SOME INVESTIGATIONS IN SPEECH ANALYSIS-SYNTHESIS SCHEMES AND THE DEVELOPMENT OF A FORMANT VOCODER FOR VOWEL SOUNDS

P.BASKARAN.

THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN ELECTRICAL ENGINEEING
TO
INDIAN INSTITUTE OF TECHNOLOGY, DELHI
1970

ACKNOWLEDGEMENTS

The author is highly grateful to Dr. P.V.Indiresan, Professor of Electrical Engineering, I.I.T. Delhi, for constant inspiration, encouragement and help throughout the period of research.

The author is thankful to Sri S.P.Dhamija of the Electrical Engineering Department, I.I.T. Delhi, for all the help received in the laboratory for experimental work.

Baskaran
(P.Baskaran)

TABLE OF CONTENTS

P R E F A C E

The present state of the art in Vocoder Technology (Technology of speech analysis-synthesis) owes its existence to the pioneering work of Homer Dudley who first developed the first vocoder (VOICE-CODER). Since the first channel vocoder was built by Dudley just before the World War II at the Bell telephone laboratories, different type of vocoders (channel vocoders, Formant vocoders, autocorrelation vocoders, pattern recognition vocoders etc) have been proposed and built. The expanding interest in the vocoder field in its earlier stages of growth, was primarily stimulated by the speech bandwidth compression offered by the vocoders, while the need for conservation of bandwidth of a speech channel was great, especially for military use. Bandwidth compression schemes using analysis-synthesis techniques, both in the frequency domain and the time domain occupied absorbing attention of researchers. The need for achieving higher and higher compression ratios, stimulated the need for achieving a better insight into the processes of speech generation and perception. Research in the process of speech production and perception has gained momentum with the use of fast and versatile digital computors, in the last two decades. Simultaneously work is progressing in the development of automatic recognition devices, and reliable man-machine communication systems. Powerful techniques of speech analysis which have been developed for use in the compression systems, have been applied for providing perceptual aids for the deaf.

With the present state of communications technology, when satellite communications, and laser transmissions, have made bandwidth compression relatively unimportant, research in the vocoder field has more significance in providing a greater understanding of the process of speech generation and perception, and the possible applications to man-machine communication systems.

It was with theses objectives, a new formant vocoder was proposed and fabricated for analysis and synthesis of vowels, at IIT Delhi.

Use of digital techniques and implementation of hardware by use of general purpose transistors (while holding promise of use integrated circuitry) are the main features of the scheme.

The basic functional block on both the formant tracker or analyser and the formant synthesizer is an electronically tunable filter. The filter is a transversal filter using a transistorised digital delay line with electronically tunable delay. Non-recursive (non-feedback) filters, are used in the analyser, while recursive (with feedback) filters are used in the synthesizer.

In the analyser, the frequencies and amplitudes of the first two formants are extracted along with the pitch or the fundamental frequency of the input speech sounds. The analyser therefore consists of two formant trackers, for tracking the first and second formant respectively, and a pitch extractor.

The formant tracker is a continuously tuned automatic tracking filter tracking the formants in real time. The identification of the formant is achieved with the aid of AFC techniques. The tracking system uses two stagger tuned electronically tunable filters which are tunable from a common control source, and having a common input (a suitably filtered speech signal). Their outputs are rectified, filtered and compared, and their difference, is used to tune the centres of the two tunable filters, such that the input spectral maximum coincides with the midpoint between the centres. To achieve tracking in real time, two artifices have been adopted. One is to use a zero crossing counter, which counts the number of zero crossings of the speech input signal, and develops a DC voltage propo- tional roughly to the frequency of spectral maximum in the input band of frequencies. The output of the zero crossing counter is fed to the formant tracker and controls it in such a way as to shift the centres of the two staggered tuned filters, to the region where the formant frequency would exist, thus making the automatic frequency controlaaction fast enough. The second is to use polyphase rectification of the outputs of the filters, for reducing the time constants of the ripple filters which act upon the rectified signals. This has been possible because the tunable filters are transversal filters using delay lines, and the digital type of delay lines offer a large number of convenient taps, to give polyphase outputs.

Another important feature of the scheme is that the formant frequency information is available in a digital form from the analyser.

The pitch tracker, uses a combination of non-linear techniques and switching circuitry to produce reliable pitch pulses, one pulse being produced, during every fundamental period of the voiced sound.

The synthesiser, uses the parmeters extracted from the analyser to synthesise the speech sounds transmitted. It uses two electronically tunable filters, and a buzz generator, which are controlled by the parameters sent from the analyser.

The complete system was tested with live speech input, and different sustained vowels, were successfully synthesized, at the output of the synthesiser.

CHAPTER :I gives a general introduction to speech sounds, their generation and perception, and the underlying redundancy in the speech signal giving scope for bandwidth reduction.

CHAPTER IIdiscusses some of the classical bandwidth compression schemes, leaving a discussion of the formant vocoders to chapter III.

CHAPTER IV discusses some of the proposals examined for the development of a new formant coding compression system leading to the evolution of the model which was finally constructed in the laboratory.

CHAPTER V discusses the development of a digital delay line.

CHAPTER VI pertains to a discussion of the performance of tunable filters derived as transversal filters from tunable delay lines.

CHAPTER VII is the longest chapter in the thesis, dealing with the formant tracker. A description and discussion of the different units constituting the formant tracker is given.

CHAPTER VIII describes the action of the pitchfinder.

CHAPTER IX discusses the action of the speech synthesizer.

CHAPTER X gives a description of the procedures followed in testing the formant tracker and synthesizer as functional units, and aligning the analyser-synthesizer system as a whole.

CHAPTER XI gives a report of the final test results on live speech input signals on the system and the conclusions drawn.

In Appendix is included a brief analysis of the AFC loop of the formant tracker and a short note on phase lock loops.