1-10-2002

# Extending waveform interpolation to wideband speech coding

C. H. Ritz
*University of Wollongong*, critz@uow.edu.au

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

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

# Extending waveform interpolation to wideband speech coding

## Abstract

This paper investigates the extension of waveform interpolation (WI) to wideband speech coding. Included is an analysis of the evolutionary behaviour of wideband speech and the consequences for WI. We highlight problems associated with direct application of the classical WI algorithm applied to wideband speech.

## Keywords

interpolation, linear predictive coding, speech coding

## Disciplines

Physical Sciences and Mathematics

## Publication Details

This paper originally appeared as: Ritz, CH, Burnett, I and Lukasiak, J, Extending waveform interpolation to wideband speech coding, IEEE Workshop Proceedings. Speech Coding, 6-9 October 2002, 32-34. Copyright IEEE 2002.

# EXTENDING WAVEFORM INTERPOLATION TO WIDEBAND SPEECH CODING

*C.H. Ritz, I.S. Burnett and J. Lukasiak*
Whisper Labs, University of Wollongong
Wollongong, Australia.
chritz@st.elec.uow.edu.au, i.burnett@elec.uow.edu.au, jasonl@elec.uow.edu.au

## ABSTRACT

This paper investigates the extension of Waveform Interpolation (WI) to wideband speech coding. Included is an analysis of the evolutionary behaviour of wideband speech and the consequences for WI. We highlight problems associated with direct application of the classical WI algorithm applied to wideband speech.

## 1. INTRODUCTION

Waveform Interpolation (WI) is a method for coding narrowband speech with high perceptual quality at low bit rates [1]. It is based on the description of speech (in practice the LP residual) as a surface formed from evolving pitch-length Characteristic Waveforms (CWs). The key to quantisation of the CWs at low bit rates is decomposition of that surface into two sub-surfaces: the Slowly Evolving Waveform (SEW) and Rapidly Evolving Waveform (REW), which represent the voiced and unvoiced speech components, respectively.

To date, most research into WI has focused on narrowband speech [1]. However, we recently proposed that WI could be used to achieve low bit rate wideband speech compression at 4kbps [2]. The focus of this paper is the evolutionary behaviour of wideband speech CWs and an investigation as to whether the decomposition methods used in classical WI are appropriate for wideband speech. We also present an analysis of the Linear Prediction (LP) requirements for wideband WI as residual characteristics are central to the success of the classical WI decomposition. For clarity, we define wideband and narrowband speech to have bandwidths of 50 Hz to 7 kHz and 300 Hz to 3.4 kHz, respectively. Wideband speech was generated by re-sampling speech extracted from the ANDOSL database [3] to 16 kHz and band-limiting to 50 Hz to 7 kHz. For comparison, narrowband speech was also generated by re-sampling the same ANDOSL speech to 8 kHz and band-limiting to the range 300 Hz to 3.4 kHz.

In Section 2 we examine the LP requirements for wideband WI, while section 3 presents an analysis of wideband CW evolution. Section 4 then considers the decomposition of the wideband CW surface.

## 2. LP FOR WIDEBAND WI

Separation of the vocal tract shape from the speech spectrum is vitally important for WI coding since it allows independent interpolation of the LP parameters and pitch [1]. Effective LP is also important for ensuring successful extraction of the residual domain CWs; in particular, minimisation of discontinuities at CW end-points is imperative if a smooth CW evolutionary surface is to be produced.

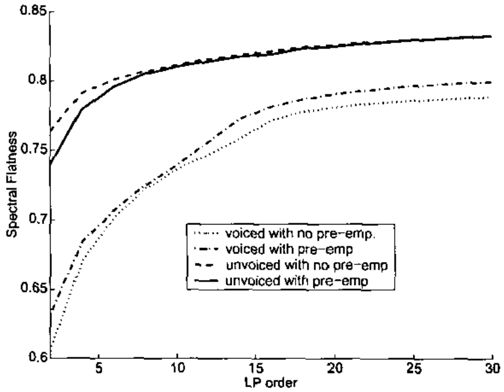

Figure 1. Spectral Flatness versus LP order for voiced and unvoiced speech, with and without pre-emphasis

In [2] we used an LP order of 20 based on the suggestions in [4] that the order should be equal to the sampling rate (in kHz) plus 4 or 5. LP orders ranging from 16 to 20 have been suggested as appropriate for wideband speech coding that is not based on split band techniques [5,6], with lower orders advantageous in terms of LP coefficient quantization. We seek here to investigate the appropriate LP order for wideband WI.

The intention of the LP in WI is to generate a spectrally flat residual for decomposition; spectral flatness is used to advantage in narrowband WI CW surface decomposition and quantisation. Thus, we evaluate the Spectral Flatness Measure (SFM) of the wideband residual for voiced and unvoiced speech at various LP orders. The SFM is defined in [4], and represents the ratio of geometric to arithmetic mean of the magnitude components of the residual spectrum. A perfectly flat spectrum will have a SFM of 1, which indicates that the LP filter has perfectly modeled the speech spectrum. The results of the SFM experiments are shown in Figure 1. For comparison, results for LP applied to pre-emphasised speech are also shown. Pre-emphasis was performed using a first order filter the form $1\text{-}0.94z^{-1}$.

The results in Figure 1 indicate that an LP order of 20 is indeed appropriate for wideband speech as there is no significant increase in spectral flatness beyond this order. The graphs also show that pre-emphasis leads to an improvement in the LP model for voiced speech. This is particularly important for wideband WI as it ensures better modeling of the formants and better separation of the wideband vocal tract shape from the speech spectrum.

0-7803-7549-1/02/$17.00 ©2002 IEEE.

## 3. EVOLUTION OF WIDEBAND CWS

In this work, CWs are extracted at 400Hz from the 20th order LP residual. This ensures that CWs are extracted at least once per pitch period for a maximum pitch frequency is 400 Hz [1]. Following extraction, the CWs are aligned by circular rotation to maximally correlate with the previous CW. Alignment is necessary to ensure a smooth evolution of the CW surface [1].

In classical WI applied to narrowband speech, decomposition is motivated by the differences in evolution of voiced and unvoiced speech. During voiced speech, the correlation between CWs decays slowly as their separation (in evolutionary time) increases. In contrast, during unvoiced speech, this correlation decays rapidly. These differences in correlation can be seen in the energy spectrum of the evolution of the CWs. We now consider the evolutionary behaviour of CWs derived from wideband speech.

### 3.1 Correlation Decay

The cross correlation was measured between CWs separated by n extraction points. To facilitate investigation of the CW evolution in different frequency bands we perform analysis in the DFT domain. The real and imaginary DFT coefficients corresponding to frequency f, of the CW extracted at time n are defined as:

$$CW_n(f) = [a_n(f) \quad b_n(f)] \tag{1}$$

The cross correlation between CWs extracted at time 0 and n can then be measured as:

$$c_n = \frac{\left|\sum_f CW_0(f) \times CW_n(f)^T\right|}{\sqrt{\sum_f CW_0(f) \times CW_0(f)^T \sum_f CW_n(f) \times CW_n(f)^T}} \tag{2}$$

where, $f \in [f1, f2]$ defines the frequency band of interest.

We used (2) to measure the correlation between CWs with $0 \leq n \leq 20$ and considered 2 frequency bands covering the 0 to 4kHz and 4 kHz to 8 kHz ranges for voiced and unvoiced speech. For comparison, results for correlation of CWs extracted from narrowband speech were also measured. These results are shown in Figure 2. This shows that the correlation decay in the 0 to 4 kHz range during voiced speech is similar for both wideband and narrowband CWs. However, the 4 kHz to 8 kHz range of wideband CWs during voiced speech demonstrates more rapid correlation decay. For unvoiced speech, the correlation decays rapidly for both narrowband CWs and both frequency bands of wideband CWs. These results indicate that compared with the 0 to 4kHz region, the 4 kHz to 8 kHz region of the wideband CWs does not exhibit slow evolution during voiced speech.

### 3.2 Evolution Bandwidth of Wideband CWs

The correlation results above suggest that the evolving spectrum of the CWs for the 0 to 4 kHz region will differ to the 4 kHz to 8 kHz region. The evolution spectrum can be described by taking a two dimensional DFT of a sequence of CWs [7]. This results in an evolutionary magnitude spectrum for each coefficient of the CWs in the frequency domain. Figure 3 shows an example of the 3-D evolutionary spectrum for a sequence of wideband CWs extracted from a section of voiced speech. The magnitude spectra of the evolutionary coefficients were averaged for eight 1 kHz frequency bands, covering the entire CW bandwidth.

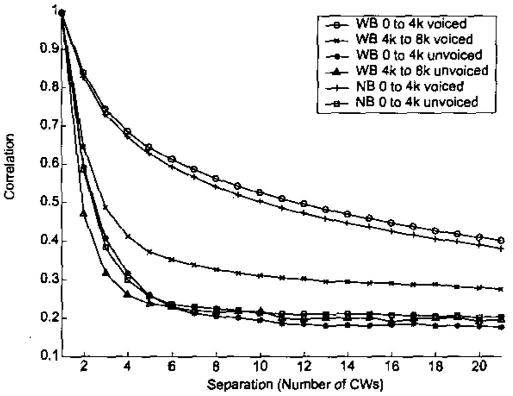

Figure 2. Cross Correlation of CWs as separation increases.

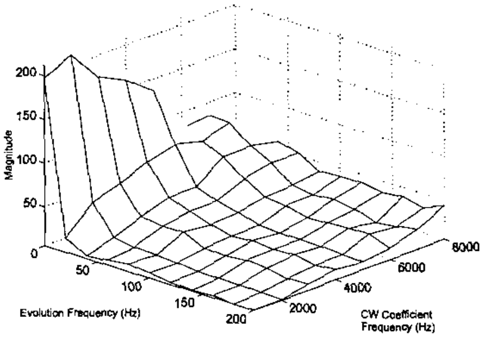

Figure 3. Evolution spectrum of wideband CWs.

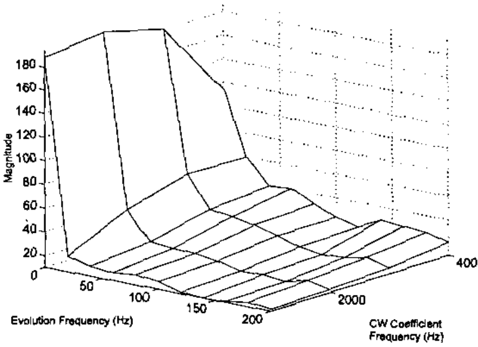

Figure 4. Evolution spectrum of narrowband CWs.

Figure 3 shows that the evolution of the low frequencies of the CW has energy concentrated at low evolution frequencies, while the evolution spectrum of the high frequencies is more evenly distributed. For comparison, Figure 4 shows the evolutionary spectrum of narrowband CWs extracted from the same section of speech as Figure 3. Comparison of the surfaces of Figures 3 and 4 indicates that the evolutionary behaviour of narrowband CWs are similar to the 0 to 4 kHz region of the wideband CWs, with energy concentrated at low frequencies. These results for narrowband CWs agree with the suggestions in [1], that most of the energy in the evolution spectrum is below 20Hz.

## 4. DECOMPOSITION OF WIDEBAND CWS

In narrowband WI, the decomposition of the CWs is usually achieved by low pass filtering the evolution of the CW surface. The low pass component forms a SEW surface while the high pass component forms the complimentary REW surface. For narrowband speech, the decomposition filter typically has a cutoff frequency of 20 Hz. Results from Section 3 suggest that the decomposition requirements for the lower half band of wideband CWs should be similar to that for narrowband CWs; however, the decomposition strategy for high frequencies should differ. Here we analyse the decomposition using the SEW-to-REW (STR) energy ratio, defined as:

$$STR = \frac{\sum_{f} |S(f)|^2}{\sum_{f} |R(f)|^2} \quad f \in [f1,f2] \qquad (3)$$

where, S(f) and R(f) are the DFT magnitudes of the SEW and REW at frequency f, respectively and *[f1,f2]* defines the bandwidth of interest. We decomposed the wideband CWs using cutoff frequencies ranging from 5 Hz to 70 Hz and the STR was measured for the 0 to 4 kHz and 4 kHz to 8 kHz frequency bands. Figure 5 shows results for both wideband and narrowband CWs generated from identical speech. For comparison, results for voiced and unvoiced speech are shown.

For the 0 to 4kHz band of wideband and narrowband CWs derived for voiced speech, the graph shows similar shaped curves. The small differences in STR can be explained by the band-limiting of the narrowband CWs to 300 Hz to 3.4 kHz as well as the differing LP filters for wideband and narrowband. An appropriate choice of cutoff frequency would thus be the frequency corresponding to the knee of the curve, since at this point most of the slowly evolving energy has been captured. It can be seen that in both cases the knee points correspond to a frequency of around 20 to 30 Hz. This confirms the choice of a 20 Hz cutoff for the decomposition filter in narrowband WI and for the 0 to 4 kHz region of wideband CWs.

In contrast, for the 4 kHz to 8 kHz region of wideband CWs derived for voiced speech, the STR curve does not show a clear knee point. This indicates that there is no clear SEW component at the high frequencies during voiced speech. In addition, this curve has a similar shape to the STR curves for CWs derived for unvoiced speech. This indicates that the 4kHz to 8 kHz region of wideband CWs derived for voiced speech have similar characteristics to CWs derived for unvoiced speech.

## 5. DISCUSSION AND CONCLUSIONS

This paper has presented an investigation into WI applied to wideband speech. It was found that 20th order LP with pre-emphasis is appropriate for wideband WI to ensure that the residual is spectrally flat. It was also found that the evolution of CWs derived for wideband speech differs across frequency bands. Low frequencies of wideband CWs display similar evolution characteristics to narrowband CWs, while high frequencies display more rapid evolution. The consequences of these results are that decomposition by low pass filtering to 20 Hz (as is commonly used in narrowband WI) is justified for the low frequencies of the wideband CWs, but not necessarily for high frequencies. These results mean that classical WI decomposition for the high frequency sections of wideband CWs may not be appropriate.

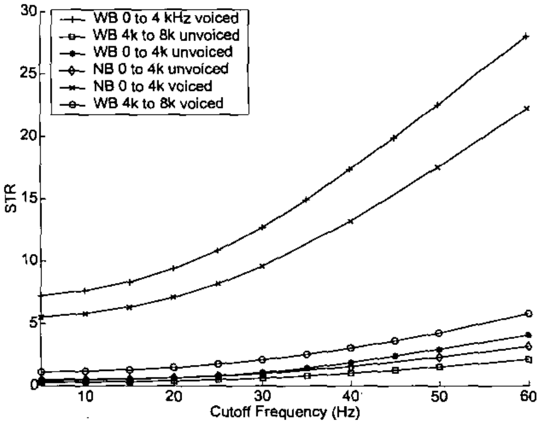

Figure 5. STR versus cutoff frequency of the decomposition filter.

An alternative to decomposition for high frequencies could be to model this section of the CW using modulated noise. A similar method was suggested in [8] for modelling high frequencies of the excitation in wideband speech coding.

**Acknowledgements**

C.H. Ritz is in receipt of an Australian Postgraduate Award and was previously supported under a Motorola (Australia) Partnerships in Research Grant.

## REFERENCES

[1] Kleijn, W.B. and Haagen, J., "Waveform Interpolation for Coding and Synthesis", in *Speech Coding and Synthesis,* pp. 175-207, Kleijn, W.B. and Paliwal, K.K., editors, Elsevier Science B.V., 1995.

[2] Ritz, C.H. and Burnett, I.S., "Wideband Speech Coding at 4 kbps using Waveform Interpolation", *Proc. DSPCS'2002,* pp. 144-148, January, 2002.

[3] Australian National Database of Spoken Language (ANDOSL), CD ROM.

[4] Markel, J.D. and Gray Jr., A.H., *Linear Prediction of Speech*, pp. 139-142, Springer-Verlag, 1976.

[5] Ragot, S., Adoul, J.-p., Lefebvre, R. and Salami, R., "Low complexity LSF quantization for wideband speech coding", *Proc. IEEE Workshop on Speech Coding*, pp. 22-24, June 1999.

[6] Lin, W., Koh, S.N. and Lin, X,. "Mixed excitation linear prediction coding of wideband speech at 8 kbps", *Proc. ICASSP'2000*, Vol. II, pp., 1137-1140, June, 2000.

[7] Tanaka, Y., and Kimura, H., "Low-bit-rate speech coding using a two dimensional transform of residual signals and waveform interpolation", *Proc. ICASSP'94*, Vol. I, pp. 173-176, April, 1994.

[8] McCree, A., Unno, T., Anandakumar, A., Bernard, A. and Paksoy, E., "An embedded adaptive multi-rate wideband speech coder", *Proc. ICASSP'2001*, Vol. 2, pp. 761-764, 2001.