II A TENTATIVE SURVEY OF THE DISTINCTIVE FEATURES

2.1 PREFATORY ACOUSTICAL REMARKS

In the sound spectrograms* the frequency-intensity pattern of speech is portrayed as a function of time. In this "running frequency analysis" the statistical properties of the speech wave are sampled within time intervals that are short compared to the duration of a phoneme. The spectrograms and the supplementary "cross sections" of intensity vs. frequency provide a source of information that may be rather confusing unless an optimal set of parameters is used in the analysis. These parameters can best be discovered by an analysis of language into distinctive features.

The speech wave may be considered as the output of a linear network; i.e., the vocal tract coupled to one or more sources. The speech wave has no other properties than those of the sources and the network. This relation may be written

$$W = T \cdot S$$

where W represents the speech wave, T the transfer function of the network, and S the source. Two simultaneous sources may be handled by superposition:

$$W = T_1S_1 + T_2S_2$$

Speech analysis shows that only a very limited number of characteristics of the source and of the transfer functions are utilized in the various languages of the world for semantic discriminations. These characteristics are described in the following paragraphs.

2.11 Properties of the Source Function Utilized in Language

2.111 Type of Source. There are basically two kinds of sources, periodic and noise sources. A periodic source is manifested by a characteristic harmonic structure in the spectrogram. A noise source, on the other hand, causes an irregular distribution of energy in the time dimension. These two sources can be simultaneously active in the production of a single phoneme.

2.112 Number of Sources. Some sounds such as [v] or [z] have two sources. One of these is located at a point of maximum stricture in the vocal tract, while the other, i.e. the so-called voice, is located at the larynx and is more or less periodic. A source which lies above the larynx in the vocal tract produces anti-resonances in the transfer function (cf. 2.122).

* The sound spectrograms to which reference is made in this report either are of the type produced by the Kay Electric Company Sonagraph or are from the book Visible Speech by Potter, Kopp, and Green (1).

2.113 Transient Effects. The manner in which the source is turned on and off is linguistically significant. We distinguish abrupt onsets and declines from smooth ones. For example, in the phoneme /ʃ̂/ as in chip we have an abrupt onset, while in /ʃ/ as in ship the onset is smooth.

2.12 Transfer Functions Utilized in Language

2.121 General Properties. In the mathematical treatment of the transmission properties of the vocal tract it has been found convenient to utilize the techniques and concepts developed for network analysis (2). One of the standard cases treated in network analysis is a lossless transmission line with no branches in parallel, where the input (source) is located at one end of the line and the output measured at the other. The intensity vs. frequency spectrum of the output of such a transmission line can be completely specified by stating the frequencies at which the output will be infinite (resonance). In network analysis it is usual to call these resonance frequencies poles.

When in a lossless line, some of these conditions are not fulfilled, e.g. the source is not located at the end of the line, then the output will deviate from that in the case discussed above: in certain frequency regions there will be no output. It is possible to view the deviation as due to an anti-resonance or zero which suppresses the energy in a given frequency region; i.e. acts like a resonance in reverse. Thus, in order to specify the intensity vs. frequency spectrum of the output of any lossless system, it is sufficient to state the frequencies of the poles and the zeros (if any).

When a system contains small losses, the responses at resonance and anti-resonance are finite. In the complex frequency notation the poles and the zeros then have two parts, one giving the frequency location of the resonance or anti-resonance, and the other specifying the amount of damping (damping constant).

The poles depend primarily on the electrical properties of the series transmission line. In the case of speech this means that the poles depend upon the configuration of the vocal tract. The zeros, on the other hand, depend primarily upon the interaction of parallel branches. In speech that would mean that they depend upon the interaction of the two resonating systems in parallel which are created either by a) the opening of a supplementary passage or by b) the non-terminal location of the source.

When the location of a zero is close to that of a pole, the zero tends to cancel the effect of the pole. As the separation between them increases, the suppression decreases.

2.122 Location of Source. In general the zeros occur at frequencies at which the impedance arising from the source in the direction opposite to the air flow is infinite. A source located at the larynx will cause no anti-resonances of importance in the transfer function. It is for this reason that we can specify vowels completely by the poles which give the frequency positions and bandwidths (damping constants) of the formants. A source located above the larynx in the vocal tract between cavities that have finite coupling to each other introduces zeros into the transfer function.

2.123 Shape of the Vocal Tract. The poles of the transfer function are primarily related to the geometrical configuration of the vocal tract and are independent of the source and its location. Calculations on the basis of x-ray data lead to substantial agreement between the poles and zeros of the predicted and measured spectra (3).

2.13 Neutral Position of the Vocal Tract

In the following, reference will be made to the neutral position of the vocal tract. This is the position of the vocal organs for producing a very open [æ] With respect to its acoustic results this articulatory position can best be approximated by a single tube closed at one end. As is well known, a tube of length L closed at one end resonates at frequencies where L is an odd multiple of one quarter wavelength. Since the average length of the vocal tract of males is about 17.5 cm, the resonances appear at approximately 500, 1500, 2500 cps. etc. The neutral position is of importance for predicting the effects on formant positions of variations in the over-all length of the vocal cavity of different individuals (3). It also serves as a reference point for the tenseness feature (cf. below 2.431).

2.14 Phoneme Boundaries

For practical purposes each phoneme can be represented by a quasi-stationary spectrum in which the transfer function is invariable with respect to time, except in the manner stated for transient effects (cf. 2.113). These transient effects, which are produced by rapid variations in the source function, may serve to delimit the individual phonemes in the chain. Rapid variations in the transfer function caused by swift changes in the position of the articulating organs also indicate the beginning or end (boundary) of a phoneme. Here, however, the minimum rate of change must be determined experimentally for each case. Rapid fluctuations in the over-all intensity of the speech wave provide an additional means for determining the location of a phoneme boundary.

2.2 FUNDAMENTAL SOURCE FEATURES

This class consists of two binary oppositions: vocalic vs. non-vocalic and consonantal vs. non-consonantal.

2.21 Vocalic vs. Non-Vocalic

Phonemes possessing the vocalic feature have a single periodic ("voice") source whose onset is not abrupt.