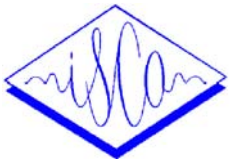

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

A NOVEL SPEECH CODING APPROACH BASED ON HALF-WAVE VECTOR QUANTIZATION[1]

Xiaoping Chen Yantao Song Tiecheng Yu

Speech Recognition Laboratory, Institute of Acoustics,
Chinese Academy of Sciences, Beijing 100080, P.R. China
Email: **cxp@speech1.ioa.ac.cn** or **tcyu@public3.bta.net.cn**

ABSTRACT

In this paper, a novel waveform coding approach based on half-wave vector quantization is presented. The input speech signal is divided into frames, each of which is classified as silence, unvoiced and voiced. The voiced category is subdivided into 64 sub-categories according to the length of each half-wave vector. The unvoiced category is subdivided into 16 sub-categories according to the average zero-crossing rate of each frame. For each type of vectors, a set of codebook is generated by using training data. Then each kind of sub-category is encoded, transmitted, and decoded with its own particular bit allocation configuration. This variable bit coding approach can operate at medium rate and provide acceptable-to-good speech quality.

10.21437/Eurospeech.1999-339

1. INTRODUCTION

It has become widely recognized that a vector, i.e., an ordered set of signal samples or parameters, can be efficiently coded by matching it with a similar pattern or a code vector in a codebook. Vector quantization (VQ) [1] is a well-established and widely used technique. It has been applied to the efficient coding of LPC parameters, pitch predictor filter parameters, gain parameters, coding of the excitation or residual signal in analysis-by-synthesis predictive coding techniques, such as VXC, CELP, and VAPC. However, the direct application of VQ to blocks of samples has been regarded a difficulty due to the wide dynamic variety of shapes and wide range of amplitude levels of speech waveform. But it is not really the case that VQ can't be directly and efficiently used to encode the original speech waveform. In this paper, we present a novel waveform coding approach based on half-wave vector quantization. In-depth coverage of this waveform coding will be presented in the following sections.

2. OVERALL CODING STRATEGY

As it is well-known that zero-crossing (ZC) rate is very important for the intelligibility of speech, while the amplitude of sample contributes mainly to the quality of speech. Therefore, in coding, if we can keep the ZC position of each half-wave unchanged, we can keep the formant that is important for understandability. As for the quality of speech, we can attain it by using variable bit rate (VBR) coding [2] configuration with considerable size of codebook. In this case, quantization can be performed directly on half-wave vector, which is determined by two consecutive ZC points. Considering the widely-varying acoustic phonetic character of the speech signal, we apply multimodal VBR coding. The input speech signal is first divided into frames, each of which is classified as silence, unvoiced and voiced according to the short-time average energy and short-time average ZC rate. The voiced category is subdivided into 64 sub-categories according to the length of each half-wave, and the unvoiced category is subdivided into 16 sub-categories according to the ZC rate of each frame. Then each kind of sub-category is encoded, transmitted, and decoded with its own particular bit allocation configuration. Fig. 1 gives the block diagram of this novel waveform coding strategy.

[1] The author gratefully acknowledges the support of K. C. Wong Education Foundation, Hong Kong.

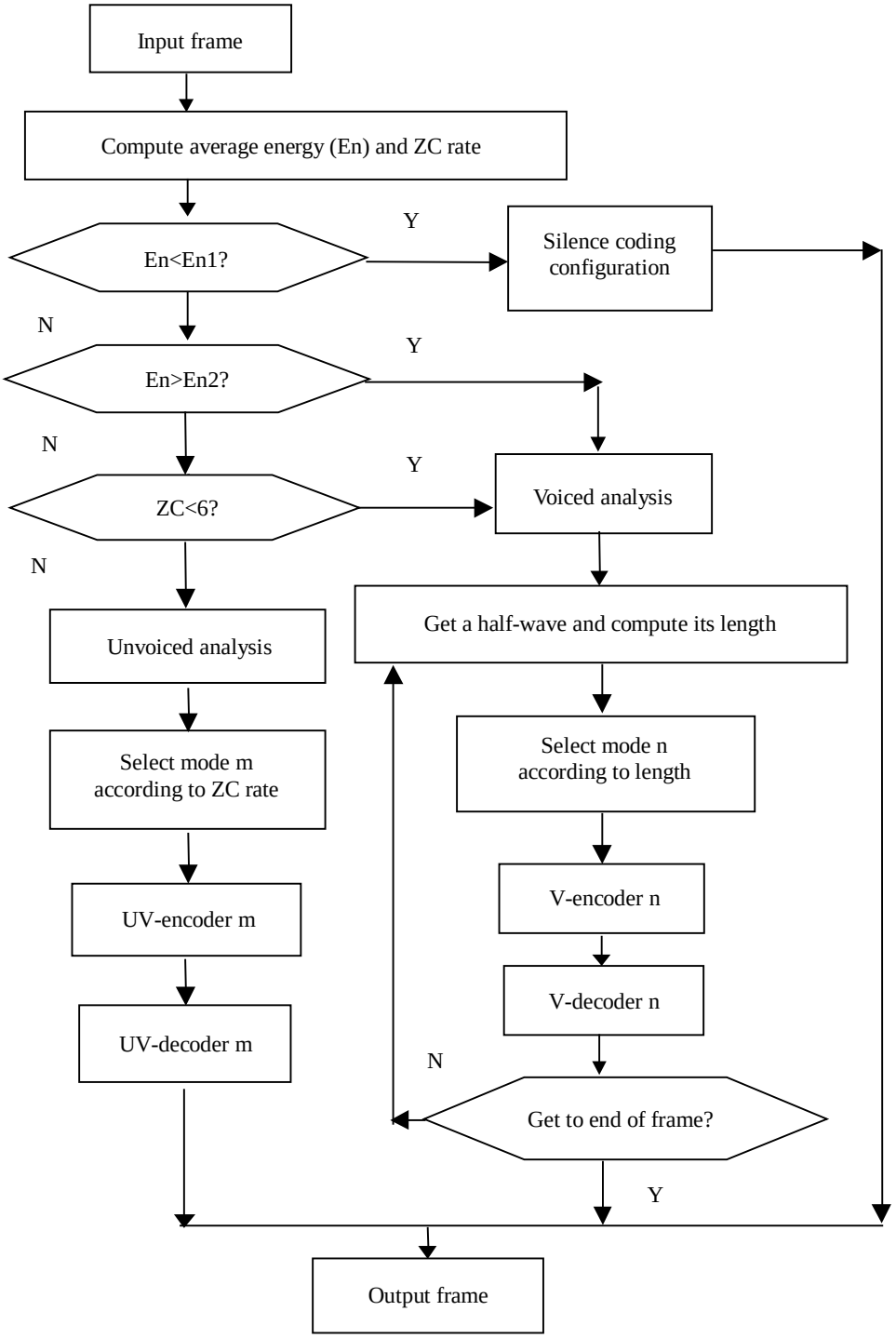

Figure 1. Block diagram of a novel waveform coding approach. (Where m is from 1 to 16 corresponding to ZC rate from 6 to 21 of input frame, and n is from 1 to 64 corresponding to the length of half-wave vector being quantized. En1 and En2 are short-time average energy thresholds, which are determined empirically.)

3. CODEBOOK CONFIGURATION

Codebook configuration contributes largely to the quality of reconstructed speech signal. As mentioned before, for

voiced category, 64 subcategories of codebook are constructed, each of which corresponds to a particular kind of half-wave. Bit allocation of each codebook depends on three factors: probability of input vector, perceptual sensitivity, and distortion evaluated by signal-to-noise-ratio (SNR). According to human auditory response curve, we know that human ear is most sensitive to frequency component between 1.2kHz and 2.5kHz. If we sample speech signal at the rate of 11kHz, then the half-waves with length between 4 and 9 fall within the sensitive area of human ear. Besides, these kinds of half-wave are more probable than the other kinds. Therefore, we allocate as more as 8 bits to their corresponding codebooks. For those with half-wave length between 10 and 40, we use 7 bits. And for those between 41 and 58 and those between 59 and 64, we apply 5 bits and 4 bits to their codebooks respectively. Here, it is worth to note that half-wave with one and two points occupy an absolutely large part in common speech waveform. But human ear is insensitive to the modification of them. For example, if we set those to zero amplitude, there's no perceptual difference between original speech and that of being modified. Therefore, we assign only one bit to them respectively.

As we know, more quantization noise is usually perceived for signals of small amplitude than for signals of large amplitude, because a louder signal can mask the quantization noise. To exploit this masking effect, we can quantize the signal samples on a logarithmic scale, rather than a linear scale. In a logarithmic scale, the step size between quantization levels becomes progressively larger with increasing amplitude. Considering the computation, we only apply logarithmic scalar quantization to unvoiced speech. For each unvoiced subcategory with different short-time ZC rate, we apply scalar quantization to construct codebooks of log value of short-time average energy. The bit allocation principle is similar to that used for voiced subcategories. Accordingly, 6 bits, 5 bits and 4 bits logarithmic quantizers are used for unvoiced frame with ZC rate within 6~12, 13~18, and 19~21 respectively.

Codebooks are designed from the training data set using the generalized Lloyd algorithm (GLA), which is identical to the Linde-Buzo-Gray (LBG) algorithm. The resulting quantizers are then evaluated using rate-distortion theory as the performance measure. Given a fixed quantization rate (R) for each quantizer, we wish to minimize the quantization distortion, D(R) [3]. The size of the codebook is often selected such that full use is made of an integer number of bits, i.e. the size is a power of 2.

4. ENCODING AND DECODING

It has become recognized that the efficient coding of a vector can be achieved by pre-storing a codebook of pre-designed code vectors. For a given input vector, the encoder then simply identifies the address, or index, of the best matching code vector by using the nearest neighbor rule. The index, as a binary word, is then transmitted to the decoder, where the corresponding code vector is reconstructed by a table-lookup from a copy of the same codebook as used in encoder. Fig.2 illustrates the basic idea of this process of coding and decoding.

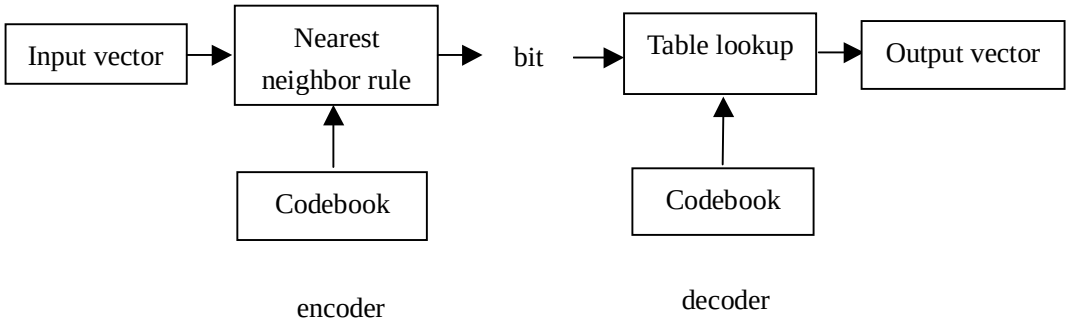

Figure 2. Block diagram of coding and decoding

5. EXPERIMENTAL RESULTS

The speech data base used in our experiments consists of 100 minutes of speech (including Chinese and English) recorded from 12 different FM radio stations which was sampled at the rate of 11kHz and the resolution of 16 bits/sample. The first 90 minutes of speech was used for training codebooks, and the last 10 minutes of speech was used for testing. Phonetic classification was performed every 2ms speech waveform. For unvoiced frame, according to its short-time average ZC rate, scalar quantization was used to construct corresponding codebook of log value of short-time average energy. For voiced frame, half-wave vectors were located first. And then according to the length of half-wave vector, corresponding codebook was constructed.

Preliminary experiments show that, the mean bit rate of voiced speech is about 2.79 bits/sample, the corresponding average compression ratio is about 6. The mean bit rate of unvoiced speech is about 0.44 bit/sample, the mean compression ratio is about 37. For voiced vector quantization, the mean SNR is over 13dB. This variable bit rate coding approach can operate at medium rate and provide acceptable-to-good speech quality.

6. CONCLUSIONS

The principle of a novel waveform coding approach based on half-wave vector quantization has been presented in this paper. Moreover, in-depth explanation of codebook configuration has been provided. We also have given the preliminary experimental results based on this coding approach. Compared to unimodal fixed-rate coding, this kind of multimode variable-rate coding offers increased flexibility to efficiently adapt the coding scheme and vector bit allocation to suit the short-term statistical character of the speech. It is particularly advantageous for voice storage, code-division multiple access wire-less networks, and packetized communication systems.

REFERENCES

[1] Gersho and R. Gray, *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.

[2] W.B. Kleijn, K.K. Paliwal (Eds.), *Speech coding and synthesis*, 1995 Elsevier Science B.V., Netherlands.

[3] Atal, V. Cuperman, and A. Gersho (Eds.), *Advances in Speech Coding*, Kluwer Academic Publishers, 1991.