

XVI. SPEECH ANALYSIS*

Prof. M. Halle
G. W. Hughes

R. Capraro

J. Emerson
P. Lieberman

RESEARCH OBJECTIVES

The process of communication normally involves a code, a transmitter and a receiver, a channel, and a signal. In speech communication the code is the language in which the utterance is spoken, the transmitter and the receiver are the speaker and the listener, the channel is normally the ambient air of the human environment, and the signal is the acoustic wave produced by the speaker and received by the listener. In the cases that are best known to communication engineers the code and the properties of the transmitter and of the receiver are completely specified. In speech communication our knowledge of these factors is still fragmentary; and it is obvious that nothing definitive can be said about the signal until more information is available on the factors that determine it; that is, on the nature of the language and the properties of the speaker and the listener.

The main problems that confront speech analysis are therefore the following: (a) What is the nature of human language, in general, and of various national languages, in particular? How does the structure of the language determine the physical properties of the utterances? (b) What are the capabilities of the human vocal apparatus as a source of speech signals? What limitations are imposed on the signal by the restricted control that man has over his vocal organs? (c) How does man perceive sounds, in general, and speech-like sounds, in particular? What are the limitations of the human organism in dealing with acoustic stimuli that serve as vehicles for the identification of messages? (d) What are the physical properties of utterances in various languages? In particular, what are the properties that serve to distinguish utterances that are different from each other?

M. Halle

A. STUDY OF INTONATION

It is well known that the intonation pattern of a word or phrase is determined by its connotation. For example, "ma" spoken plaintively differs from "ma" spoken in a commanding tone. Indeed, the very concept of "tone" implies that intonation patterns contain significant information.

In a proposed system of analysis the sentence intonations are described by means of four numbers, called "pitch levels," and three symbols, called "terminals." This notation, when placed on the representation of any sentence, enables a reader to reproduce its intonation pattern correctly.

Lists of isolated words and phrases were read by two informants who were able to read the pitch level notation. Their responses were tape-recorded, and the intonation patterns were identified by a trained observer who used the proposed pitch level system. The recorded speech waves were then displayed on a dual-beam oscilloscope, together with the rectified, smoothed envelope, and photographed with an electrically driven motion picture camera at 400 inches per minute. The glottal excitation frequency and

*This work was supported in part by the National Science Foundation.

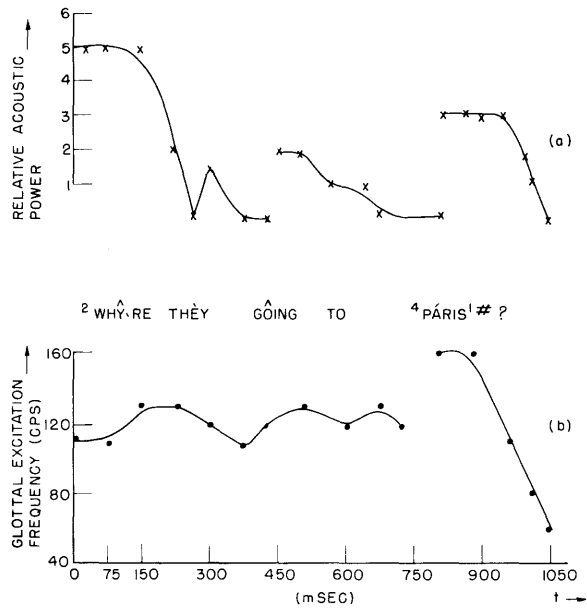


Fig. XVI-1. Glottal excitation frequency and relative acoustic power of an isolated sentence. (a) Relative acoustic power. (b) Glottal excitation frequency.

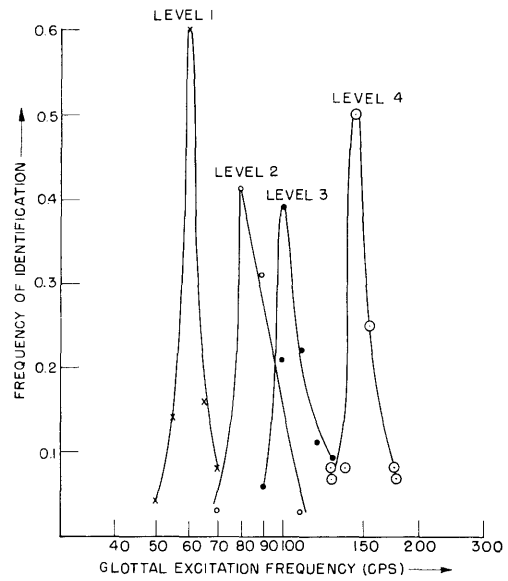


Fig. XVI-2. Frequency of identification of glottal excitation frequencies with pitch levels for informant A.

(XVI. SPEECH ANALYSIS)

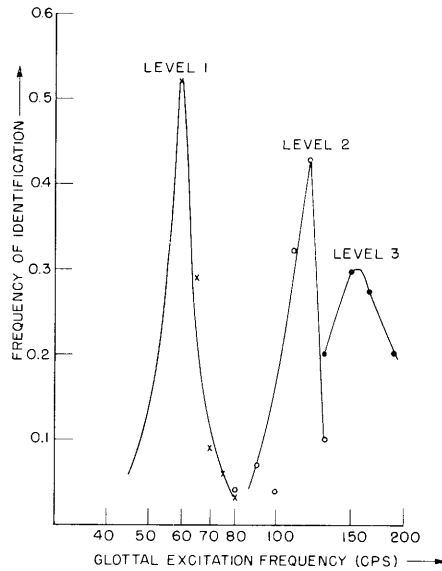


Fig. XVI-3. Frequency of identification of glottal excitation frequencies with pitch levels for informant B.

total audio power could then be measured by viewing the processed film on a calibrated microfilm reader. (See Fig. XVI-1.)

Figures XVI-2 and XVI-3 present the distribution of glottal excitation frequency within each pitch level identification for informants A and B. Both informants group their pitch levels in a fairly distinct way. The spread in the groupings (increase in variance) as we go to higher glottal excitation frequencies is due in part to the visual measurement of the pitch period from the film, since the error in frequency measurement increased at higher frequencies, from approximately 2.5 per cent at 60 cps to 10 per cent at 260 cps. Pitch level 4 could not be accurately determined for informant B, as we did not obtain enough samples in this initial experiment.

Transitions of one octave in normal speech, e. g., from level 1 to level 2 for subject B, were quite common, and changes of two octaves were not uncommon. These transitions are probably not consciously noted in ordinary speech, since speech does not consist of pure steady-state or quasi steady-state tones.

The exact nature of the terminal effects is not quite clear at this time. What is obvious is that there is no absolute change in the glottal excitation frequency corresponding to the various aurally perceptible changes in pitch that often occur at the end of a phrase; for example, when a question is asked or the emphatic form is used.

Further studies will examine the correlation of the glottal excitation frequency with pitch for a wider class of informants. There is some possibility that the pitch levels are associated with other factors, particularly in the case of level 4. The terminal conditions seem to be of at least two types: those in which a slow change in the glottal

(XVI. SPEECH ANALYSIS)

pitch frequency appears to be the significant factor and those in which the rate of change is significant. The data support the view that pitch level is directly related to the glottal excitation frequency. The correlation between pitch level and intensity is, on the contrary, extremely weak. Further work continues.

P. Lieberman, R. B. Lees