

2001

Low rate WI SEW representation using a REW-implicit pulse model

J. Lukasiak

University of Wollongong, jl01@uow.edu.au

I. Burnett

University of Wollongong, ianb@uow.edu.au

Publication Details

This article was originally published as: Lukasiak, J & Burnett, I, Low rate WI SEW representation using a REW-implicit pulse model, IEEE Signal Processing Letters, August 2001, 8(8), 228-230. Copyright IEEE 2001.

Low rate WI SEW representation using a REW-implicit pulse model

Abstract

Reducing the bit rate of waveform interpolation speech coders while maintaining the perceptual quality has been the focus of a great deal of research. This letter proposes a new method of slowly evolving waveform (SEW) quantization specifically targeted at low rate coding. The proposed method uses a pulse model whose parameters are implicitly contained in the quantized rapidly evolving waveform (REW) parameters, thus requiring no bits for transmission. Results indicate no degradation in perceptual speech quality when compared to that of the existing SEW quantization method. This retention of perceptual quality is in spite of a 12% reduction in the overall coder bit rate.

Keywords

low rate speech coding, slowly evolving waveform, waveform interpolation

Disciplines

Physical Sciences and Mathematics

Publication Details

This article was originally published as: Lukasiak, J & Burnett, I, Low rate WI SEW representation using a REW-implicit pulse model, IEEE Signal Processing Letters, August 2001, 8(8), 228-230. Copyright IEEE 2001.

Low Rate WI SEW Representation Using a REW-Implicit Pulse Model

J. Lukasiak, *Student Member, IEEE*, and I. S. Burnett, *Member, IEEE*

Abstract—Reducing the bit rate of waveform interpolation speech coders while maintaining the perceptual quality has recently been the focus of a great deal of research. This letter proposes a new method of slowly evolving waveform (SEW) quantization specifically targeted at low rate coding. The proposed method uses a pulse model whose parameters are implicitly contained in the quantized rapidly evolving waveform (REW) parameters, thus requiring no bits for transmission. Results indicate no degradation in perceptual speech quality when compared to that of the existing SEW quantization method. This retention of perceptual quality is in spite of a 12% reduction in the overall coder bit rate.

Index Terms—Low-rate speech coding, slowly evolving waveform (SEW), waveform interpolation (WI).

I. INTRODUCTION

THE WAVEFORM interpolation (WI) paradigm proposed by Kleijn [1] is the focus of much current research in speech coding circles. The WI paradigm involves first linear predictive (LP) filtering the input speech. The residual signal is then separated into pitch cycles (known as characteristic waveforms (CW) [1]) and these are used to form a two-dimensional (2-D) waveform, which evolves on a pitch synchronous nature. To maximize the smoothness of the 2-D surface the individual pitch length segments are aligned when constructing the surface. This 2-D waveform is then decomposed into Slowly evolving and Rapidly evolving waveforms (SEW/REW). The SEW and REW are down sampled and quantized separately in the encoder. The decoder reconstructs the SEW and REW via interpolation before recombining them. Synthesized speech is produced by converting the reconstructed 2-D surface back to a one-dimensional (1-D) signal and passing this signal through the linear predictive synthesis filter.

The current research involving WI can be broadly grouped into distinct categories these being a) low rate speech coding, and b) perfect reconstruction allowing waveform coding. A commonality in much of the present research involves identifying better means of representing the SEW to achieve improved perceptual quality [2], [3]. The original WI coder [1] operating at approximately 2.4 kbps quantizes and transmits only the SEW DFT magnitude values. A phase model is used

in the decoder to reconstruct the SEW. New methods proposed for better representation of the SEW waveform include directly quantizing the SEW DFT phase in an analysis by synthesis structure (AbyS) [2] and critically sampling and warping to a constant length to achieve perfect reconstruction [3]. These methods report improved perceptual quality but at the expense of increased complexity and bit rate.

We propose a new method of SEW transmission specifically targeted at low rate coding. The method proposed uses a pulse model whose parameters are implicitly contained in the quantized REW parameters, to represent the SEW.

This letter is organized as follows. Section II introduces the new SEW quantization technique and also defines the existing scheme. Section III compares the perceptual performance of the new technique with the existing SEW quantization method. Finally the major points are summarized in Section IV.

II. LOW RATE SEW QUANTIZATION

A. New Quantization Method

A new scheme to quantize and reconstruct the SEW is proposed. This method uses a new pulse modeling mechanism for reconstruction of the SEW waveform together with phase matching when recombining the REW and SEW waveforms.

The use of a pulse model for the SEW results in no bits being used for transmission of this parameter. The model used is based on the Zinc function which is defined in the discrete time domain as [4]

$$z(n - \lambda) = A \operatorname{sinc}(n - \lambda) + B \operatorname{cosec}(n - \lambda) = \begin{cases} A & n - \lambda = 0 \\ \frac{2B}{(n - \lambda)\pi} & n - \lambda = \text{odd} \\ 0 & n - \lambda = \text{even.} \end{cases} \quad (1)$$

The zinc model has been found to be superior in modeling the LP residual and is widely used in time domain analysis by synthesis schemes where the parameters A , B , and λ are selected to minimize the error between the pulse and the residual signal [4]. To allow the zinc pulse to be conveniently used in the WI structure, the pulse was translated to the frequency domain via the DFT. This allows straightforward interpolation between adjacent pulses of different lengths via zero padding. For low rate coding, the parameters A , B and λ cannot be transmitted due to bit rate constraints. It was found that speech of high perceptual quality could be produced by setting λ to zero. This forces the zinc pulse to be wholly positive and also places the pulse peak at the beginning of the frame. Positioning the pulse at the beginning of each frame is equivalent to removing the SEW's linear

Manuscript received October 11, 2000. This work was supported in part by an Australian Postgraduate Award (Industry) and a Motorola (Australia) Partnerships in Research grant, and Motorola, Inc., Australia. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Y. Shoham.

The authors are with Whisper Laboratories, TITR, University of Wollongong, Wollongong, NSW, Australia 2522 (e-mail: jl01@ouw.edu.au).

Publisher Item Identifier S 1070-9908(01)06117-X.

phase component, which is acceptable in WI as the linear phase is already modified by aligning the waveforms in the encoding stage prior to down sampling. The values of A and B are set according to

$$A = B = 1 - \frac{1}{N} \sum_{i=0}^{N-1} |\text{REW}(i)|. \quad (2)$$

Equation (2) uses the assumptions that the LP residual spectrum is flat and that spectrum has been normalized to unity value previously in the coding process by the removal of the gain term. Using these assumptions and the properties of the inverse discrete Fourier transform it can be shown that the resultant value of (2) is equal to the height of a single sample time domain pulse that would exhibit the required flat magnitude spectrum. While the zinc pulse consists of a large initial pulse followed by further impulses whose amplitudes decrease rapidly with time, informal perceptual testing has shown that setting the height of the initial pulse equal to the height of a single impulse, as calculated by (2), produces good results.

The method of deriving an implicit SEW pulse from the REW allows the pulse height to be dynamically varied according to the magnitude of the REW. Thus, for sections with a high noise content the pulse is small and vice versa. Forcing the pulse to be positive and of fixed position may appear to be suboptimal in modeling the SEW. However, initial perceptual testing indicated a preference for this configuration over a pulse of variable position and polarity.

B. Existing WI SEW Quantization [1]

The DFT magnitude coefficients for the lowest 800 Hz are quantized using a seven-bit Vector quantizer. The remaining DFT magnitude coefficients are calculated from the reconstructed REW using the fact that above 800 Hz, the overall residual signal magnitude spectrum may be considered flat. The SEW phase spectrum is not transmitted but is set equal to that of a predetermined model. The bit rate required for this method is 280 bps for the frame size used.

C. Recombination of the SEW and REW

In [1] the synthesized CW is generated by adding the Fourier series coefficients representing the SEW and REW. However, due to the prior discarding of both the SEW and REW phase information, this method of recombination does not ensure constructive rather than destructive recombination. Also it is possible that the REW may introduce an extra pulse into the reconstructed section. One option for improving the recombination procedure is to time align the entire SEW and REW waveforms. This method works well for the SEW quantized using the existing scheme. However, for the new zinc pulse scheme time aligning before recombination does not work well and produces hiss in the synthesized speech. This is due to the fact that the zinc pulse detailed in Section II-A exhibits a large initial impulse followed by further impulses of decreasing magnitude. This characteristic makes the maximum correlation criteria for time alignment unreliable as the correlation due to the subsequent impulses can tend to subtract from the correlation for

TABLE I
MOS TEST RESULTS

	Standard SEW	Pulse Model
MOS Score	3.45	3.44
95% confidence level	0.13	0.13

the initial impulse, thus causing the waveforms to align incorrectly. To achieve good reconstruction with the zinc pulse SEW, a method that matches the REW phase to that of the SEW below 800 Hz was developed. This ensures the phase of the reconstructed section is matched to the SEW phase below 800 Hz and is a combination of the ratio of SEW to REW phase above this figure [5]. Matching the low frequency phase where the SEW magnitude is dominant ensures that these low frequencies are aligned in the time domain and thus produce a degree of temporal masking around the SEW pulse. This masking removes the hiss from the reconstructed signal.

To determine the individual effect of the phase matching method, it was tested with the existing SEW quantized waveforms. Informal listening tests found that when compared to time alignment, the phase matching method resulted in neither an improvement nor degradation in the perceptual quality of the reconstructed speech.

III. EXPERIMENTAL RESULTS

A. Coder Configuration

The structure of the coder is as detailed in [1]. The coder allocates 26 bits for the LSF parameters, ten bits for the power, six bits for the pitch and eight bits for the REW waveform per frame, with a frame size of 25 ms. The eight bits allocated to the REW are used to represent the REW magnitude spectrum with random phase used in the decoder to reconstruct the REW. The REW quantization scheme used is the same as that detailed [1].

The overall bit rate of the base coder is 2 kbps. When the SEW is quantized using the existing method, 280 bps must be added to this value.

B. Subjective Test Results

For testing purposes, the coder detailed in Section III-A was used to code eight input speech sentences (four male, four female) from the TIMIT database using both of the SEW representations. Mean opinion score (MOS) testing was carried out using 24 untrained listeners each using Sony headphones. The results are shown in Table I.

The MOS scores indicate that the pulse model SEW representation produced no degradation in the quality of the synthesized speech when compared to the existing method. This is despite the pulse model requiring no bits for transmission compared to 280 bps for the existing method.

This indicates that for low rate WI coding, attempting to coarsely quantize and preserve the shape of the SEW with the limited number of bits available, offers no perceptual

improvement when compared to representing the SEW with a fixed shape zinc model that is smoothly evolving and thus easily interpolated.

IV. CONCLUSION

The bit rate saving and preservation of perceptual quality offered by the new pulse model representation of the SEW offers a significant advantage for low rate WI coding. The result is a 12% reduction in the overall coder bit rate. Alternately, the coder bit rate could be maintained and the extra bits used to more accurately quantize other parameters such as the LSF or the REW. As the parameters of the pulse model are inferred in the REW waveform, a more accurate representation of this waveform should also produce an improvement in the SEW quality, thus offering two contributions toward improving the overall quality of the coder.

In conclusion, the results suggest that to achieve the best tradeoff for bit rate and perceptual quality, WI should either

quantize the SEW very accurately requiring a higher bit rate as in [2] and [3] or opt for a parametric representation that is smoothly evolving and thus easily interpolated. Attempting to coarsely maintain the SEW shape using a limited bit count appears to offer little benefit.

REFERENCES

- [1] W. B. Kleijn and J. Haagen, "Waveform interpolation for coding and synthesis," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. New York: Elsevier, 1995, pp. 175–207.
- [2] O. Gottesman, "Dispersion phase vector quantization for enhancement of waveform interpolative coder," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing '99*, Phoenix, AZ, 1999.
- [3] N. R. Chong, I. S. Burnett, and J. F. Chicharo, "Adapting waveform interpolation (with pitch spaced subbands) for quantization," in *Proc. IEEE Workshop Speech Coding*, Poorvoo, Finland, 1999, pp. 96–98.
- [4] D. J. Hiotakakos and C. S. Xydeas, "Low bit rate coding using an interpolated zinc excitation model," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing '94*, vol. 3, Adelaide, Australia, 1994, pp. 884–997.
- [5] H.-G. Kang and D. Sen, "Phase adjustment in waveform interpolation," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing '99*, vol. 1, Phoenix, AZ, 1999, pp. 261–264.