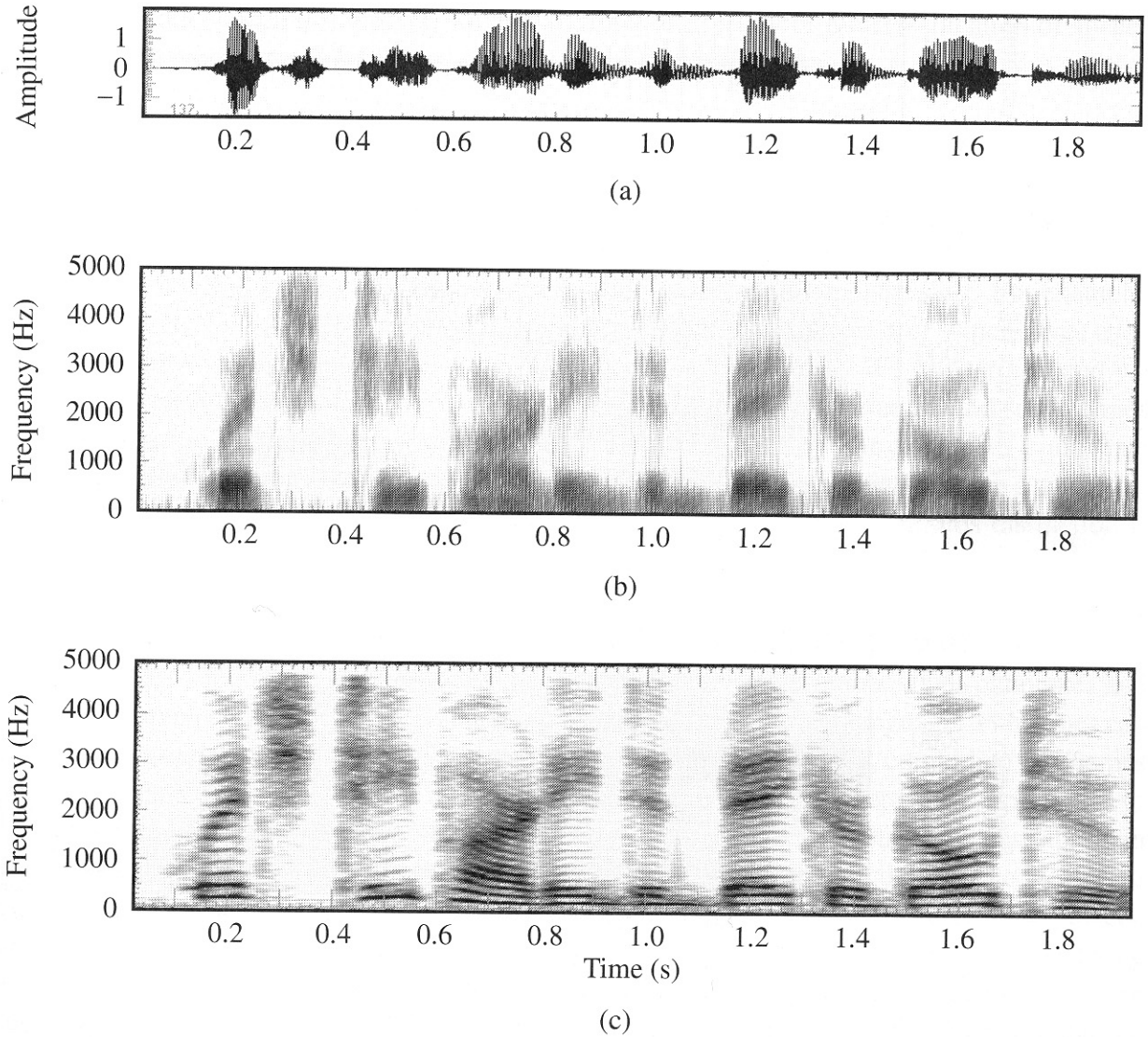

Figure 3.15 Comparison of measured spectrograms for the utterance, "Which tea party did Baker go to?": (a) speech waveform; (b) wideband spectrogram; (c) narrowband spectrogram.

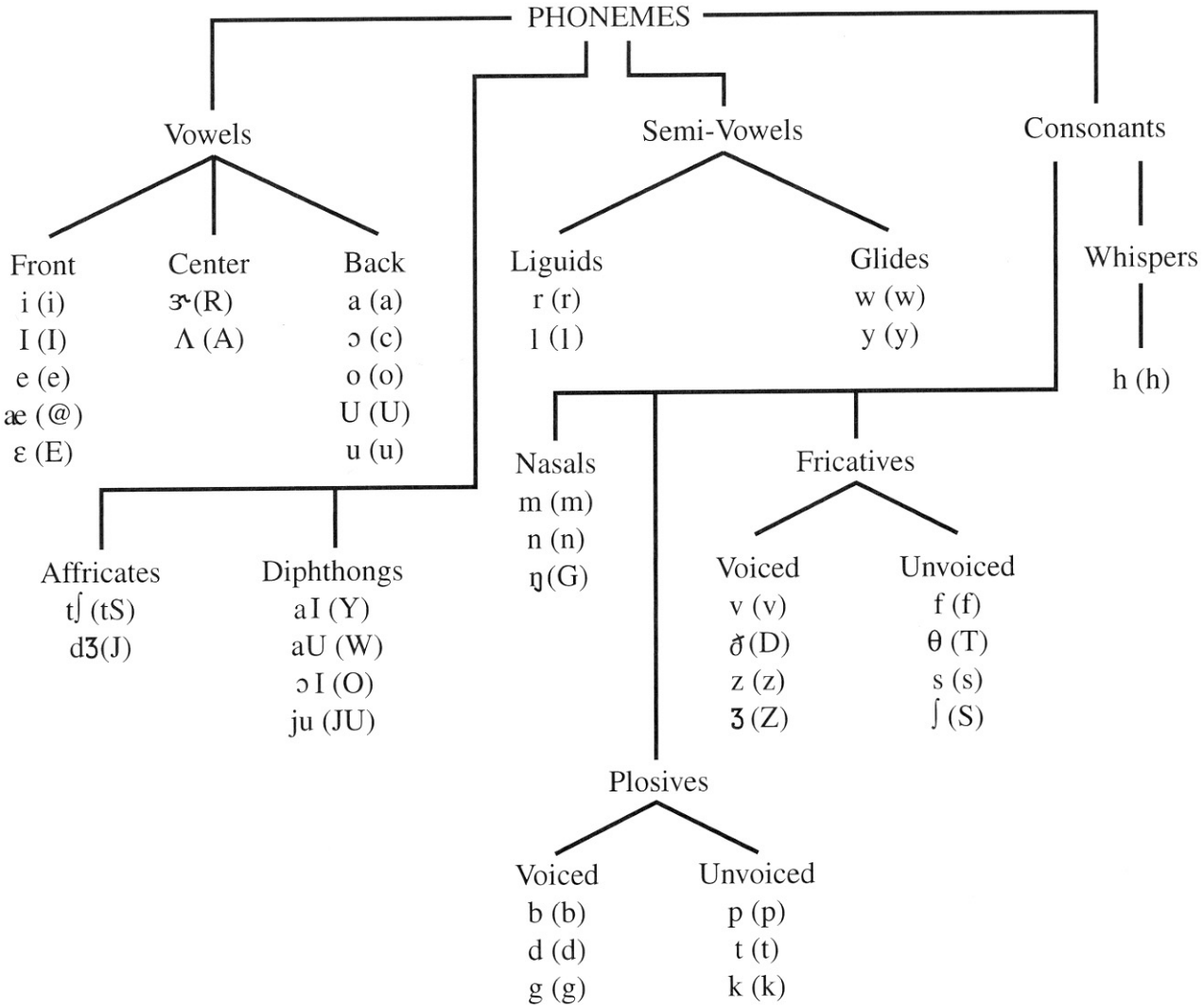

Figure 3.17 Phonemes in American English [6],[32]. Orthographic symbols are given in parentheses to the left of the International Phonetic Alphabet symbols.

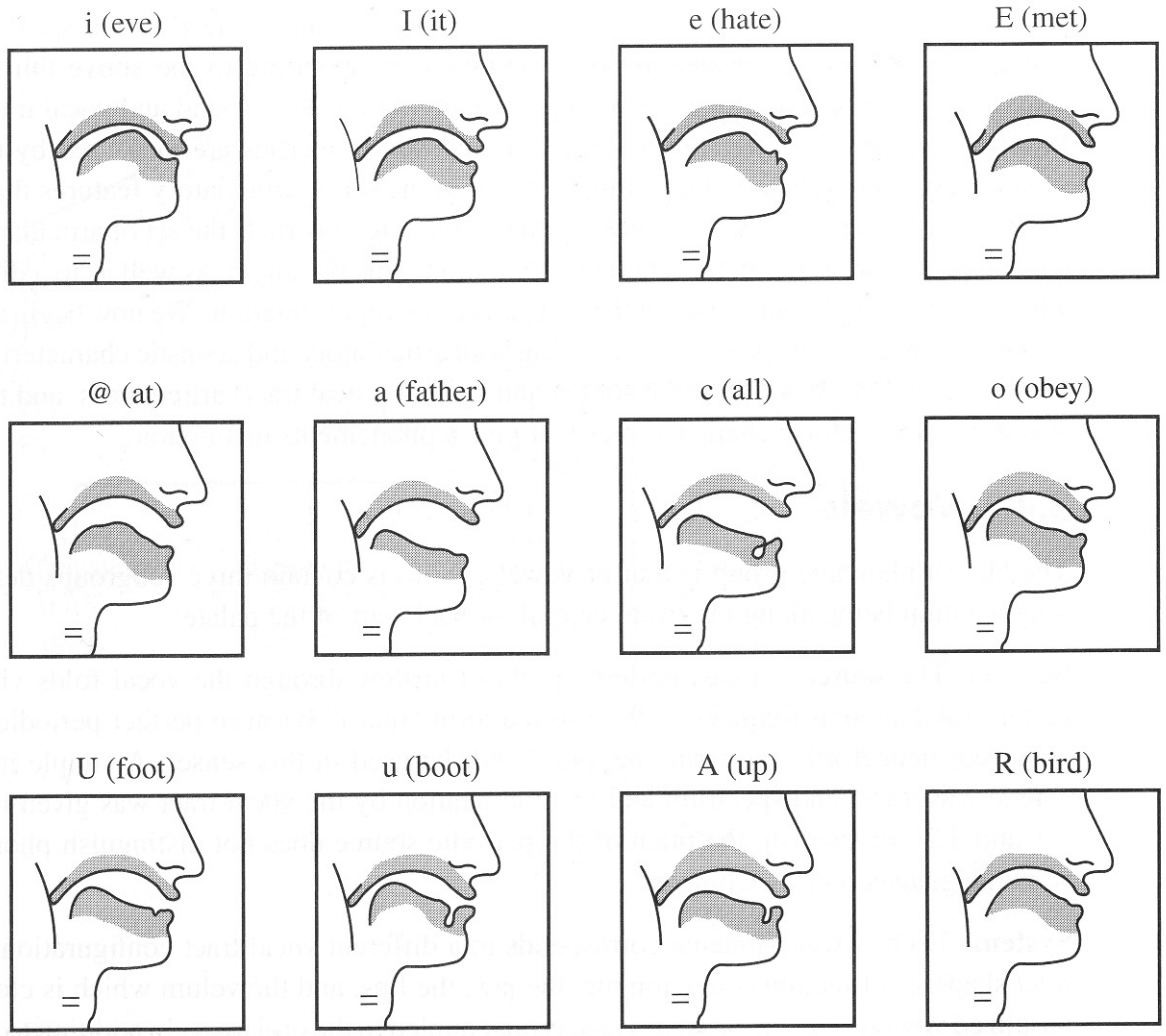

Figure 3.18 Vocal tract profiles for vowels in American English. The two horizontal lines denote voicing.

SOURCE: R.K. Potter, G.A. Kopp, and H.G. Kopp, *Visible Speech* [31]. ©1966, Dover Publications, Inc. Used by permission.

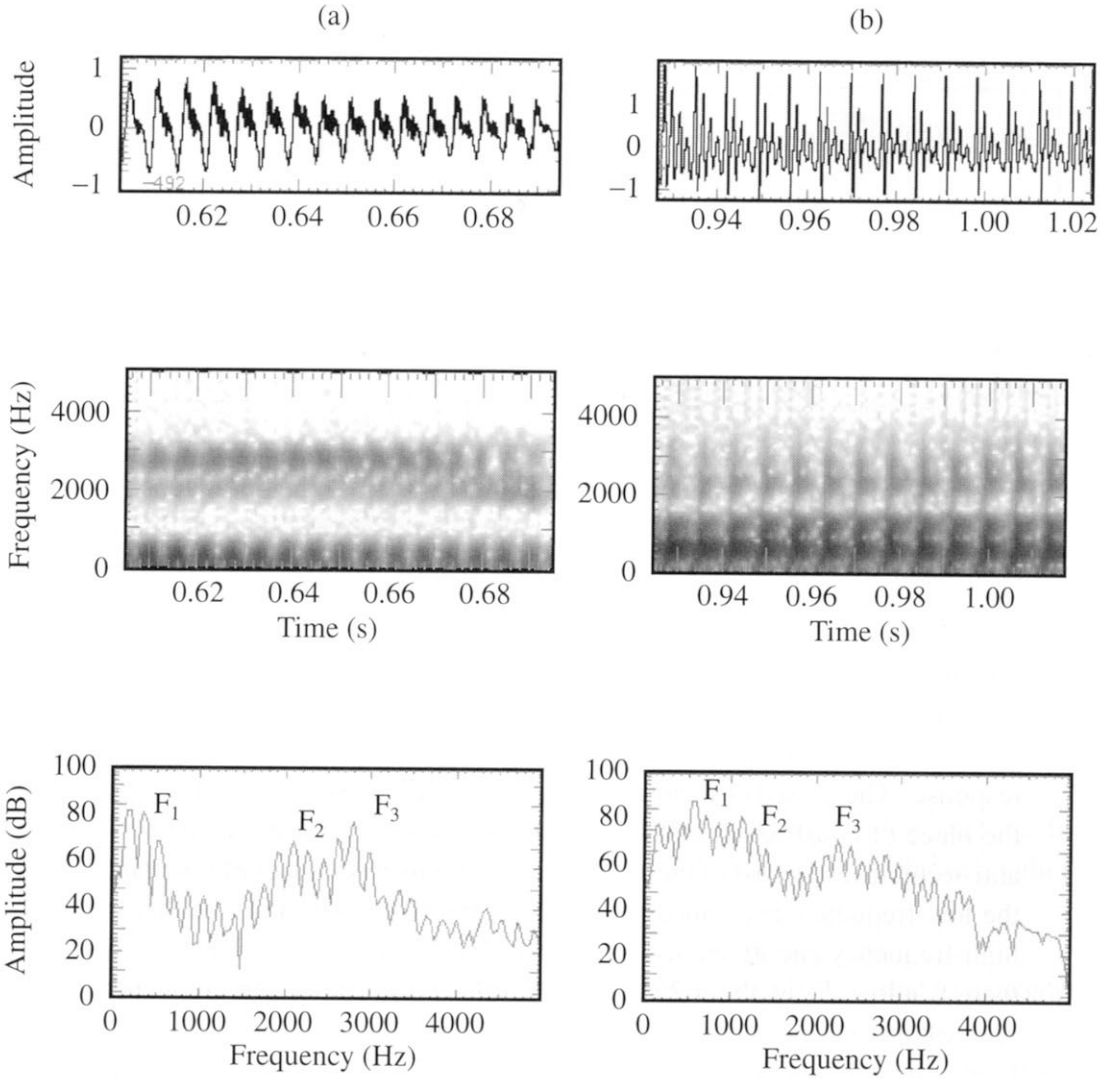

Figure 3.19 Waveform, wideband spectrogram, and spectral slice of narrowband spectrogram for two vowels: (a) /i/ as in "eve"; (b) /a/ as in "father." The first three formants F_1, F_2, and F_3 are marked on the spectral slices.

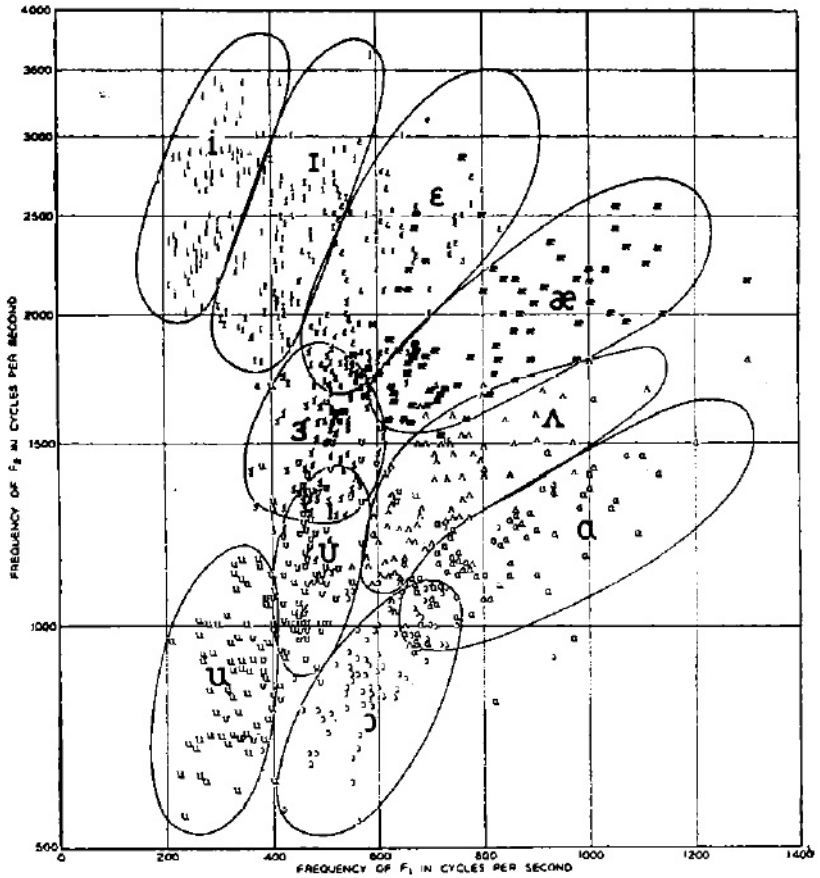

FIG. 8. Frequency of second formant *versus* frequency of first formant for ten vowels by 76 speakers.

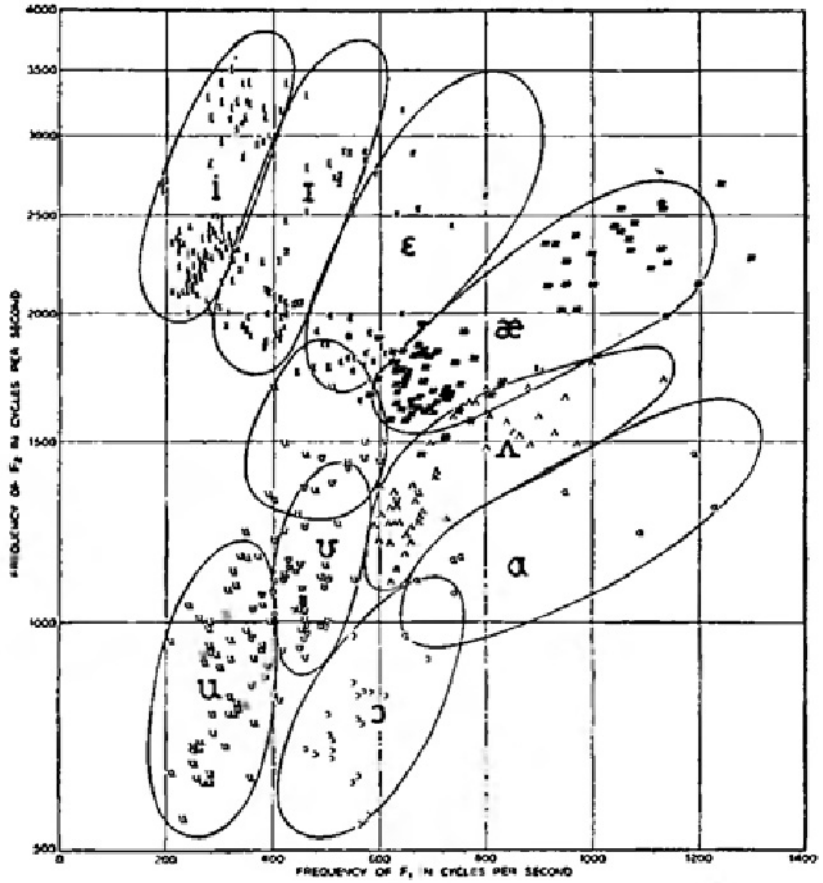

FIG. 9. Frequency of second formant *versus* frequency of first formant for vowels spoken by men and children, which were classified unanimously by all listeners.

(Peterson and Barney, JASA, 1952)

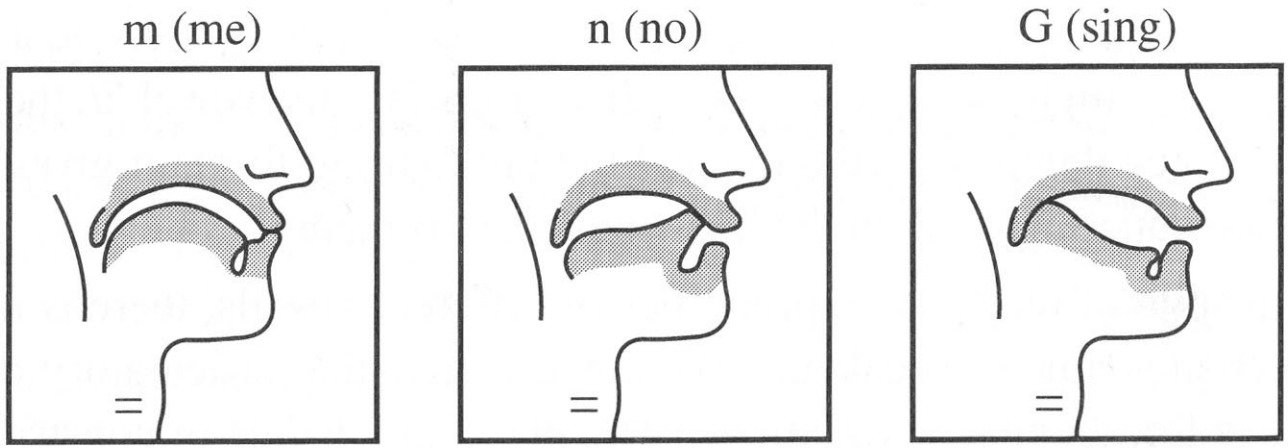

Figure 3.20 Vocal tract configurations for nasal consonants. Oral tract constrictions occur at the lips for /m/, with the tongue tip to the gum ridge for /n/, and with the tongue body against the palate near the velum for /ng/. Horizontal lines denote voicing.

SOURCE: R.K. Potter, G.A. Kopp, and H.G. Kopp, *Visible Speech* [31].

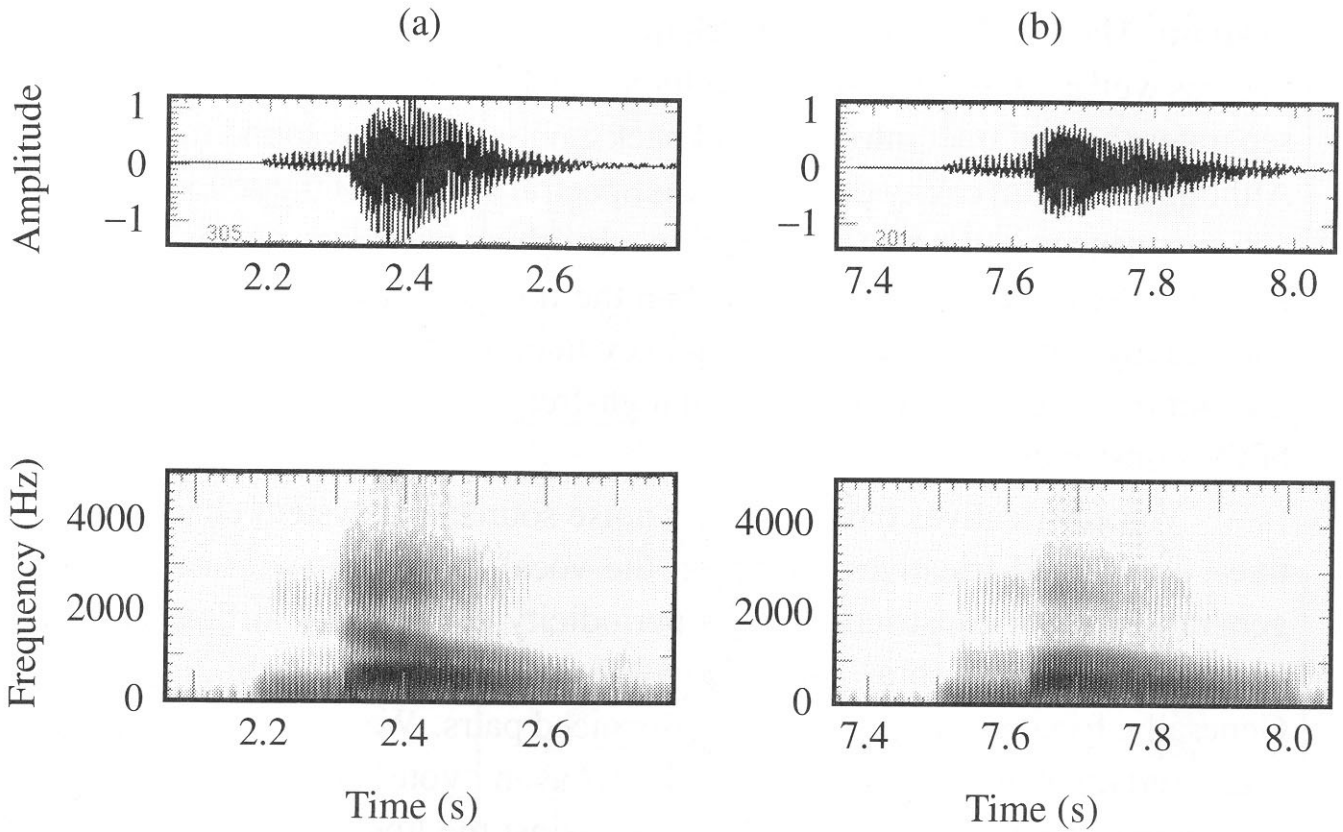

Figure 3.21 Wideband spectrograms of nasal consonants (a) /n/ in "no" and (b) /m/ in "mo."

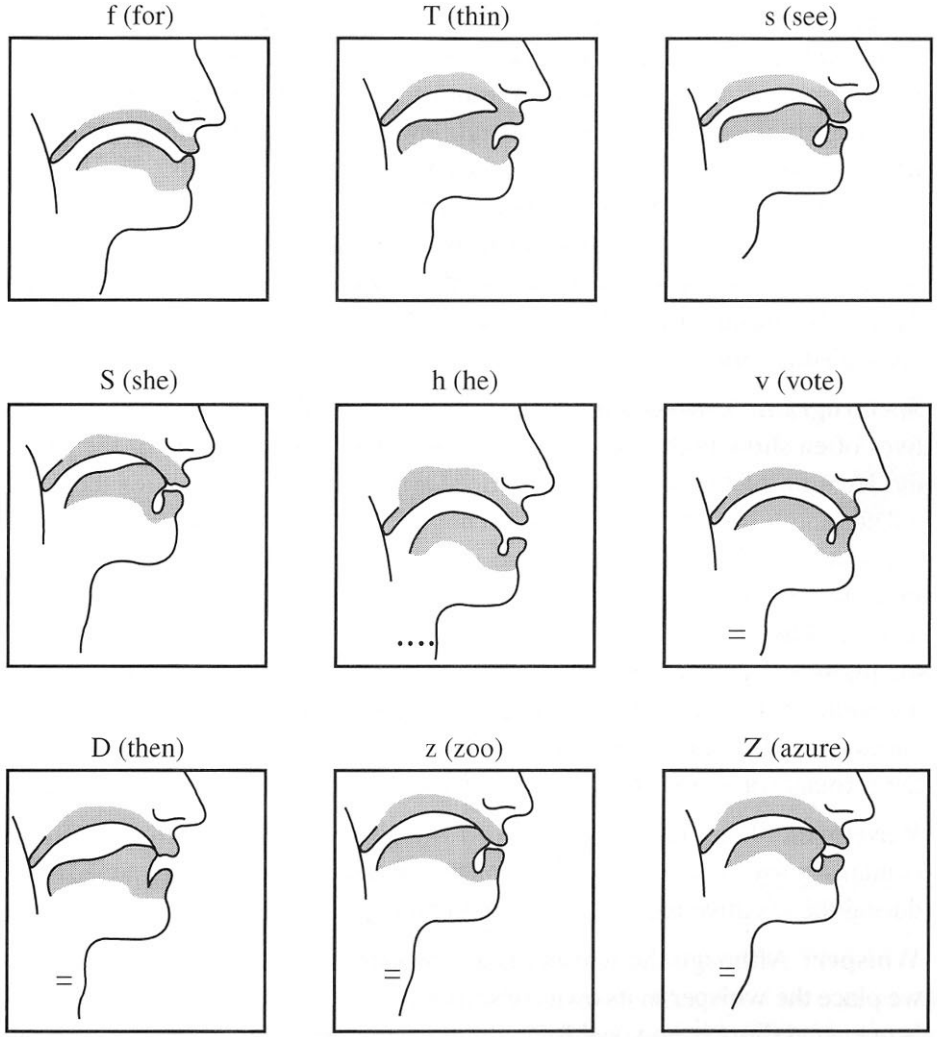

Figure 3.22 Vocal tract configurations for pairs of voiced and unvoiced fricatives. Horizontal lines denote voicing and dots denote aspiration.

SOURCE: R.K. Potter, G.A. Kopp, and H.G. Kopp, *Visible Speech* [31]. ©1966, Dover Publications, Inc. Used by permission.

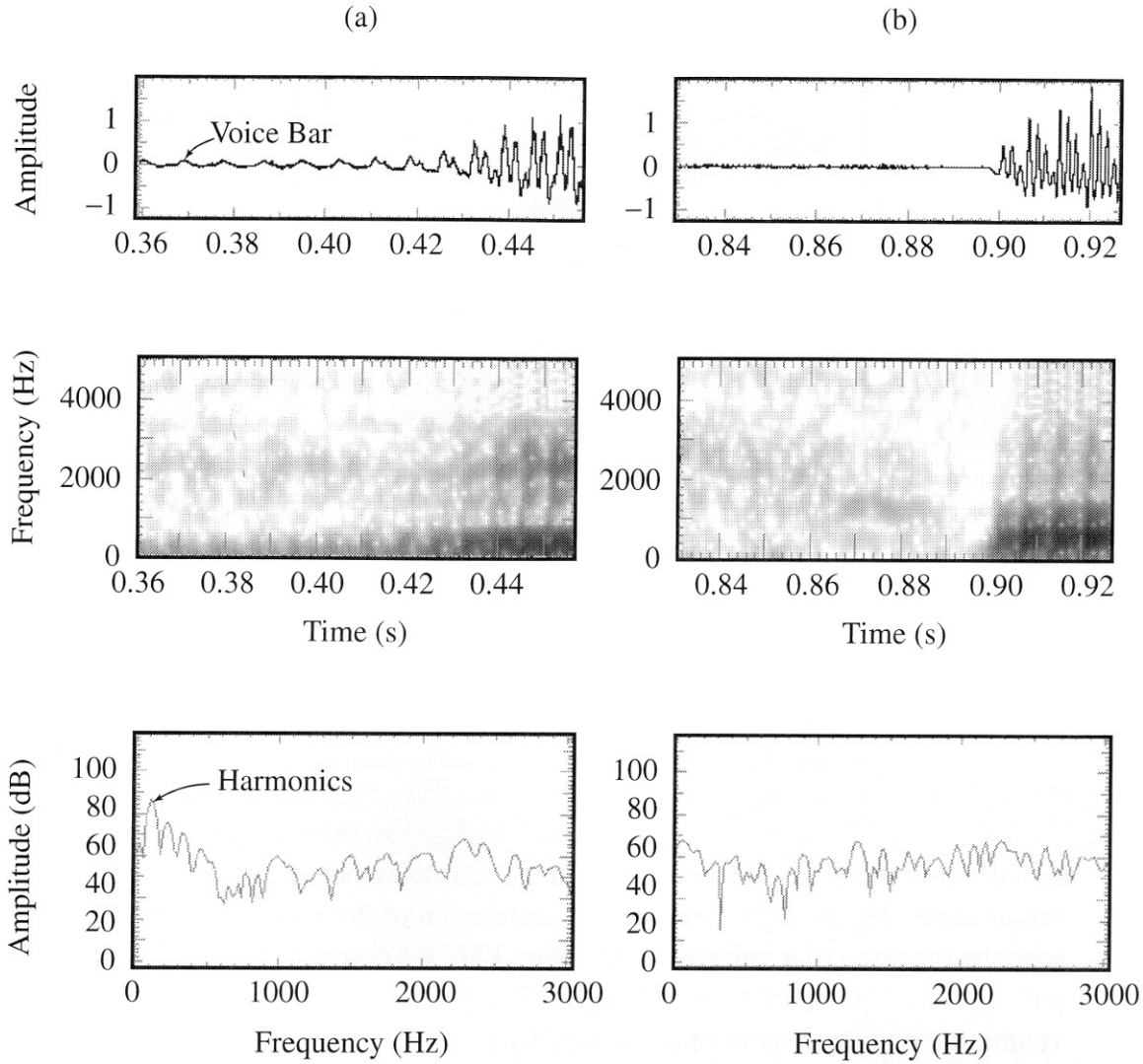

Figure 3.23 Waveform, wideband spectrogram, and narrowband spectral slice of voiced and unvoiced fricative pair: (a) /v/ as in "vote"; (b) /f/ as in "for." Spectral slices taken in fricative regions over a 20-ms window.

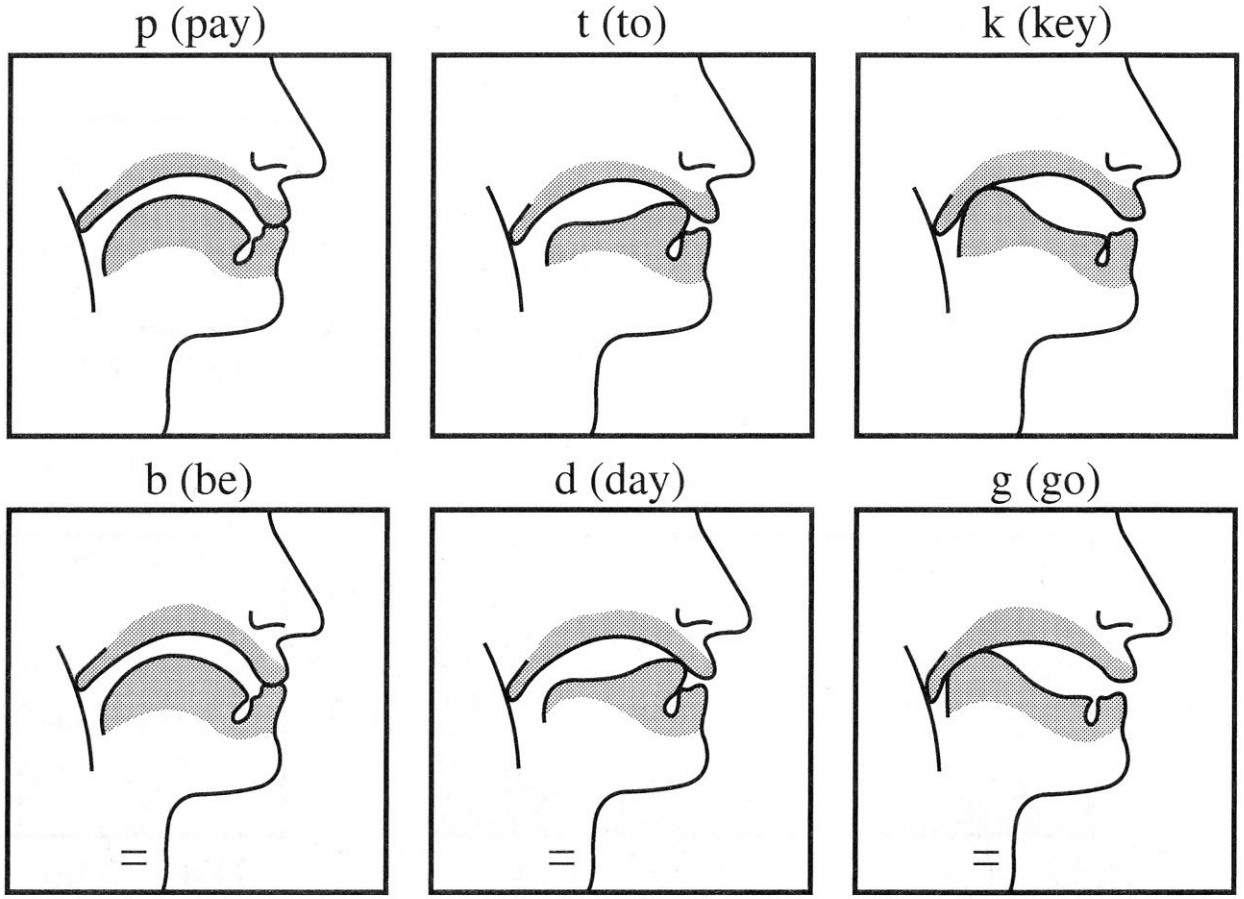

Figure 3.24 Vocal tract configurations for unvoiced and voiced plosive pairs. Horizontal lines denote voicing.

SOURCE: R.K. Potter, G.A. Kopp, and H.G. Kopp, *Visible Speech* [31]. ©1966, Dover Publications, Inc. Used by permission.

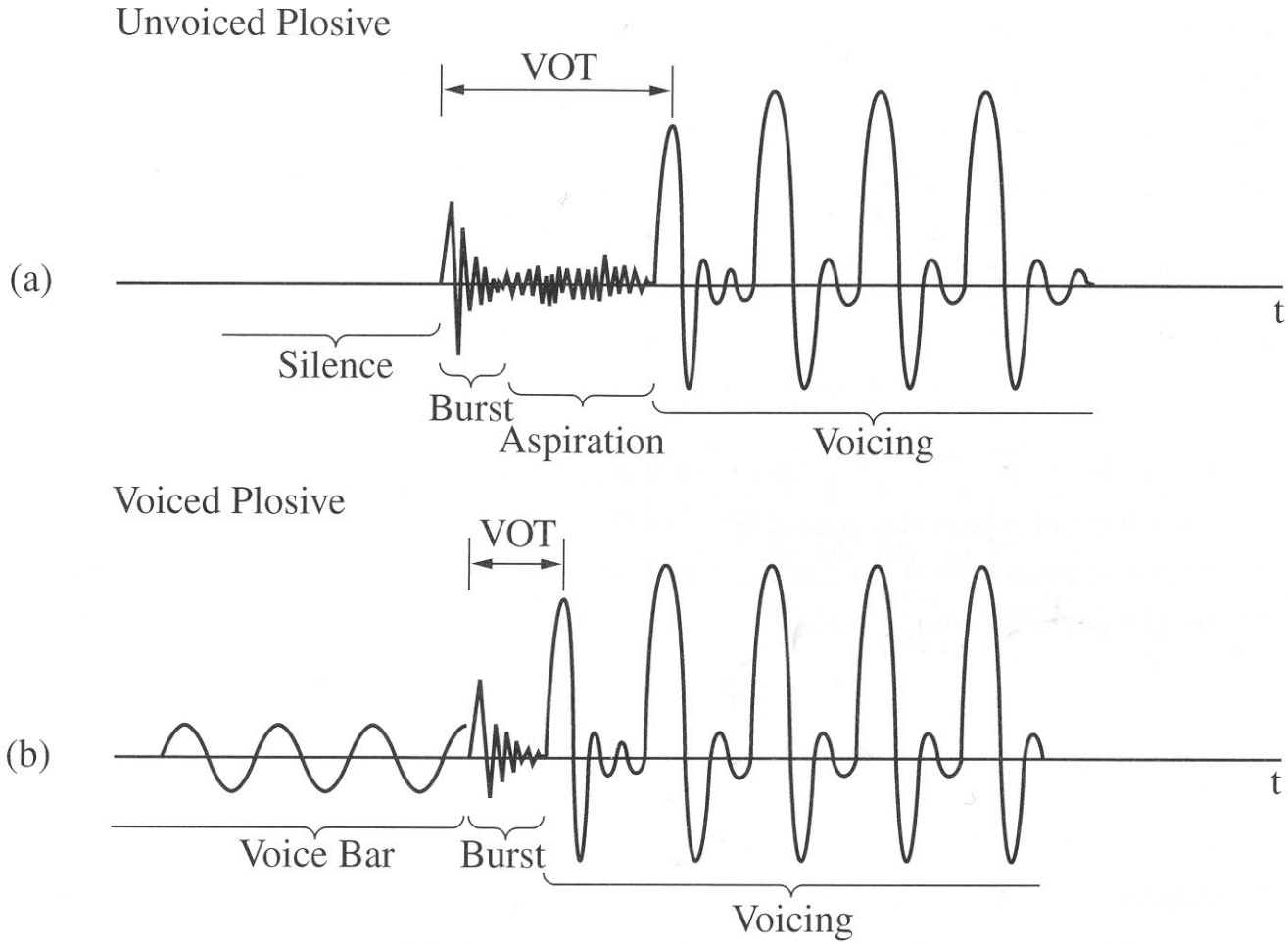

Figure 3.25 A schematic representation of (a) unvoiced and (b) voiced plosives. The voiced onset time is denoted by VOT.

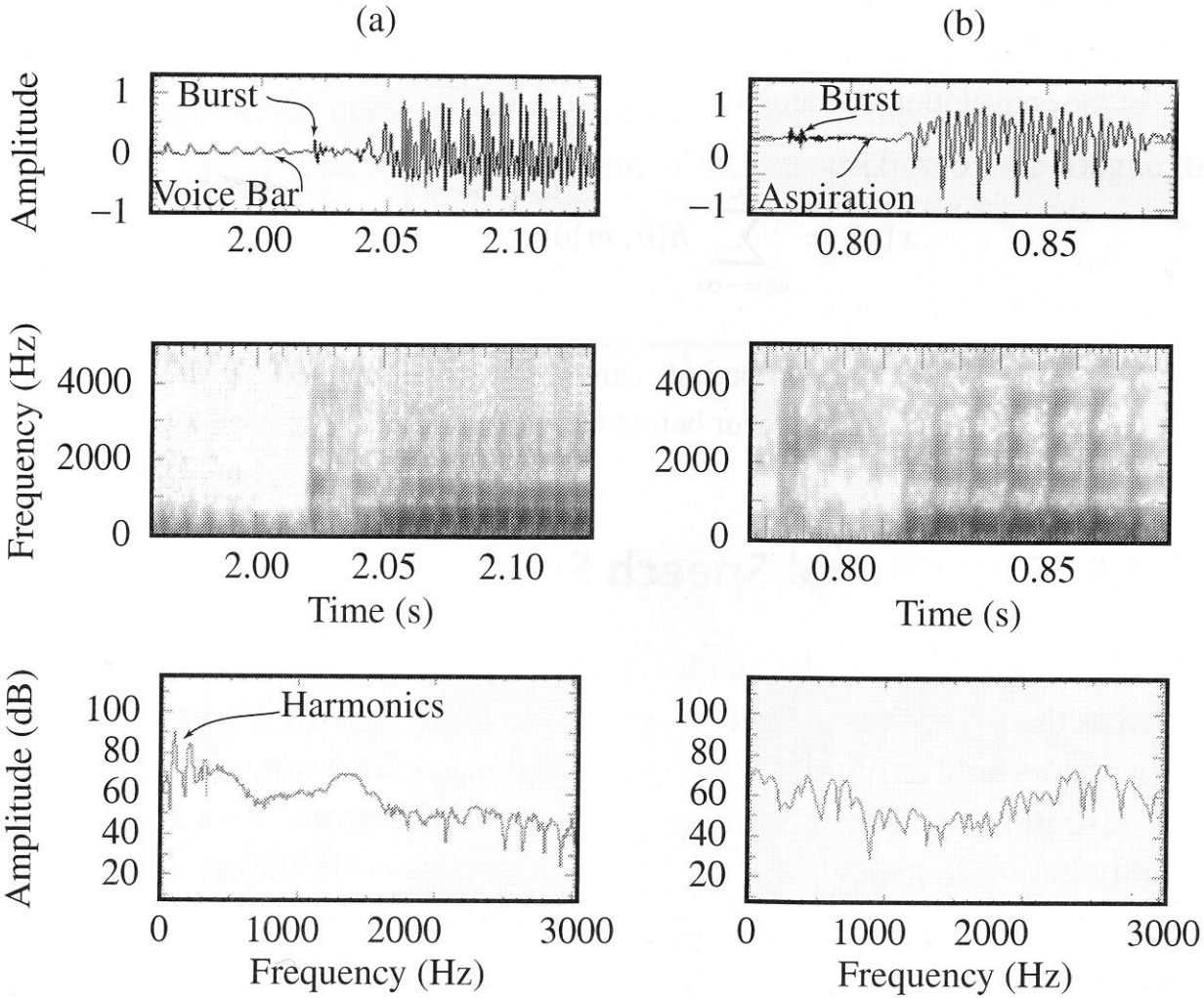

Figure 3.26 Waveform, wideband spectrogram, and narrowband spectral slice of voiced and unvoiced plosive pair: (a) /g/ as in "go"; (b) /k/ as in "key." Spectral slices are taken in burst regions over a 40-ms window in (a) and a 25-ms window in (b).

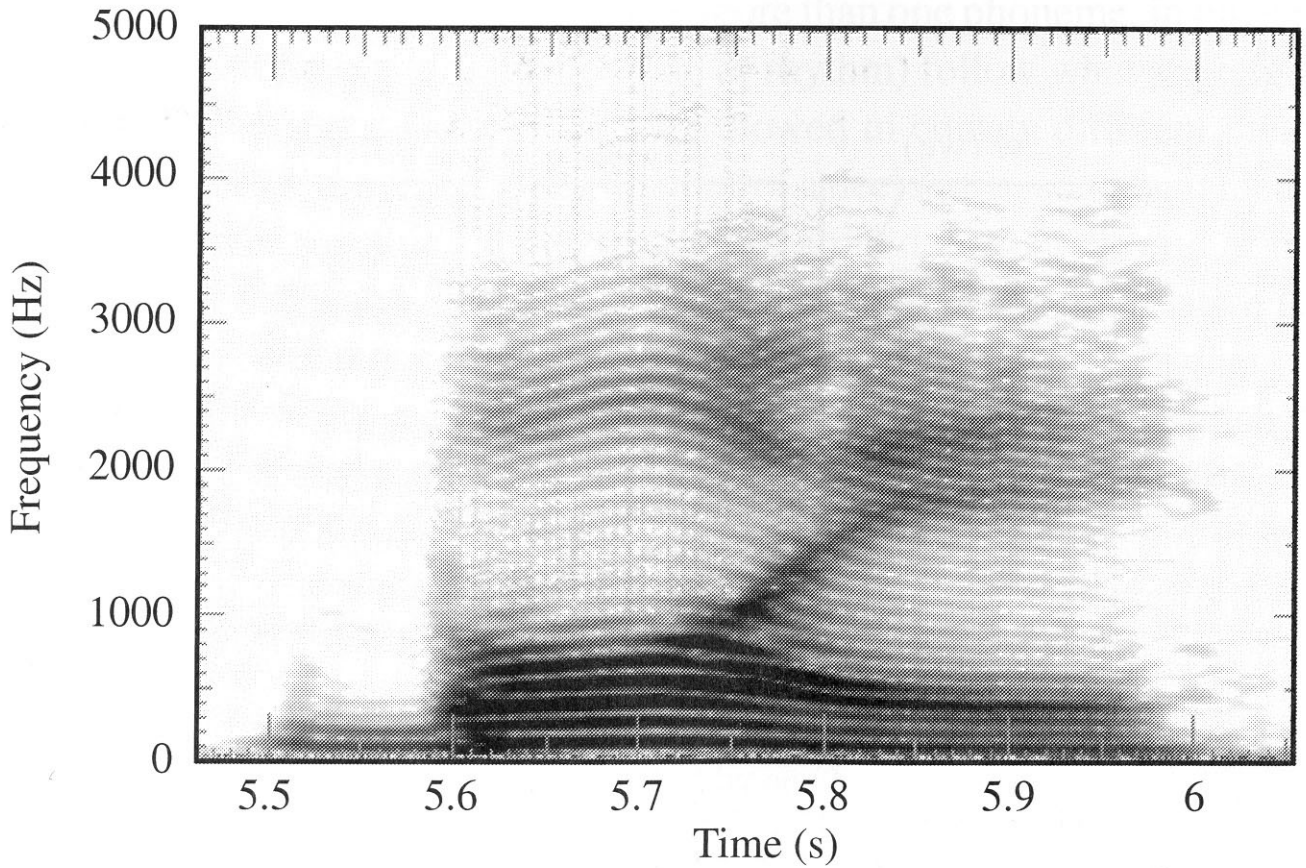

Figure 3.27 Narrowband spectrogram example for the diphthong /O/ in "boy."

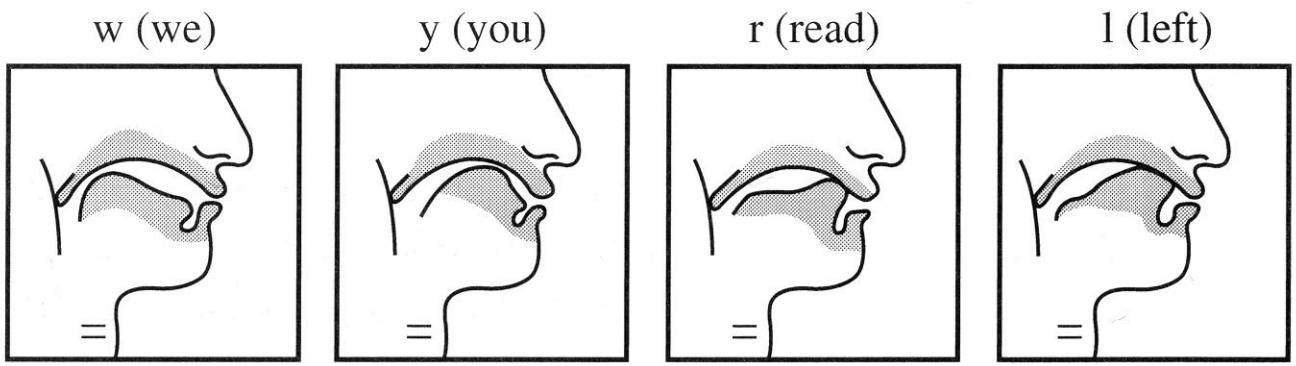

Figure 3.28 Configurations of glides and liquids. Horizontal lines denote voicing.

SOURCE: R.K. Potter, G.A. Kopp, and H.G. Kopp, *Visible Speech* [31]. ©1966, Dover Publications, Inc. Used by permission.

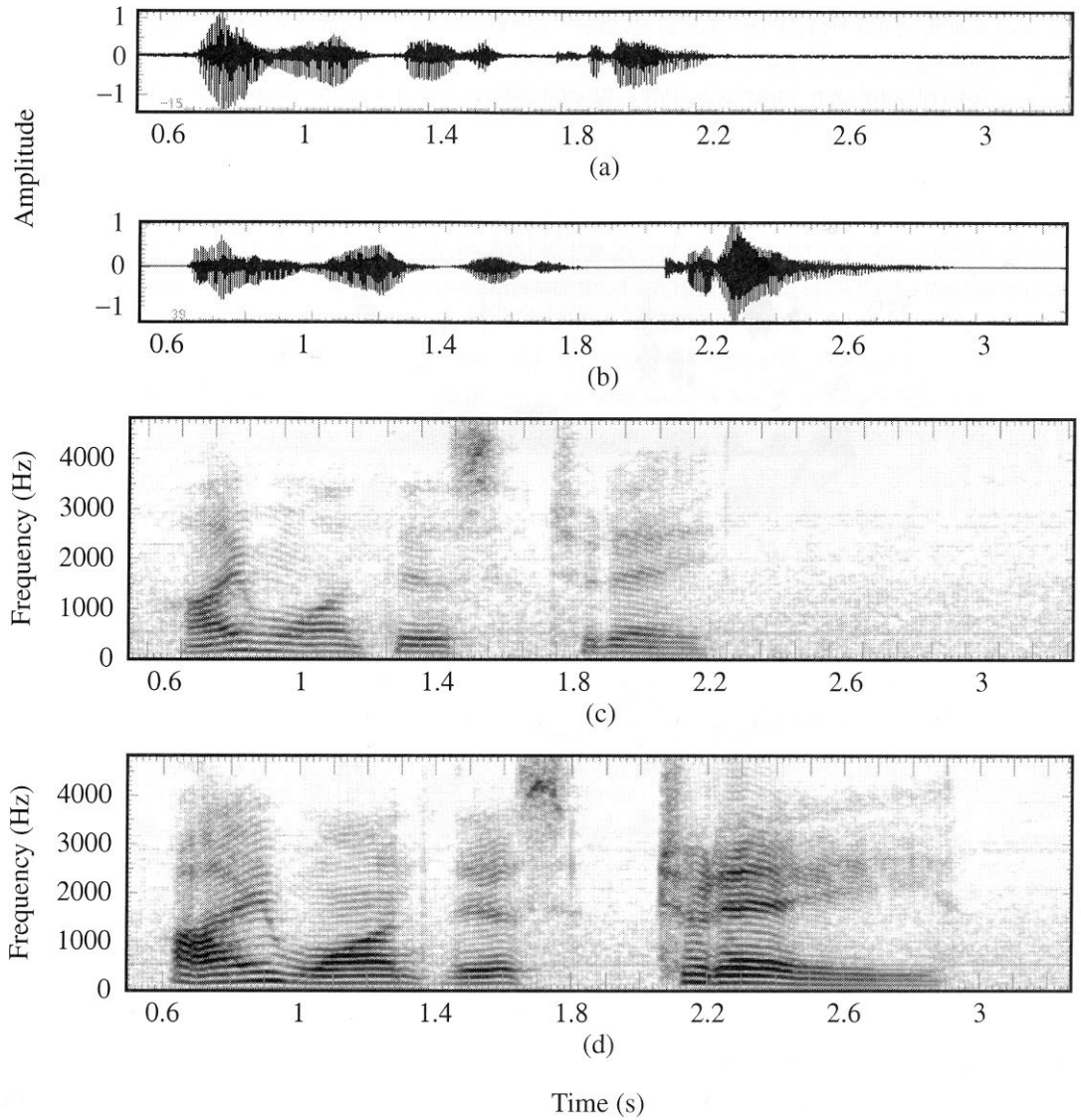

Figure 3.29 Comparison of "Please do this today," where "today" is spoken in a normal and stressed style: (a) waveform of normal; (b) waveform of stressed; (c)–(d) spectrograms of (a)–(b).

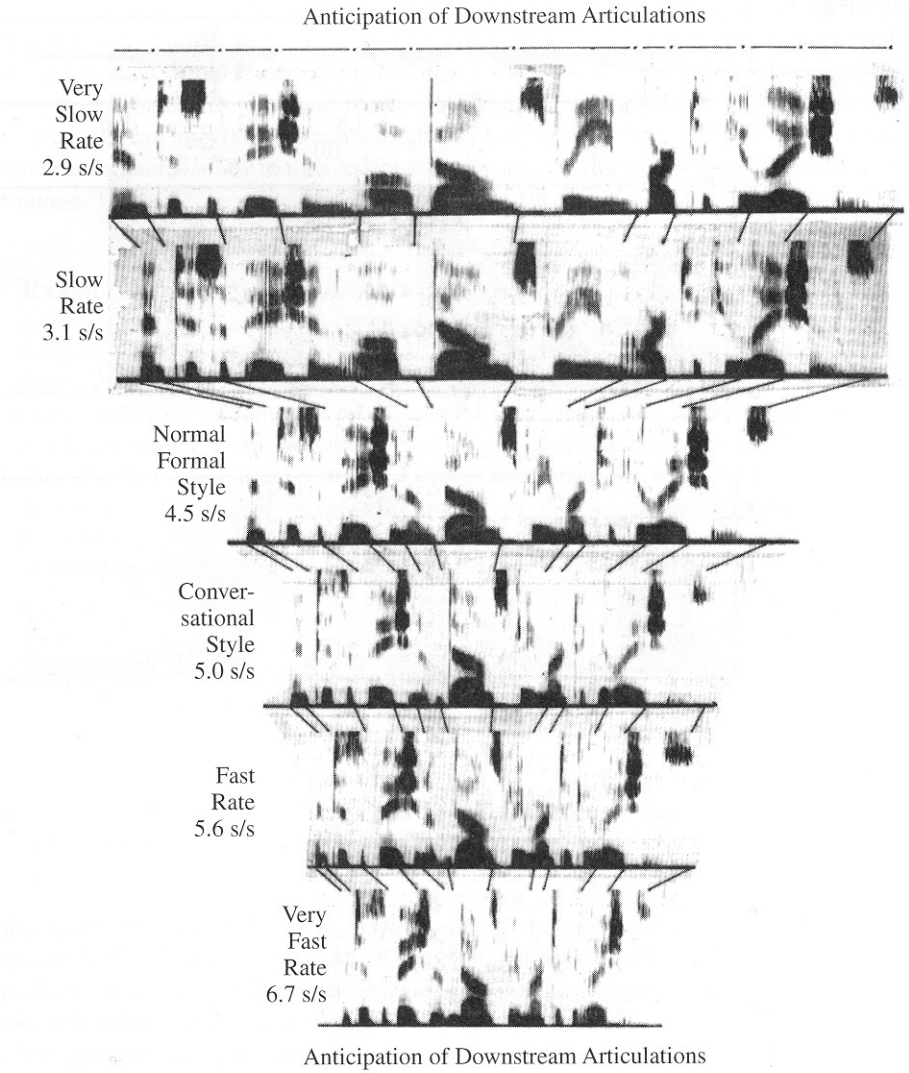

Figure 3.30 Spectrograms of utterance "anticipation of downstream articulators" at different rates of articulation (slow to fast) by Pickett [29]. Articulation rates are estimated in syllables per second (s/s). The lines drawn between successive spectrograms are synchronized with the same points within the utterance and show the extent of compression across different articulation rates.

SOURCE: J.M. Pickett, *The Sounds of Speech Communication* [29]. ©1980. Reprinted by permission of Allyn and Bacon.