

# A

---

## Acoustic Phonetics

---

Acoustic phonetics is the study of the acoustic characteristics of speech. Speech consists of variations in air pressure that result from physical disturbances of air molecules caused by the flow of air out of the lungs. This airflow makes the air molecules alternately crowd together and move apart (oscillate), creating increases and decreases, respectively, in air pressure. The resulting sound wave transmits these changes in pressure from speaker to hearer. Sound waves can be described in terms of physical properties such as cycle, period, frequency, and amplitude. These concepts are most easily illustrated when considering a simple wave corresponding to a pure tone. A cycle is a sequence of one increase and one decrease in air pressure. A period is the amount of time (expressed in seconds or milliseconds) that one cycle takes. Frequency is the number of cycles in one second, expressed in hertz (Hz). An increase in frequency usually results in an increase in perceived pitch. Amplitude refers to the magnitude of vibrations, with larger vibrations resulting in greater peaks of pressure (greater amplitude), which usually result in an increase in perceived loudness.

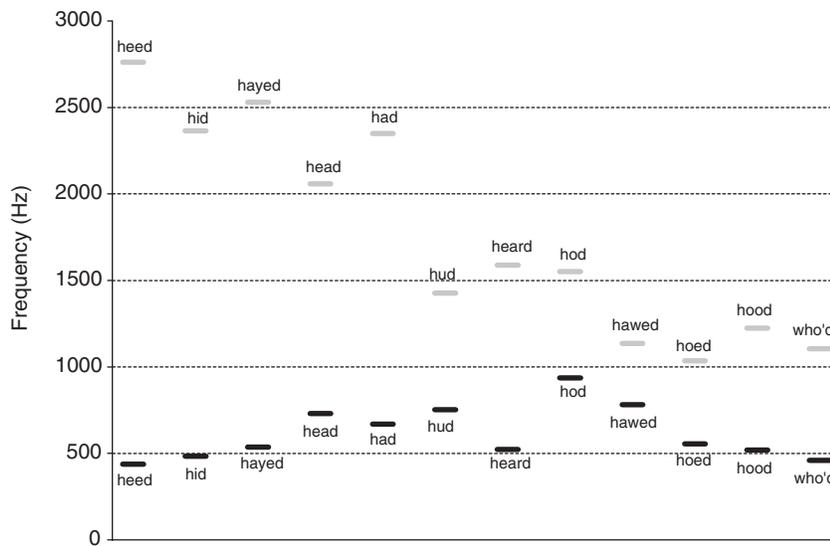
Unlike pure tones, which rarely occur in the environment, speech sounds are complex waves with combinations of different frequencies and amplitudes. However, as first stated by the French mathematician Fourier (1768–1830), any complex wave can be described as a combination of simple waves. A complex wave has a regular rate of repetition, known as the fundamental frequency ( $F_0$ ). Changes in  $F_0$  give rise to differences in perceived pitch, whereas changes in the number of constituent simple waves and their

amplitude relations result in perceived differences in timbre or quality.

Fourier's theorem enables us to describe speech sounds in terms of the frequency and amplitude of each of its constituent simple waves. Such a description is known as the spectrum of a sound. A spectrum is visually displayed as a plot of frequency vs. amplitude, with frequency represented from low to high along the horizontal axis and amplitude from low to high along the vertical axis.

The usual energy source for speech is the airstream generated by the lungs. This steady flow of air is converted into brief puffs of air by the vibrating vocal folds, two muscular folds housed in the larynx. The dominant way of conceptualizing the process of speech production is in terms of the source-filter theory, according to which the acoustic characteristics of speech can be understood as a result of a source component and a filter component. The source component is determined by the rate of vocal fold vibration, which in turn is affected by a number of factors, including the rate of airflow and the mass and stiffness of the vocal folds. The rate of vocal fold vibration directly determines the  $F_0$  of the waveform. The mean  $F_0$  for adult women is approximately 220 Hz, and approximately 130 Hz for adult men. "In addition to their role as properties of individual speech sounds,  $F_0$  and amplitude also signal emphasis, stress, and intonation."

For speech, the source component itself has a complex waveform, and its spectrum will typically show the highest energy at the lowest frequencies and a number of higher frequency components that



**Figure 1.** Frequencies of the first two formants of 12 vowels of American English, averaged across 48 adult female speakers. F1 is in black, F2 is in gray. (Source: Hillenbrand et al., *J. Acoust. Soc. Am.*, 1995)

systematically decrease in amplitude. This source component is subsequently modified by the vocal tract above the larynx, which acts as the filter. This filter enhances energy in certain frequency regions and suppresses energy in others, resulting in a spectrum with peaks and valleys, respectively. The peaks in the spectrum (local energy maxima) are known as formant frequencies. The lowest-frequency peak is known as the first formant, or F1, the next lowest is F2, and so on. The vocal tract filter is determined by the size and shape of the vocal tract and is therefore directly affected by the position and movement of the articulators such as the tongue, jaw, and lips.

Vowels are typically characterized in terms of the location of the first two formants, as illustrated in Figure 1 for the vowels of American English. For a given speaker, each vowel typically has a unique formant pattern. However, variation in vocal tract size among speakers often leads to a degree of formant overlap for different vowels.

Consonants can also be described in terms of their spectral properties. These sounds are produced with a complete or narrow constriction in the vocal tract, essentially creating a vocal tract with two sections: one behind and the other in front of the constriction. The length of the section in front of the constriction is one of the primary determinants of the spectra of these sounds. The longer this section (i.e. the farther back the constriction), the lower the frequency at which a concentration of energy occurs. For example, consonants like *k* and *g*, which are produced at the back of the mouth, are typically characterized by a concentration of energy between approximately 1,500 and 2,500 Hz, whereas more anterior consonants like *t* and *d* typically have a concentration of energy above 3,000 Hz. Similarly, the sibilants [ʃ, ʒ] produced in the middle of the mouth

have major energy around 2,500 to 3,500 Hz, whereas the more anterior ones [s, z] have major energy well above 4,000 to 5,000 Hz. However, in the case of consonants with a constriction toward the very front of the vocal tract, the extremely short section in front of the constriction does not result in clearly defined spectra. As a result, bilabial [b, p] and labiodental [f, v] consonants are described as having diffuse spectra, without any clear concentration of energy.

From a linguistic point of view, a detailed description of speech sounds in terms of their frequency, in addition to amplitude and duration, can elucidate the factors that shape sound categories and determine phonological processes both within and across languages. In addition, acoustic phonetic analysis may serve to quantify atypical speech patterns produced by nonnative speakers or speakers with specific speech disorders.

## References

- Asher, R., and Eugenie Henderson (eds.) 1981. *Towards a history of phonetics*. Edinburgh: Edinburgh University Press.
- Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague: Mouton.
- Flanagan, James L. 1965. *Speech analysis synthesis and perception*. Berlin: Springer-Verlag.
- Hardcastle, William, and John Laver (eds.) 1997. *The handbook of phonetic sciences*. Oxford and Cambridge, MA: Blackwell.
- Helmholtz, Hermann von. 1877. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: F. Vieweg; as *On the sensations of tone as a physiological basis for the theory of music*, translated by Alexander J. Ellis, London: Longmans, Green, 1885.
- Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97(5). 3099–111.
- Jakobson, Roman, Gunnar Fant, and Morris Halle (eds.) 1952.

- Preliminaries to speech analysis. Cambridge, MA: MIT Press.
- Johnson, Keith. 1997. Acoustic and auditory phonetics. Oxford and Cambridge, MA: Blackwell.
- Kent, Raymond D., Bishnu S. Atal, and Joanne L. Miller (eds.) 1991. Papers in Speech Communication: speech production. Woodbury, New York: Acoustical Society of America.
- Ladefoged, Peter. 1962. Elements of acoustic phonetics. Chicago and London: The University of Chicago Press, 2nd edition, 1996.
- Lehiste, Ilse. 1970. Suprasegmentals. Cambridge, MA: MIT Press.
- Lieberman, Philip, and Sheila E. Blumstein. 1988. Speech physiology, speech perception, and acoustic phonetics. Cambridge: Cambridge University Press.
- Stevens, Kenneth N. 1998. Acoustic phonetics. Cambridge, MA: MIT Press.

ALLARD JONGMAN