Proc. Biennial Symposium Commun. (Kingston, ON), pp. 107-110, May 1994

A Pitch Pulse Evolution Model for a Dual Excitation Linear Predictive Speech Coder

Jacek Stachurski[1] and **Peter Kabal**[1,2]

[1] Department of Electrical Engineering
McGill University
Montreal, Quebec H3A 2A7

[2] INRS-Télécommunications
Université du Québec
Verdun, Quebec H3E 1H6

Abstract

This paper introduces a new technique to model the excitation waveform for a linear predictive speech coder. The target application is high quality speech coding for rates near 4 kb/s. Our pitch pulse evolution model decomposes the excitation into two separate but simultaneous signals: the evolving pitch pulse component and the unvoiced, noise-like contribution. A number of formulations for decomposing the excitation waveform are suggested.

1. Introduction

Low-bit rate coding of speech using a code-excited linear prediction model (CELP) [1] has become popular in recent years. Much effort has been devoted to the efficient coding of the excitation vector and the linear prediction (LP) coefficients. The human ear is particularly sensitive to small changes in speech periodicity. Poor coding of the excitation for voiced segments is to a large part responsible for the degradation of speech quality with decreased bit rate. A number of techniques has been applied to address the problem for CELP-like coders (see the later review). Sinusoidal coding [2] uses an alternative model. Some of the recent work can be viewed as an attempt to combine the two approaches, namely time domain and frequency domain coding. Time-frequency interpolation (TFI) [3] seems to be a promising direction for low-bit rate speech coding. This study focuses on a technique which can be used within conventional CELP or in TFI — the difference being the domain in which the excitation vector components are quantized.

Our focus will be on coding of the excitation vector. This excitation models the residual after LP analysis. While the LP analysis can be coupled with the excitation modelling as suggested later, for the preliminary results discussed here, the LP analysis is carried out on the original speech waveform (sampling rate 8 kHz) in the conventional manner every 20 ms (10-th order analysis; Hamming window of length 240 samples; shifted in steps of 160 samples; interpolated in the LSF domain every 40 samples).

In TFI, the residual after linear prediction is transformed to the frequency domain using a discrete Fourier transform with a length keyed to the pitch period. In this way, successive transforms are lined up harmonic-by-harmonic. Differential coding is applied to the amplitudes of the harmonics. This differential coding has the effect of reinforcing the underlying pitch waveshape, but with noise contributions from both the present and the previous pitch waveforms.

Our goal is to identify the underlying pitch waveshape in a more robust manner. Observing successive pitch waveforms, one can see the evolution, though the waveform is often obscured by noise components that tend to be different for each pitch signal. We have developed a pitch pulse evolution (PPE) model for this purpose. It consists of two parts. For voicing there is a pitch pulse shape that evolves slowly. These pitch pulse waveshapes may overlap if the pitch interval is small enough. Superimposed on the pitch waveform is an unpredictable component — the unvoiced, noise-like part of the signal. In the context of TFI, the pitch waveshape has to be coded in both amplitude and phase, while the coded unvoiced noise-like component need not reproduce the phase; random phase is adequate for good reproduction. The phase response of the pitch signal can be decomposed into two components. The linear phase component simply shifts the pitch waveform into the correct position. In effect, transmitting the pitch period is an incremental approach to determining the position of the pitch waveform. If the remaining phase is zeroed, the pitch waveform is maximally peaky. One can then attribute the second component of the phase response to dispersal of the pitch pulse.

2. Improving the Quality of Voiced Speech

Consider determining a canonical waveshape based on a relatively long history of pitch pulses. In the case of speech produced by an idealized model, the waveshape is that of the glottal pulse. Initially we did try to extract such a pulse, to determine a suitable long-term amplitude and phase spectrum. The resulting pitch pulse was found to be quite oscillatory, yet highly correlated from pulse to pulse. This then suggested that any significant smoothing must be done pulse-to-pulse, not within a pulse. Our failure to extract a pitch pulse which resembles the classical smooth glottal pulse depicted in text books is not surprising. Isolating the glottal pulse is a notoriously difficult problem; the interaction of unmodelled elements, as well as non-linear factors lead to pitch pulses which tend to ring from the abrupt change at the glottal closing time [4, pp. 341–342]. Our PPE approach does not presuppose any particular shape for the pitch pulses, only that the underlying pitch pulses have some form of continuity from one instance to another.

In speech coding, one finds diverse implementations which all attempt to improve the coding of voiced speech. There are lessons to be learned in hindsight from these implementations. Prototype waveform interpolation [5] attempts to model the pitch pulse waveforms. Waveforms of relatively distant pitch pulses are extracted and intermediate pulses are interpolated from these prototypes. This approach does not fully use the actual intermediate pulses to identify appropriate prototypes.

This research was supported by a grant from the Canadian Institute for Telecommunications Research under the NCE program of the Government of Canada.

Shoham [6] demonstrated an improvement in speech quality if the noise-like part of the excitation is reduced below the level that would be chosen to minimize the (weighted) mean-square error. During voiced speech, the pitch synthesis filter supplies most of the excitation. Decreasing the unvoiced component enhances the periodicity of the waveform. However in CELP, the pitch synthesis filter can only shift old segments of the waveform into the present frame. This means that this approach produces periodic speech, but with an inadequate mechanism to adapt the pitch pulse shape and suppress noise on the pulse shape. In fact the noise is just made periodic so that it is no longer perceived as noise-like, but merely an error in spectral weighting of the pitch harmonics.

In pitch sharpening [7], a single tap recursive comb filter is used to filter the pitch contribution. The filter coefficient is adaptive as in Shoham's scheme. The essential difference from Shoham's constrained excitation approach is that the reduced gain is used only for updating the pitch contribution, while the "optimum" gain is retained for reconstructing the waveform.

A complementary approach is that of harmonic noise weighting [8] using a multi-tap pitch filter to weight the error. This weighting has the effect of steering the stochastic codebook contribution to one which matches the pitch waveform harmonics, while ignoring the noise contribution between harmonics. Wang and Gersho [9] moved the pitch filter into the feedback loop to make it more effective.

3. Pitch Pulse Evolution

For the purposes of aligning the pitch waveforms, it may be better to think of the waveforms in the time domain, rather than the frequency domain, though, of course, either approach can be used. The pitch waveforms will be considered to be vector entries in a table. We can now apply prediction and filtering on these entries, processing the corresponding components in the vectors. This filtering can be considered to be a generalization of comb filtering. However, one should note that the pitch interval changes between successive entries in the table, so that the comb filter has time-varying lags.

Our pitch pulse evolution scheme has the following features.

(i) The approach does not sharply categorize speech into voiced and unvoiced. The proportion of the two components changes with time as shown schematically in Fig. 1. The pitch pulse waveform can be frozen or adapted very slowly during unvoiced segments. This means that a pitch pulse waveform is available for coding perceptually important voicing onsets.

(ii) We decompose the overall residual waveform into predictable and non-predictable components for separate coding. The receiver uses prediction to determine the pitch pulse waveform based on the information that it has available. The transmitter has more information available (the present and possibly the future). It does filtering to produce an estimate of the pitch pulse waveform. The difference between the prediction and the estimate can be coded for transmission to the receiver. The unpredictable part (roughly the stochastic codebook contribution) models the unvoiced contribution.

(iii) The alignment of the vectors in the table can be done to sub-sample resolution to improve the performance. The reconstruction at the receiver may be done at normal resolutions or via interpolation of the pitch interval.

(iv) The filtering of the time-aligned array of waveforms can be viewed as 2-D filtering. Very general filters could be specified. Our early experiments suggest that simple 1-D filters, akin to comb filters, that operate on corresponding vector elements are very effective. However, some coupling between nearby vector elements (giving a true 2-D filter) may also be useful for smoothing.

(v) In PPE the pulse shape can be decoupled from its final multiplicative contribution to the excitation waveform. If fact our later formulations are based on the prediction and filtering of the normalized signals. Separate quantization of the gain and the shape can be used.

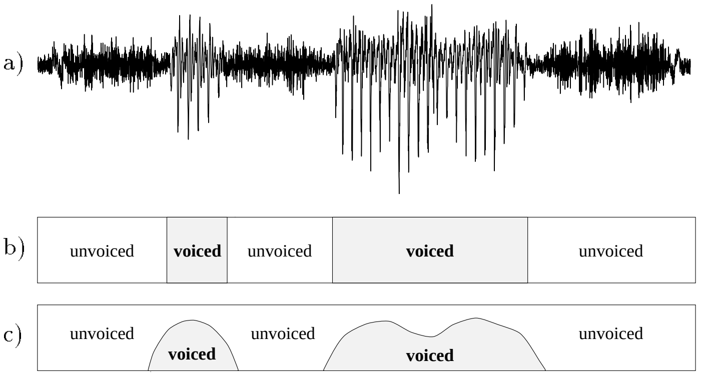

Fig. 1 (a) A noisy speech signal (b) traditional voiced/unvoiced division (c) voiced/unvoiced decomposition in the PPE dual excitation model

PPE also allows for a number of new approaches which may further improve the performance.

(i) The pitch waveforms can be separated, even though for short pitch intervals they overlap in time. Thus once we have identified a pitch pulse waveform, the tail of this waveform which overlaps the next pitch waveform can be subtracted out. This removes changes in the apparent pitch waveform which occur due to changes in the pitch interval, even if the underlying pitch pulse waveform is constant.

(ii) The LP analysis can be modified to minimize the error between the residual and the target pitch pulse waveform.

4. Decomposition and Estimation

A general formulation of the PPE method is developed in this section. We adopt the notation in which ˜ marks coded vectors available at both the transmitter and the receiver. Vectors obtained in the process of prediction or filtering are marked with ˆ.

A pitch length vector of the residual corresponding to the time instant i is denoted as $\mathbf{u}^{(i)}$ and its coded equivalent as $\tilde{\mathbf{u}}^{(i)}$. The past coded pitch vectors corresponding to the slowly evolving pitch pulse shape are written as $\tilde{\mathbf{v}}^{(i-1)}$, $\tilde{\mathbf{v}}^{(i-2)}$, ... All vectors are assumed to be properly aligned.

The new pitch vector can be predicted from the past values of $\tilde{\mathbf{v}}$ and $\tilde{\mathbf{u}}$ according to some prediction function $\mathcal{F}_p$:

$$\hat{\tilde{\mathbf{v}}}^{(i)} = \mathcal{F}_p\,(\,\tilde{\mathbf{u}}^{(i-1)}, \tilde{\mathbf{u}}^{(i-2)}, \ldots,\ \tilde{\mathbf{v}}^{(i-1)}, \tilde{\mathbf{v}}^{(i-2)}, \ldots). \qquad (1)$$

The same prediction can be performed in both the transmitter and the receiver with the function $\mathcal{F}_p$ fixed or adaptive.

The transmitter has access to more information, namely the uncoded versions of $\mathbf{u}$ (including the past, present and possibly the future ones if delay is permissible) and $\hat{\mathbf{v}}$ (unquantized past estimates). It can therefore form a better estimate of the present value of $\mathbf{v}$, according to some filtering function $\mathcal{F}_f$:

$$\hat{\mathbf{v}}^{(i)} = \mathcal{F}_f\,(\,\ldots, \mathbf{u}^{(i+1)}, \mathbf{u}^{(i)}, \mathbf{u}^{(i-1)}, \ldots,\ \hat{\tilde{\mathbf{v}}}^{(i)}, \hat{\mathbf{v}}^{(i-1)}, \ldots). \qquad (2)$$

Function $\mathcal{F}_f$ can also use vectors $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ directly, although the use of $\hat{\tilde{\mathbf{v}}}^{(i)}$ seems sufficient.

We define a vector $\mathbf{d}^{(i)}$ representing the unpredicted drift of the pitch vector $\mathbf{v}^{(i)}$, and a vector $\mathbf{n}^{(i)}$ representing the combined factors of the unvoiced, aperiodic part of $\mathbf{u}^{(i)}$ and the background noise of the signal so that

$$\mathbf{d}^{(i)} = \hat{\mathbf{v}}^{(i)} - \hat{\tilde{\mathbf{v}}}^{(i)}\,, \quad \mathbf{n}^{(i)} = \mathbf{u}^{(i)} - \tilde{\mathbf{v}}^{(i)} \tag{3}$$

with the quantized versions $\tilde{\mathbf{d}}^{(i)}$, $\tilde{\mathbf{n}}^{(i)}$ used in

$$\tilde{\mathbf{v}}^{(i)} = \hat{\tilde{\mathbf{v}}}^{(i)} + \tilde{\mathbf{d}}^{(i)}\,, \quad \tilde{\mathbf{u}}^{(i)} = \tilde{\mathbf{v}}^{(i)} + \tilde{\mathbf{n}}^{(i)}. \tag{4}$$

Note that with this formulation $\mathbf{n}^{(i)}$ accounts also for the quantization noise of $\tilde{\mathbf{d}}^{(i)}$.

In general the transmitter performs both operations, $\mathcal{F}_p$ and $\mathcal{F}_f$. The receiver performs $\mathcal{F}_p$ and based on the transmitted information reconstructs the waveform.

Many formulations of $\mathcal{F}_p$ and $\mathcal{F}_f$ are possible. We will describe two based on the linear filtering paradigm and one based on an error minimization criterion.

In the following presentation we will separate the gain and the shape of each vector so that $\mathbf{u} = \gamma\underline{\mathbf{u}}$, $\mathbf{v} = \beta\underline{\mathbf{v}}$, and $\mathbf{n} = \alpha\underline{\mathbf{n}}$ where vectors marked with an underscore are normalized to unit energy. We also simplify the notation by dropping the time index i. The superscript (i) is omitted and $(i-k)$ is replaced by a subscript k. From now on, parameter k is assumed to be in the range $1, \ldots, N$. We have $\mathbf{u} = \mathbf{u}^{(i)}$, $\mathbf{u}_k = \mathbf{u}^{(i-k)}$ and $\mathbf{u}_k = \beta_k\underline{\mathbf{u}}_k$.

All the vectors are considered as column vectors and we define $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_N]$, $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_N]$, $\mathbf{b} = [\beta_1 \cdots \beta_N]^T$, so for example $\tilde{\underline{\mathbf{V}}} = [\tilde{\underline{\mathbf{v}}}_1 \cdots \tilde{\underline{\mathbf{v}}}_N]$. We also use a vector normalizing function $\mathcal{N}(\cdot)$ defined as $\mathcal{N}(\cdot) = (\cdot)/\left[(\cdot)^T(\cdot)\right]^{1/2}$.

4.1. Estimation Based on Linear Filtering

In the first method we evaluate

$$\boldsymbol{v}_p = a_{p_1}\tilde{\underline{\mathbf{v}}}_1 + a_{p_2}\tilde{\underline{\mathbf{v}}}_2 + \ldots + a_{p_N}\tilde{\underline{\mathbf{v}}}_N \tag{5}$$

$$\hat{\tilde{\beta}} \;=\; a_{\beta_1}\tilde{\beta}_1 + a_{\beta_2}\tilde{\beta}_2 + \ldots + a_{\beta_N}\tilde{\beta}_N \tag{6}$$

and

$$\boldsymbol{v}_f = a_{f_0}\underline{\mathbf{u}} + a_{f_1}\hat{\underline{\mathbf{v}}}_1 + \ldots + a_{f_N}\hat{\underline{\mathbf{v}}}_N \tag{7}$$

$$\hat{\beta} \;=\; \mathbf{u}^T\hat{\underline{\mathbf{v}}} \tag{8}$$

where a_{p_k}, a_{β_k} and a_{f_k} are fixed scalar parameters of our prediction/filtering operations.

Vectors $\boldsymbol{v}_p$ and $\boldsymbol{v}_f$ have to be normalized after each calculation because (5) and (7) do not guarantee that the result is of unit energy. The use of the vector $\underline{\mathbf{u}}$ in (7) will effectively scale the underlying $\underline{\mathbf{v}}$ into a lower energy value because $\underline{\mathbf{u}}$ may contain a sizeable noise component. We should therefore rescale $\underline{\mathbf{u}}$ by $1/\left(\underline{\mathbf{u}}^T\mathcal{N}(\boldsymbol{v}_p)\right)$ to compensate for the noise energy. This is equivalent to replacing a_{f_0} with $a_{f_0}/(\underline{\mathbf{u}}^T\mathcal{N}(\boldsymbol{v}_p))$. The inaccuracy of the estimation of the noise in $\underline{\mathbf{u}}$ is not very important because of the later normalization of $\boldsymbol{v}_f$.

Write $\mathbf{a}_p = [a_{p_1} \cdots a_{p_N}]^T$, $\mathbf{a}_\beta = [a_{\beta_1} \cdots a_{\beta_N}]^T$ and $\mathbf{a}_f = [a_{f_1} \cdots a_{f_N}]^T$, so that

$$\boldsymbol{v}_p = \tilde{\underline{\mathbf{V}}}\mathbf{a}_p\,, \;\; \hat{\tilde{\beta}} = \tilde{\mathbf{b}}^T\mathbf{a}_\beta\,, \;\; \hat{\beta} = \mathbf{u}^T\mathcal{N}(\boldsymbol{v}_p) \tag{9}$$

$$\boldsymbol{v}_f = \frac{a_{f_0}}{\underline{\mathbf{u}}^T\mathcal{N}(\boldsymbol{v}_p)}\underline{\mathbf{u}} + \hat{\underline{\mathbf{V}}}\mathbf{a}_p\,. \tag{10}$$

We have

$$\mathcal{F}_p:\; \hat{\tilde{\mathbf{v}}} = \hat{\tilde{\beta}}\,\mathcal{N}(\boldsymbol{v}_p), \quad \mathcal{F}_f:\; \hat{\mathbf{v}} = \hat{\beta}\,\mathcal{N}(\boldsymbol{v}_f). \tag{11}$$

With this formulation the noise component $\mathbf{n}$ as well as $\mathbf{d}$, the difference between $\hat{\tilde{\mathbf{v}}}$ and $\hat{\mathbf{v}}$, must be coded. Coding $\mathbf{d}$ is necessary for the pitch pulse evolution at the receiver; with $\tilde{\mathbf{d}}$ available, $\tilde{\mathbf{v}}$ can be formed for the next iteration. The update rate for $\tilde{\mathbf{d}}$, however, may be lower than the one for $\mathbf{n}$.

In the second method we initially write

$$\boldsymbol{v}_p = a_{p_{01}}\tilde{\underline{\mathbf{u}}}_1 + a_{p_1}\tilde{\underline{\mathbf{v}}}_1 + \ldots + a_{p_N}\tilde{\underline{\mathbf{v}}}_N \tag{12}$$

$$\hat{\tilde{\beta}} = a_{\beta_1}\tilde{\beta}_1 + a_{\beta_2}\tilde{\beta}_2 + \ldots + a_{\beta_N}\tilde{\beta}_N \tag{13}$$

$$\boldsymbol{v}_f = \boldsymbol{v}_p\,, \quad \hat{\beta} = \mathbf{u}^T\mathcal{N}(\boldsymbol{v}_f). \tag{14}$$

With a similar argument as in the first case, we now have

$$\boldsymbol{v}_p = \frac{a_{p_{01}}}{\tilde{\underline{\mathbf{u}}}_1{}^T\tilde{\underline{\mathbf{v}}}_1}\tilde{\underline{\mathbf{u}}} + \tilde{\underline{\mathbf{V}}}\mathbf{a}_p \tag{15}$$

with $a_{p_1} = 0$ if we want vector $\tilde{\underline{\mathbf{u}}}_1$ to replace the previous estimate $\tilde{\underline{\mathbf{v}}}_1$. This gives again $\mathcal{F}_p$ and $\mathcal{F}_f$ as in (11).

With this method the drift of the pitch pulse shape does not have to be coded because the information about its evolution can be extracted from the coded vector $\tilde{\underline{\mathbf{u}}}$. In this case $\hat{\underline{\mathbf{v}}} = \hat{\tilde{\underline{\mathbf{v}}}}$.

The original TFI scheme can be viewed as a special case of this formulation with only a single non-zero coefficient $a_{p_{01}}$. It does not benefit from the additional smoothing available.

4.2. Estimation Based on Error Minimization

Given a set of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$ the predicted estimate of the underlying pitch pulse shape, $\hat{\mathbf{v}}$, should minimize the sum of the norms of the error (noise) vectors $\mathbf{n}_1, \ldots, \mathbf{n}_N$, where

$$\mathbf{n}_k = \mathbf{u}_k - \hat{\mathbf{v}}\,. \tag{16}$$

We allow the pitch pulse energy to vary in each of its contributions to the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$. We can rewrite (16)

$$\mathbf{u}_k = \hat{\beta}_k{}'\,\hat{\underline{\mathbf{v}}} + \alpha_k\,\underline{\mathbf{n}}_k. \tag{17}$$

Here $\hat{\beta}_k{}'$ is different from $\hat{\beta}_k$ of the previous section because it belongs to the time update i as opposed to the update $i-k$. A new set of $\hat{\beta}_k{}'$ is associated with each time i and in fact $\hat{\beta}_{0_k}{}'$ would correspond to the $\hat{\beta}_k$ of time $(i-k)$. Note that $\hat{\mathbf{v}} = \hat{\beta}_0{}'\hat{\underline{\mathbf{v}}}$, which is simply written as $\hat{\mathbf{v}} = \hat{\beta}\,\hat{\underline{\mathbf{v}}}$.

We let

$$\hat{\beta}_k{}' = \mathbf{u}_k{}^T\hat{\underline{\mathbf{v}}}\,. \tag{18}$$

From (17) and (18) it follows that the pitch pulse and the noise component are orthogonal, $\underline{\mathbf{n}}_k{}^T\underline{\mathbf{v}} = 0$, and the error energy, $\alpha_k{}^2$, is equal to

$$(\mathbf{u}_k - \hat{\beta}_k{}'\hat{\underline{\mathbf{v}}})^T(\mathbf{u}_k - \hat{\beta}_k{}'\hat{\underline{\mathbf{v}}}) = \mathbf{u}_k{}^T\mathbf{u}_k - \hat{\beta}_k{}'^2. \tag{19}$$

The minimization of the noise energy, equivalent to the minimization of $\sum_k \alpha_k{}^2$, is now equivalent to the maximization of $\sum_k \hat{\beta}_k{}'^2$.

In matrix form (18) becomes $\mathbf{U}^T\hat{\underline{\mathbf{v}}} = \hat{\mathbf{b}}'$. We want to solve

$$\max_{\|\hat{\underline{\mathbf{v}}}\|=1} \|\mathbf{U}^T\hat{\underline{\mathbf{v}}}\| \tag{20}$$

which is the L_2 norm or maximum singular value of $\mathbf{U}$, σ_1. Vector $\hat{\underline{\mathbf{v}}}$ is the first right singular vector of $\mathbf{U}^T$ (corresponding to the singular value σ_1) [1].

We introduce normalization and weighting of the vectors $\mathbf{u}_k$. The former will deemphasize vectors with larger energy (they may

[1] Vector $\hat{\underline{\mathbf{v}}}$ is also the eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{U}\mathbf{U}^T$.

have a strong noise component) while the latter can assign more importance to the most recent ones which hopefully carry better approximation of the current vector $\underline{\mathbf{v}}$. Now

$$\mathcal{F}_p : \qquad \hat{\tilde{\mathbf{v}}} = \hat{\tilde{\beta}}\, \hat{\tilde{\underline{\mathbf{v}}}}, \quad \hat{\tilde{\beta}} = \tilde{b}'^T \mathbf{a}_\beta$$

$$\hat{\tilde{\underline{\mathbf{v}}}} = \underset{\|\mathbf{V}\|=1}{\text{argmax}} \|\mathbf{A}_p \underline{\tilde{\mathbf{U}}}^T \mathbf{v}\| \tag{21}$$

$$\mathcal{F}_f : \qquad \hat{\mathbf{v}} = \hat{\beta}\, \hat{\underline{\mathbf{v}}}, \quad \hat{\beta} = \mathbf{u}^T \hat{\underline{\mathbf{v}}}$$

$$\hat{\underline{\mathbf{v}}} = \underset{\|\mathbf{V}\|=1}{\text{argmax}} \|\mathbf{A}_f \begin{bmatrix} \underline{\mathbf{u}}^T \\ \underline{\mathbf{U}}^T \end{bmatrix} \mathbf{v}\| \tag{22}$$

where matrices $\mathbf{A}_p$, $\mathbf{A}_f$ have weighting coefficients on their diagonals and zeros elsewhere.

With this method, the information about the pulse shape evolution is extracted at the receiver from the vectors $\tilde{\underline{\mathbf{u}}}$, so that coding of $\mathbf{d}$ is not necessary.

5. Results

To test the validity of these ideas we performed a number of experiments to track the underlying pitch waveform. The three formulations were implemented and tests were conducted to find a good set of the predictor and the filter coefficients. None of the vectors was coded. An example of the simulation data is shown in Fig. 2. The analysis was performed according to (9)–(11). The columns reflect a time evolution of the observed vectors. The following are shown: the aligned residual vector $\mathbf{u}$, the prediction $\hat{\tilde{\mathbf{v}}}$, the estimate $\hat{\mathbf{v}}$ and the error $\mathbf{n} = \mathbf{u} - \hat{\mathbf{v}}$. The orthogonal error between the aligned, consecutive residual waveforms is displayed in the last column. In a conventional differential approach this is the signal which would be coded. Note the relatively large energy of the vectors $\mathbf{u}_k - \beta_{opt}\mathbf{u}_{k-1}$ compared with the energy of the $\mathbf{n}$ vectors. This indicates, along with the general lack of correlation between vectors $\mathbf{n}$, that our PPE scheme is efficiently decomposing the pitch waveform. Also note that filtering along each $\mathbf{v}$ vector is not appropriate. The rapid changes in the time waveform seem to be part of the underlying pitch waveform.

The second filter-based method performs slightly worse (larger energy and more vector to vector correlations of $\mathbf{n}^{(i)}$). This is not surprising since this method does not use the pitch-pulse-shape change information $\mathbf{d} = \hat{\mathbf{v}} - \hat{\tilde{\mathbf{v}}}$, $\hat{\mathbf{v}} = \hat{\tilde{\mathbf{v}}}$. It is difficult to evaluate the third method if no coding is performed because the function $\mathcal{F}_p$ operates solely on the coded vectors $\tilde{\underline{\mathbf{u}}}$.

It is clear that the smooth evolution of the pitch pulses depends on smooth changes in the LP analysis parameters, for if different pitch waveforms are processed by different LP filters unnecessary pulse-to-pulse variations may be introduced. We believe that coupling the LP analysis to the evolving pitch waveform, as suggested earlier, will help in this regard.

Which method turns out to be the most advantageous can only be determined when they are embedded in a complete speech coder. This is our next task.

References

[1] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Predictive (CELP): High Quality Speech at Very Low Bit Rates," *Proc. IEEE Int. Conf. ASSP*, pp. 937–940, 1985

[2] R. J. McAulay and T. F. Quatieri, "Speech Analysis / Synthesis Based on a Sinusoidal Representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 744–754, Aug. 1986.

[3] Y. Shoham, "High-Quality Speech Coding at 2.4 to 4.0 kbps Based on Time-Frequency Interpolation," *Proc. IEEE Int. Conf. ASSP* (Minneapolis), pp. 167–170, April 1993.

[4] J. R. Deller, Jr., J. G. Proakis and J. H. L. Hansen, *Discrete-time Processing of Speech Signals*, Macmillan, 1993.

[5] W. B. Kleijn, "Methods for Waveform Interpolation in Speech Coding," *Digital Signal Processing*, Vol. 1, pp. 215–230, Jan. 1991.

[6] Y. Shoham, "Constrained stochastic excitation coding of speech at 4.8 kb/s", in *Advances in Speech Coding*, pp. 339–348, Kluwer, 1991.

[7] T. Taniguchi, M. Johnston and Y. Ohta, "Pitch-sharpening for perceptually improved CELP, and the sparse-delta codebook for reduced computation", *Proc. Inf. Conf. Acoust. Speech and Signal Processing* (Toronto), pp. 241–244, May 1991.

[8] I. A. Gerson and M. A. Jasiuk, "Techniques for improving the performance of CELP type speech coders", *Proc. Int. Conf. Acoust. Speech and Signal Processing* (Toronto), pp. 205–208, May 1991.

[9] S. Wang and A. Gersho, "Improved excitation for phonetically-segmented VXC speech coding below 4 kb/s", *Proc. IEEE Globecom'90* (San Diego), pp. 946–950, Dec. 1990.

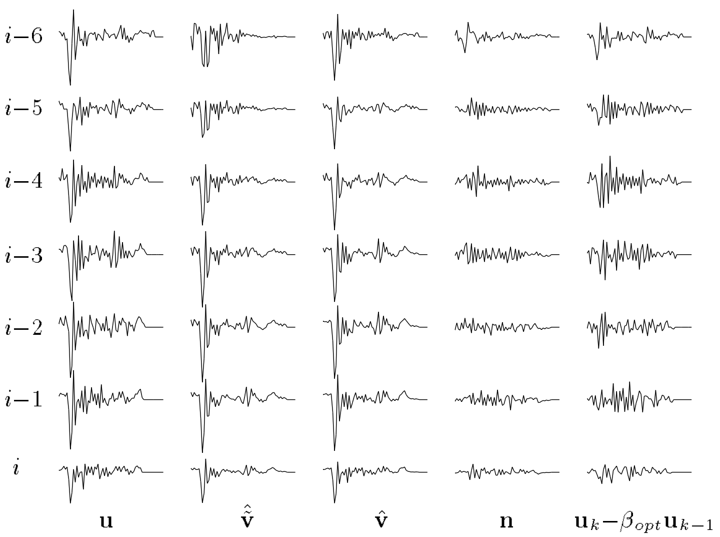

Fig. 2 The evolution of the pitch pulse waveform. Across the rows: the aligned residual vector $\mathbf{u}$, the predicted vector $\hat{\tilde{\mathbf{v}}}$, the estimate $\hat{\mathbf{v}}$, the error $\mathbf{n}$, and the orthogonal error between $\mathbf{u}_k$ and $\mathbf{u}_{k-1}$. The time index i runs down the columns.