# New Techniques for Multi-Prototype Waveform Coding at 2.84kb/s

*I. S. Burnett and G. J. Bradley*
*Department of Electrical and Computer Engineering,*
*University of Wollongong, NSW, Australia*

## ABSTRACT

This paper describes new techniques for Prototype Waveform (PW) coding at coding rates as low as 2.84kb/s. The algorithm produces good communications quality speech with significant improvements over previously reported PW coding schemes. In Multi-Prototype Waveform (MPW) coding, prototypes are extracted at 2.5ms intervals. No distinction is necessary between voiced and unvoiced speech since the normalised LP residual prototypes are coded as a combination of a noise vector and smoothly evolving pitch pulse. At low bit rates it is unnecessary to explicitly code the underlying pulse shape - the relative magnitude of the rapidly evolving noise vectors is sufficient to describe the level of periodicity in each prototype. Prototypes are quantised using either an open or closed-loop scheme with similar results. Quantised prototypes are interpolated by continuous phase interpolation of Fourier coefficients in the discrete frequency domain.

## 1. INTRODUCTION

In this paper we describe and discuss significant advances in the design of prototype waveform coders [1][2][3], an early example of which was discussed in our previous paper [3]. Prototype Waveform (PW) coders represent the speech as a series of characteristic waveforms. These are extracted at typically 40Hz and then interpolated to reconstruct the speech. In previous work [3], PW coding was restricted to usage in voiced speech and typically a reduced complexity CELP coding scheme was used for unvoiced speech sections. Such an approach has the disadvantage of making a discrete voiced/unvoiced decision which can lead to non-linearity and distortion in transition regions [4]. The coders described in this paper use a new prototype model proposed by Kleijn [5] which provides a unified PW structure for both voiced and unvoiced speech (and also background noise [5]).

The unified coding scheme also leads to the removal of significant redundancy when coding prototypes; this has led to the adoption of the multi-prototype scheme which extracts many more prototypes per frame than in previous coders [1][2]. Our Multi-Prototype Waveform (MPW) coder extracts prototypes at a rate of 400Hz, significantly increasing the rate of change (evolution bandwidth [5]) in the characteristic waveforms which the coder can accommodate.

The overall result of this increased resolution is improved tracking of speech dynamics which makes the coding of signals with low periodicity practical. Previous PW coders, with low evolution bandwidths, failed to accommodate such signals, leading to the adoption of the switched coding schemes.

## 2. MULTI-PROTOTYPE WAVEFORM CODING

The new MPW coder has a number of significant differences from the previously reported PW coders [1][2][3]; these can be separated primarily into the areas of Prototype Extraction and Quantisation. The basic structure of the MPW coding algorithm is shown in the block diagram of Fig. 1. While the details of extraction and quantisation methods varied, this structure was used throughout this work.

### 2.1 Prototype Extraction

For correct extraction of prototypes it is necessary to detect the pitch of the input speech frame (200 samples at 8kHz). The pitch of each frame is computed using a derivation of the autocorrelation technique [6]. To ensure a 'smooth pitch track' the deviation of the pitch over each frame is restricted. Prototypes are then extracted from the LP residual every 20 samples with the period of each prototype taken to be the linear interpolation of the pitch between its frame update points. This contrasts with the single prototype extracted for each frame in references [1][2] and makes the previously reported extraction technique [2] inappropriate.

Interpolation of the residual to increase prototype selection [2] is rendered unnecessary by the significant decrease in prototype extraction interval. In the new technique, a small number of candidate prototypes around each extraction point are assessed on the basis of minimising the squared error between the endpoints of the chosen prototype. To maximise tracking of the speech dynamics, it was found essential to limit the choice of prototypes to a narrow region surrounding the extraction point. This also avoids 'double' occurrences of transitory events in the input speech appearing in the synthesised output. Without quantisation these prototypes were found to give near-transparent speech quality when aligned, interpolated and passed through the LP synthesis filter.

A significant difference between MPW coding and the previous work is the extraction of prototypes regardless of the voiced/unvoiced nature of the speech. While this removes the switching problems of previous techniques [3][4], it can lead



0-7803-2431-5/95 $4.00 

to undesirable periodicity during the coding of unvoiced speech sections. This has been avoided by ensuring that prototypes derived from unvoiced speech regions are encoded with a 'period' of a minimum of 4 prototype update intervals (80 samples).

## 2.2 Prototype Quantisation

Previous prototype coders [3] quantised the prototypes using frequency or time domain quantisation of the total residual prototype characteristics. Kleijn's recent paradigm [5] decomposes each prototype into a slowly evolving waveform (SEW) and a rapidly evolving waveform (REW). The REW is a noise-like waveform, which may be quantised purely as a spectral magnitude vector. The SEW and REW are formed by respectively, low and high pass filtering the DFT coefficients of the progressively time-aligned [1] prototypes. In the DFT domain the time alignment operation can be shown to be the determination of $\theta'$ such that the cross-correlation of equation (1) is maximised:

$$\theta = \underset{\theta'}{argmax} \sum_{k=0}^{\tau_m - 1} \mathrm{Re}\left[ \mathbf{P}_m(k) \mathbf{P}^*_{m-1}(k) e^{j2\pi k\theta'} \right] \qquad (1)$$

where $\tau_m$ is the pitch period of the mth and current prototype, which has DFT coefficients $\mathbf{P}_m(k)$.

We have also investigated an alternative definition of SEW in which the SEW is simply computed as the mean of the current 'frame' of prototypes. This definition has been found to be advantageous when the pitch detector is unreliable and produces undetected doubling or halving effects.

Both SEW definitions lead to decomposition into an underlying pitch pulse and a noise-like waveform. Since prototype energy is normalised, the REW spectral magnitude can be used to describe both the noise-like behaviour and level of periodicity in the prototype waveform. The evolution of SEW and REW waveform magnitudes from our current coder are shown in Fig. 2 (a) and (b).

Two prototype quantisation schemes based on the REW/SEW paradigm have been tested. Both coding schemes operate on perceptually weighted residual prototypes with normalised energy; the related gain term being transmitted twice per frame. In both cases, the REW is explicitly, but coarsely, quantised and the level of periodicity in the prototype derived directly from the REW.

## 2.3 Open-Loop Quantisation

This scheme is open-loop and uses simple vector quantisation of the REW. The magnitude of the overall REW spectrum is used as the basis of quantisation. It has been shown [5], and can be seen from Fig. 2 that there is a clear linkage between periods of high energy in the REW and low energy in the SEW (and vice versa). The REW is chosen from a selection of 16 spectral vectors which are distributed between various mean magnitude levels. The magnitude (and linked vector) chosen is used as a basis for the SEW magnitude (since the REW and SEW contributions are inherently linked in the normalised prototype), allowing a single vector selection to be made. The SEW is then reconstructed from a pulse shape (based on the current pitch period) and magnitude. Only the single REW vector is then required for transmission.

Good results were obtained using a codebook of 16 REW vectors (with related REW and SEW magnitudes). The REW vectors are essentially noise-like spectra and were chosen on the basis of recorded REWs from the speech database.

## 2.4 Closed-Loop Quantisation

The closed-loop scheme selects an REW/SEW vector combination on the basis of spectral magnitude comparison with the normalised prototype. This requires the use of a linked codebook structure. The 16 REW vectors (similar to those used in the open-loop scheme) are of differing mean magnitude and spectral shape, allowing a linked SEW magnitude to be derived for each REW vector. The SEW/REW combination (ie. the quantised prototype) is then created as the summation of the noise-like REW and the SEW pulse (the length of which is based on the current prototype period). It is this combination which is compared to the overall normalised prototype for purposes of SEW/REW quantisation.

The REW spectral magnitude vectors are weighted so as to suppress low frequency components. These were found to 'interfere' with the low frequency, periodic SEW structure of the prototype. This caused unacceptable distortion in the output speech due to the loss/corruption of the pitch-pulse structure.

One advantage of the closed loop coding method is a significant reduction in the computation required by the REW/SEW quantisation scheme. While this method utilises the concept of decomposing the normalised PW into noise-like and pitch-pulse waveforms, it does not require the complex filtering operations involved in explicitly computing the SEW and related REW.

A further, perceptual, advantage is that basing the decision process on the combination of SEW and REW avoids undesirable interactions between the quantised SEW pulses and fluctuations in the quantised REW noise. This is due to the inherent consideration of the SEW shape in the closed-loop quantisation approach.

## 2.5 Speech Reconstruction

In both quantisation schemes discussed above, the residual prototype is reconstructed solely on the basis of a quantised gain term, the encoded REW magnitude and the related SEW periodicity. The spectral magnitude vector representing the REW is used to modulate a noise source, the output of which is added to the SEW pulse to reproduce the prototype. Improved output speech quality was achieved by time aligning the modulated REW contribution with the previous quantised prototype. This ensures smooth evolution of the REW contribution between successive prototypes and also serves to prevent 'destructive interference' between REW and SEW contributions. For each frame, five prototypes are linearly interpolated to reconstruct the excitation, which is then filtered by the LP synthesis filter. The continuous interpolation between

successive prototypes is performed using the following algorithm: The 'pitch' period of successive prototypes evolves such that the period at a given interpolation point, i, between prototypes is given by:

$$\tau_i = \alpha_i P_m + (1-\alpha_i) P_{m-1} \quad \text{for } 0 \le \alpha_i \le 1 \qquad (2)$$

where $\alpha_i$ is a linear parameter expressing progression between the two prototypes in time.

For the phase of the prototypes to evolve smoothly between updates the phase at each sample, i, progresses as:

$$\varphi_i = \varphi_i + \frac{2\pi}{\tau_i} \qquad (3)$$

The continuously interpolated excitation over the interpolation interval L, between the previous and mth prototype is then found as:

$$e_m(i) = 2\,\mathrm{Re} \sum_{k=0}^{\tau_i - 1} \left[\alpha_i P_m(k) + (1-\alpha_i) P_{m-1}(k)\right] e^{j k \varphi_i} \quad \text{for } i = 0, \ldots\ldots, L-1 \qquad (4)$$

This DFT domain interpolation can be regarded as a progressive linear adjustment of the Fourier basis functions' phases. This successfully caters for changes in pitch and hence prototype length.

### 2.6 Bit Allocation

The MPW coder uses the following bit allocation:

| Parameter | No. of Bits | Total bits / frame |
|---|---|---|
| LSFs | 34 | 34 |
| Pitch | 7 | 7 |
| REW/ prototype | 4 | 20 |
| Gain | 5 | 10 |
| **Total** | - | **71** |

This is equivalent to a bit rate of 2.84kb/s when using a 200 sample frame at 8kHz. The LSFs are currently quantised using an adapted Federal Standard 1016 quantiser, primarily for reasons of computational load. Further reductions in bit-rate could be made using Split-VQ techniques [7].

## 3. RESULTS AND CONCLUSIONS

An overall bit rate of 2.84 kb/s has been achieved with the MPW coder. Tests were performed on a speech database of 30 mixed male/female speakers and preliminary listening tests place the coded speech close to that of 4.8 kb/s CELP algorithms. The coder generates more natural sounding speech than the harsh results obtained from low rate CELP implementations and shows improved tracking of speech dynamics compared to previous PW coders [1][2][3].

The current MPW coder also highlights a number of critical areas and issues relevant to this type of coder. In particular the tracking of the pitch (or essentially the optimum prototype repetition rate) is an essential component of the MPW coding technique.

A smooth, accurate pitch track is an essential component - the break down of the pitch track in transitional voiced/unvoiced speech or in noise is currently the main obstacle to improved performance. A second problem with the current pitch tracking is the delay incurred by the requirement for 'future' pitch information (which allows linear interpolation between updates).

It is also unclear as to whether the SEW/REW decomposition is optimum. While it appears that the resulting vectors are essentially independent, it is feasible that other prototype decompositions may give similar or higher performance.

This work shows that Multi-Prototype Waveform coding is a suitable technique for low bit rate coding. The adoption of a closed loop quantisation scheme leads to significant reductions in the algorithm's complexity.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the assistance of W.B. Kleijn of AT&T Bell Laboratories in supplying a pre-print of reference [5].

This work was performed under the auspices of the Switched Networks Research Centre at the University of Wollongong.

## REFERENCES

[1] W. B. Kleijn, "Encoding Speech Using Prototype Waveforms," *IEEE Trans. Speech and Audio Processing*, Vol. 1, pp. 386-399, 1993.

[2] W. B. Kleijn, "Continuous Representations in Linear Predictive Coding," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Toronto, pp. II 201 - II 204, 1991.

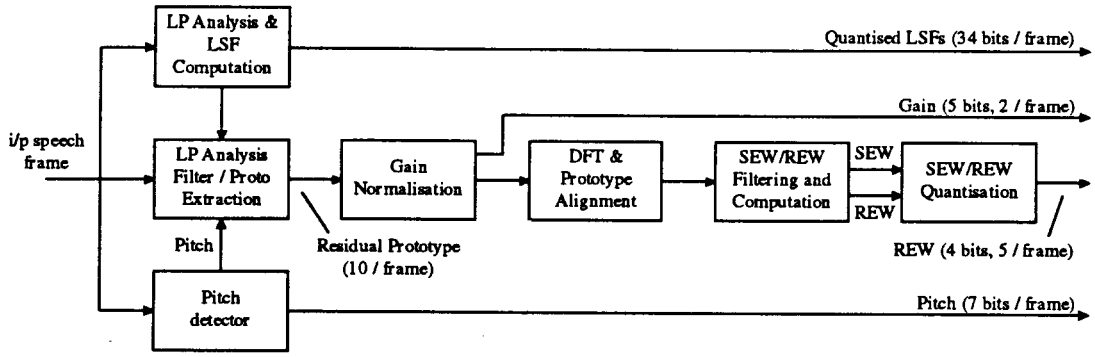
[3] I. S. Burnett, R. J. Holbeche, "A Mixed Prototype Waveform / CELP Coder for sub 3kb/s", *Proc. Int. Conf. Acoust. Speech Signal Process.*, Minneapolis, pp. II 175 - II 178, 1993.

[4] G. Kubin, B. S. Atal & W. B. Kleijn, "Performance of Noise Excitation for Unvoiced Speech," *Proc. IEEE Workshop on Speech Coding for Telecommunications*, pp. 35-36, 1993.

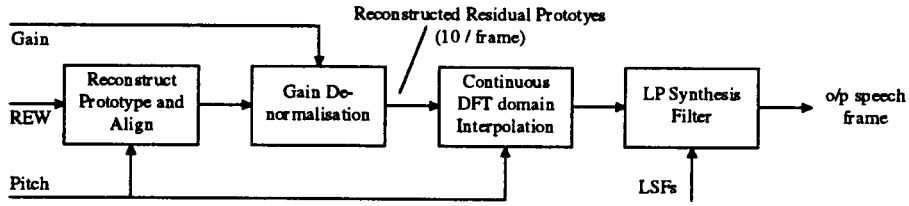
[5] W. B. Kleijn, J. Haagen, "Transformation and Decomposition of the Speech Signal for Coding," *IEEE Sig. Proc. Letters*, Vol. 1, No. 9, pp. 136-138, Sept. 1994.

[6] J. J. Dubnowski, R. W. Schafer and L. R. Rabiner, "Real-Time Digital Hardware Pitch Detector," *IEEE Trans. on Acoust., Speech and Signal Proc.*, Vol. 24, No. 1, pp. 2-9, Feb. 1976.

[7] K. K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame," *Proc. Int. Conf. Acoust., Speech and Signal Proc.*, pp. 661-664, May 1991.




(a) MPW Encoder




(b) MPW Decoder

**Fig. 1:** Block Diagrams of the Multi-Prototype Waveform (MPW) coder and decoder structures.

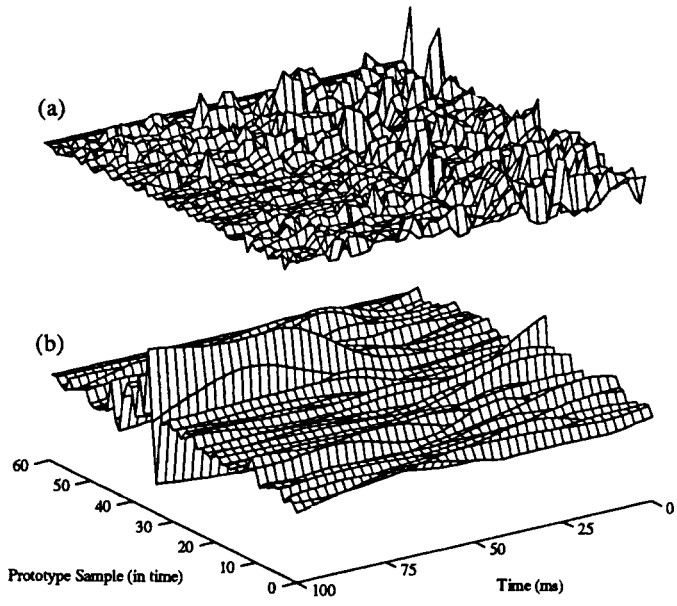



**Fig. 2:** The evolution of the REW (a) and the SEW (b) at the beginning of the utterance "strong" spoken by a male speaker. The nature of the SEW as the underlying pitch pulse shape and the REW as the noise-like content of the prototype can be seen. The dominance of the REW in 'unvoiced' and the SEW in 'voiced' sections of speech is also clear.