PHASE QUANTIZATION BY PITCH-CYCLE WAVEFORM CODING

IN LOW BIT RATE SINUSOIDAL CODERS

Soledad Torres
e-mail: martor@tel.uva.es
ETSI Telecomunicación
Universidad de Valladolid
SPAIN

F. Javier Casajús-Quirós
e-mail: javier@gaps.ssr.upm.es
ETSI Telecomunicación
Universidad Politécnica de Madrid
SPAIN

ABSTRACT

A new phase coding algorithm is introduced in this paper, which works in the pitch-cycle waveform domain. It provides accurate phase coding at low bit cost. Its performance is analyzed inside a multiband excitation coder with improved onset representation. In this context, the introduction of original phase information by means of the proposed coding algorithm provides noticeable quality improvement without increasing the total bit rate of the coder.

1. INTRODUCTION

In low bit rate sinusoidal coding, original phases of the harmonics are substituted in the decoder by predicted or random ones [1, 2, 3]. The lack of original phase information causes noticeable degradation, specially in low pitched speakers. The addition of phase parameters to the model seems to be incompatible with low rates, since they need to be accurately coded in order to provide a significant improvement of quality.

It has been shown, however, that proper coding of the vowel onset characteristics allows better reconstruction of the voiced sounds [4]. In particular, the use of the original initial phases of the onset harmonics significantly reduces reverberation in synthetic speech, in spite of the fact that on the foregoing frames traditional prediction/random techniques are used. Onset phase coding can be done without increasing the bit rate, by substituting the mixed voiced-unvoiced spectral information of the frame containing the end of the unvoiced segment and the beginning of the voiced one [4]. Proceeding this way, an important number of bits are available for coding the phases of a single onset frame.

In order to develop an efficient algorithm to quantify this onset phase information, it is convenient to first define a suitable model for the phase spectrum. This topic is discussed in section 2. Next, a coding algorithm based on pitch-cycle waveform coding is introduced. Section 4 presents some practical considerations, together with performance results, obtained with a Multiband Excitation (MBE) coder enhanced with the proposed phase coding algorithm. Finally, the paper is concluded in section 5.

2. PHASE SPECTRUM MODELING

The most general model which can be applied to speech phase spectrum is based on the linear and separable speech production model. Following this approach, we can represent the short time phase of the mth harmonic component of a speech frame j as

$$\phi_m(j) = \psi_m(j) + \theta_m(j) \qquad (1)$$

where $\psi_m(j)$ and $\theta_m(j)$ represent the excitation and the vocal system components, respectively. The excitation of the vocal filter is usually approximated by a quasi-periodic sequence of pulses in voiced speech segments. This sequence can be expressed in terms of a summation of harmonic sinusoids which add coherently at the pitch pulse onsets. Thus, their phases are locked and evolve linearly with frequency, being predictable from frame to frame as

$$\hat{\psi}_m(j) = \psi_m(j-1) + \left[\omega_0(j-1) + \omega_0(j)\right]\frac{mS}{2} \qquad (2)$$

where ω_0 is the estimated fundamental frequency and S is the window shift.

Concerning the vocal system phase component, several approaches have been made in the sinusoidal coding context. Low bit rate multiband coders [1, 2, 3] simply add a random component to (2) to provide the required phase dispersion. In Sinusoidal Transfer Coding (STC) [6] the phase model is defined on a basis of a minimum phase assumption for the vocal tract. Although this assumption has proven to be reasonably effective, it is not entirely true, as long as the glottal pulse contribution to the vocal system phase is clearly non minimum. In fact, our experiments show that the speech quality may be improved by providing more accurate phase information, specially in unvoiced to voiced transition (onset) frames [4].

Work supported by National Project TIC96-0956-C04-03

Our proposal implies the computation of a minimum phase approach to $\theta_m(j)$, $\hat{\theta}_m(j)$, from the magnitude spectrum of the vocal system. The employed algorithm is the proposed in [6], which makes use of the discrete cepstral coefficients.

This minimum phase component is then added to the excitation component $\psi_m(j)$, providing a prediction to the actual phase spectrum. The residual error of this prediction is then coded in the pitch cycle waveform domain, as explained in next Section.

3. PITCH-CYCLE WAVEFORM CODING

Pitch-cycle waveform (PCW) coding is a promising technique which consists of modeling voiced speech by means of the mean pitch pulse waveform of the speech frame. The application of PCW to multiband coding was proposed in [5] as a method to simultaneously code spectral amplitudes and phases. In our approach, pitch-cycles will be employed to code phase information alone in onset frames, thus decoupling it from spectral envelope coding, which can be performed by several well known strategies.

Let us define our coding domain as a *phase-only PCW,* computed as a sum of sinusoids, as many as harmonics are present in the speech frame, with unit amplitudes and somehow defined phases. The length of our PCW, in samples, is the integer part of the pitch period of the speech frame. Phase modeling and coding take place in this domain and follow the block diagram of figure 1.

First, the measured harmonic phases in the actual frame, ϕ_m, are represented by the following PCW

$$PCW_O(i) = \sum_{m=1}^{L} \cos\left(2\pi \frac{mi}{P_I} + \phi_m\right) \qquad (3)$$
$$i = 0, 1, \ldots, P_I - 1$$

where L is the number of harmonics in the frame and P_I the integer part of the pitch period. The minimum phase approximation to the system phase is in turn represented by

$$PCW_{Min}(i) = \sum_{m=1}^{L} \cos\left(2\pi \frac{mi}{P_I} + \hat{\theta}_m\right) \qquad (4)$$
$$i = 0, 1, \ \ldots, P_I - 1$$

where $\hat{\theta}_m$ is the mth harmonic sample of the minimum phase spectrum of the vocal system. In order to consider (4) as an approximation to (3), the linear term ψ_m corresponding to the excitation must be eliminated from (3). This operation is equivalent to temporal waveform alignment in the PCW domain and is performed by finding the shift I that maximizes the circular cross-correlation of both PCWs. The difference between the I-shifted version of PCW_O and PCW_{Min} constitutes the residual waveform

$$PCW_{Res}(i) = PCW_O(i \oplus I) - PCW_{Min}(i) \qquad (5)$$

where the symbol $\oplus$ refers to circular shift.

A first obstacle in the quantization of the residual PCW arises from its variable length, as it depends on the pitch period of the speech frame. As we are working in a temporal waveform domain, an easy way to overcome this problem is to perform a length conversion on PCW_{Res}. Interpolation to its maximum possible length, equal to the maximum pitch period allowed by the coder, ensures no degradation is introduced at this point.

It is also convenient to reduce non-relevant information in PCW_{Res} before it is quantized. Previous work on phase perception [7] shown that original phase information can only be distinguished under 1.5 kHz. Elimination of non-relevant high frequency components of PCW_{Res} can be performed at the earliest steps of the coding process, by limiting the summations in (3) and (4) to the harmonics below 1.5 kHz.

Finally, the interpolated residual waveform is vector quantized. Figure 2 shows an example of the entire process, including the recovered waveform at the receiver. An inversion of the codification algorithm takes place there: from the received spectral envelope, the decoder reconstructs the minimum phase spectrum, and then PCW_{Min}, which added to the quantized residual waveform gives a rotated version of the original PCW, $\tilde{PCW}_o$. Harmonic phases under 1.5 kHz are recovered making use of the expression [4]

$$\theta_m = -\arctan \frac{\sum_{i=0}^{P_I-1} \tilde{PCW}_0(i) \sin\left(2\pi \frac{mi}{P_I}\right)}{\sum_{i=0}^{P_I-1} \tilde{PCW}_0(i) \cos\left(2\pi \frac{mi}{P_I}\right)} \qquad (6)$$

and those over 1.5 kHz are directly obtained from the sampling of the minimum phase spectrum (see figure 4).

It is worth noting that in the explained phase coding and decoding process, the lineal component is lost unless the shift I is also quantized and sent to the decoder. Otherwise, this component can also be recovered at the decoder following simple prediction techniques like (2).

4. IMPLEMANTATION AND PERFORMANCE

The performance of the proposed phase coding algorithm has been analyzed by including it in a low bit rate MBE coder with improved onset representation [4]. The improvement consists basically of a careful representation of onset temporal characteristics, among them the phases of the just born harmonics. In order to test the phase coding scheme, the rest of the coder parameters are left unquantized.

A database of approximately 30 minutes of continuous spanish speech from 6 different speakers (3 men and 3 women) has been analyzed with the coder. The obtained onset phase information served to train the vector quantizer for the residual PCWs. In order to measure the distortion introduced by the quantizing process, we define a quadratic distance between pitch-cycle waveforms A and B of length L as

$$d(A,B)=\frac{1}{L}\sum_{i=1}^{L}[A(i)-B(i)]^2 \qquad (7)$$

Another 30 s database has been processed and the corresponding onset phase information quantized with codebooks of different sizes. The distortion results, computed as the mean value of the distance measure defined above over the second database, are shown in table 1.

16 codewds.	32 codewds.	64 codewds.
3.172	2.865	2.658

Table 1: quantization distortions of residual PCWs for several codebook sizes

The performance of the overall coding process has been compared with the classical scheme employed in low rate MBE coders such as the IMBE [2], where no original phase information is sent to the decoder. Three new PCWs are built using the following phase information:
a) original onset phases, which is used as a reference;
b) onset phases recovered after pitch-cycle coding with a 64 codevectors quantizer;
c) the phases the IMBE would use to reconstruct the onset frame.

To isolate the comparison results from time shifts, the linear phase component (2) is eliminated by rotating PCWs b) and c) to the point of maximum circular cross-correlation with a) previously to the comparison. The obtained results are shown in table 2 in terms of mean distortion over the second database.

IMBE	PCW coding
2.197	0.231

Table 2: mean phase distortion of IMBE and PCW phase coding

Informal listening tests carried out to this moment show that high pitched speakers are almost insensible to phase manipulation, while naturalness of low pitched speakers can be significantly improved by introducing original onset phase information. Our experiments also show that the proposed phase quantization scheme is transparent even for low pitched speakers if a codebook of 64 codevectors is employed.

5. CONCLUSIONS

A new phase coding algorithm has been developed which works in the pitch-cycle waveform domain. The advantages of this representation of phase information have been found to be:

1) the elimination of the linear phase component is performed via circular cross-correlation of PCWs, which is an easy and reliable operation

2) as the number of phases to be quantized varies from frame to frame, a length conversion algorithm seems convenient. Pitch pulses are easier to interpolate and decimate than phase spectra

3) the residual pitch-cycle has a noisy waveform which can be successfully vector quantized, with a relatively small amount of bits (6-7 bits)

The performance of the proposed scheme has been objectively demonstrated in the context of onset quantization in a low bit rate MBE coder. It provides a close representation of onset phases achieving a significant quality improvement.

6. REFERENCES

[1] A. Das, A. Gersho, *Enhanced multiband excitation coding of speech at 2.4 kb/s with phonetic classification and variable dimension VQ*, Proc. of EUSIPCO 94, pp 943-946.
[2] Digital Voice Systems, "Inmarsat-M Voice Codec, Version 2", *Inmarsat-M specification,* Inmarsat, London, February 1991.
[3] G. Yang, H. Leich, "High-quality harmonic coding at very low bit rates", Proc. of ICASSP 94, pp I.181-I.184.
[4] Torres-Guijarro, M.S. and Casajús-Quirós, F.J., *Improved transient representation and quantization for sinusoidal speech coders*, Proc. of EUROSPEECH 95, pp 681-684.
[5] H. Yang, S.-N. Koh and P. Sivaprakasapillai, *Enhancement of multiband excitation (MBE) by pitch-cycle waveform coding*, Electronics Letters, Sept. 1994, Vol.30, No.20, pp 1645-1646.
[6] R. J. McAulay and T. F. Quatieri, *Low-rate speech coding based on the sinusoidal model*, Advances in Speech Signal Processing, ed. Marcel Dekker, 1992.
[7] J. Marques, I. Trancoso, A. Abrantes, *Harmonic coding of speech: an experimental study*, Proc. EUROSPEECH'91, pp 235-238

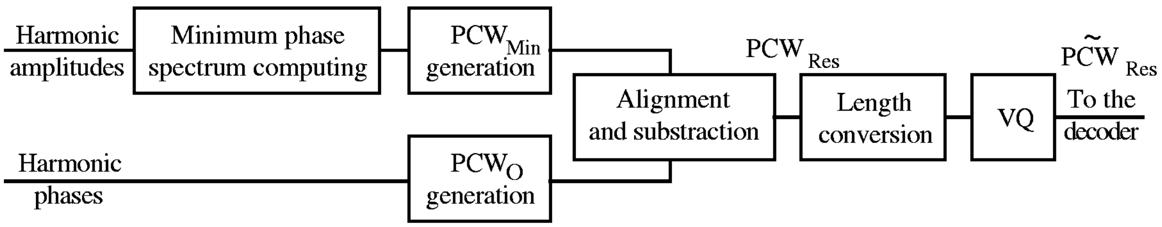
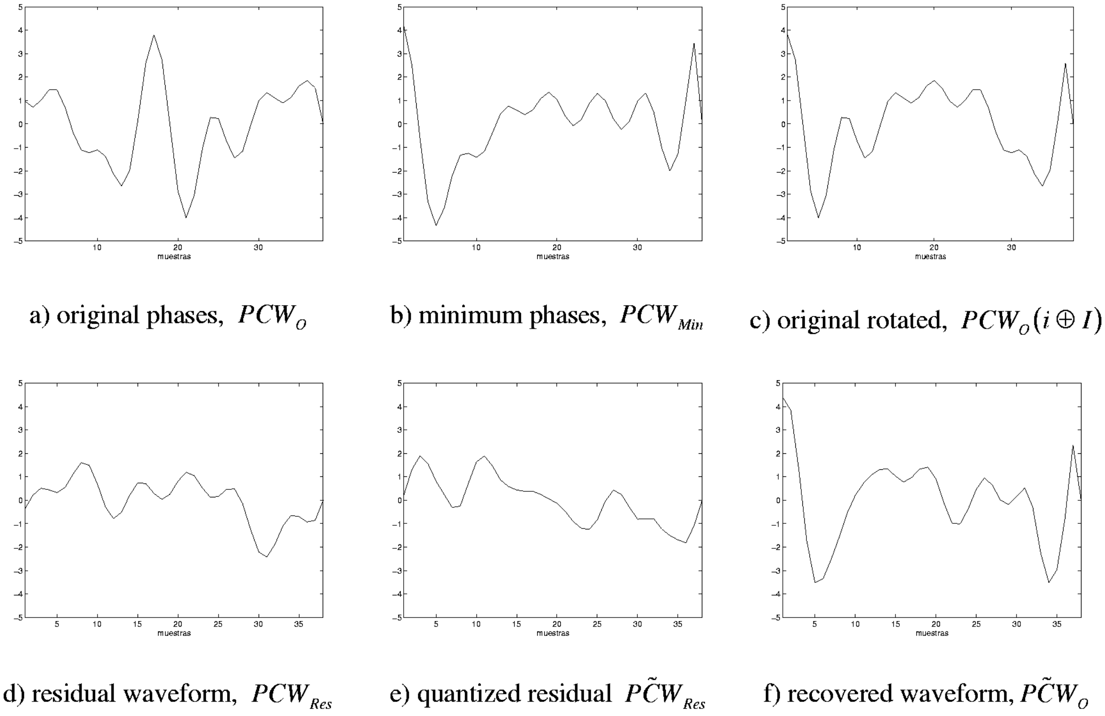


Figure 1: block diagram of the phase coder



a) original phases, PCW_O

b) minimum phases, PCW_{Min}

c) original rotated, $PCW_O(i \oplus I)$

d) residual waveform, PCW_{Res}

e) quantized residual $P\tilde{C}W_{Res}$

f) recovered waveform, $P\tilde{C}W_O$

Figure 2: pitch-cycle waveform coding example

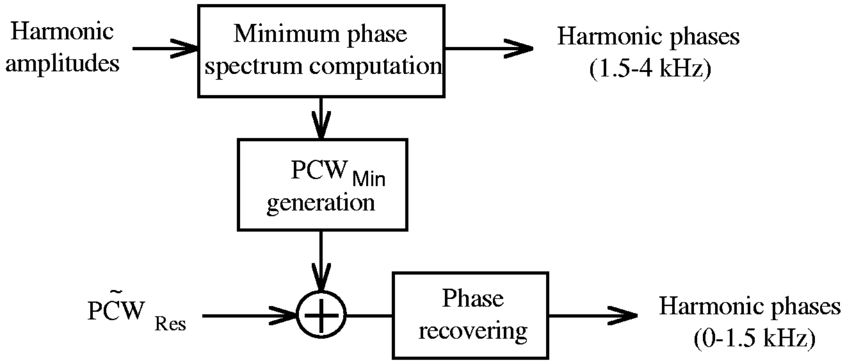

Harmonic amplitudes
Minimum phase spectrum computation
Harmonic phases (1.5-4 kHz)
PCW_{Min} generation
$P\tilde{C}W_{Res}$
Phase recovering
Harmonic phases (0-1.5 kHz)

Figure 3: block diagram of the phase decoder