

XV. SPEECH ANALYSIS

Prof. M. Halle
G. W. Hughes
J. -P. A. Radley

INVESTIGATION OF CONSONANTS

This quarter was devoted to a study of the energy spectra of fricative and affricate consonants. As in the work previously reported, the purpose was to see what information can be obtained from the spectrum for the identification of the various sounds of speech.

Fricative and affricate consonants are produced by forcing air through a constriction narrow enough to generate turbulence. Fricatives have a gradual onset; affricates have a sudden, stop-like onset. Examples of the former are /s/ as in "see," /v/ as in "veal," /ʃ/ as in "shoe"; examples of the latter are /tʃ/ as in "choke" and /dʒ/ as in "joke."

The procedure was in general like that followed in our investigation of the stop consonants (see the Quarterly Progress Report of April 15, 1954). The samples studied were gated portions, 50 msec in length, of syllables spoken by a male and a female speaker. The gate was set after examination of the whole syllable on an oscilloscope. The gated sample was passed through a Hewlett Packard model 300 A wave analyzer with a passband that was substantially flat over 300 cps. The filtered signal was passed through a squarer, integrated, and measured. Between 1 kc/sec and 10 kc/sec, points at intervals of 500 cps were measured. A measurement was also taken at 600 cps and at any maxima or minima that did not coincide with the standard measurement points. Some sample spectra are shown in Fig. XV-1.

The most striking thing observed was the consistency with which all /s/ and /z/ sounds measured had peaks in the high-frequency region, at 6 kc/sec or above. No other sounds showed peaks in this region for our female speaker; however, for our male speaker /f/ and /v/ sounds also had peaks in the same region. In the latter instance, it was impossible to find any criterion that would distinguish the spectra of /f/ and /v/ from those of /s/ and /z/. We noticed, however, that /f/ and /v/ were of lower overall intensity than /s/ and /z/.

The following rule was formulated to separate /s/ and /z/ sounds (and /f/ and /v/ for our male speaker) from the others: Locate the peak of the spectrum. If the peak is at 6 kc/sec or above and at the same time exceeds all lower maxima by 4 db, the sound is /s/ or /z/ (or, for our male speaker, /f/ or /v/).

To identify the remaining sounds the following measurements had to be taken:

1. The average power level in the region between 600 kc/sec and 1500 kc/sec.
2. (a) The average power level of a 1 kc/sec band centered at the peak, if the peak is between 2 kc/sec and 4 kc/sec; (b) the average power level between 2 kc/sec and

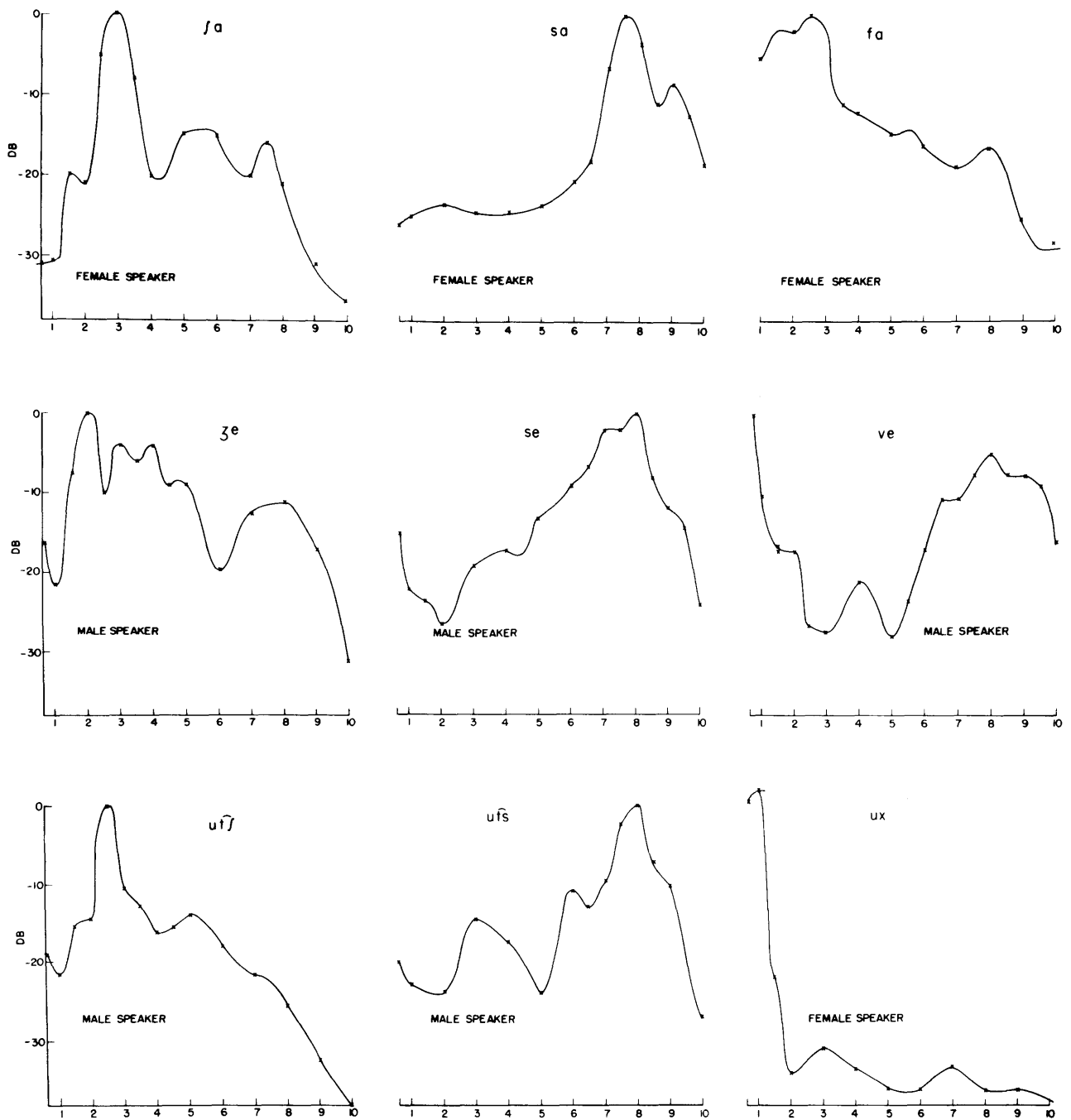


Fig. XV-1

Spectra of fricative and affricate consonants. Spectra were obtained by measuring a 50-msec portion of the consonantal part in the syllables indicated in the upper right-hand corner of the graphs. Numbers on abscissas indicate frequency in kilocycles per second.

3 kc/sec, if the peak is below 2 kc/sec; (c) the average power level between 3 kc/sec and 4 kc/sec, if the peak is above 4 kc/sec.

For voiced sounds 10 db was subtracted from measurement 1 in order to eliminate the contribution of the voicing component.

Compact (1) consonants were marked by large differences between measurements 1 and 2. Diffuse (1) consonants (in this instance /f/ or /v/) gave small values for the difference. A predominance in the 600 cps to 1500 cps region identifies the grave compact consonants (1) whereas a predominance in the 2 kc/sec to 4 kc/sec region identifies acute compact consonants.

Of 116 syllables that were subject to the second set of criteria, correct identifications were obtained in 101 of the cases as follows: compact diffuse, 16 out of 19 cases; compact grave, 61 out of 68 cases; /f/ or /v/, 24 out of 29 cases.

M. Halle, G. W. Hughes

References

1. For an explanation of these terms see R. Jakobson, C. G. M. Fant, and M. Halle, Technical Report No. 13, Acoustics Laboratory, M. I. T., May 1952.