Improved Characteristic Waveform Decomposition and Novel Bit Reduction Scheme for WI Coders

KEUNSEOK CHO, HEESIK YANG, SANGBAE JEONG, MINSOO HAHN
School of Engineering
Information and Communications University
119, Munjiro, Yuseong-gu, Daejeon, 305-732
REPUBLIC OF KOREA

Abstract: In this paper, we proposed an improved SEW/REW decomposition method with pitch-dependent phase generation and a noble VBR scheme for bit rate reduction in waveform interpolative coders. The proposed decomposition is performed in the magnitude domain to reduce spectral distortions. The phase of the characteristic waveforms is generated after classifying the signal into silence, unvoiced and voiced speech using the pitch value. The proposed VBR scheme is achieved by substituting white Gaussian noises with the excitation signal of silence and unvoiced speech and allocating bit rates variably. Experimental results show that our proposed algorithm achieves the improved speech quality while reducing the required bit rate compared to the conventional methods.

Key-Words: Characteristic waveform decomposition, Excitation signal, Speech coders, Variable bit rates, Waveform interpolation

1 Introduction

These days, the focus on the research of speech coders is to achieve high quality speech at low bit rates. The waveform interpolation (WI) coder has a hybrid structure that uses the filter model plus the excitation signal as a waveform. By the merit about achieving high quality speech at low bit rates, the WI coder can be used as a speech coder in digital wireless communication and can be used to construct low-capacity speech synthesizers. In the WI coder, the characteristic waveforms (CWs), i.e., the pitch length segments of the residual signal are used. By taking advantage of the differences in human perception between a slowly evolving waveform (SEW) and a rapidly evolving waveform (REW) extracted by the decomposition of the CW, the separate quantization of them is offered to get high coding efficiency [1]. The SEW characterizes voiced speech and the REW represents noise-like unvoiced speech [2]. Many researches on the speech quality improvement and the bit rate reduction have been executed. One of the past researches on the generation of the REW phase is based on the SEW/REW energy ratio for the improvement of the speech quality [3]. Source-controlled variable bit rate (SC-VBR) was proposed for the WI coders to reduce the required bit rates. The SC-VBR scheme is performed by the allocation of the variable bit rate according to the types of speech, that is, voiced, unvoiced, silence, and transition [4].

In conventional WI coders, the phase information is not transmitted to the decoder for the bit rate reduction. So, the SEW phase is generated by the fixed phase and the REW phase, by the random phase. The decomposition of the CW in the discrete time Fourier series (DTFS) domain cannot preserve the magnitude response of the excitation signal. The speech quality degradation can be caused by the non-transmission of the phase information and the variation of the magnitude components by the CW decomposition. In this paper, we propose a modified SEW/REW decomposition method with pitch-dependent phase generation and a noble VBR scheme for the speech quality improvement and the required bit reduction. To evaluate the performance, the perceptual evaluation of speech quality (PESQ) is used.

This paper is organized as follows: Section 2 describes the CW decomposition and phase generation method in conventional WI coders. Section 3 discusses the proposed SEW/REW decomposition scheme with pitch-dependent phase information generation and a novel VBR scheme. Section 4 and 5 describes the experimental results and the conclusion of this paper, respectively.

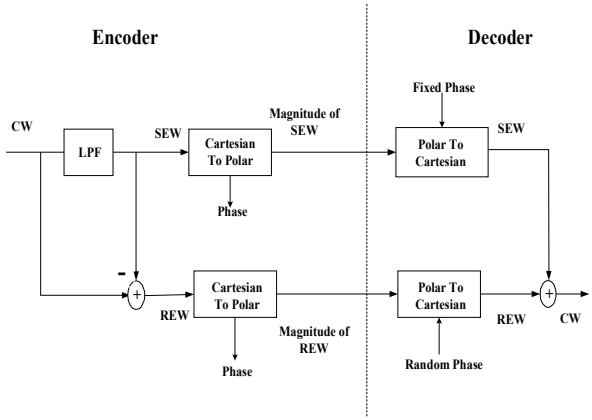

Figure 1. Block diagram of SEW/REW decomposition and phase generation in conventional WI coder

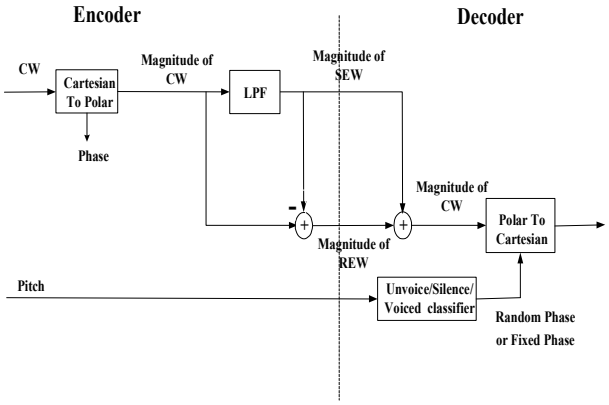

Figure 2. Block diagram of modified SEW/REW decomposition with pitch-dependent phase generation

2 Conventional methods

In the WI coder, a speech signal is represented as a sequence of evolving pitch-length CWs. In the conventional WI coder, CWs are extracted 8 times per frame. The extracted CW is transformed by the corresponding DTFS coefficients. Figure 1 shows the process of the SEW/REW decomposition and the phase generation in the conventional WI coder. The CW is split into the SEW and the REW in the DTFS domain [1]. The SEW is extracted by the low-pass filtering of the CW. The REW is extracted after subtracting the SEW from the original CW. Due to the decomposition of the SEW and the REW, they have magnitude and phase information, respectively. In the parameter quantization, their magnitude information is used only to reduce bit rates. Because the transmitted SEW and REW magnitudes are extracted after the CW decomposition in the DTFS domain, they cannot preserve the original magnitude of the CW.

At the decoder, the transmitted SEW and REW magnitudes are converted into DTFS coefficients using a fixed and a random phase contour, respectively. Then, a phase-distorted CW is reconstructed [1][3]. In human auditory systems, the magnitude information is more important than the phase one. But, the conventional decomposition cannot restore the original magnitude response of the excitation signal because it is performed in the DTFS domain.

3 Proposed methods

In the conventional WI coder, the SEW/REW decomposition is performed in the DTFS domain and the SEW and REW phases are generated by the pre-defined and random values, respectively. This process can cause some speech degradation at the decoder side. In order to solve the problems, the methods which apply an enhanced SEW/REW decomposition with pitch-dependent phase generation is proposed. Also, a novel VBR scheme dependent on the kinds of speech is proposed.

3.1 Modified SEW/REW decomposition

The proposed SEW/REW decomposition is performed in the magnitude value rather than in the complex value of the DTFS coefficients. In contrast with the conventional WI coder, the magnitude of the DTFS is extracted from the CW in our proposed method firstly. The magnitude spectrum is computed as (1).

$$|\mathrm{X(k)}| = \sum_{k=1}^{\lfloor P/2 \rfloor} \left(\sqrt{A_k^{\,2} + B_k^{\,2}} \right) \tag{1}$$

Where A_k and B_k are the DTFS coefficients and P is the pitch value. The magnitude and the phase information can be extracted from the CW. But, only the magnitude information is quantized and transmitted. The low pass filter in our proposed decomposition is identical to that of the conventional WI coder. The SEW magnitude is extracted by the low-pass filtering of the CW. The REW magnitude is extracted by subtracting that of SEW from that of CW. Using this decomposition scheme, the magnitude of the original CW is preserved by simply adding the magnitudes of the SEW and the REW at the decoder. At the encoder, each magnitude of the SEW and the REW is quantized and transmitted. Then, the transmitted SEW and REW magnitudes are simply added to reconstruct the CW which has the same magnitudes to its original one at the decoder. Figure 2 shows the process of the

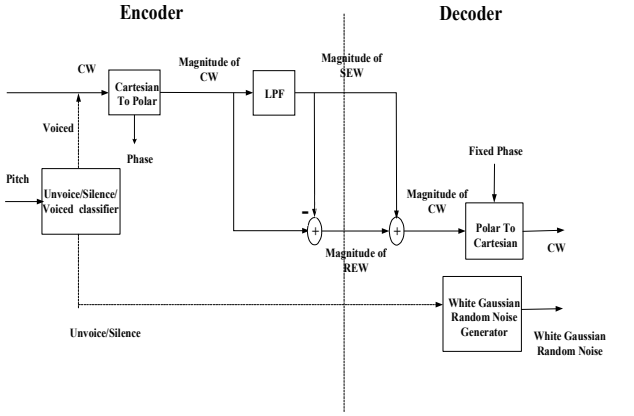

Figure 3. Block diagram of a modified SEW/REW decomposition and a novel VBR scheme

modified SEW/REW decomposition and the pitch-dependent phase generation in our proposed WI coder. Then, the CW is reconstructed by the combination of the generated phase and the transmitted magnitude information at the decoder.

3.2 Pitch-dependent phase generation and novel VBR scheme

The transmitted magnitudes using the proposed decomposition can reconstruct the CW efficiently by combining the phase information at the decoder. Generally, fixed phase and random phase values are used. But, in contrast with the conventional WI coder, they are generated according to the types of transmitted signals which are determined by the pitch values. The range of the pitch value is 20 ~ 145 at 8 kHz sampling rate. The minimum pitch value of 20 is assigned to silence and unvoiced speech at the encoder. If the received pitch value is 20 at the decoder, the frame is classified as the unvoiced or silence segment. In this case, the CW is reconstructed with the random phase. If the pitch value is not 20, then the frame is classified as the voiced segment and the CW is reconstructed by combining its magnitudes with the pre-defined fixed phase. Because the original phase of silence and unvoiced speech is not important, the random phase is applied to them. At the encoder, because a CW is aligned by the adjacent CWs, the phases of successive CWs are almost same in voiced intervals. So, the fixed phase is used in the voiced segment. By this method, the magnitude is preserved at the decoder. The residual signals in unvoiced and silence segment are similar to white Gaussian noises. So, if the pitch value is 20, then the power normalized noise signals are substituted. This method is performed by allocating variable bit rates according to the types of speech. Figure 3 is for the processes of the proposed SEW/REW decomposition and the novel VBR scheme in our WI coder. By using the novel pitch-based VBR scheme, the bit rate can be reduced.

4 Experiments and Results

For the performance evaluation of the proposed method, the PESQ is used as an objective measure whose score approximates the well-known mean opinion score (MOS) [5]. 100 Korean dialogue sentences recorded by 10 speakers are used as the test data. The duration of each sentence is about 10 ~ 20 seconds while the sampling rate is 8 kHz.

Table 1 shows the experimental results about the conventional, the previous REW phase generation and the proposed method. The results show that the PESQ score of the proposed method is higher than that of the conventional one. We can also see that the speech quality can be improved by the proposed SEW/REW decomposition better than the conventional one. This is mainly because our algorithm preserves the magnitude spectrum. Although a novel VBR scheme is added to the modified decomposition, the PESQ is almost same. So, better speech quality and more bit-rate reduction are possible with our proposed scheme. Using the proposed schemes we can improve 0.32 of the PESQ score. Table 1 shows the comparison with the previous REW phase generation method [3]. In [3], the total phase of the transmitted CW is calculated using a weighted sum of fixed and random values. The weight is calculated by the energy ratio of

Table 1. PESQ scores of conventional, previous REW phase generation and proposed method

Method		PESQ Score
Conventional Decomposition + Phase Generation		3.043
Conventional Decomposition + Previous REW Phase Generation Method		3.049
Novel Decomposition	+ Pitch-dependent Phase Generation	3.374
	+ Novel VBR Scheme	3.368

Table 2. PESQ results of SC-VBR-based and proposed VBR scheme-based WI

Method	PESQ Score
Source Controlled Variable Bit-rate (SC-VBR) based on WI	3.042
Modified Decomposition + Novel VBR Scheme	3.368

the SEW to the REW. The proposed pitch-dependent phase generation scheme is also better than the power ratio method.

Table 2 summarizes the results of the SC-VBR scheme in [4] and the proposed novel VBR one with the modified decomposition method. The results also confirm that our proposed novel VBR is better than the SC-VBR. The bit allocation of the proposed VBR scheme for each parameter is presented in Table 3. If the mode of an analysis frame is silence or unvoiced, we can get the reduction of 47 bits per frame. The transmitted pitch value of 20 is for silence or unvoiced. For those intervals, the excitation signal is substituted with a sequence of white Gaussian noises. The results of the bit rate reduction of the SC-VBR and the proposed VBR are about 14.8 % and 6.7 % respectively. Although the bit rate reduction of the SC-VBR is higher than that of the proposed VBR, more reduction is possible with our method if other parametric properties of the speech types which are adopted in the SC-VBR are additionally utilized.

Table 3. Bit allocation of proposed VBR scheme method

	Mode (bits)	
Parameters	Silence/Unvoiced	Voiced/Transition
Power	8	8
Pitch	-	7
LSFs	30	30
SEW (magnitude)	-	28
REW (magnitude)	-	12
Modes	1	1
Bit Allocation	39	86
Bit rate (bit/s)	1950	4350

5 Conclusions

In this paper, we proposed an improved SEW/REW decomposition and a novel VBR scheme to improve the speech quality and to reduce required bit rates. The efficiency of the proposed methods is confirmed by the experimental results. The speech quality is improved by the proposed method. The PESQ score is increased by 0.34 compared to the conventional WI scheme. In addition, the required bit rate is decreased by 6.7% using the proposed novel VBR scheme. As our future works, pitch detection based on an analysis-by-synthesis sense and the application of the current work to the high quality speech compression for corpus-based text-to-speech systems will be studied.

References:

[1] *Eddie L. T. Choy, "Waveform Interpolation Speech Coder at 4kb/s," MS. Thesis, McGill University*

[2] *Tammi M., Heikkinen A., Saarinen J., "On methods for perfect reconstruction WI speech coding with preprocessing," Speech Communication, Vol.38, NO.3, pp.305-320 November 2002.,*

[3] *Kihyun Choo, Namsoo Kim, "4.0 kb/s Waveform Interpolation coder", Korean Signal Processing Conference, Vol.12, NO.1, pp.323-326, 1999*

[4] *F. Plante and B.M.G. Cheetham, D. Marstom, P.A. Barrett, "Source controlled variable bit-rate speech coder based on waveform interpolation," Proc. ICSLP, pp.848-851, 1998*

[5] *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, ITU-T Recommendation P.862, February 2001.*