on Fundamentals of Electronics, Communications and Computer Sciences

1. Introduction

The waveform interpolation (WI) coder has offered a relatively good speech quality at very low bit rates compared with CELP-based coders. Interpolation based rate reduction was also of great interest for use in communication systems where bandwidth or memory constraints are important for good speech quality. Though these qualities and the bit rate of a WI code are not dependent on the quantization of the parameter vectors, the variation in the speech quality is due to the quantization of the parameter vectors. Therefore, a method for quantizing the speech quality of low bit rate coders with a low rate and appropriate quantization schemes for each parameter vector requires further investigation. A possible approach may be to use the VBR scheme for the interpolation scheme for other coders.

One possible target of VBR schemes is to remove the speech redundancy that originates from the slowly varying characteristics of speech signals [1]. A VBR scheme in the WI coder is not well addressed [2], and it is called the variable interpolation scheme [3]. The VBR coder determines the input speech segments that are used for the interpolation, encoded and transmitted using a variable block size of the energy ratio of each frame, and the energy ratio between consecutive frames. In spite of the powerful classification capability of the computational complexity of the SC-VBR, two different scales are set to be quantized. Firstly, the interpolation factors are used to decide the quantization levels. Secondly, the speech coder VBRs are used to decide the scale levels. Finally, the speech quality can be improved by using the combined parameter of the two different scales. In this way, the variable block size is used by using the variable block size to quantize the energy ratio of each frame. The speech coder VBRs are used to decide the quantization levels. Finally, the variable block size is used by using the combined parameter of the two different scales. In this way, the variable block size is used by using the combined parameter of the two different scales. In this way, the variable block size is used by using the combined parameter of the two different scales. In this way, the variable block size is used by using the combined parameter of the two different scales. In this way, the variable block size is used by using the combined parameter of the two different scales.
LETTER

New Variable-Bit-Rate Scheme for Waveform Interpolative Coders

Heesik YANG, Student Member, Sangbae JEONG, and Minsoo HAHN, Nonmembers

SUMMARY In this paper, we propose a new variable-bit-rate speech coder based on the waveform interpolation concept. After the coder extracts all parameters, the amounts of distortions between the current and the predicted parameters, which are estimated by extrapolation using the past two parameters, are measured for all parameters. A parameter is not transmitted unless the distortion exceeds the preset threshold. At the decoder side, the non-transmitted parameter is reconstructed by extrapolation with the past two parameters used to synthesize signals. In this way, we can reduce 26% of the total bit rate while maintaining the speech quality degradation below the 0.1 perceptual evaluation of speech quality (PESQ) score.

Key words: waveform interpolation, VBR or multimode

1. Introduction

The waveform interpolation (WI) coder has shown a relatively good speech quality at very low bit rates compared with CELP-based coders. Nevertheless, further bit-rate reduction without quality degradation is indispensable due to the bandwidth or memory limitation and the customer demand for good speech quality.

The speech quality and the bit rate in WI coder primarily depend on the quantization of the parameters. Thus one approach to guarantee a good speech quality at low bit rates may be to design an appropriate quantization scheme for each parameter. Another approach may be to use the variable-bit-rate (VBR) or multimode scheme for other coders.

One possible target of VBR schemes is to remove the source redundancy that originates from the slowly varying property of speech signals [1]. A VBR scheme in the WI coder has been proposed [2], and it is called the source controlled-VBR (SC-VBR). The SC-VBR classifies input speech segments into silence, voiced, unvoiced and transition using a voice activity detector (VAD), the energy ratio at each frame, and the cross-correlation between consecutive frames. In spite of the powerful classification capability and the computational simplicity of the SC-VBR, two drawbacks can be pointed out. Firstly, the classification fully depends on decision rules and a decision error will induce the uncontrollable degradation of reconstructed speech. In particular, when the slowly evolving waveform (SEW) and the rapidly evolving waveform (REW) are not perfectly decomposed, the degradation would be unpredictable. Secondly, the SC-VBR is basically an open-loop algorithm and the prefixed threshold of decision rules cannot cover the variations of speakers and environments.

In this paper, we propose a parameter-based VBR with the WI scheme, which is focused on maximally removing the interframe redundancy and is more robust to input speech or channel situations than the SC-VBR [2]. Section 2 describes the WI codec and Sect. 3 concerns the proposed WI coder with a VBR scheme. Experiments and results are presented in Sect. 4, and conclusions are given in the final section.

2. Waveform Interpolation

The main target of the WI approach is to retain the speech quality at a very low bit rate even with the inevitable additional complexity [3]. The decompositions of a pitch cycle waveform into the SEW and the REW enables effective quantization of the residual signal.

The baseline WI coder is designed for wide-band speech. The frame size is set to 20 ms and the linear prediction (LP) analysis window is set to 30 ms. The line spectral frequencies (LSFs) converted from LP coefficients (LPCs) are quantized using a split vector quantizer. Characteristic waveforms (CWs) are extracted 8 times per frame from the LP residual signal. The extracted CWs are then power-normalized and decomposed into the REWs and the SEWs by time-domain FIR low-pass filtering with a cutoff frequency of 25 Hz. Time normalization and alignment are carried out to convert a CW into a two-dimensional signal. We use the split VQ for the CW with four-dimensional conversion vector quantization (DCVQ). This means CWs are categorized into 4 classes according to their lengths and all CWs in each class are normalized to have a unique representative length.

3. Proposed VBR Scheme

Speech signals are slowly varying with time. This slow evolution of the signals can be easily observed in voiced regions. On the contrary, transition or unvoiced regions show rapid evolution which makes difficult to take advantage of interframe correlations. Our proposed scheme includes an attempt to find the interframe correlations at the parameter level to maximally reduce the interframe redundancy regard-