A SPEECH COMPRESSION CODING ALGORITHM BASED ON HALF-WAVEFORM

Song Yantao, Zhou Jianlai, Yu Tiecheng
Speech Recognition Lab, Chinese Academy of Sciences, Institute of Acoustics,
P.O.BOX 2712, LAB 5, Beijing 100080,P.R.CHINA
Tel: +86-010-62553842, Email: {syt, zjl, tcyu@speech1.ioa.ac.cn}

ABSTRACT

This paper discusses a new kind of Speech Signal Compression Coding Algorithm based on Half-waveform. According to different characteristics of different speech signal parts, before we encode the Speech Signal, we segment Speech Signal into three kinds of segments: Silence segment, Unvoiced sound segment, Voiced sound segment. As such we encode each kind of speech segment and allocate different bit rate to each kind of speech segment to save the channel sources by different principles. Then we can get these advantages: low bit rate, high compression ratio, high quality of reestablished speech signal.

Key Words: Half-waveform, Vector Quantization (VQ), Speech Compression Coding.

1. PREFACE [2]

Generally speech coding algorithms can be divided into two kinds: Speech Waveform Coding Algorithm, Speech Parameter Coding Algorithm. The advantages of the former algorithms are: simplicity, high quality of reestablished Signal and anti—noise. But its compression ratio is low, it need high transmissibility. This kind of algorithm can not satisfy the requirement when transmissibility is not high.

Speech Parameter Coding Algorithm can achieve high compression ratio and low bit rate. But there are many shortages, such as: bad quality of reestablished Signal, loss of the nature of the speaker, bad naturality etc. And the parameter coding algorithms is sensitive to the environment noise. So this kind of algorithm can not satisfy the requirement of high quality.

This paper presents a new kind of Speech Coding algorithm based on Half-waveform, it encodes Speech Signal according to this characteristics of parts of Speech Signal to save the bit rate. It can achieve good quality of reestablished Signal at low bit rate. So this algorithm can partly overcome the shortages of the above two kinds of algorithms.

2. BASIC THOUGHT OF THE ALGORITHM AND SYSTEM DESCRIPTION [4]

BASIC THOUGHT:

According to the features of Speech Signal, Speech Signal can be divided into three kinds: Silence, Unvoiced sound, Voiced sound. In Silence part of Speech Signal, there is only a little information. So we can allocate very little bit to Silence part of Speech Signal. Basically unvoiced sound is a kind of noise, if we can keep the features of the noise changeless, then we can keep the quality of unvoiced sound changeless. The features of unvoiced sound are short-time zero cross rate and short-time average magnitude. So first we classify the short-time unvoiced sound by short-time zero cross rate, then we quantify these sorts of short-time unvoiced sound by short-time average magnitude to keep the short-time zero cross rate and average magnitude of unvoiced sound changeless. For voiced sound, F0 and low formant components are the most important and are basically determined by zero cross distribution. If we can keep the F0 and low formant components changeless, then we can keep the satisfied quality of voiced sound signal. This means that we should keep zero cross distribution changeless. So we can utilize the waveform vector quantization to reach the goal. If we use fixed frame length as the unit of

VQ(Vector Quantization)[2], because of the levity of voiced sound signal, we can not keep the zero cross position changeless and achieve the satisfied result of VQ. So we should take the half-waveform as the unit of VQ to keep zero cross positions changeless, as length of half-waveforms is variable, then length of vector is changed with the half-waveform. At the meantime, if the SNR of all kind half-waveform is good enough, then we can keep the reestablished waveform of voiced sound exactly close to original waveform to keep high quality of voiced sound.

According to the thought of this algorithm, we process the voiced and unvoiced sound separately to get the separate subcodebooks. Then we integrate all subcodebooks into a codebook. At last we can encode and decode speech signal according to codebook.

ALGORITHM DESCRIPTION:

This system is composed of two parts: Vocoder, Decoder. System architecture is showed as the below figure 1:

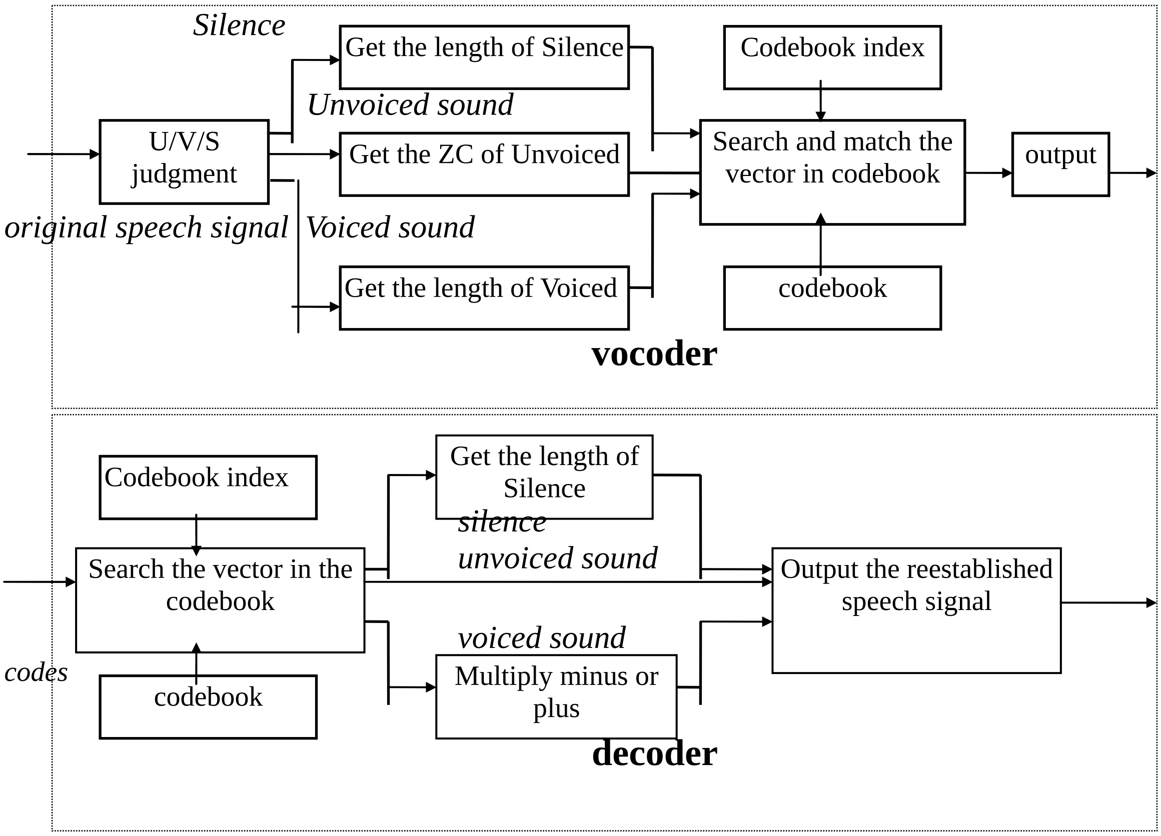

Figure 1, system architecture

The procedures of encoding and decoding:

1, segment the original speech signal:

Take the short-time zero cross rate and short-time average magnitude as the criterions to segment the speech signal into three kinds [1]: Silence, Unvoiced sound and Voiced sound. This segment method is not necessarily precise, as the demand for segment is not very high. Because of the system demand and algorithm features, we just hope that voiced sound will not be segmented to unvoiced sound, voice sound will not be segmented to silence. We set the threshold biased to reach this demand.

2, encoding process:

It is described as the below:[3][4]

(1) Silence: because there is not nearly any information in silence, we can express the silence with value 0. So we can take one code to indicate silence, another code to represent length of silence. Then Silence part is presented very well.

(2) Unvoiced sound: We divide unvoiced sound into frames of fixed length, only one code to express one frame of unvoiced sound. We classify frames of unvoiced

sound by its short-time zero cross rate, then search one vector in this kind of subcodebook which short-time average amplitude is most closed to this original frame of unvoiced sound, then output the code of the vector.

(3) Voiced sound: First extract half-waveforms from Voiced sound part of speech signal and get the length of the half-waveform which will be encoded immediately. Then according to match principle, search the most matched vector in the kind of subcodebook in which length of vectors is the same as this half-waveform. At last output the code of the vector.

Match principle: the best matched vector to one half-waveform is waveform of this vector which is most similar to the half-waveform, it means the distance between the vector and the half-waveform is shortest. The formula to calculate the distance is:

$$D = \sum_{i=1}^{L} \left| x_i - y_i \right|$$

Formula 1

L is the length of half-waveform, x_i, y_i are respectively i-th components of the half-waveform being coded and vector code in codebook.

At the mean time, the experiment that we had done before shows: if we reverse the polarity of any segment speech signal, human hearing will not be affected. And in speech signal half-waveforms appear as the sequence: positive half-waveform and negative half-waveform appear alternately. According to this feature of speech signal, we do not allocate any code to indicate polarity of half-waveform, we convert all half-waveform to positive half-waveform to encode and decode. For the first half-waveform of voiced sound, we allocate the polarity randomly. For the later half-waveform, we allocate the polarity to half-waveforms alternately. It can save a little bit rate.

3.decoding process is described as the below:

First input codes are analyzed, if the code represent silence, then we continue to get the next code, the next code represents the length of silence, then we output the length of value 0 as this segment of speech signal. If the code belong to unvoiced sound code, then we get the corresponding frame of unvoiced sound from codebook as this segment of speech signal. If the code belongs to voiced sound code, then we get the corresponding vector from codebook, and endow the polarity with this vector as this segment of speech signal.

Figure2 shows comparison between original speech signal and reestablished speech signal. The above is original speech signal, the below is the reestablished speech signal.

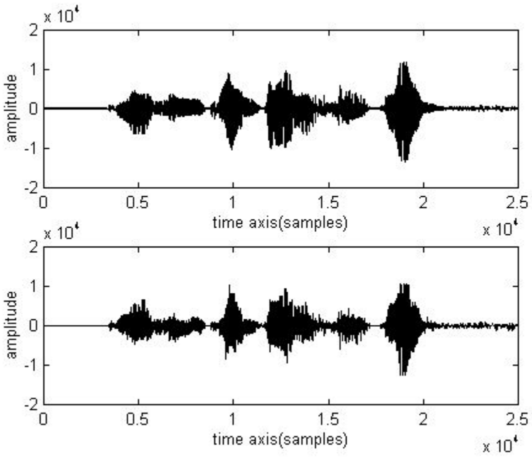

Figure 2, Comparison between original speech signal and reestablished speech signal

3. CODEBOOK GENERATION

Codebook generation is the most important part of this algorithm; we need to process the unvoiced sound and voiced sound to generate subcodebooks. For silence sound, we do not need to process it; it is presented by value 0.

First we sample a lot of raw speech signal datum, and segment it into three part: Silence, Unvoiced sound, Voiced sound by simple segment algorithm discussed in the above.

Generation of unvoiced sound subcodebook: we classify raw unvoiced sound signal into many kinds of unvoiced sound frames according to short-time zero cross rate, and quantize average amplitude of every kind of unvoiced sound frames. Then we get some average amplitudes for every kind of unvoiced sound frames, for every average amplitude, we select one unvoiced sound frame which average amplitude is closest to the average amplitude as one vector of subcodebook. At last we integrate all vectors of this kind of unvoiced sound frames to get one unvoiced subcodebook. One by one, we achieve all kind of subcodebook.

Generation of voiced sound subcodebook: we classify raw voiced speech signal into many kinds of half-waveform according to length of half-waveform, then process separately these kinds of half-waveforms by vector quantization. Here we use k-means LBG [2] algorithm. At last we integrate all kinds of voiced sound to achieve the voiced subcodebook.

After having all kind of subcodebook, we integrate all subcodebooks to get the codebook.

4. EXPERIMENTS AND CONCLUSION[3]

The features of this Speech Signal Compression Coding Algorithm: high compression ratio, low bit rate, satisfied quality of reestablish speech signal etc. At the meantime, because length of half-waveform is not fixed, so the bit rate is variable. And people can not decide the mean bit rate of speech signal. According to tests or calculation, the mean bit rate of Voiced Sound Segment , Unvoiced sound segment ,the Voice and Unvoiced segment are about: 4 bit/sample, 0.591 bit/sample, 1.18bit/sample respectively; the mean compression ratio are about: 4, 27.1, 13.6 respectively; the bit rate of Silence segment is very little, can be ignored. The subject hearing test was performed, the MOS of the algorithm is about 3.5(informal small-scale test), there exist a little noise in reestablished speech signal, but it can keep nature of speakers very good, and the quality of reestablished speech signal is also very good. From waveform comparison, when the SNR of VQ is precise enough, we can conclude that reestablished waveform is nearly the same as original waveforms. From the comparison of frequency analysis, we can know that components of low frequency and F0 (below 3kHz) is not changed on the whole, high frequency component is changed partly, but it don’t affect human hearing, because human hearing is mainly decided by low frequency component. From the above, we conclude that the algorithm have high compression ratio and low bit rate, so it can be used diffusely in Speech Signal Communication or local digital storage in the near future.

REFERENCES:

1,L.R.Labiner, R.W.Shafer <<Digital Processing of Speech Signal>>, Prentice-Hall, Inc., 1978

2,Yang Xingjun, Chi Huisheng, <<Digital Processing of Speech Signal>>(Chinese), Electrical Industrial Press, 1995

3,M.S thesis of Song Yantao, << Advanced Research about a new kind of Speech Signal Compression Coding Algorithm based on Waveform>>, Chinese Academy of Sciences, 1998/7/5

4,Song Yantao, <<a new kind of speech signal compression coding method based on waveform>>, The fifth Chinese man-machine communication conference, 1998