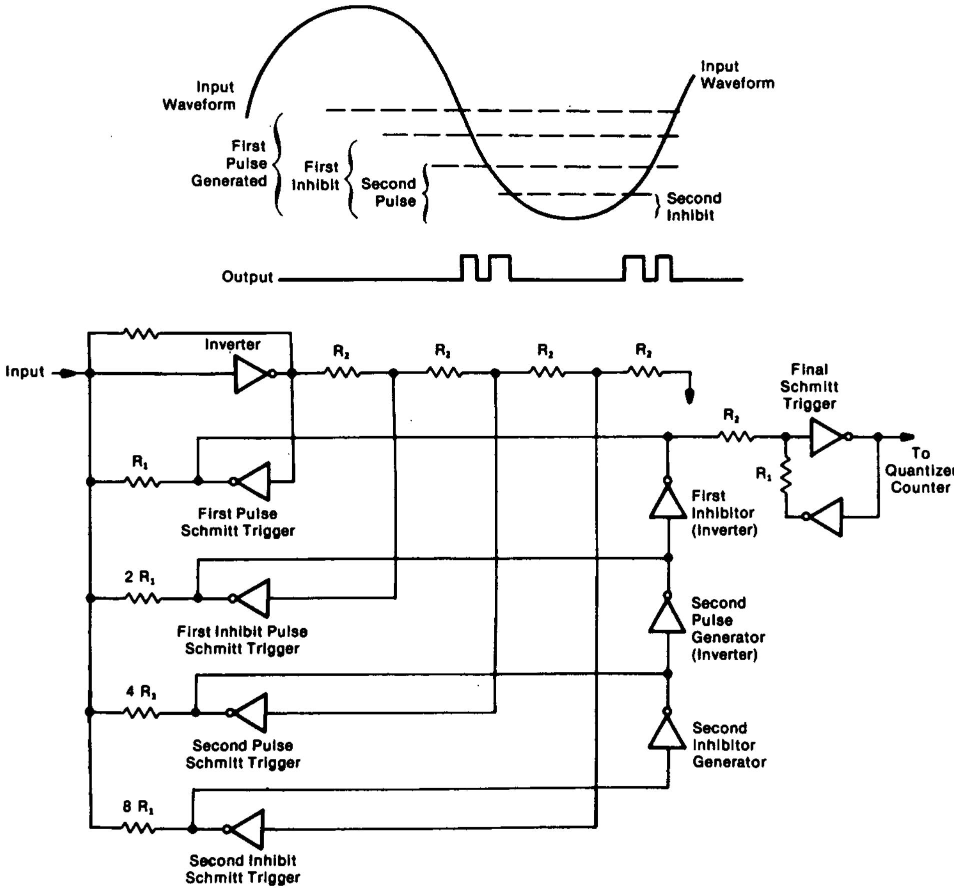
NASA TECH BRIEF
Goddard Space Flight Center

Techniques for Decoding Speech Phonemes and Sounds: A Concept

The human voice is a very complex communications system. Speech is produced by a combination of glottal harmonics and air-noise sounds. Several methods have been attempted to decode human speech in an effort to obtain some form of communication between man and machine and to help deaf people. To date, studies using zero-crossing detectors or spectrum analysis in speech recognition have had limited success.

Recently, two new techniques have been studied. Both involve the conversion of speech sounds into machine-compatible pulse trains. In one method a voltage-level quantizer is used. The quantizer produces a number of output pulses proportional to the amplitude characteristics of vowel-type phoneme waveforms. The second technique involves logic operations. Pulses produced by the quantizer of the first speech formants are compared with pulses



Quantizer Operation

(continued overleaf)

produced by the second formants. This yields better separation in distinguishing the features of vowel-type phonemes.

The quantizer scheme is illustrated using a simplified waveform input. If only this frequency is present, the automatic gain control (AGC) would make the waveform full amplitude. As the rising waveform reaches a voltage level above the zero average, one approach is to turn on the final Schmitt trigger. As the waveform achieves another height, a Schmitt trigger fires, causing an inhibit of the first inverter output. This alternating operation can be repeated as often as needed, producing pulses proportional to amplitude. As the amplitude of the sine wave drops, the alternating operations continue producing more pulses proportional to changes in amplitude.

Thus the number of pulses produced per unit time is proportional to frequency and amplitude. Therefore if two frequency components are fed through the AGC amplifier, there would be a smaller number of pulses produced since the sine wave input into the quantizer would have a smaller amplitude. The result is that the quantizer resolves the resonances of vowel-type speech better than zero-crossing limiters do, but it does not produce the vast amount of information that the spectrum analysis creates.

Many experts in the field believe the major frequency components of vowels are very close together. The second decoding technique helps to widen or accentuate the differences between the vowels and compensates for mouth size differences of speakers. The quantizer output pulses feed a counter which is frozen by the second formant counter if the second formant produces 16 pulses or more. The result is that fewer quantizer pulses get counted than previously. The output produced is the ratio of these two formant frequencies multiplied by 16. If the second formant counter does not produce 16 pulses within the sample period, then basically the first formant quantizer is counted proportional to its frequency.

The third formant should be counted by the same technique as the first formant was counted. This compensates for many errors created by speech waveforms. The result will be two sets of six bits which could easily feed read-only memories to define the the phoneme spaces. This phoneme space allows decoding of continuous speech to make directly any code for phoneme characters needed by the user.

Note:

Requests for further information may be directed to:

Technology Utilization Officer
Goddard Space Flight Center
Code 704.1
Greenbelt, Maryland 20771
Reference: TSP75-10086

Patent status:

This invention is owned by NASA, and a patent application has been filed. Inquiries concerning non-exclusive or exclusive license for its commercial development should be addressed to:

Patent Counsel
Goddard Space Flight Center
Code 204
Greenbelt, Maryland 20771

Source: Donald C. Lokerson
and Howard G. Holby
Goddard Space Flight Center
(GSC-11898)

Categories: 02 (Electronics Systems)
05 (Life Sciences)