of a successively time-compressed utterance spoken at different rates. The top spectrogram is of the utterance spoken deliberately slowly, the third spectrogram corresponds to the normal rate of speech, and the last to the fastest rate of speech. We would expect the consonants to be typically time-compressed less than the vowels because of greater limits to consonant compressibility. Plosive consonants require complete vocal tract closure and certain timing relations between the burst and onset of voicing, while fricative consonants require a suitable vocal tract constriction and also durational constraints. Pickett looked at the effect of articulation rate change on short and long vowels, and plosive and fricative consonants. Going from the slowest to the normal rate, both vowels and consonants were compressed by about 33%, more or less independently of the class. On the other hand, in going from normal to the fastest rate, the short vowels were compressed by 64% and the long vowels by 43%, while the plosive consonants were compressed by 36% and the fricative consonants by 26%.

3.6 Speech Perception

In this section, we first briefly describe the acoustic properties of speech sounds that are essential for phoneme discrimination by the auditory system, i.e., the acoustic aspects of the speech sound that act as *perceptual cues*. Our motivation is the goal of preserving such properties in speech signal processing.[14] We then take a glimpse at perceptual models of how the discriminatory features may be measured and processed by the listener.

3.6.1 Acoustic Cues

We are interested in acoustic components of a speech sound used by the listener to correctly perceive the underlying phoneme. In previous sections, we touched upon a number of these acoustic cues. Here we review and elaborate on these cues and give some additional insights. We look first at vowels and then at consonants.

Vowels — Formant frequencies have been determined to be a primary factor in identifying a vowel. We have already seen that the listening experiments of Peterson and Barney approximately map the first two formants (F_1 and F_2) to vowel identification. Higher formants also have a role in vowel identity [29]. As we will see in Chapter 4, because formant frequencies scale with tract length, we would expect the listener to normalize formant location in doing phoneme recognition. In fact, there is evidence that such normalization is performed making relative formant spacings essential features in vowel identification [36]. Another factor in vowel perception is nasalization, which is cued primarily by the bandwidth increase of the first formant (F_1) and the introduction of zeros [16]. As we pointed out earlier, however, unlike other languages, vowel nasalization is not used in the American English language to aid in phoneme discrimination.

Consonants — Consonant identification depends on a number of factors including the formants of the consonant, formant transitions into the formants of the following vowel, the voicing (or unvoicing) of the vocal folds during or near the consonant production, and the relative timing

[14] We are also interested in acoustic properties that convey speaker recognizability, some of which were described throughout this chapter.