XVI. SPEECH ANALYSIS*

Prof. M. Halle
G. W. Hughes

RESEARCH OBJECTIVES

Various branches of science treat of the acoustical properties of speech. Physical acoustics views speech as special vibratory events propagating in a medium; psychophysics, as a particular class of auditory stimuli perceived by man; phonetics, as a sequence of discrete events in time (speech sounds, phones, phonemes, and so on) serving a communicative function.

Our interest in speech is that of the phonetician; hence the methodological demand that each utterance shall always be perceived as a sequence of individual "sounds" is fundamental to our research activity. Our aim is to find the cues in the speech wave that make it possible to transcribe utterances in terms of a small number of discrete symbols; that is, in terms of some kind of alphabet.

Most investigators have tacitly or overtly studied speech in terms of discrete events. This assumption, however, is not usually central in their work because other requirements, such as ease of instrumentation, the necessity for bandwidth economies in transmission, and so forth, take precedence. The importance of carrying through a complete investigation based on a consistent view of speech as a chain of discrete events can perhaps be appreciated if it is realized that the successful completion of such a research program would provide a firm physical foundation for the science of linguistics, shed light on a number of complex problems in the psychology of perception and, finally, fill in the gaps in the most promising models of the communication process that postulate the existence of a code unit smaller than the smallest meaningful entity.

M. Halle

A. ACOUSTICAL PROPERTIES OF STOP SOUNDS

The stop sounds /p/ as in "pay," /t/ as in "ate," /k/ as in "key," /b/ as in "bee," /d/ as in "aid," and /g/ as in "gay" are characterized by being produced with a rapid closure and/or opening of the vocal tract. During the period of closure, the vocal cords may or may not vibrate; if they do, we have a "voiced" stop; for example, /b/ as in "bay"; if they do not, we have an "unvoiced" stop; for example, /p/ as in "ape."

In discussing stop consonants it is customary to consider two aspects separately: the burst and the transition of the formants in the adjacent vowels. There are two reasons for this. First, there are instances in which only a burst or only a formant transition is present. For example, we distinguish between "whisps," and "whisks," although the stops /p/ and /k/ are not adjacent to any vowel and hence there can be no talk of formant transitions. On the other hand, in imploded stops like the /p/ and /k/ in the normal pronunciation of the words "apt" and "act" there are only single bursts but /pt/ and /kt/ are perceived, respectively. Second, in a visual presentation, such as in an oscillogram or sonagram of a stop sound pronounced with both burst and formant transition, the burst and formant transition are easily distinguishable as separate events in time. The seeming independence of these two cues and the relative ease with which they can be distinguished suggest that they be studied separately.

*Supported in part by National Science Foundation.

In spoken English it is usual to distinguish three classes of stop consonants, each associated with a different point of closure: /p/ and /b/, where the closure is made at the lips; /t/ and /d/, where the closure is made behind the teeth; and /k/ and /g/, where the closure is made somewhere along the palate. The long range objective of our studies is to find a set of measurements on the acoustical stimulus that will permit us to classify it as belonging to one of the three classes just mentioned: /p/ /b/, /t/ /d/, or /k/ /g/.

In the past quarter, investigations were conducted on stops produced with a clearly defined burst and in position adjacent to a vowel. A master tape was prepared of a number of speakers, male and female, reading a list of short English words. Each sample was re-recorded on a tape loop and subjected to two separate investigations: first, the burst was gated out, its energy density spectrum measured, and some gross spectral properties noted; second, the transitions in the vowel, if any, were studied by means of Sonagraph records.

In Fig. XVI-1 a number of typical spectra of stop consonants is shown. The two top rows present spectra that show clearly the salient features of the burst. Unlike the other stops, the /k/ and /g/ have clearly defined peaks in a central region of the spectrum. The former are distinguished by the fact that /p/ and /b/ have essentially falling spectra with increasing frequency, while /t/ and /d/ have spectra that are either rising or more or less flat in the region below 6 kc/sec. The essential point to be considered is the relatively large amount of high-frequency energy in /t/ and /d/ and the absence of it in /p/ and /b/.

In the bottom row we have presented a few spectra that deviate. We have a spectrum of /p/ with a clearly marked peak in a central region that we would have to classify as a /k/, and we have a spectrum of /k/ without any clearly marked peak, that looks, therefore, quite similar to a /t/. The /k/ spectrum was taken from the word "kid," in which the /k/ is produced quite far in the front of the mouth fairly near the point of articulation of /t/.

On the basis of the detailed spectra we devised a number of measurements for determining certain gross spectral characteristics that correlate with the three classes /p-b/, /t-d/, /k-g/. We attempted to determine the presence of a peak in a central region and the contribution of the upper frequency bands to the total energy. Presence of a peak in the central region would class the sound as /k/ or /g/; the absence of a peak would be taken as an indication that the sound is not /k/ or /g/. The sounds lacking centrally located peaks could then be classified either as /p/ and /b/ or /t/ and /d/ by estimating the contribution of the high frequencies. The sounds /t/ and /d/ had a considerable amount of energy in the upper regions, /p/ and /b/ did not. The application of our criteria to a larger body of material is in progress.

The study of the transitions has been completed and results will be reported in the

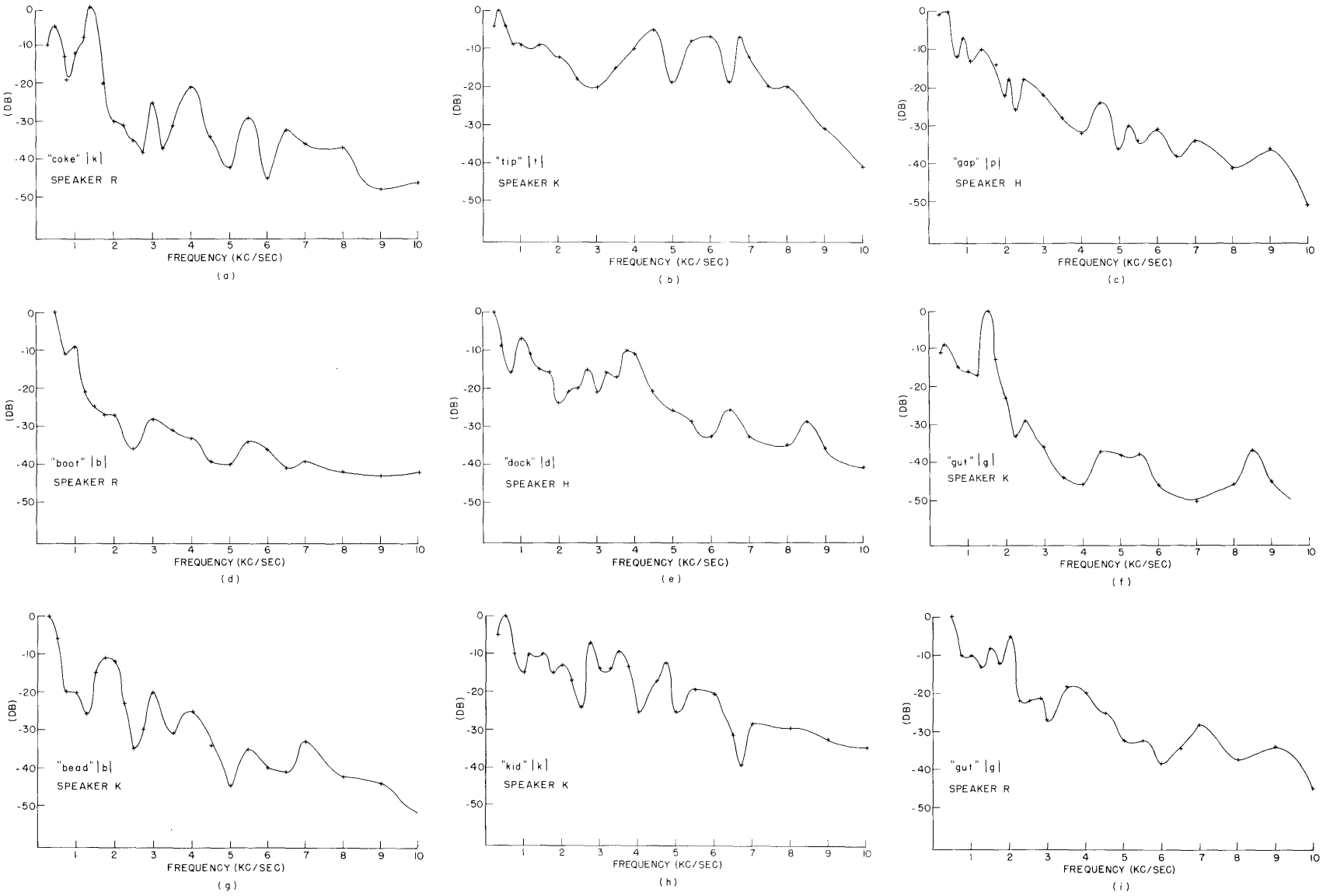

Fig. XVI-1. Typical spectra of stop consonants: (a-c) unvoiced stops; (d-f) voiced stops; (g-i) deviants.

forthcoming Master's thesis of J. -P. A. Radley.

We are conducting a series of perceptual tests in which the two major cues, the burst and the transition, are isolated and presented to a group of listeners and their responses correlated with various acoustical features of the stimuli.

M. Halle, G. W. Hughes