AN IMPROVED DECOMPOSITION METHOD FOR WI USING IIR WAVELET FILTER BANKS

N. R. Chong, I.S.Burnett, J. F. Chicharo

Whisper Laboratories,
University of Wollongong,
NSW, Australia

ABSTRACT

In this paper, we present an alternative characteristic waveform (CW) decomposition mechanism for the Waveform Interpolation (WI) paradigm based on the Pitch Synchronous Wavelet Transform (PSWT). In this technique, IIR filters replace the conventional FIR filters of the PSWT, offering computational and spectral magnitude performance advantages, in addition to significant delay reductions.

Previously, the PSWT has only incorporated filter banks with slowly reacting FIR wavelet filters. While these filters possess the desirable properties of linear phase, and design simplicity, a large delay is incurred which increases exponentially with increasing resolution. The progression to IIR filter banks gives rise to a multi-resolution decomposition mechanism, beneficial for real-time applications, such as speech coding, where delay is an important issue.

1. INTRODUCTION

Waveform Interpolation (WI) allows the efficient compression of signals by exploiting the nature of human speech perception [1]. Since voiced and unvoiced speech have been shown to possess fundamentally different quantisation requirements for perceptually accurate reconstruction, separation of these components enables high coding efficiency. In the current WI decomposition, the signal evolution is decomposed into two waveform surfaces, one characterising voiced speech (slowly evolving waveform (SEW)) and a second representing unvoiced speech (rapidly evolving waveform (REW)). However, we have shown [2] that the simple filtering used in this approach is not adequate to gain good separation of the quasi-periodic and noise-like components of speech.

Recently, we proposed the Pitch Synchronous Wavelet Transform (PSWT) [3] as an alternative decomposition method for WI [4]. The Pitch Synchronous Wavelet decomposition can to be easily applied to the WI paradigm due to its similarity to the current SEW/REW decomposition. In both methods, pitch-length segments of speech are extracted to form a 2-D representation of the signal, followed by filtering of the signal evolution.

The multi-resolution wavelet decomposition results in the isolation of the periodic trend, as well as characterising the aperiodic behaviour at several scales. The bit allocation for the decomposed surfaces can be flexible, allowing preferential quantisation to achieve high quality performance.

Previously, the filter banks of the PSWT have been restricted to FIR wavelet filters [3][4]. Application of these filter banks to the WI paradigm has revealed that the pitch synchronous wavelet decomposition is well suited to the task of multi-resolution analysis of the evolving signal. But, while these filters possess the desirable property of linear phase, and simplicity of design, a large delay is incurred. This delay increases exponentially with increasing resolution, due to the tree-structure of the transform. Hence, use of the PSWT with FIR filters for multiple decomposition levels, while favourable for speech storage or image processing, is impractical for real-time speech processing.

Here we present the PSWT using infinite impulse response (IIR) filters as an improved decomposition mechanism.

The perfect reconstruction properties of biorthogonal wavelets, the dynamic capture achieved from sharper filters and, in particular, the substantial delay reduction, render the wavelet decomposition using IIR filters very appealing. These filter banks contrast the previously proposed FIR QMF banks to render a practical CW decomposition mechanism.

The outline of the paper is as follows: Section 2 gives a brief overview of the PSWT. Two IIR Wavelet filter bank designs and their application to the PSWT are described in Section 3, with quantisation techniques in Section 4. Results and comparisons are discussed in Section 5.

2. THE PSWT

The PSWT operates in the frequency evolution domain. Whereas the Discrete Wavelet Transform (DWT) correlates dilated and translated versions of a unique analysing wavelet with the input signal (conventional frequencies), the PSWT correlates the wavelet with the evolutionary signal. The evolutionary signal, $v_q(k)$, can be expressed by

$$v_q(k) = \sum_{n,m} V_{n,m,q} \varphi_{n,m}(k) \qquad \ldots(1)$$

where,

$$V_{n,m,q} = \sum_k v_q(k) \varphi_{n,m}(k) \qquad \ldots(2)$$

is the wavelet transform of $v_q(k)$, $\varphi_{n,m}(k)$ is the analysing wavelet, index n=1,2,…,N represents scale, m=0,1,2,…,M represents time shift, and q=0,1,…P(k)-1, where P(k) holds pitch information.

In order to perform the PSWT, the one-dimensional speech (or residual) signal must first be transformed into a two-dimensional surface by extracting characteristic waveforms (also known as prototypes) and aligning them to maximise the cross-correlation between adjacent waveforms. To perform these procedures in real-time, we adapt the PSWT to maintain a fixed sampling rate as opposed to the critical sampling proposed in [3]. Thus, prototypes are oversampled as in WI, This guarantees a fixed rate of parameters which is appropriate for fixed frame-rate encoding.

The PSWT is similar to the operation carried out in the SEW/REW decomposition, involving simple filtering of the evolution of the DFT coefficients. This similarity enables the wavelet decomposition to be easily applied to the WI paradigm. The main difference here is that the filters used are perfect reconstruction wavelets, dilated and shifted to form a non-uniform filter bank. Hence, deviations from periodicity are detected, isolating the periodic trend and at the same time characterising the aperiodic behaviour at several scales.

Biorthogonal wavelets possess linear phase and can decompose and reconstruct an input signal with no aliasing or distortion. This is achieved by designing the generating filters to have mirror symmetry.

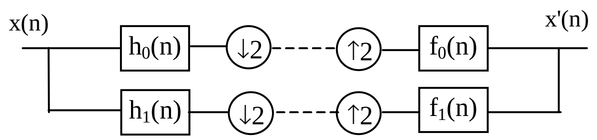

Figure 1. Maximally decimated 2-channel filter bank.

In order to cancel aliasing, the filters are related as follows:

$$F_0(z) = H_1(-z) \qquad \ldots(3)$$

$$F_1(z) = -H_0(-z) \qquad \ldots(4)$$

where H_0 and H_1 are the Analysis (or decomposition) lowpass and highpass filters respectively, and F_0, F_1 are the Synthesis (or reconstruction) lowpass and highpass filters respectively.

3. IIR FILTER BANKS

The main disadvantage of using FIR filters in the PSWT decomposition is the large delay incurred. Due to the nature of the decomposition, this delay increases exponentially with increasing resolution. Delays of seven frames could easily be incurred with a three-level decomposition using low-order filters. Obviously, this is far from desirable for real-time applications such as speech coding. Another important factor when dealing with tree-structured filter banks is the frequency separation between the bands. Poor separation can result in extremely unfavourable reconstructed signals after quantisation, due to the inability to cancel out aliased components.

In the proposed technique, IIR filters replace the conventional FIR filters of the PSWT to reduce the delay imposed on the system. IIR filters also offer computational and spectral magnitude performance advantages. The progression to IIR filters gives rise to a much more difficult filter design procedure, but the benefits of improved filter roll-off and delay are substantial.

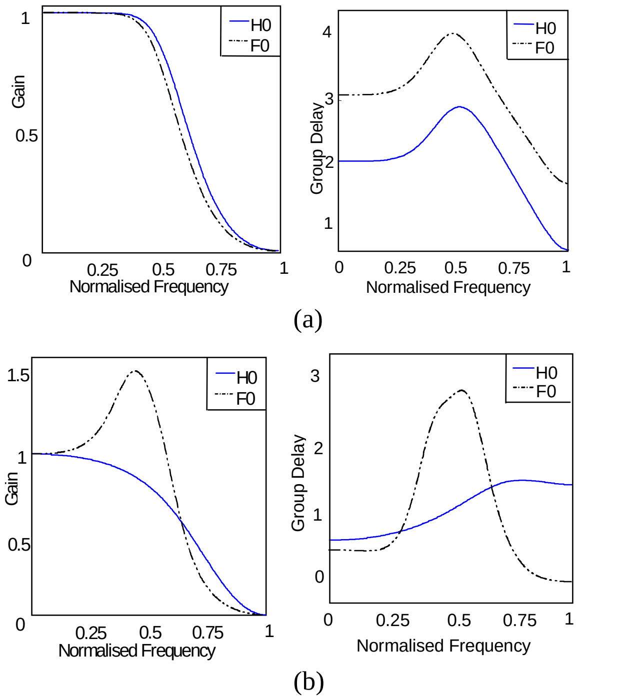

Figure 2. Magnitude Responses (left) and Group Delays (right) for the lowpass analysis (H_0) and synthesis (F_0) filters for (a) Design Method 1 and (b) Design Method 2.

Most of the literature on IIR Quadrature Mirror Filter (QMF) banks give solutions for filter banks possessing causal, unstable synthesis filters. These can be implemented as stable, anti-causal filters which is beneficial for image coding, however again inappropriate for speech coding. The necessary filter criteria include causality, stability, and perfect reconstruction. This means that the filter banks must be biorthogonal. In addition, linear phase was desirable. Two IIR filter designs were applied and compared.

3.1 Design Method 1

In this method, the filter bank is designed from the prototype filters by transformation of variables. Here, we consider only the simplest IIR transformation function and aim to reduce the filter overshoot. Here, we use the filters of Example 1, in [5]. The resulting analysis and synthesis decomposition filters in the bank are of fourth and eighth order respectively. Their magnitude responses are shown in Figure 2(a) along with the associated group delays. These filters experience a combined group delay of 5 samples.

The scaling and wavelet functions for both the analysis and synthesis filters are displayed in Figure 3. While we do not use these functions directly, but rather deal with the digital filter coefficients, it gives an understanding of the iterative filtering operation on the signal by depicting the wavelet characteristics such as smoothness and regularity.

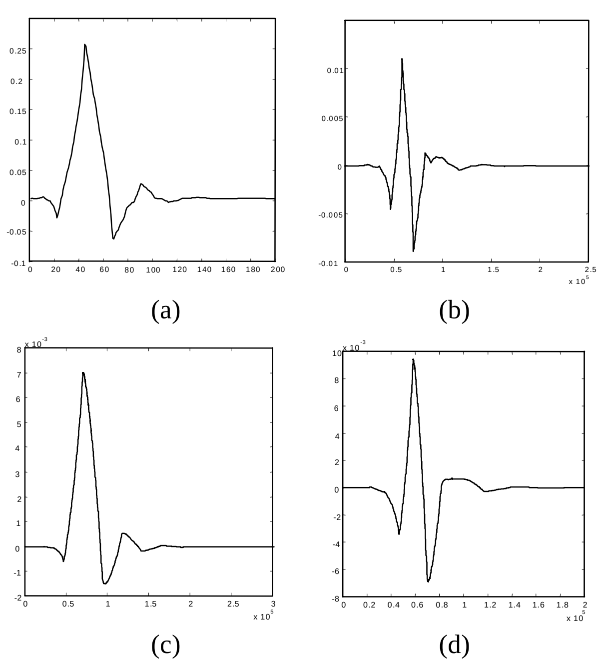

Figure 3. Scaling and wavelet functions of design method 1, after 15 iterations. (a) Analysis scaling function (b) Analysis wavelet function (c) Synthesis scaling function (d) Synthesis wavelet function.

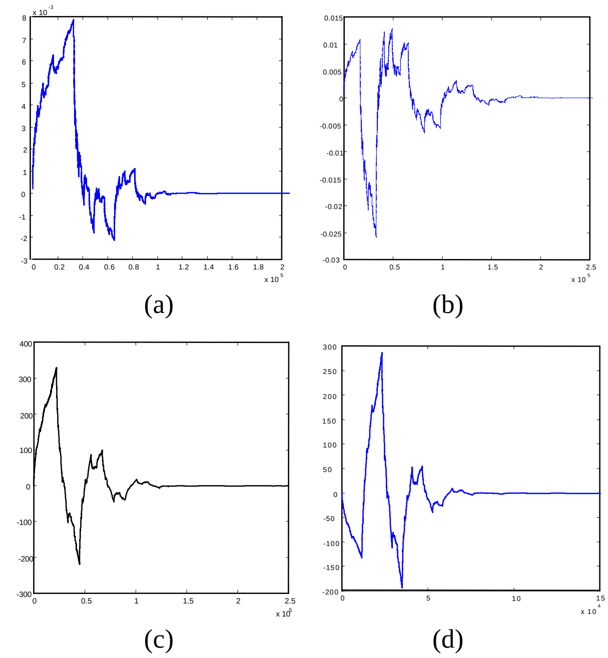

Figure 4. Scaling and wavelet functions of design method 2 after 15 iterations. (a) Analysis scaling function (b) Analysis wavelet function (c) Synthesis scaling function (d) Synthesis wavelet function.

Filter banks resulting from the design of higher order transformation functions dramatically increase group delay. Thus, they are not adopted in the current work.

3.2 Design Method 2

The second causal, stable IIR filter design procedure followed was that of Basu *et al.*[6] in which a complete parameterisation of the solution is outlined. In this method, many intermediate parameters can be independently defined. We use a second order, lowpass half-band Bessel filter as the prototype filter in this comparative study, chose polynomial $r = 5z^2 + 4z + 1$, and let p and q be 2nd and 3rd order respectively.

The magnitude response and group delay for the lowpass filters are shown in Figure 2(b). Figure 4 depicts the analysis and synthesis scaling and wavelet functions. The scaling functions and wavelets of Figure 4 are irregular which is likely to be disadvantageous for analysing many signals.

4. QUANTISATION OF THE SURFACES

The transmission frequency of each surface is defined by the PSWT due to the decimation. Each decomposed surface is sent at a rate corresponding to its sampling frequency. The surface with the lowest resolution can be transmitted once per frame, the next lowest twice per frame, and the next four times per frame. This contrasts the current WI decomposition method in which no transmission rate for the SEW and REW is apparent to provide the most accurate reconstruction.

The PSWT decomposition exhibits potential for higher and variable rate WI coding since bit allocation for the surfaces can be flexible, allowing a more accurate description of perceptually important scales. Scales that are relatively insignificant perceptually may be upsampled to a common frequency and combined or alternatively omitted. In this case, the first decomposition level highpass surface was seen to not add significant perceptual detail to the waveform and was thus not transmitted.

The extraction of pitch-length waveforms in Waveform Interpolation to exploit the periodicity of the signal results in the need to quantise variable length waveforms. Rather than use a multi-codebook approach, where a separate codebook is used for each possible dimension, a Variable Dimension Vector Quantisation (VDVQ) [7] scheme was applied to quantise the individual surfaces. In VDVQ, the variable dimension vector is extended to a fixed length. Codebooks are then trained on these sequences. During the codebook search, the fixed-length codebooks are sub-sampled and compared with the variable-length input vector. This technique overcomes the difficulty of quantising characteristic waveforms whose lengths vary with the pitch period.

VDVQ was used to quantise the magnitudes of the surfaces. In a similar manner to the SEW/REW decomposition more bits are required to quantise the lowpass magnitude surface than the highpass magnitude surfaces.`

The phase spectrum of the lowpass output is quantised using a phase model derived from natural speech. Random phase is used for the highpass phase spectra.

The effects of quantisation of the surfaces are, of course, more significant with IIR filters than with FIR filters, adding extra

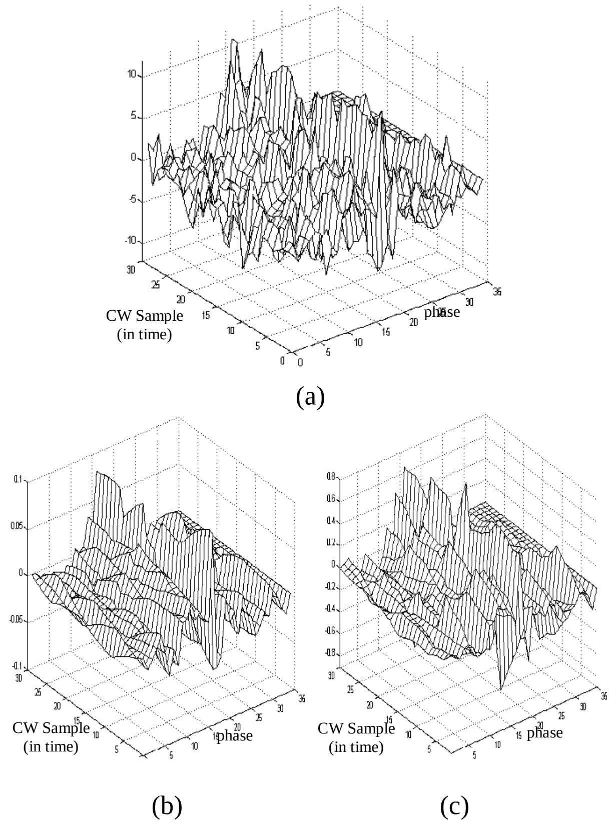

Figure 5. Comparison of the 3rd decomposition level lowpass surface. (a) CW surface (b) Lowpass output of PSWT with filters of design method 1. (c) Lowpass output of PSWT with filters of design method 2.

complexity. In order to reduce the swirling distortion or reverberation of the output speech, we need to maintain the phase relationships between the surfaces. The disadvantage of the IIR approach is that while phase is not perceptually important in WI itself, here the inter-surface phase relationships require accurate transmission.

5. RESULTS

Comparison of Figures 2(a) and 2(b) shows that the filters of Design Method 1 have a much more appealing frequency response with far less overshoot and much sharper roll-off. While both methods ensure perfect reconstruction of the signal in the unquantised case, the filters of method 1 produce less distortion when the magnitude and phase for each of the decomposed surfaces is quantised. On the other hand, the filters of design method 2 have the advantage of having a very low delay - only 1 sample for the analysis/synthesis pair (compared to a delay of 7 samples for a FIR filter of the same order). In a three-level decomposition, this results in a total of only 7 samples of delay in the evolution domain (1 frame delay when extracting 8 prototypes per frame).

Figure 5 shows the output of the third level lowpass decomposition filter for both filter design methods. While the underlying shape has been separated in both cases, the filters of design method 1 tend to reduce the surface evolution bandwidth. This allows more efficient quantisation.

6. CONCLUSION

The PSWT using IIR filters presented here addresses the main problem of the PSWT with FIR QMF banks, this being large incurred filter delays. In addition, the improved roll-off characteristic of the IIR filters results in a more defined separation of the periodic and aperiodic fluctuations by faster tracking of the dynamic aspects of the evolutionary surfaces. While the filters of design method 1 display many benefits, the further delay reductions of the filters of design method 2 are a significant advantage.

Quantisation of the surfaces is the most challenging issue in this technique. Both phase and magnitude are far more sensitive to quantisation errors than in the case for FIR filters. Best results are obtained when the phase relationships are substantially maintained between the highpass surfaces.

7. ACKNOWLEDGEMENTS

Miss N.R. Chong is in receipt of an Australian Postgraduate Award (Industry) and is the recipient of a Motorola (Australia) Partnerships in Research Grant. Whisper Laboratories is funded by Motorola, and the Australian Research Council.

8. REFERENCES

[1] W.B. Kleijn, J. Haagen, “Waveform Interpolation for Coding and Synthesis”, in Speech Coding and Synthesis, edited by W. B. Kleijn and K.K. Paliwal, Elsevier 1995.

[2] N.R.Chong, I.S.Burnett, J.F.Chicharo, M.M.Thomson, “The Effects of Noise on the Waveform Interpolation Speech Coder”, *Proceedings of IEEE TENCON'97*, Brisbane, Australia, Dec., vol. 2, pp609-612, 1997

[3] G.Evangelista, “Pitch-Synchronous Wavelet Representations of Speech and Music Signals”, *IEEE Trans. Sign. Process.*, vol. 41, no. 12, pp. 3313-3329, 1993

[4] N.R.Chong, I.S.Burnett, J.F.Chicharo, M.M.Thomson, “Use of the Pitch Synchronous Wavelet Transform as a New Decomposition Method for WI”, *Proc. Int Conf. Acoustics, Speech, Sig. Processing*, Seattle, USA, vol. 1., pp 513-516, May 1998.

[5] D. Tay, "Design of Causal, Stable, IIR Perfect Reconstruction Filter Banks using Transformations of Variables", *IEEE Int. Symp. Circuits and Systems*, vol. 4, pp. 2425-2428, June 1997.

[6] S.Basu, C-H.Chiang, H-M.Choi, "Wavelets and Perfect Reconstruction Subband Coding with Causal Stable IIR Filters, *IEEE Trans. Circuits and Systems II*, vol.42, no. 1, pp. 24-38, 1995.

[7] Das, A., Rao, V. R., Gersho, A., "Variable-Dimension Vector Quantisation", IEEE Signal Processing Letters, vol.3, no.7, pp. 200-202, July 1996.