Spectral Shaping — Under certain conditions, the relation between a glottal airflow velocity input and vocal tract airflow velocity output can be approximated by a linear filter with resonances, much like resonances of organ pipes and wind instruments. The resonance frequencies of the vocal tract are, in a speech science context, called *formant frequencies* or simply *formants*. The word "formant" also refers to the entire spectral contribution of a resonance so we often use the phrases "formant bandwidth" and "formant amplitude" (at the formant frequency). Formants change with different vocal tract configurations. With different vowels, for example, the jaw, teeth, lips, and tongue, are generally in different positions. Panel (a) of Figure 3.10 shows the tongue hump high in the front and back of the palate (upper wall of mouth), each position corresponding to different resonant cavities and thus different vowels.

The peaks of the spectrum of the vocal tract response correspond approximately to its formants. More specifically, when the vocal tract is modeled as a time-invariant all-pole linear system then, as we will see in Chapter 4, a pole at $z_o = r_o e^{j\omega_o}$ corresponds approximately to a vocal tract formant. The frequency of the formant is at $\omega = \omega_o$ and the bandwidth of the formant is determined by the distance of the pole from the unit circle (r_o). Because the poles of a real sequence typically occur in complex conjugate pairs (except for the case of a pole falling on the real axis), only the positive frequencies are used in defining the formant frequencies, and the formant bandwidth is computed over positive frequencies using, for example, the definitions of bandwidth in Chapter 2. Under the linear time-invariant all-pole assumption, each vocal tract shape is characterized by a collection of formants. Because the vocal tract is assumed stable with poles inside the unit circle, the vocal tract transfer function can be expressed either in product or partial fraction expansion form:

$$\begin{aligned} H(z) &= \frac{A}{\prod_{k=1}^{N_i}(1 - c_k z^{-1})(1 - c_k^* z^{-1})} \\ &= \sum_{k=1}^{N_i} \frac{\tilde{A}_k}{(1 - c_k z^{-1})(1 - c_k^* z^{-1})} \end{aligned} \tag{3.2}$$

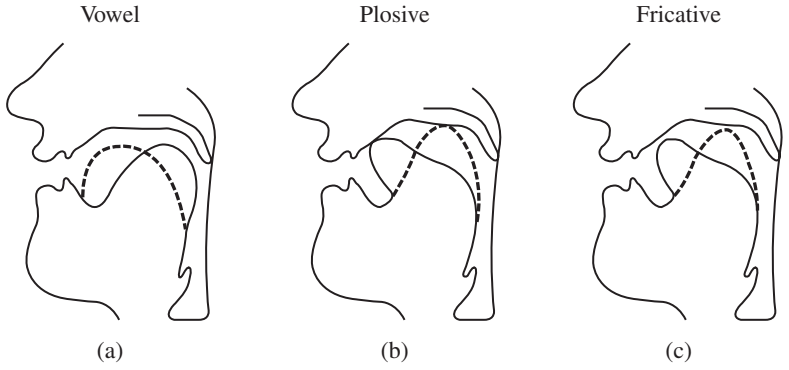

Figure 3.10 Illustration of changing vocal tract shapes for (a) vowels (having a periodic source), (b) plosives (having an impulsive source), and (c) fricatives (having a noise source).