# ENHANCEMENT OF IMPROVED MULTI-BAND EXCITATION (IMBE) USING A NOVEL METHOD TO ENCODE SPECTRAL AMPLITUDES

Haiyun Yang Soo-Ngee Koh Sivaprakassaipillai P

School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 2263

ABSTRACT - A novel method using pitch-cycle waveform (PCW) to encode the spectral amplitudes of the multi-band excitation (MBE) model is proposed in this paper. With the new method, both the magnitudes and phases of the spectra are transmitted, instead of only magnitudes, as in the case of the Improved MBE (IMBE). The perceptual weighting filter is also included in the encoding procedure. Through our experiments, it is found that the new method significantly improves the perceptual quality of the MBE decoded speech, especially for the case of male speech.

## 1. INTRODUCTION

The multi-band excitation (MBE) model proposed by Griffin and Lim (Griffin & Lim, 1985) can achieve high quality speech at the bit rate ranging from 4.8 to 8 kbps with low computation complexity (Griffin & Lim, 1988). The improved MBE (IMBE) has been adopted by the Inmarsat as the speech coding algorithm for satellite communications (Digital Voice Systems Inc., 1991). However, there is room for improvement of the IMBE model because the perceptual quality of the decoded speech of IMBE is still far from transparent. Firstly, the method of quantization of the spectral magnitudes of the voiced speech signals is not perceptually optimal, especially for the case of male speech which has many more harmonics than female speech. Secondly, in the IMBE coder, the original values of the phases are separated from the magnitudes and they are not transmitted and used in the synthesis. Through our experiments, it is found that the perceptual quality of the decoded speech can be significantly improved if the original phase information is used at the synthesis stage. However this is true only if the phase information is quantized accurately. A coarsely quantized phase leads to almost the same speech quality as that of the IMBE model which uses synthesized phase values.

To overcome these two weaknesses of IMBE, a new method based on the coding of pitch cycle waveforms (PCW) is proposed. It preserves and transmits the spectral amplitudes indirectly. With the estimated magnitudes and phases of the fundamental frequency and its harmonics, a PCW is generated. The PCW is then encoded by a method similar to the well known code excitation linear prediction (CELP) method. The perceptual weighting filter is also included in the closed-loop to encode the PCW. In the next section, the relation between the spectral amplitudes and PCW is given. Sections 3 describes the coding of the PCW. The quantization of corresponding parameters is presented in Section 4. Some test results and discussion are presented in Section 5. The final section concludes this paper.

## 2. RELATION BETWEEN SPECTRAL AMPLITUDES AND PCW

In the IMBE coder, the short-term spectrum is split into several bands, each of which encompasses the fundamental frequency or one of its harmonics. The voiced/unvoiced (v/uv) decision is made for each group of three bands. The magnitude and phase of each voiced band can be estimated using the method described in the Inmarsat IMBE algorithm (Digital Voice Systems Inc. 1991). The magnitude of each unvoiced band is also obtained in the way presented in the Inmarsat IMBE, while

its corresponding phase is generated randomly. Let $M_i$ and $\theta_i$ denote the magnitude and phase of the $i$-th band. The PCW of the current frame is then generated by

$$w(k) = \sum_{i=1}^{L} M_i \cos\left(2\pi \frac{ik}{p_{in}} + \theta_i\right) \qquad k = 0, 1, ..., p_{in} - 1 \tag{1}$$

where $L$ is the number of harmonics and $p_{in}$ is the integer part of the refined pitch period $p$.

The PCW $w(k)$ can be encoded using the method to be described in the next section. With the reconstructed PCW $\hat{w}(k)$, the magnitude and phase of all bands can be re-generated by :

$$\hat{M}_i = \frac{2}{p_{in}} \left\{ \left[ \sum_{k=0}^{p_{in}-1} \hat{w}(k) \cos\left(2\pi \frac{ik}{p_{in}}\right) \right]^2 + \left[ \sum_{k=0}^{p_{in}-1} \hat{w}(k) \sin\left(2\pi \frac{ik}{p_{in}}\right) \right]^2 \right\}^{\frac{1}{2}} \tag{2}$$

$$\hat{\theta}_i = -\arctan \frac{\sum_{k=0}^{p_{in}-1} \hat{w}(k) \sin\left(2\pi \frac{ik}{p_{in}}\right)}{\sum_{k=0}^{p_{in}-1} \hat{w}(k) \cos\left(2\pi \frac{ik}{p_{in}}\right)} \tag{3}$$

where $i$=1, 2, ..., $L$. It is easy to prove that $\hat{M}_i$ and $\hat{\theta}_i$ are equal to $M_i$ and $\theta_i$ respectively if $\hat{w}(k) = w(k)$. Therefore, the coding of PCW instead of $M_i$ and $\theta_i$ ensures that the synthesized spectra are very close to the original spectra. Also, the coding of PCW implies that the magnitudes of the harmonics are not quantized independently of one another and hence the use of a perceptual weighting filter can be exploited to improve the subjective quality of the decoded speech. Finally, the phase information is not lost when the PCWs are quantized with reasonable accuracy. All these advantages make the new method superior to the existing approach which encodes the spectral magnitudes directly.

## 3. PCW CODING

The PCW should be efficiently quantized to achieve high quality decoded speech. Since a speech signal is close to a quasi-stationary process, the PCWs of neighbouring frames of voiced speech are very similar, in general. We can therefore categorize frames into two classes. The first class, designated as "related frames", consists of frames whose PCWs are considered to be similar to those of their previous frames. The remaining frames belong to the second class and are designated as "unrelated frames". In our simulation, the decision whether to consider two consecutive frames as similar, is based on the rate of change of pitch period. If the change of the pitch period between the current frame and the previous frame is less than 10%, then the current frame is considered to belong to the class of related frames. Otherwise, it is considered as an unrelated frame. This measure works well and no extra bits are needed to transmit the related/unrelated information.

### 3.1 Encoding of PCWs of Related Frames

Figure 1 shows the block diagram of the system used to encode the PCWs of the related frames. Only one codebook is included and no quantizers are shown for simplicity. The coding of the PCWs of the unrelated frames is simpler than that described in Figure 1 because the similarity between the neighbouring frames is not considered. Therefore, only the encoding of the PCWs of the related frames is described in detail in this paper.

In Figure 1, the PCW of the current frame $w_0(k)$ is first generated using Eq.(1) and repeated 3 times. Then, the LPC spectrum $A_0(z)$ of the current PCW is computed. The synthesized PCW of the previous frame $\hat{w}_{-1}(k)$ is "unweighted" with its LPC spectrum $\hat{A}_{-1}(z)$, where ^ denotes the synthesis parameter. The "length conversion" block adjusts the length of the PCW while preserving its shape. Therefore $\hat{w}_{-1}^{[1]}(k)$ has the same length as $w_0(k)$. The "aligned" block makes $\hat{w}_{-1}^{[2]}(k)$ match $w_0(k)$ as much as possible by circulating $\hat{w}_{-1}^{[2]}(k)$, which is the output of the perceptual weighted synthesis filter $1/A_0(z/\gamma)$. The circulating index $i$ is transmitted for synthesizing $\hat{w}_0(k)$ in the receiver. After alignment,

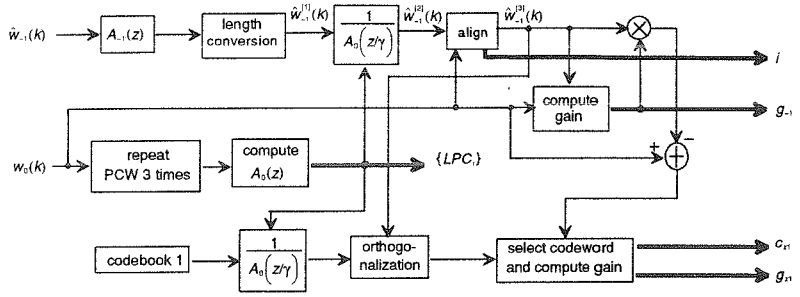


Figure 1 Coding of PCW based on one codebook
(Subscripts 0 and -1 denote the current and previous frames' PCWs respectively.
w(k) is the PCW. A(z) is LPC spectrum.)

the optimal gain $g_{-1}$ is computed. The error signal $e_0(k)$ between two neighbouring frames, given by the equation

$$e_0(k) = w_0(k) - g_{-1} \cdot \hat{w}_{-1}^{[3]}(k) \tag{3}$$

is then coded in a way similar to CELP. The "orthogonalization" block makes the weighted codevectors orthogonal to the previous estimation in order to maintain optimization in the sequential selection of the codebook indices. The two codebooks are used to code the PCW. One consists of a bandlimited single pulse. The other is a stochastic codebook. The four parameters to be transmitted are the index and gain from the first codebook ($g_{x1}$ and $c_{x1}$), and the index and gain ($g_{x2}$ and $c_{x2}$) from the second codebook.

3.2 Synthesis of PCWs of Related Frames

The synthesis of the PCWs for the related frames is presented in Figure 2. It is obvious that the synthesized PCW of the related frame comes from three parts, namely the previous frame and the two codebooks.

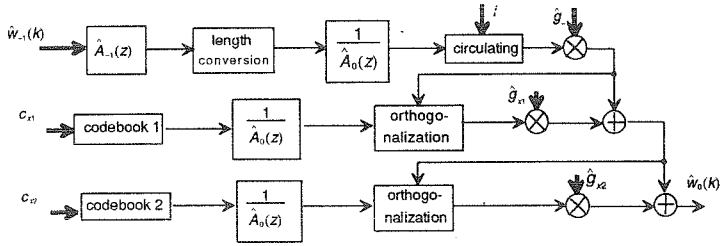


Figure 2 Synthesis of PCW

## 4. QUANTIZATION OF PARAMETERS OF PCW

From Figure 1, the parameters of a PCW for the related frames are the circulating index ($i$), the ten LPC coefficients, the two indices ($c_{x1}$ and $c_{x2}$) and the three gains ($g_{-1}$, $g_{x1}$ and $g_{x2}$). For the unrelated frames, the parameters are the same as those of the related frames but without the circulating index and the gain $g_{-1}$. The bit allocation for PCW coding is given in Table 1. The other parameters (pitch period and v/uv decisions) included in the IMBE model are quantized in the same way with the same bits as in the case of the Inmarsat IMBE.

Table 1 Bit allocation for PCW coding

| parameters | $i$ | LPC | $c_{x1}$ | $c_{x2}$ | $s_0$ | $p_{-1}$ | $p_{x1}$ | Total |
|---|---|---|---|---|---|---|---|---|
| related frames | 6 | 30 | 7 | 8 | 5 | 7 | | 63 |
| unrelated frames | 0 | 36 | | | 6 | NA | 6 | |

### 4.1 Quantization of LPC

The LPC coefficients are first converted into line spectral pairs (LSP). As shown in Table 1, the LSP parameters are quantized with different numbers of bits for different frames. In the case of unrelated frames, the LSP parameters are directly quantized with 36 bits using a group of scalar non-uniform quantizers because 36 bits are sufficient to achieve transparent quality for LPC parameters based on scalar quantization (Paliwal & Atal, 1993). In the other case, there are only 30 bits allocated for the quantization of LSP parameters. Therefore, a more efficient hybrid vector-scalar quantizer (Moriya & Honda, 1986) is designed and used.

### 4.2 Quantization of Gains

To efficiently quantize the three gains and to improve the robustness of the proposed system, the three gains are not directly quantized. Let $\hat{s}_{-1}$, $s_{x1}$ and $s_{x2}$ be the energy of the previous synthesized PCW $\hat{w}_{-1}(k)$ and the two excitations respectively. Then the three energy parameters will be given by

$$\hat{s}_{-1} = \sum_{k=0}^{P_{ip}-1} \left[\hat{w}_{-1}(k)\right]^2 , \; s_{x1} = \sum_{k=0}^{P_{ip}-1} \left[x1(k)\right]^2 \text{ and } s_{x2} = \sum_{k=0}^{P_{ip}-1} \left[x2(k)\right]^2 \tag{4}$$

where $x1(k)$ and $x2(k)$ represent the excitations of the two codebooks respectively. Because of the "orthogonalization" block in Figure 1, the energy of the synthesized current PCW, denoted by $\hat{s}_0$, can be expressed by

$$\hat{s}_0 = \sum_{k=0}^{P_{ip}-1} \left[\hat{w}_0(k)\right]^2 = g_{-1}^2 \cdot \hat{s}_{-1} + g_{x1}^2 \cdot s_{x1} + g_{x2}^2 \cdot s_{x2} \tag{5}$$

We next define two terms $p_{-1}$ and $p_{x1}$ as follows :

$$p_{-1} = \frac{g_{-1}^2 \cdot \hat{s}_{-1}}{\hat{s}_0} \quad \text{and} \quad p_{x1} = \frac{g_{x1}^2 \cdot s_{x1}}{\hat{s}_0} \tag{6}$$

Then, using Equations (5) and (6), the three gains can be expressed by another three parameters, $\hat{s}_0$, $p_{-1}$ and $p_{x1}$, as follows :

$$g_{-1} = \sqrt{\frac{p_{-1} \cdot \hat{s}_0}{\hat{s}_{-1}}} , \; g_{x1} = \sqrt{\frac{p_{x1} \cdot \hat{s}_0}{s_{x1}}} \quad \text{and} \quad g_{x2} = \sqrt{\frac{(1 - p_{-1} - p_{x1}) \cdot \hat{s}_0}{s_{x2}}} \tag{7}$$

Therefore, instead of quantizing the three gains directly, the new parameters $\hat{s}_0$, $p_{-1}$ and $p_{x1}$ are quantized. The energy of the current synthesized PCW $\hat{s}_0$ is quantized by a 6-bit logarithmic non-uniform quantizer. The pair ($p_{-1}$ , $p_{x1}$) is vector quantized with an 7-bit codebook.

The gain parameters of the unrelated frames can be considered as a special case of the related-frames when the first gain $g_{-1}$ is set to zero. In this case, Eqs.(7) are rewritten as

$$g_{x1} = \sqrt{\frac{p_{x1} \cdot \hat{s}_0}{s_{x1}}} \quad \text{and} \quad g_{x2} = \sqrt{\frac{(1-p_{x1}) \cdot \hat{s}_0}{s_{x2}}} \tag{8}$$

Thus the two gains, $g_{x1}$ and $g_{x2}$, are replaced by $\hat{s}_0$ and $p_{x1}$, which are scalar quantized as shown in Table 1.

The technique used to quantize the three gains is not only efficient but also improves the robustness of the system in the presence of transmission errors. This is because once the energy of the current PCW $\hat{s}_0$ is received correctly, the decoded speech cannot be far away from the original speech even if the received values of the pair ($p_{-1}$, $p_{x1}$) are wrong (Gerson & Jasiuk, 1990).

4.3 Quantization of Circulating Index and Two Excitation Indices

The circulating index $i$ and the excitation index $c_{x1}$ are coded with 6 and 7 bits respectively. The excitation index $c_{x2}$ of the stochastic codebook is coded with 8 bits.

4.4 Training

The speech signals in the TIMIT corpus CD-ROM are used to train the scalar and vector quantizers. The TIMIT speech signal is low-pass filtered by a 240-order FIR filter and down-sampled to 8 kHz. A total of 434 speech utterances, spoken by 105 females and 112 males, make up the training data set. The total duration of the data set is about 20 minutes after removing the silence segments. Both the scalar and vector quantizers used in this paper are trained using the LBG method (Linde, Buzo & Gray, 1980) and the training data set.

## 5. EXPERIMENTAL RESULTS

The new method proposed above was compared to the 4.15 kbps IMBE, which uses 63-72 bits to quantize the spectral magnitudes. In our experiment, eight speech utterances (four female and four male, total of about 20 seconds), which were not included in the training set, were used to assess the performance of the new coder. The average signal to noise ratio (SNR) of the PCW for 8 test utterances was 10.20 dB, while the average SNR of the FS1016 4.8 CELP for the same utterances was 7.30 dB. Although the SNR of PCWs and the SNR of the decoded speech waveform are two different measures, the 10.20 dB average SNR of PCW implies that the encoding of PCWs presented above is quite efficient. The quantization SNR of the 4.15 kbps IMBE is also computed. The average value for the test utterance is 13.53 dB. Although this value is about 3.3 dB higher than that of the PCW average SNR, the perceptual quality of the decoded speech using PCW is significantly better than that synthesized by IMBE. This implies that the direct encoding of the spectral magnitudes in IMBE is not perceptually optimal. In our informal listening tests, it was also found that the decoded speech of the Inmarsat IMBE has some hoarseness for the decoded speech, which was due to the omission of the original phase information. The decoded speech of IMBE with PCW sounded significantly clearer. Therefore the phase information is still important for high quality decoded speech.

## 6. CONCLUSION

In this paper, coding of the PCW is proposed for conveying the spectral amplitudes obtained by the multi-band analysis. Efficient coding of the PCW is achieved by using a coding method similar to CELP with a bandlimited single pulse codebook and a stochastic codebook. The decoded speech quality is improved significantly compared to that produced by the 4.15 kbps IMBE coder at the same bit rate.

REFERENCES

Atal B. S and Schroeder M. R. (1984), Stochastic coding of speech at very low bit rates, Proc Int. Conf. Commun., pp.1610-1613.

Digital Voice Systems Inc. (1991), Inmarsat-M voice codec, ver.3.0, August

Gerson Ira A. and Jasiuk Mark A. (1990) Vector sum excited linear prediction (VSELP) speech coding at 8 kbps, IEEE Int. Conf. Acoust., Speech, Signal Processing, pp.461-464, Albuquerque, NM, USA.

Griffin D. W. and Lim J. S. (1985), A new model-based speech analysis/synthesis, IEEE Int. Conf. Acoust., Speech, Signal Processing, pp.513-516, Tampa, Florida, USA.

Griffin D. W. and Lim J. S. (1988), Multiband excitation vocoder, IEEE Trans. Acoust., Speech, Signal Process. ASSP-36, pp.1223-1235.

Linde Y., Buzo A. and Gray R. M. (1980) An algorithm for vector quantizer design, IEEE Trans. Commun., vol. COM-28, pp.45-95.

Moriya T and Honda M. (1986) Speech coder using phase equalizationand vector quantization, IEEE Int. Conf. Acoust., Speech, Signal Processing, Tokyo, Japan, pp.1701-1704.

Paliwal K.K. and Atal B. S. (1993), Efficient vector quantization of LPC parameters, IEEE Trans. Speech and Audio Process. vol.1 no.1 January.