17-9-2000

Exploiting simultaneously masked linear prediction in a WI speech coder

Jason Lukasiak
University of Wollongong, jl01@ouw.edu.au

I. Burnett
University of Wollongong, ianb@uow.edu.au

Exploiting simultaneously masked linear prediction in a WI speech coder

Abstract

This paper uses a method of incorporating simultaneous masking into the calculation of a linear predictive filter (SMLPC) as the front end to a 2 kbps waveform interpolation (WI) speech coder. A modification to the masking threshold calculation used in SMLPC is proposed. This modification improves the performance of SMLPC in noise like sections by placing greater emphasis on strongly voiced speech. MOS test results reveal that the modified SMLPC improved the perceptual quality of the WI coder. The improvement is significant for female speakers whilst the quality for male speech is virtually unchanged. This result conflicts with previous results reported for SMLPC where only male speech was improved. The change is attributed to the modification of the masking threshold and confirms that adapting the masking threshold according to the pitch of the speech will allow SMLPC to remove more perceptually important information from all input speech than standard LPC.

Keywords

filtering theory, interpolation, linear predictive coding, speech coding, speech intelligibility, telecommunication equipment testing, vocoders

Disciplines

Physical Sciences and Mathematics

Publication Details

This paper originally appeared as: Lukasiak, J & Burnett, I, Exploiting simultaneously masked linear prediction in a WI speech coder, Proceedings, IEEE Workshop on Speech Coding, 17-20 September 2000, 11-13.

EXPLORING THE CHARACTERISTICS OF ANALYTIC DECOMPOSITION OF SPEECH SIGNALS

J. Lukasiak, I.S. Burnett

Whisper Laboratories, TITR
University of Wollongong
Wollongong, NSW, Australia, 2522

ABSTRACT

This paper investigates the properties of analytic transformation of speech into envelope and phase functions. The envelope is shown to evolve slowly with the pitch of the input speech, whilst the phase consists of two components; one evolving slowly with pitch and another that exhibits a more rapid evolution. We investigate decomposing the phase component further using two distinct methods: a) Filtering of the phase in the pitch evolutionary direction and b) Performing a second analytic decomposition of the phase into secondary envelope and phase components. To examine the characteristics of the pitch cycle evolution, the analytic transform is employed in a Waveform Interpolation (WI) coding structure. The two phase decompositions are then analysed with particular emphasis on quantisation sensitivity and the required transmission rate. Results indicate that the analytic decomposition may offer a degree of scalability to speech coders, especially when employed in coders that exploit pitch evolution such as Waveform Interpolation (WI) [3]

1. INTRODUCTION

An Analytic signal is a complex signal that contains only positive frequencies. It is associated with a real signal by the removal of the real signals negative frequencies and doubling the value of its positive frequencies. The analytic signal directly links a signal with its envelope and phase(/instantaneous frequency) as shown in equation 1[1].

$$x(n) = s(n) + j\hat{s}(n) = E(n)e^{j\Phi(n)}$$

$$E(n) = |x(n)| = \sqrt{s(n)^2 + \hat{s}(n)^2} \qquad (1)$$

$$\Phi(n) = \tan^{-1}\left(\frac{s(n)}{\hat{s}(n)}\right)$$

Where x(n) is the analytic signal, s(n) is the input speech, $\hat{s}(n)$ is the Hilbert transform [1] of s(n), E(n) is the envelope and $\Phi(n)$ is the phase.

This property provides a straight forward means of decomposing a signal into two separate signals that exhibit distinctly different characteristics. The removal of the negative frequencies also allows the sampling rate of the analytic signal to be half that of the input signal without causing aliasing.

Whilst analytic signals are widely used in such areas as time frequency signal analysis, spectral analysis and many others (see references of [2]) , their use in the coding of speech has not been widely reported.

This paper investigates the characteristics of the analytic representation of a speech signal. This investigation takes the envelope and phase associated with the analytic signal and analyses such parameters as time evolution, frequency analysis, sensitivity to quantisation and the required transmission rate.

Initial results show that the envelope evolves slowly with the pitch of the input speech, whilst the phase is made up of two components one of which also evolves slowly with pitch whilst the other exhibits a more rapid evolution. To explore these pitch synchronous characteristics the analytic transform was employed in a Waveform Interpolation (WI) [3] coding structure. The phase component was then also further decomposed to separate the slowly and rapidly evolving components.

The paper is organized as follows. In section 2 the analytic transform is defined. Section 3 examines the characteristics of speech transformed via an analytic transform and reports the results of employing the analytic transform in a WI coder. Finally the major points are summarised in section 4.

2. Calculation of Discrete time Analytic signals

The analytic representation of a discrete time signal may be calculated in the time domain via the Hilbert transform or in the frequency domain by removing the negative frequencies and doubling the value of the positive frequencies [1]. The time domain approach can involve the use of a filtering mechanism which approximates the Hilbert transform. The frequency domain approach has been adopted for this paper and is defined for a sequence of length N, where N is even as [2]:

$$X(n) = \begin{cases} S(0) & for\ n = 0 \\ 2S(n) & for\ 1 < n < \frac{N}{2} \\ S\left(\frac{N}{2}\right) & for\ n = \frac{N}{2} \\ 0 & for\ \frac{N}{2} + 1 < n < N - 1 \end{cases} \qquad (2)$$

Where S(n) is the N point DFT of the input speech s(n). The time domain analytic signal is then calculated by performing an N point Inverse DFT of $X(n)$. The above procedure results in a frequency spectrum that contains only positive frequencies. Also as can be seen in equation 1 the real part of the analytic signal is equal to the input speech signal.

3. Analytic analysis of speech

3.1 Decomposition of the speech

An input speech file was transformed to its envelope and phase functions using the analytic decomposition detailed in equations 1 and 2. The envelope and phase were then plotted against the input speech as shown in Figure 1. Figure 1 indicates that for a section of voiced speech the envelope and phase waveforms both evolve with the pitch of the input waveform. This characteristic indicates a level of redundancy in the envelope and phase waveforms.

To examine the extent of this redundancy the speech was processed in a WI coding structure. The WI structure was selected as it extracts pitch length segments of speech (characteristic waveforms) and uses these to exploit redundancy of a pitch evolutionary nature. Due to problems with WI extracting characteristic waveforms (CW) in the speech domain (as detailed in [3]) the extraction was performed on the LP residual. The CW were then decomposed into their envelope and phase using equations 1 and 2.

The evolution of the CW envelope and phase for a section of female speech is shown in Figures 2 and 3 respectively. These figures indicate that the pitch evolutionary characteristics shown in Figure 1 continue to apply to speech in the residual domain although the pitch is more noise like due to the removal of the minimum phase component by the linear predictive filter. Figure 2 shows that the envelope evolves very slowly, particularly for voiced speech, thus indicating a large level of redundancy. Examining figure 3 reveals that the phase exhibits an underlying slow evolution with a more rapidly evolving component superimposed. This characteristic indicates that further decomposition of the phase would separate these components and thus allow them to be quantised individually.

Two separate methods were used to achieve a further decomposition of the phase waveform, these being filtering in the evolution direction (similar to SEW/REW decomposition in standard WI [3]) and a second analytic decomposition into an envelope and phase function. Both of these methods achieved a good degree of seperation and their characteristics are further analysed in section 3.2.

3.2 Sensitivity to Quantisation noise

Section 3.1 details three distinct methods of decomposition these being:

1) Envelope and Phase.

2) Envelope, lowpass phase and highpass phase.

3) Envelope, Phase envelope and Phase phase.

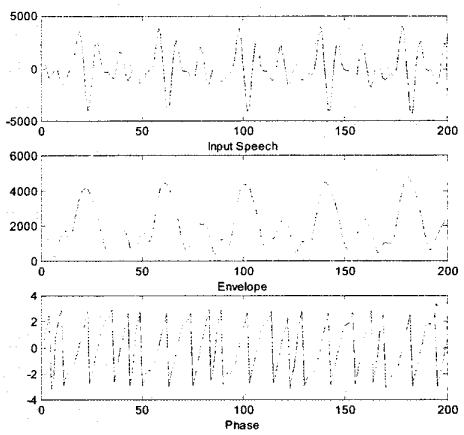

Figure 1: Analytic representation of speech.

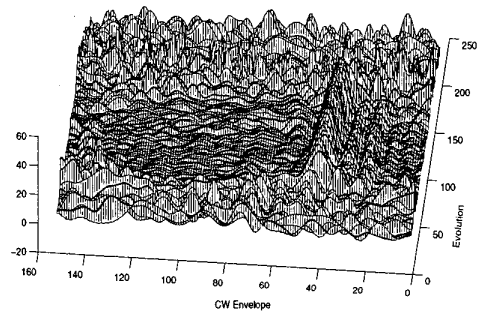

Figure 2: Evolution of the envelope

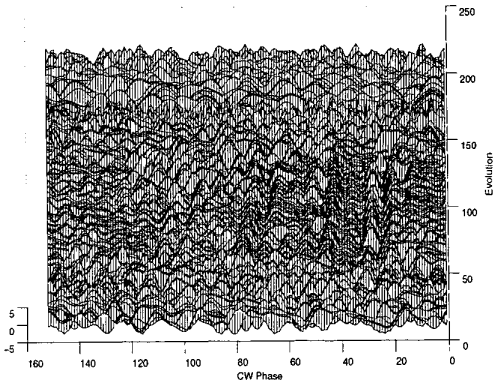

Figure 3: Evolution of the Phase (viewpoint changed from Figure 2 to maximize information content).

The maximum tolerable quantisation noise of each component in both the time and frequency domains was determined and the results are shown in Table 1. The results are expressed as a percentage of the maximum value for each parameter.

The results indicate that all the parameters exhibit a reasonable immunity to quantisation noise. The envelope, which is required for all of the decompositions, requires moderate accuracy in the time and DFT magnitude domains whilst the DFT phase could be coded very coarsely. These characteristics combined with the slow evolution of the envelope indicate that the envelope would be suitable for variable dimension vector quantisation (VDVQ)[4] in the time domain or vector quantisation of the DFT magnitude with a phase model used to reproduce the DFT phase.

Both the lowpass and highpass parameters resulting from filtering the phase offer improved robustness to quantisation whilst decomposing the phase with a second analytic transform generates parameters that are slightly more sensitive to quantisation noise than the original phase.

3.3 Transmission rates of the Parameters

Ten CW's were extracted per 25ms frame at a rate of 400Hz. The CW's were decomposed using each of the methods detailed in 3.1. The evolutionary bandwidth of each parameter was determined by lowpass filtering the parameter in the evolutionary direction before reconstructing the signal. The results are shown in Table 2. The results indicate that the envelope and phase envelope could be down sampled by a factor of 10 before transmission and then reconstructed via interpolation in the decoder without causing distortion, whilst phase and phase envelope require full bandwidth for undistorted reconstruction.

Perceptual testing of the synthesized speech shows that the envelope contains most of the intelligibility of the speech. Transmitting only 1 envelope per frame with phase set to zero produces highly intelligible speech. The speech however sounds quite robotic.

Comparing the analytic and filtering methods of decomposing the phase, showed that the analytic method provides a means of increasing the naturalness of the speech as the transmission rate of the parameters is increased. Whereas whilst the filtering method does increase naturalness, the speech was scratchy unless the phase was transmitted without down sampling.

Transmitting only the down sampled envelope and phase envelope, with the Phase phase set to zero produced clear speech of better quality than if only the envelope was sent. The naturalness of the speech was then increase as the rate of transmission of the phase phase was increased.

4. CONCLUSION

An analytic decomposition of speech has been investigated. This investigation revealed that both the envelope and phase parameters associated with the analytic decomposition evolve with the pitch of the input speech. This property makes the analytic decomposition an ideal candidate for use in a WI type coder which exploits the pitch evolution of speech to achieve compression.

	Tolerable quantisation noise as % of maximum		
	Time domain	DFT magnitude	DFT phase
Envelope	4	4.2	25
Phase	6.98	10.5	12.5
Low pass Phase	9.13	15.42	16.67
High pass Phase	6.54	10.82	16.67
Phase Envelope	10.3	5.18	16.67
Phase Phase	4.36	3.82	5.56

Table 1-Sensitivity to quantisation noise

	Normalised Bandwidth
Envelope	0.1
Phase	1
Phase Envelope	<0.1
Phase Phase	1

Table 2- Evolutionary bandwidth of parameters

The characteristics of the analytic signal when employed in a WI paradigm were then investigated. These results indicate that decomposing the signal on a pitch synchronous nature using an analytic transform produces envelope and phase parameters that exhibit distinctly different quantisation and evolutionary properties. These properties indicate that both parameters exhibit good immunity to quantisation noise. The envelope is also very slowly evolving and suitable for down sampling. The phase exhibits a slowly evolving component superimposed with a more rapidly evolving component. Performing a second analytic transform on the phase appears to successfully separates these components. This finally decomposes the signal into envelope, phase envelope and phase phase parameters. Analysis of these parameters indicates that the method offers a degree of scalability to the coding structure by varying the transmission rate of the parameters.

Whilst some characteristics of the analytic transform such as added complexity have not been addressed in this paper, the results obtained indicate that further investigation of the method is warranted.

5. ACKNOWLEDGEMENTS

J.Lukasiak is in receipt of an Australian Postgraduate Award (Industry) and a Motorola (Australia) Partnerships in research Grant. Whisper Laboratories is funded by Motorola and the Australian Research Council.

6. REFERENCES

[1] A.V. Oppenheim and R.W. Schafer, Discrete time Signal processing, Prentice-Hall, New Jersey, 1989. p685

[2] S.L. Marple," Computing the discrete time analytic signal via FFT", Conf. Rec. of 31st Asilomar conf on signals, systems and computers, vol. 2,pp.1322-1325,1998.

[3] W.B. Kleijn and J. Haagen, "A speech coder based on decomposition of characteristic waveforms", Proc. ICASSP 95, Vol. 1, pp.508-511, 1995.

[4] A. Das, A.V.Rao, A. Gersho, "Variable dimension vector quantisation", IEEE sig. Proc. lett., Vol.3, No.7, pp200-3, July 1996.