



A NEW GLOTTAL LPC METHOD OF LOW COMPLEXITY FOR SPEECH ANALYSIS AND CODING

Paavo Alku, Unto K. Laine

Helsinki University of Technology
Otakaari 5 A, SF-02150 Espoo, FINLAND

ABSTRACT

A new straightforward method to compute glottal pulses from speech samples is presented. In the method the human speech production mechanism is modelled with three functional blocks: the source, the tract and the lip radiation. After cancelling the effect of the vocal tract and the lip radiation glottal pulses close to the natural shape are obtained. The method can be effectively applied to speech coding. Improvements of this new method as compared to conventional LPC-coders are discussed. Coding results of long vowels and short words consisting of vowels are presented.

1. INTRODUCTION

Linear predictive coding (LPC) is a speech processing method which has been intensively used for many applications during the last two decades. LPC-theory has been studied and reported in many references (5). Today LPC can be regarded as a "low level building block" in many applications. LPC has some useful features that make its use highly favoured. Especially in speech coding the spectral whitening effect of linear prediction is taken advantage of. Estimation of the spectrum of a stochastic process is also possible with LPC. Mathematically LPC is a straightforward method which is easy to be implemented in real time applications.

Even though linear predictive coding is so frequently used we know that it is not the best possible way to model human speech production. As a result of linear predictive analysis separate parts of the speech production mechanism are combined to be modelled by only one LPC-filter (Fig. 1a). To synthesize speech this IIR-filter is excited with either an impulse train (voiced utterances) or noise (unvoiced utterances).

The acoustic theory of human speech production is based on three separate elements: The vocal tract is modelled with one filter, the lip radiation effect is modelled with a differentiator and the source is cascaded with these two filters (Fig. 1b).

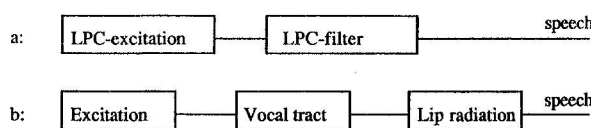


Fig. 1. a: Model used in conventional LPC
b: Model used in our method

Especially in the case of voiced utterances the model used in linear prediction results in an excitation that is totally different from the real physical excitation of the vocal tract, the glottal pulses. This can clearly be seen from Fig. 2, where the excitation of the LPC-model, the residual, and the physical excitation, the glottal wave, are shown.

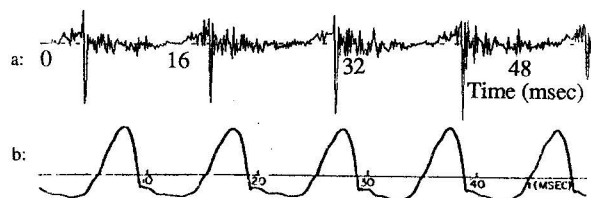


Fig. 2. a: Residual of conventional LPC
b: A typical glottal wave (b)

To use linear predictive analysis in low bit rate speech coding seems to be a natural method for many researchers. Instead of trying to improve the overall modelling strategy attention is paid mainly to the problem of coding the residual, a signal that in fact does not exist in the real physical system. Various kinds of coding methods have been developed with this strategy. The authors don't want to underestimate all this work: good quality speech coders based on the conventional method have been implemented during the last years. However, we strongly believe that there is also another approach. By using a model with a more truthful correspondence to real human speech production low bit-rate speech coders of high quality can be realized. Two noticeable improvements are achieved with this new method. Firstly, the shape of the excitation signal to be

10.21437/Eurospeech.1989-183

quantized is smooth which makes its coding easier. Secondly, the transfer function of the vocal tract can be estimated with higher accuracy than in conventional LPC.

In this paper we present a new glottal-LPC method to compute the real excitation of the vocal tract, the glottal waveform. The application of the method to speech coding is also discussed. Our method is based on the idea of modelling the human speech production mechanism with the abovementioned three separate functional blocks (Fig. 1b). With this procedure a close estimate to the real glottal wave can be obtained. Our method is also quite straightforward in comparison to other glottal-LPC methods (2).

This paper is organized as follows. The next section shortly describes the new method to estimate the true glottal pulses. (A detailed introduction can be found in (1).) Results of the method are studied in section 3, where applications to speech coding are emphasized.

2. METHOD

The most difficult problem in the glottal-LPC methods is the separation of the given speech spectrum into two parts: the vocal tract filter part and the glottal source part. Traditionally we have to know the source characteristic beforehand in order to be able to separate it from the given speech signal. Because the source characteristics of different speech signals vary significantly the separation can not be based on any hypothesis about the type of the source.

More relevant *a priori* knowledge is available about the acoustics of the vocal tract filter. The volume velocity transfer function of the vocal tract is known to be approximately of an all-pass type excluding the local resonances (the formants). Based on this fact we designed for our method an adaptive pre-emphasis filter, which is marked by *block 4* in Fig. 3. With this FIR-filter the contribution of the source is eliminated. As a result a signal of all-pass type is obtained which makes the accurate estimation of the vocal tract filter possible.

The adaptive pre-filter is computed with two consecutive LPC-analysis. The idea is to estimate the glottal contribution by computing "the envelope of the envelope" for the spectrum of the speech signal. After the effect of the glottis is eliminated the vocal tract filter can be solved with conventional linear predictive coding (*block 5* in Fig. 3).

The final source-filter separation is done by inverse filtering the original speech signal through the vocal tract inverse filter. Finally, the glottal volume velocity pulses are obtained by

cancelling the lip radiation effect (a differentiation) by integrating the inverse filter output.

The authors have implemented the method on a Symbolics Lisp machine using the QuickSig signal processing environment (3). The analysis of a signal which is composed of 1024 samples takes a few seconds. Using the latest digital signal processors this procedure can be easily implemented to run in real time.

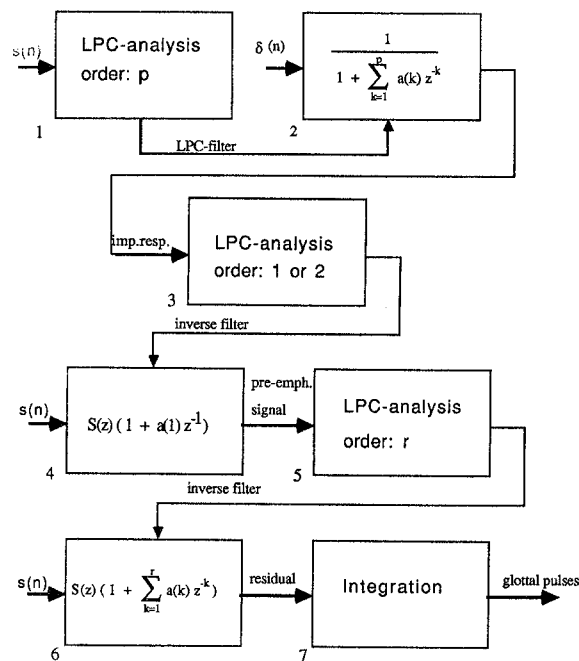


Fig. 3. Block diagram of the method

3. RESULTS

3.1 General

One application of the developed method is the analysis of the glottal waveform. In particular we believe that this procedure could serve as a tool for phoniatrics to study the glottal excitation of the vocal tract. A significant benefit in our method is the low-complexity of this procedure that makes its hardware implementation inexpensive. The method does not need any mask at the lip opening which means that the speech can be produced in normal circumstances. Our method is also fully automated which makes it easy to use.

A typical glottal pulseform obtained with our method can be seen in Fig. 4a. The spectrum of the vocal tract LPC-filter (*block 5* in Fig. 3) is described in Fig. 5a. As a general observation we can verify that the waveform of Fig. 4a is quite close to the natural shape of the glottal pulse. By especially keeping in mind the simplicity of the procedure the resulting pulseform can be considered satisfactory.

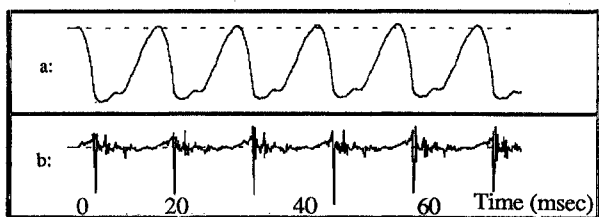


Fig. 4. a: Glottal wave computed with our method from the vowel /oe/

b: Residual of conventional LPC from the vowel /oe/

3.2 Speech Coding

3.2.1 Introduction

Another application for the above described glottal LPC-method is speech coding. Concerning speech coding our method has three main improvements compared with conventional LPC. Firstly, the glottal waveform obtained with our glottal LPC-procedure is of a smooth shape. Quantization of a signal with this shape is much easier than the quantization of the residual of conventional LPC. This is illustrated by Fig. 4 where residual signals of glottal LPC and conventional LPC computed from the same vowel are shown. Another improvement in our method deals with the LPC-filter that has to be transmitted. In the glottal LPC-method this filter is the model of the vocal tract itself while in conventional LPC it is the combined filter for the whole speech production mechanism. Thus in our method resonances of the tract can be modelled with better accuracy with the same size of LPC-filter. In other words filter of smaller size can be transmitted. This effect can be seen in Fig. 5 where the upper curve shows the spectrum of a 10th order vocal tract LPC-filter and the lower curve the spectrum of a conventional 12th order LPC-filter both computed from the same signal. A third improvement concerns noise shaping. Quantization noise in our method is very effectively shaped by the spectrum of the signal.

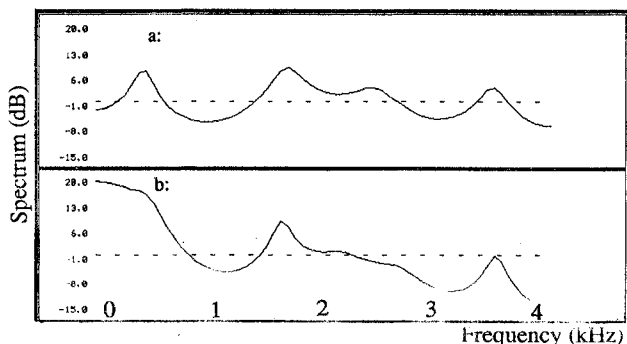


Fig. 5. a: 10th order vocal tract filter for glottal-LPC

b: 12th order conventional LPC-filter

3.2.2 Coding

In the following section two different cases of coding are studied. First we discuss the case where the signal to be coded is a long vowel. The speech was recorded with an almost phase linear microphone in an anechoic chamber. In the second case we used as speech material short words from the Finnish language consisting of vowels. The signal was in the latter case recorded with a microphone which was not phase linear in lower frequencies. The recording was also done in a noisy environment. In both of the two cases the signal bandwidth was 4 kHz.

The coding method used is the same for both cases. The glottal waveform was first computed with the method described in the block diagram of Fig. 3. The pitch period was then estimated from this waveform using autocorrelation. With the pitch information as a time basis local maxima of the pulseform were determined. After this a synthetic glottal pulse was matched to each of the obtained pulses. As a synthetic glottal pulse we used a third order polynomial suggested by Dennis Klatt (Fig. 6) (4).

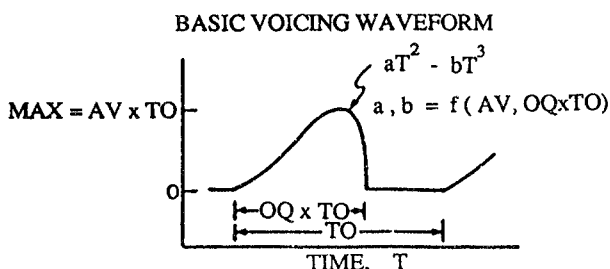


Fig. 6. Synthetic glottal pulse

Case 1.

When a long /a/-utterance pronounced by a male speaker was analysed results shown in Fig. 7 were obtained. Fig. 7a describes the original signal and 7b the coded one. The final bit rate depends on the way the parameters (the vocal tract LPC-filter and Klatt's synthetic glottal pulse) are quantized. Even though all the parameters are quantized with conventional scalar quantization a quality very close to the original one can be obtained at bit rates of about 4 kbit/s.

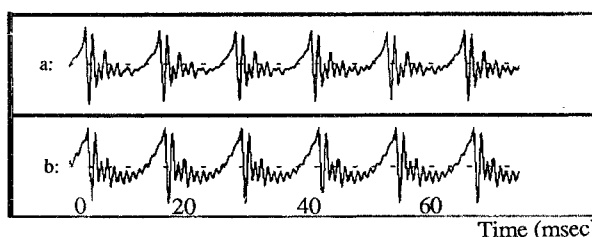


Fig. 7. a: Original vowel

b: Reconstructed vowel

Case 2.

As an example of the analysis of short words a Finnish word /yoe/ was used. In this case a very slight deterioration of the quality of the reconstructed signal was heard. The main reason for this was the poor phase response of the microphone used. As a consequence of the phase distortion the obtained pulseform is different from the natural shape. This in turn deteriorates the quality of the reconstructed signal because the obtained pulses are replaced with the synthetic pulse that was chosen to properly match the undistorted natural glottal wave.

4. DISCUSSION

A new straightforward glottal LPC-method was described. When a model with a more exact correspondence to the acoustics of speech production is used certain improvements can be obtained. The basic idea in this research is to model human speech production with three separate processes. The separation of the different parts of the speech production system is performed with three conventional LPC-analysis. As a result glottal pulses close to the natural shape are obtained.

This method can be effectively applied in vowel coding. The main improvement is the shape of the glottal residual. In comparison to the residual of conventional LPC we now have a smooth wave that is easy to be parameterized. Long vowels and diphthongs can be coded at low bit rates with almost original quality.

There are two reasons that make applications of the above described procedure slightly impractical. Firstly, we have dealt in this paper only with the coding of vowels. Certainly in "real" speech coders also other utterances must be taken into account. Secondly, in AD-conversion we have used low pass filtering. For example in standard telecommunications PCM-equipment speech is band pass filtered to a bandwidth of 0.3-3.4 kHz.

Simulation of the coding method that takes into account the above mentioned problems is under development. In the case of fricatives we have used noise excitation. Band pass filtering on the other hand distorts significantly the obtained pulses. Instead of using the ideal synthetic pulse of Fig. 6 we have studied two possibilities. The first way to code the glottal pulses of band pass filtered speech is to use a small sized codebook of distorted glottal pulse prototypes. The second way is to use Lagrange interpolation in parameterization of the glottal wave. Preliminary results show that high quality can be obtained with both of these principles.

ACKNOWLEDGEMENT

The authors wish to thank Professor Matti Karjalainen and Lic. Tech. Toomas Altosaar for their help.

References

- (1) P. Alku, U. K. Laine, "A New Glottal LPC Method for Voice Coding and Inverse Filtering", Proc. ISCAS 1989, Portland, Oregon, pp. 1831-1884, 1989.
- (2) P. Hedelin, "High Quality Glottal LPC-Vocoding", Proc. ISCAS 1986, Tokyo, pp. 465-468, 1986.
- (3) M. Karjalainen, T. Altosaar, P. Alku, "QuickSig - An Object Oriented Signal Processing Environment", Proc. ICASSP 1988, New York, pp. 1682-1685, 1988.
- (4) D. H. Klatt, "Software for a Cascade/Parallel Formant Synthesizer", Journal of the Acoustical Society of America, vol.67, pp. 971-975.
- (5) J. Makhoul, "Linear Prediction: A Tutorial Review", Proc. IEEE, vol. 63, no. 4, pp. 561-580, Apr. 1975.
- (6) A. E. Rosenberg, "Effect of Glottal Pulse Shape on the Quality of Natural Vowels", Journal of the Acoustical Society of America, vol. 45, no. 2, 1971.