A Closed-loop Multimode Variable Bit Rate Characteristic Waveform Interpolation Coder

Jing Wang, Jingming Kuang, and Shenghui Zhao

Research Center of Digital Communication Technology, Department of Electronic Engineering, Beijing Institute of Technology, Beijing

wangjing@bit.edu.cn

Abstract. A variable bit rate characteristic waveform interpolation (VBR-CWI) speech codec with about 1.86kbps average bit rate which combines closed-loop multimode techniques is presented in this paper. Each kind of characteristic waveform (CW) surface is regarded as only rapidly evolving waveforms (REWs), only slowly evolving waveforms (SEWs) or mixed REWs plus SEWs in different cases of CWs evolving performance. A cost criterion based on weighted signal-to-noise (WSNR) value in the spectral domain is used to make the mode selection. Experiments show that the proposed closed-loop multimode VBR-CWI coder has reduced the average bit rate markedly and improved the synthesis speech quality to some extent compared to the original fixed bit rate coder. Further research can be done in order to have a more accurate perceptual objective quality measurement instead of WSNR and there is also need to pay attention to computational complexity of closed-loop method in real-time applications.

Keywords: Closed-loop multimode; Variable bit rate; Cost criterion; Waveform interpolation.

1 Introduction

As we can see, there has been an increasing interest in developing Variable Bit Rate (VBR) coder which is able to reduce average bit rate by exploiting the nature of speech [1]. The usual Fixed Bit Rate (FBR) coders continuously transmit at the maximum bit rate needed to assure a given speech quality for the worst-case speech frames where the entropy is very high. Speech coders can be designed to give each frame only the number of bits needed while maintaining a desired average number of bits per frame. Generally, there are two techniques to design VBR coder including open-loop phonetic classification method and closed-loop multimode method [2]. The first method comparably has a lower computational complexity and needs a reliable classification approach. The second method is highly complex but has an advantage in that the modes that constitute the final coder are selected in a way related to how good they code the speech signal. It has two important problems including construction of the different modes and proper objective decision of when to use which mode.
In recent years low bit rate speech coders with high quality for transmission at rates below 4kbps have received much attention. Waveform Interpolation (WI) speech coder proposed by W. B. Kleijin has been shown to provide high quality speech at low bit rates [3] [4]. In Characteristic Waveform Interpolation (CWI) speech coder, Pitch Cycle Waveforms (PCWs) extracted from the linear prediction residual signal characterize the evolution of pitch cycles together with a phase track. The key to quantization of Characteristic Waveforms (CWs) at low bit rates is decomposition of that surface into Slowly Evolving Waveforms (SEWs) and Rapidly Evolving Waveforms (REWs), which represent the voiced and unvoiced speech components respectively. This decomposition is motivated by human perception and results in high coding efficiency. However, FBR-CWI speech coder doesn’t consider the different performance of different types of CWs representation. One promising way to bring WI coder to higher quality at lower bit rate is to adopt the VBR scheme.

In this paper, we try to design a VBR-CWI low bit rate speech coder based on closed-loop multimode techniques. Different types of characteristic waveform surface are regarded as only REWs, only SEWs or REWs plus SEWs for different performance of CWs. All the three modes run in parallel and the quantized and reconstructed CWs of all modes are compared using cost criterion based on weighted signal-to-noise ratio (WSNR) in spectral domain to decide which mode will be used finally.

2 Multimode CWI Coder

2.1 CWs Representation of FBR-CWI Coder

In our FBR-CWI speech coder, the input narrowband speech is segmented into 20ms frames. For each speech frame, standard linear predictive analysis is made to extract 10th order predictive coefficients and to get residual signal. The LPC parameters are converted to LSF parameters which are quantized with 20 bits Predictive Split Vector Quantization (PSVQ) technique. Characteristic waveforms are extracted at fixed time intervals (extraction rate is 400Hz in this paper) from the residual based on the pitch information and are represented with Discrete Time Fourier Series (DTFS)

\[ s(n,\phi) = \sum_{k=0}^{P(n)} \left[ A_k(n) \cos(k\phi) + B_k(n) \sin(k\phi) \right], \]

where \( A_k \) and \( B_k \) are the DTFS coefficients and \( P(n) \) is the pitch (one CW length). Each extracted waveform is aligned with a cyclic-shift to maximize its correlation with the preceding aligned waveform. After the CWs are extracted and aligned, their powers are then normalized. Thus the CWs are modeled as the sums of harmonic SEWs and noisy REWs. Typically the SEWs surface is formed by filtering the CWs evolving surface along the time axis using a 17-taps linear-phase and non-causal lowpass FIR filter at a cutoff frequency 25Hz (or equivalently) [5]. Lowpass
filtering the CWs in the time-domain is equivalent to lowpass filtering their DTFS coefficients using the following formula

\[
\tilde{A}_k(n) = \sum_{i=1}^{Lsf} A_i (n - iL_p) H_p(i) \\
\tilde{B}_k(n) = \sum_{i=1}^{Lsf} B_i (n - iL_p) H_p(i) \\
\quad \text{for } k = 1, 2, ..., \left\lfloor \frac{P(n)}{2} \right\rfloor, (2)
\]

where \(L_{sf}\) is the extraction interval, \(H_p\) is the impulse response of the lowpass filter.

The REWs can be found by subtracting the SEWs from the CWs. The sequences of SEWs and REWs are downsampled to 100Hz and 200Hz update rate respectively and their spectral parameters are quantized using different resolution. REW amplitude spectrum is described with Variable Dimension Vector Quantization (VDVQ) [6] technique with random phase. SEW spectrum is described with VDVQ by fixed phase (obtained from male voice with a very low fundamental frequency) and is split into three non-overlapping subbands: 0-1kHz, 1k-2kHz and 2k-4kHz, each subband is vector quantized separately.

### 2.2 Multimode Representations of CWs

As to the conventional CWI coder, it is very important to maintain a proper balance between the SEW and the REW energy in the reconstructed speech. An imbalance of the SEW-to-REW power causes the output to sound buzzy and noisy [5]. One obvious drawback of conventional performance of CWI is the background buzz artifact that mainly occurs in noise-like segments caused by excessive SEW energy by the decomposition of nonideal filtering CWs. We have found that different types of CWs can use different SEW-REW representations which can remove most buzz and noise artifacts from the synthesized speech and can enhance the perceptual quality of the CWI coder.

If the extracted characteristic waveforms evolve very slowly (e.g. voiced segment), they can be regarded as only SEWs. In this case of original FBR coder, the decomposed SEWs will have much more energy than REWs. In fact, there is no need for CWs to be decomposed and CWs are directly downsampled and quantized using the method of processing SEWs in FBR coder. On the contrary, the CWs will be regarded as only REWs if they evolve very rapidly and look like noise signal (e.g. unvoiced segment). For the speech segments that are neither stationary nor noisy and the decomposed SEWs and REWs have comparative energy (e.g. onsets or transitions), they will be represented with both SEWs and REWs like the original CWs representation of FBR-CWI coder. Therefore the number of bits required for each type of input signal varies widely by the different CWs representation (i.e. multimode representations) and coding bits can be saved when encoding extremely slow and extremely rapid evolving CWs. Also we have found that if the multimode representation of CWs can be performed very well, the synthesized speech quality will be improved with fewer buzzy or noisy artifacts.
In our multimode variable rate CWI coder, non-speech segments are firstly detected before Linear Prediction (LP) analysis and are represented separately using Bark-band perceptual noise model [7]. Therefore there are four kinds of coding modes including 3 kinds of CWs representation modes as is shown in Table 1. Note that here the mode names of REWs and SEWs are different from that of FBR-CWI coder and stand for the different evolving performance and coding forms of current extracted CWs. For example, as to mode 2, when the extracted CWs evolve very slowly, the SEWs are just the original CWs and can be represented with 2 waveforms per frame, and there are no REWs component.

<table>
<thead>
<tr>
<th>Mode flag</th>
<th>Mode name</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Non-speech</td>
<td>Noise modeling</td>
</tr>
<tr>
<td>1</td>
<td>REWs</td>
<td>4 REWs per frame</td>
</tr>
<tr>
<td>2</td>
<td>SEWs</td>
<td>2 SEWs per frame</td>
</tr>
<tr>
<td>3</td>
<td>Mixed CWs</td>
<td>4REWs+2SEWs per frame</td>
</tr>
</tbody>
</table>

In this paper, we try to use closed-loop multimode techniques to design variable bit rate CWI coder. In closed-loop multimode CWs representation, a trial multimode representation and reconstruction of the current extracted CWs is performed with each mode. Then a proper objective measure of performance is computed, comparing the coded CWs of each mode with the original; the best mode is selected, and its data is finally transmitted.

3 Closed-loop Multimode Variable Rate Design

3.1 Closed-loop Multimode Scheme

In our closed-loop multimode CWI coder, mode selection is made by testing the overall coding performance of each mode and selecting the one yielding the best result. Performance is generally assessed by a perceptual objective measure which, for each mode, compares the original unquantized CWs with the reconstructed quantized CWs produced by that mode. The mode producing the best perceptual quality is selected. This paper uses WSNR-based cost criterion to be the perceptual objective measurement. By minimizing the cost function, the proper mode of CWs representation is decided by closed-loop scheme.

In order to design an efficient variable bit rate CWI coder, a simple Voice Activity Detector (VAD) module before LP analysis is used to have a speech/non-speech classification. Non-speech frame is represented by Bark-band perceptual noise model [7]. The frame spectrum is constructed by piecewise constant magnitude across each Bark band with uniform random phases. For one analysis frame, 16 Bark-band spectrum estimates form into 16-dimensional vector using 10 bits Split Vector Quantization (SVQ). Also this noise model is suitable to code noise-like unvoiced
segment which has very low energy but is perceptually important. For the mobile environment, the design of a VAD is complicated by the high level of acoustic noise coming to the telephone. This paper mainly considers noise free environment, so the two features ‘frame energy and short-term zero-crossing ratio’ are good enough to achieve VAD function. If the input frame is active speech, the closed-loop mode selection is processed by representing and reconstructing CWs as only REWs, only SEWs or REWs plus SEWs. The 3 modes of CWs representation and quantization run all in parallel. Afterwards, the coder modes are evaluated, and only parameters of the best modes are kept for the final transmission. In order to avoid the abrupt mode jumping between mode1 and mode2, an additive process is needed to set the beginning and the end of mode 2 to be mode3.

The closed-loop multimode VBR-CWI coder consists of four kinds of coding strategy (mode0-3) adapted for different modes. Fig. 1 presents an overview of the proposed multimode scheme of the VBR-CWI encoder with which different CWs representations are decided by closed-loop method.

![Fig. 1. The simple scheme of multimode VBR-CWI encoder.](image)

The bit allocation and coding rate of each mode are shown in Table 2. The coding rate of FBR-CWI coder is 3.75kpbs and corresponds with that of mode3 excluding mode information. Mode0 uses 10 bits for the noise spectrum information. Through multimode processing, the averaging coding rate of VBR-CWI coder will be much lower than that of FBR-CWI coder.
Table 2. Bit allocation for each mode.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mode0</th>
<th>Mode1</th>
<th>Mode2</th>
<th>Mode3</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSFs</td>
<td>/</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Pitch</td>
<td>/</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Gain</td>
<td>/</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>SEW</td>
<td>/</td>
<td>0</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>REW</td>
<td>/</td>
<td>12</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Mode</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bits/frame</td>
<td>12</td>
<td>49</td>
<td>65</td>
<td>77</td>
</tr>
<tr>
<td>kb/s</td>
<td>0.6</td>
<td>2.45</td>
<td>3.25</td>
<td>3.85</td>
</tr>
</tbody>
</table>

3.2 Weighted Objective Measure

The quality measures of different modes are calculated between the original and quantized variable dimension vectors using perceptual weighted SNR in spectral domain [8]. The weighted SNR is calculated by averaging the WSNR values for individual vectors obtained through

\[
WSNR = 10 \log_{10} \left( \frac{x^T W x}{(x - \hat{x})^T W (x - \hat{x})} \right) [dB],
\]

where \( x \) and \( \hat{x} \) denote the original and the quantized spectral vector of one CW, respectively. The elements \( w_k \) of the diagonal weighting matrix \( W \) are computed by evaluating Equation (4) at multiples of pitch frequency, i.e. at \( z = e^{j2\pi k/P} \), where \( P \) is the corresponding pitch period in samples.

\[
w(z) = \frac{1}{K} \left( \frac{\sqrt{G(z/\gamma)}}{A(z)} \right), \; 0 < \gamma_z \leq 1,
\]

where \( K \) is the number of harmonics, \( G \) is the power of the corresponding residual waveform, and \( A(z) \) denotes the 10th order LP polynomial. The weighting parameters are set to \( \gamma_1 = 0.9 \), and \( \gamma_2 = 0.6 \).

By experiments, we have found that in the case of slowly evolving waveforms (mostly stationary voiced segments) the average WSNR of mode2 (SEW-representation) is mostly much higher than that of the other modes, but in the case of rapidly evolving waveforms (mostly noise-like speech) it is similar among the three modes although most of the mode1 (REW-representation) conditions perform a little better. For the case that mode3 performs a little better than mode2, the CWs also can be represented by only SEWs and in this case it’s better to select mode2 than mode3 in order to get lower coding rate. The human ear can tolerate a lower WSNR in the regions of rapidly evolving waveforms. Hence sometimes the proposed WSNR may fail to decide the proper mode in a perceptually meaningful manner, mainly because of its reliance on an objective measure which is not enough to represent the real perceived quality. For the above reason, a cost criterion based on average WSNR is
employed to differentiate the CWs modes, only keeping the modes with best performance.

3.3 Cost Criterion

The fundamental property of the cost criterion is to reward high speech quality at a low bit rate and to penalize high rate to favor efficient coding [9]. Mode selection is performed to detect the minimum cost function. The method is similar to rate-distortion optimization and our cost function is defined as

$$J_i = D_i + \lambda R_i = \lambda R_i - \text{WSNR}_i,$$  \hspace{1cm} (5)

where $J$ stands for coding cost of one mode, $R$ is the coding rate of that mode, WSNR stands for the average weighted spectral SNR of all characteristic waveforms. The penalty parameter $\lambda > 0$ has been chosen to a value that minimizes the risk of bad selection. The number $i$ equals to 1, 2 or 3 which corresponds to the three modes 1, 2 or 3 defined as Table 1. The coding rate of each mode is shown in Table 2 ($R_1=2.45\text{kbps}$, $R_2=3.25\text{kbps}$, $R_3=3.85\text{kbps}$).

We mainly focus on the two cases that easily generate bad selection of the coding modes and will affect the coding quality and efficiency badly. The first one is the case that correct mode decision must be made between mode 1 and mode 2 when the WSNR of mode 2 is higher than that of mode 1. The second one is the case that correct mode decision must be made between mode 2 and mode 3 when the WSNR of mode 3 is higher than that of mode 2. The penalty parameter is estimated as follows.

- **<1>** WSNR of mode 2 is higher than that of mode 1.
  The correct decision is mode 1, i.e., $J_1 < J_2 \Rightarrow$

  $$\lambda > \frac{\text{WSNR}_2 - \text{WSNR}_1}{R_2 - R_1} \geq \max \left( \frac{\text{WSNR}_2 - \text{WSNR}_1}{R_2 - R_1} \right) = \frac{\text{mode1}_\text{Thr}_{21}^{\max} R_2 - R_1}{R_2 - R_1}. \hspace{1cm} (6)$$

  The correct decision is mode 2, i.e., $J_1 < J_2 \Rightarrow$

  $$\lambda < \frac{\text{WSNR}_2 - \text{WSNR}_1}{R_2 - R_1} \leq \min \left( \frac{\text{WSNR}_2 - \text{WSNR}_1}{R_2 - R_1} \right) = \frac{\text{mode2}_\text{Thr}_{21}^{\max} R_2 - R_1}{R_2 - R_1}. \hspace{1cm} (7)$$

- **<2>** WSNR of mode 3 is higher than that of mode 2.
  The correct decision is mode 2, i.e., $J_2 < J_3 \Rightarrow$

  $$\lambda > \frac{\text{WSNR}_3 - \text{WSNR}_2}{R_3 - R_2} \geq \max \left( \frac{\text{WSNR}_3 - \text{WSNR}_2}{R_3 - R_2} \right) = \frac{\text{mode2}_\text{Thr}_{32}^{\max} R_3 - R_2}{R_3 - R_2}. \hspace{1cm} (8)$$

  The correct decision is mode 3, i.e., $J_2 < J_3 \Rightarrow$

  $$\lambda < \frac{\text{WSNR}_3 - \text{WSNR}_2}{R_3 - R_2} \leq \min \left( \frac{\text{WSNR}_3 - \text{WSNR}_2}{R_3 - R_2} \right) = \frac{\text{mode3}_\text{Thr}_{32}^{\max} R_3 - R_2}{R_3 - R_2}. \hspace{1cm} (9)$$

From above equations, the estimated penalty parameter is limited in the range,
where $Thr_{ij}$ is the tolerated threshold that stands for maximum or minimum WSNR difference between mode $i$ and mode $j$ in the case of either mode (potentially WSNR of mode $i$ is higher than that of mode $j$). The WSNR values of different CWs representations (presuming that WSNR is zero when mode0) are shown in Fig.1 (b). By statistical observation of each case, $Thr_{ij}$ can be set as follows,

$$\begin{align*}
    \text{mode1}_{_\text{Thr}_{21}^{_{\text{max}}} < \lambda < \text{mode2}_{_\text{Thr}_{21}^{_{\text{max}}}}} & \quad R_2 - R_1 \\
    \text{mode2}_{_\text{Thr}_{32}^{_{\text{max}}} < \lambda < \text{mode3}_{_\text{Thr}_{32}^{_{\text{max}}}}} & \quad R_3 - R_2
\end{align*}$$

By statistical observation of each case, $Thr_{ij}$ can be set as follows,

$$\begin{align*}
    \text{mode1}_{_\text{Thr}_{21}^{_{\text{max}}} \doteq 1\ [dB], \text{mode2}_{_\text{Thr}_{32}^{_{\text{max}}} \doteq 2.5\ [dB]};} \\
    \text{mode2}_{_\text{Thr}_{21}^{_{\text{max}}} \doteq 0.5\ [dB], \text{mode3}_{_\text{Thr}_{32}^{_{\text{max}}} \doteq 1.5\ [dB]}.}
\end{align*}$$

So the tolerated range of penalty parameter is about $1.25<\lambda<2.5$. In this range, many experiments have been done to select the optimal $\lambda$ in order to have well-done decision and to get high perceptual quality. This paper sets $\lambda$ to 1.85 as a tradeoff. The example of mode selection result is shown in Fig. 2 (a).

![Normalized Amplitude](a)

![WSNR/DB](b)

**Fig. 2.** Multimode selection (Mode0: non-speech; Mode1: REWs; Mode2: SEWs; Mode3: mixed CWs). (a) Mode selection result; (b) WSNR values of different CWs representations.
4 Results and Discussion

4.1 Average Coding Rate

We used 16 clean speech files for the objective and subjective testing from NTT-AT standard Chinese speech database (4 males and 4 females pronouncing). Each speech segment has a 8s duration and converted to 8 kHz sampling rate. By statistical experiments, the percentages of different modes for the whole test data were 51.75% for mode0, 8.77% for mode1, 30.33% for mode2 and 9.16% for mode3. In this condition, the average bit rate of the closed-loop multimode variable bit rate CWI coder is 1.86kbps and for the active speech segments the average bit rate is about 3.22kbps. With a cost criterion as above, it is possible to control the average rate of the coder by changing the factor $\lambda$ in Equation (5). If $\lambda$ is increased, the cost of a high rate is increased, and the average rate will decrease, and vice versa. Meanwhile, the coder shows different perceptual performance while the bit rate is changed with the cost factor.

4.2 Objective Quality Assessment

As to objective quality assessment of the whole speech, many experiments have shown that the method Perceptual Evaluation of Speech Quality (PESQ) [10] have a high correlation with many different subjective experiments and gives an objective listening quality mean opinion score (MOS). In order to use PESQ to the asynchronous WI coder, the input reference signal of PESQ is set to be the unquantized result through WI scheme and the distorted signal is set to be the quantized one. PESQ MOS results from the closed-loop multimode VBR-CWI coder are compared to the fixed bit rate one in Table 3. Note that here the MOS values don’t stand for the real ones of subjective listening tests. It is shown that the objective speech quality is improved by the multimode selection and the cost criterion indeed reacts on the rate-distortion tradeoff.

<table>
<thead>
<tr>
<th>Test Speech</th>
<th>FBR-CWI</th>
<th>VBR-CWI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>3.2689</td>
<td>3.3510</td>
</tr>
<tr>
<td>Male</td>
<td>3.1580</td>
<td>3.3126</td>
</tr>
<tr>
<td>Whole</td>
<td>3.2134</td>
<td>3.3318</td>
</tr>
</tbody>
</table>

4.3 Informal Listening Tests

We carried out subjective A-B comparison tests using 10 listeners (5 males and 5 females) to assess the performance of the proposed 1.86kbps closed-loop multimode VBR-CWI coder, original 3.75kbps FBR-CWI coder and standard 4.8kbps FS1016
CELP coder. A and B stand for the codec pairs to be compared. The presentation of A and B is random and the listener will make a quality partial result i.e. prefer A or B. The statistic results are as shown in Table 4 and Table 5.

Table 4. Comparison of FBR-CWI and FS1016CELP

<table>
<thead>
<tr>
<th>Test Speech</th>
<th>Prefer FS1016</th>
<th>Prefer FBR-CWI</th>
<th>No-Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>25.0%</td>
<td>48.8%</td>
<td>26.3%</td>
</tr>
<tr>
<td>Male</td>
<td>32.5%</td>
<td>37.5%</td>
<td>30.0%</td>
</tr>
<tr>
<td>All</td>
<td>28.8%</td>
<td>43.1%</td>
<td>28.1%</td>
</tr>
</tbody>
</table>

Table 5. Comparison of VBR-CWI and FBR-CWI

<table>
<thead>
<tr>
<th>Test Speech</th>
<th>Prefer FBR-CWI</th>
<th>Prefer VBR-CWI</th>
<th>No-Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>31.3%</td>
<td>35.0%</td>
<td>33.8%</td>
</tr>
<tr>
<td>Male</td>
<td>28.8%</td>
<td>31.3%</td>
<td>40.0%</td>
</tr>
<tr>
<td>All</td>
<td>30.0%</td>
<td>33.1%</td>
<td>36.9%</td>
</tr>
</tbody>
</table>

Through informal listening tests, we have found that for most speech samples the cost criterion performs well and the closed-loop multimode VBR-CWI coder performs far better than the standard 4.8kbps FS1016 CELP coder and is a little better than the original 3.75kbps FBR-CWI. There are still some noisy or buzzy artifacts for a little part of speech segments that affect the whole quality of test speech. The problem is likely to be due to the inadequacy of the WSNR as a fidelity criterion for those segments. Further research has found that the usage of parameter ‘SEW-to-REW power ratio’ will help to exclude the bad mode to some purpose. If the ratio is lower, the CWs perform rapidly and mode1 is most probable; if it is higher, mode2 is most probable. Additionally, more accurate perceptual objective measurement of speech quality should be investigated to make the mode selection more robust.

4.4 Problem Discussion

The biggest and most formidable problem for closed-loop multimode coders remains the unavailability of an adequate objective speech quality measure. It is difficult to find a good speech quality or distortion measure for low bit rate coder assessment which is both reliable and easy to incorporate into the real-time coding process. To find an objective quality measure that accurately estimate subjective quality is a challenging task.

Complexity is also an important issue in closed-loop mode-selection schemes. The computational cost of such schemes can be excessive due to the fact that each frame of speech needs to be coded by all coding modes. If typical characteristic features of each speech class are found and good speech classification method is chosen, open-loop multimode scheme can also be considered to design VBR-CWI coder for some real-time applications.
5 Conclusion

A closed-loop multimode variable bit rate characteristic waveform interpolation speech coder is presented which applies different CWI coding structures tailored to different CWs modes. The mode selection is done with a cost criterion based on WSNR measurement. The VBR-CWI coder delivers reconstructed speech at an average rate of 1.86 kbps with a natural quality which appears to be better than the original FBR-CWI coder for most of the test data. The criterion is very simple and effective to select proper mode to some purpose, but it is not perfect because the preliminary listening tests also expose a spot of artifacts in one or two speech segments. Further research can be done to choose an objective quality measure that estimate subjective quality more accurately. For real-time applications, a robust open-loop multimode technique can be also considered.

References