audiograms and the acoustics of speech

As shown in the previous chapter, the speech patterns of English vary considerably in intensity, frequency and duration - all three dimensions of sound. It is the interaction of these three dimensions of sound that gives speech patterns their phonetic identities. There is no absolutely right way to produce speech sounds so, although there are differences between one talker and another, a wide range of dialects and accents can be understood.

To recapitulate the points made in Chapter 2 for the purposes of the present discussion, the most powerful sounds in speech are the central vowels - those made with the largest oral cavities and the widest opening of the mouth. The weakest sounds are those that restrict the breath flow maximally - the fricatives /ʃ/, /s/, and unvoiced th. Those with the highest frequency are these same fricative sounds, all of which have components up to and beyond 15,000 Hz. The longest and lowest-pitched speech patterns are those associated with prosody - the fundamental voice frequency and components of rhythm, intonation, and stress that collectively are known as suprasegmentals. The vowels are longer than most consonants, and their lower and stronger (F₁) components fall below 1000 Hz. The higher (F₂ and F₃) components of vowels fall mainly between 1000 and 3500 Hz. The consonants are spread widely across the frequency range
of speech, and vary greatly in duration and intensity. The shortest sounds in speech are the stops that occur at the end of syllables or words, such as up in the phrase up there.

The overall intensity of conversational speech normally varies within a range of about 30 dB from the quietest consonant (unvoiced th) to the loudest vowel (aw as in the word bawl). This amount of acoustic variation is known as the dynamic range of speech. Speech sounds within this dynamic range vary not just in relation to each other, but according to whether the words in which they occur are stressed or unstressed. For example, the two words for and four are pronounced similarly when said one by one. However, in a sentence, such as I waited there for four hours, the word four is likely to be stressed (emphasized). Stress in words results in more than just a change in intensity.

At a distance of about two yards (an average distance for conversation), the quietest sounds in everyday speech are about 30 dB greater than audiometric zero and the loudest, about 60 dB. Intensity is the most variable dimension of speech. In the absence of reflection and reverberation, sound increases in intensity by about 6 dB with every halving of distance, and decreases by the same amount with every doubling of distance, between the talker (or other source) and the listener. When a mother is talking to her hearing-impaired child at home and in the course of quiet play (as in the sketch presented as Figure 3-5) the acoustic conditions are at their most favorable. This is an important point and one that will be taken up again later on in relation to hearing aid use.

The overall frequency range of speech does not vary a great deal and, because the broad frequency characteristics of speech sounds are known, it is possible to relate certain aspects of speech quite closely to vertical bands drawn on an audiogram. By doing this one can relate speech reception to individual children's plotted hearing levels, and be better able to specify what sort of speech features may or may not be detected if they are fitted with appropriate hearing aids.

As suggested in the previous section on audiograms, it is possible to say that if there is no hearing in a certain frequency band one can predict that a certain range of sounds will not be heard. We can now add that if sound patterns of a certain frequency, as specified below, occur at levels 10 dB or so better than individuals' aided thresholds in that frequency range, those patterns will be detected. However, sensorineural hearing impairment may not only filter sounds, it may also distort them. If this is the case, then speech patterns as specified below may be
detected, but those having components in that frequency range may not be discriminated, identified, or comprehended because of such distortion.

**figure 3-5.** A mother talking to her young hearing-impaired child under optimal learning conditions: through one-to-one interaction; by following the child's interests; and by ensuring the best possible listening opportunities (being on the child's level, at close quarters, and in the favorable acoustic environment provided by the home). Such strong visual focus is rarely necessary when useful residual hearing exists, and maximum use is made of it.

**speech components and frequency bands**

In the following paragraphs the components of speech that fall in various frequency bands will be specified. Each band will be centered on one of the octave intervals represented on an audiogram form (125, 250, 500, 1000 Hz, and so on), and extend by a half octave in each direction. Such octave bands are shown by the vertical bands of shading shown on the the audiogram form presented as Figure 3-6. The lighter (banana-shaped) area in Figure 3-6 depicts the approximate intensity of
significant speech components as they occur in the writer's conversational speech at a distance of two yards. This area may be considered as the CLEAR zone - the zone in which conversational level elements in the acoustic range of speech occur. The approximate intensity levels of the vowels oo, ah, and ee, and the consonants sh and /s/ are shown in the frequency bands where most of their energy is to be found.

![Diagram](image)

**Figure 3-6.** An audiogram form showing octave bands centered on the audiometric frequencies (vertical bands of shading), and the CLEAR zone in which the Conversational Level Elements in the Acoustic Range of speech are to be found. The approximate center frequency and intensity of the formants of each sound used in The Five-Sound Test as produced by the writer at a distance of two yards are depicted in the CLEAR zone.

Many speech sounds share common properties. For example, /m/, /n/, and ng share nasality, all fricatives, such as /f/, /s/, and sh, share
turbulent breath stream, and all plosives, such as /p/ and /b/, produce bursts as breath is released. By and large, such common components fall within the same, or within adjacent bands of frequency. This being so, one can think of speech as having relatively few significant acoustic component features, rather than many individual speech sounds. To think this way is to be concerned with the similarities among speech sounds rather than with their differences.

To assess whether common components of sounds are present or absent in a child’s production or perception of speech can be a more economic way of carrying out an evaluation than by focusing on the presence or absence of the speech sounds (phonemes) themselves. For example, one can listen to a child’s speech in order to determine whether particular components, such as nasality or fricative turbulence, are present or absent. Once the significant components of speech that are causing the individual some difficulty have been specified, one can then find ways to provide cues that make them easier to hear, see or feel. For example, all the nasal sounds have a low frequency murmur, and all the fricatives have a significant amount of high frequency energy. If a child is neither perceiving nor producing certain of these components, then one must first question whether these speech sounds are being appropriately amplified or if other sensory information is adequate over the frequency range in which they occur. The speech components (rather than the speech sounds) that are of primary importance in speech, and that fall within each frequency band, are listed in Figure 3-7. To stress the importance of vowel-to-consonant and consonant-to-vowel formant energy, formant transitions in this list will be specified as $T_1$, $T_2$, and $T_3$ components - they are found in the frequency range over which the first, second, and third formants move as we talk.

**the intensity levels of significant speech components**

The intensity levels of the components of speech listed in Figure 3-7 vary considerably from one frequency band to another. The several examples that follow serve to indicate how extensive these differences are, and how important it is to take them into account when providing amplification for hearing impaired individuals.

The octave band centered on 1000 Hz contains the central vowels, which are more intense, but no more important, than any other vowels (see Figure 3-7). They happen to be so because the mouth is more widely
Figure 3.7. A chart showing the components of speech that fall within the octave bands centered on the audiometric frequencies 125 to 8000 Hz.
opened for vowels like aw and ah than for other vowels like oo and ee. In producing the latter sounds, the shape of the vocal tract and the height of the tongue obstruct breath flow and thus reduce their intensity. Anyone who has tried to tape-record words containing these and other vowels at a consistent level is familiar with their differences in intensity. Also, the harmonics of voice decrease in intensity as they become higher in frequency. This, in turn, causes speech sounds to be weaker in the octave band centered on 2000 Hz than in the octave band centered on 500 Hz.

Several cues that lie within each frequency range also vary in intensity. For example, the nasal formants are quieter than the harmonics of voiced sounds that are present in the octave bands centered on 250 and 500 Hz. The second formants of the front vowels are less intense than the fricative turbulence associated with both the sh and ch sounds in the octave band centered on 2000 Hz. The fricative turbulence of /s/ in the octave band centered at 4000 Hz is stronger than that of the /θ/, and of the unvoiced th that occurs in the same frequency range. Fricative turbulence associated with voiced affricates and voiced fricatives is also less intense, though no less significant, than that of their unvoiced counterparts. In short, the intensity levels of the significant components of speech do not correspond all that closely with the average intensity levels across the speech range.

As a first step in selecting amplification for severely and profoundly hearing-impaired children, the writer measured the levels of significant components in his own conversational speech at a distance of two yards. They occurred at levels that, when entered on an audiogram, created the banana-like shape shown as a shaded area across the audiogram form presented in Figure 3-6. This CLEAR zone does not correspond closely with the average intensity levels that occur in the various frequency bands across the long-term speech spectrum, because the components that are really important in speech discrimination and spoken language reception are less intense than the average levels of sounds at the high and low ends of the speech frequency range.

**the five-sound test**

It is useful to check on children’s hearing before setting out to teach them, by using the information contained in Figure 3-6. The approximate center frequencies of the formants of the three vowels oo, ah, and ee, and centers of energy in the consonants sh and /s/ are shown at their