NEW METHODS OF ANALYSIS IN SPEECH ACOUSTICS

Hisashi Wakita, Speech Communications Research Laboratory, Inc., 806 West Adams Boulevard, Los Angeles, California 90007, U.S.A.

Introduction

The recent development in digital techniques has brought substantial innovations to methods and techniques for acoustical analysis of speech sounds. The advantages of using digital computers over the conventional analog techniques are that the analysis processes can be repeated precisely and that the control of parameters is relatively easy. The use of a digital computer also permits the processing of a large amount of data within a relatively short period of time with satisfactory accuracy. Because of the above advantages, digital techniques are playing a more and more important role in speech research. This paper, thus, concerns primarily the recent digital techniques in speech analysis, particularly the linear prediction method, with special attention to its advantages and disadvantages, and also the limitations involved in the techniques.

Analysis Techniques

Among the current digital techniques for speech analysis, the linear prediction method (LP method) is predominantly used by many researchers. The LP method is suited for digitally processing the speech data to extract some acoustic parameters (Wakita, 1976).

a) Formant Analysis

The LP method assumes a simple speech production model

which consists of an excitation source and a transmission system. The excitation source is assumed to be an impulse generator, and thus the transmission system includes a glottal shaping filter, a vocal tract filter, and a lip radiation filter. Since the vocal tract filter does not assume a nasal tract, the LP model primarily assumes the production of voiced nonnasal sounds. Analysis by the LP method is an attempt to match a speech segment to the above ideal model so that the error between the matched model and the ideal one becomes minimum on a least mean square error criterion. The transmission system, thus optimally determined, is represented by either a set of predictive coefficients or a set of reflection coefficients. A smooth spectral envelope can be obtained by applying a Fourier transform to a given set of predictive coefficients. The formant frequencies can be obtained either by searching for peaks in the spectral envelope or by precisely computing the roots of the polynomial of predictive coefficients. By the above procedure, the formant frequencies are fairly accurately estimated. The formant bandwidth or amplitude can also be computed, but its accuracy is sometimes erroneous. Like other methods for estimating formant frequencies, the LP method does not solve the problem of estimating formant frequencies for speech sounds of very high pitch. This problem rather inherently exists in the speech signal, and thus it is rather intrinsic to any method.

b) Detection of Fundamental Frequency

In the LP analysis, after the vocal tract characteristics are extracted from the speech signal, the information on the fundamental frequency still remains in the residual. The

periodicity of the speech signal can thus be extracted from the residual signal by the autocorrelation method. Besides the LP method, there are various methods for extracting the fundamental frequency. Most of them have a high performance and have their advantages and disadvantages (Rabiner et al., 1976). The choice of a method depends upon the purpose, speakers, and recording conditions.

c) Other Topics

The reflection coefficients obtained by the LP analysis have been shown to give an acoustic tube representation of the transmission system in the LP model. Thus, if some appropriate preprocessing is applied to a speech segment to eliminate the excitation source characteristics and the effect of the lip radiation load, a realistic area function of the vocal tract is expected to be recovered from acoustic analysis of the speech waves (Wakita, 1979). Although the precise determination of the vocal tract shapes by acoustic analysis of the speech waves is difficult, a fairly good approximation to them is expected.

In an attempt to estimate the glottal characteristics, the actual vocal tract characteristics can also be estimated from a portion of a voiced sound during which the glottis is closed. After the vocal tract characteristics thus estimated are eliminated from the speech signal, the glottal volume velocity waves are recovered from the residual signal.

This trend of extracting some articulatory parameters from the speech waves stimulated the development of other types of articulatory models to which the speech signal is directly transformed (Atal, 1974).

Conclusion

Based on the above acoustic analysis of speech sounds, the LP method has various potential applications to many areas of speech research. The method will be a powerful tool to investigate the interrelationships between articulation and its acoustic characteristics with the aid of, results from other direct physiological measurements. This would contribute to a more complete articulatory model for understanding speech production as well as to a better speech synthesizer. Application of the techniques to speech feature extraction and segmentation will eventually make the automatic transcription of speech sounds possible.

References

Atal, B. S. (1974):"Towards determining articulator positions from the speech signal," Preprints of the 1974 Stockholm Speech Communication Seminar, Vol. 1, 1-9.

Rabiner, L. R., M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal (1976):"A comparative performance study of several pitch detection algorithms," IEEE ASSP, ASSP-24, 5, 399-418.

Wakita, H. (1976):"Instrumentation for the study of speech acoustics," in Contemporary Issues in Experimental Phonetics, N. J. Lass (ed.), 3-40, New York: Academic Press.

Wakita, H. (1979):"Estimation of vocal tract shapes from acoustical analysis of the speech waves: its status of the art," IEEE ASSP (to be published).