Decoder of the 48 kbps Audio Coder based on Segmental Sinusoidal Model

Florentinus Budi Setiawan
Teknik Elektro Universitas Katolik Soegijapranata, Semarang 50236,
email: fbudisetiawan@yahoo.com, f_budi_s@unika.ac.id.

Abstract

The recent technology of the television is implementation the digital system over transmitter, transmission channel and receiver. Therefore, television system in Indonesia is dominated by analog system, that less flexible for system development and communication on digital communication system. So that, it need actions to develop the digital broadcasting system to replace the analog broadcasting system. Audio signal information with high quality would help the television audience to increase the perception of the information displayed. Transmission channel capacity will become limited, while the need of channel communication is increased. The limited channel capacity endorse to save all aspect on the telecommunication system. There are some research on audio signal coding to obtain the lower bit rate for transmission channel usage saving. The research aim is coding the audio signal on the low bit rate for saving the channel communication usage for digital television broadcasting. The research will be done is develop an audio signal coder on the low bit rate with the suitable decoder. The encoder and the decoder development is implemented using C++ software. The hardware for the simulation process consists of microphone, digital signal processor and the personal computer equiped with sound card for audio signal acquisition.

Keywords : audio, coding, compression, sinusoidal,

1. Introduction

The limited channel capacity endorse to save all aspect on the telecommunication system. There are some research on audio signal coding to obtain the lower bit rate for transmission channel usage saving. The audio signal is processed, so that the redundant component can be decreased, then we obtain the simple size of information that reliable to be transmitted. The research aim is coding the audio signal on the low bit rate for saving the channel communication usage for digital television broadcasting. The research will be done is develop an audio signal coder on the low bit rate with the suitable decoder. The decoder consists of the parameters detector, signal synthesizer, and the periodic signal generator. The research results contribution for digital broadcasting is the method developing for decreasing the audio signal rate. So that the communication channel usage can be saved.

2. Segmental Sinusiodal Model

Sinusoidal model proposed by Almeida et.al and McAulay[1]. Almeida's approximation is implemented by finding correlation of harmonic phase between the consecutive frame of signal. In the other hand, McAulay uses mixed-voicing methods, that phase of voiced signal is picked up from spectral envelope, under minimum phase assumption[2]. In this condition, unvoiced phase is random. Spectral envelope is defined with linear prediction coefficients. Audio signal can be approximated by segmental sinusoidal model. A segment of audio signal from a maximum peak to the minimum consecutive peak can be approximate as a cosine signal from 0 to π. Then, from a minimum peak to the maximum peak can be approximate as a cosine signal from π to 2 π. The sinusoidal signal approximation is obtained by finding the maximum and minimum peaks on the observation frame. Maximum *i*-th peak is denoted by *p(i)* and the minimum *i*-th peak is denoted by *v(i)*. The *p(i)* is the maximum peak is located before the minimum peak *v(i)*, so that the reconstructed signal from the maximum peak to the minimum peak can be formulated as[3-4]:

$$s_{pv}(n) = a_0 + \sum_{i=1}^{k} a_1 \cos\left(\frac{(n - n_{p(i)})\pi}{n_{v(i)} - n_{p(i)}}\right) \quad (2)$$

Where :

$$a_0 = \frac{p(i) + v(i)}{2} \quad (3)$$

$$a_1 = \frac{p(i) - v(i)}{2} \quad (4)$$

The a_0 and a_1 are the Fourier coefficients for the DC components and the first harmonic. Then the minimum peak to the maximum peak can be reconstructed by the following formula

$$s_{vp}(n) = a_0 - \sum_{i=1}^{k} a_1 \cos\left(\frac{(n - n_{v(i)})\pi}{n_{p(i+1)} - n_{v(i)}}\right) \quad (5)$$

Where :

$$a_0 = \frac{v(i) + p(i+1)}{2} \quad (6)$$

$$a_1 = \frac{p(i+1) - v(i)}{2} \quad (7)$$

In a frame is consists of *2k* segments of reconstructed signal contain of *k* cosine signals and *k* negative cosine signals. The Eq.(2) and Eq(5) are the clipped Fourier series. The higher order of the Fourier series is reduced into zero in order to simplify the coding process for the lower rate. These equations means that the higher frequency is reduced, so that only the DC-offset and the first Fourier coefficient is passed into decoder.

Audio could be heard by ears because of fluctuation from one value to the other value of air pressure. Fluctuation of the air pressure would result peaks that contain of maximum and minimum value on certain time interval. Signal characteristic with peak and valley (minimum peak) could be use as a model to approach audio signal form. Part of signal which contain interval between one peak to the consecutive peak could be represented by one sinusoidal signal segment [4]. Peak to peak pattern was significant to represent the level of signal periodicity. Level of periodicity was very important for human hearing perception, especially voiced signal which have the most energy of audio signal.

Audio signal could be analyzed by segmental peak to peak with sinusoidal model approach. On sinusoidal transformation method, signal was modeled by harmonic over certain frame. Some of highest peak over spectra from Fourier Transform were taken to represent the estimated signal. New method on this paper is approaching process in the time domain. Process will start by marking the maximum and minimum peak. If in a certain time interval (frame), there are consist of i maximum peak and i minimum peak, denoted as p(i) and v(i). Audio signal is quantized based on peaks value and distance between two consecutive peaks with segmental sinusoidal model. Period length quantization can be reduced in size by application of codebook or look-up table. Codebook is design with respect on the statistical characteristics of the coded signal. The method to finding the code vectors is training the large number of signal vector.

In period length quantization, codebook is design based on the quantization of period length value of the large number of audio signal, in order to obtain the accurate code vector to minimize the distortion of period length. In this paper, we explain the design of codebook model based on the quantization of period length. Audio signal is fetched every 20 ms (160 samples) before separated into voiced and unvoiced. This signal will be coded into segmental sinusoidal model with bit allocation is 960 bits for 20 ms in order to obtain the rate of 48000 bit per second. Audio signal can be coded into segmental sinusoidal model for every 960 samples. If the length of pitch period is more than 960 samples, it is applied the 2:1 decimation process. So the maximum number of maximum and minimum peaks is no more than 25.

Based on the experimental results, the period length of quantization is vary from one to 45, as showed at the fig. 2 Distribution of the quantization of the period is decreased for increasing of the length. More than 90 % the value of the quantization is less than 10.

The unvoiced signal contributes less than five of the period length quantization. The voiced signal contributes the length of period quantization more than five. The quantization of the length of period into codebook is reduced the quality of reconstructed signal.

The number of period length quantization for a block of signal with 20 ms length is varied. The coding process is more optimal if the quantization is coded into blocks with variation of the number index of codebook number of period length quantization that less than five is dominant. Variation the number of period length quantization is vary from one to 45, so that it needs 6 bit for each value.

The length of period is quantized into six bit (64 probability of quantization value) based on the higher value which is 45. the value of quantization is coded into codebook with vary in length because of variation of kp for a block with 20 ms length.

3.Signal Reconstruction

Audio signal encoder at 48 kbps is designed in several block and algorithm. Detail of the encoder is shown in fig. 3.

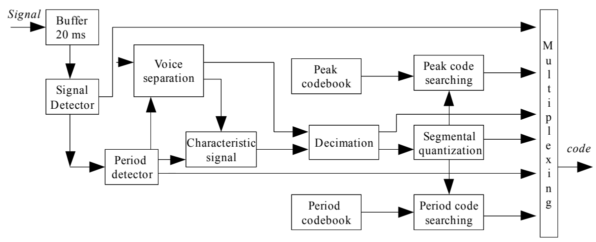

Fig 1. Encoder

The encoder contains existing signal detector, windowing process, and pitch detector. The next blocks are voiced and unvoiced classificator, and sinusoidal based coder. There are some operation mode of the encoder system depends on the kind of signal to obtain the high performance of coding system [5-11]. The operation modes consist of two operation mode, that is silent operation mode and signal operation mode. The signal operation mode consists of vibrating mode operation and non-vibrating mode operation. Input signal is audio signal in 16-bit PCM format at 48 kHz frequency sampling. The first block is signal buffer with 20 ms length. The next block is existing signal detector. Then the 20 ms signal will be detected its pitch period width. Based on pitch period information, signal would be classified into vibrating and non-vibrating signal. If it is less than 960 samples, the signal in buffer is called as vibrating signal. Then, if it is more than 960 samples, it is called as non-vibrating signal. The next process is depend on the kind of signal. For vibrating signal (voiced), characteristic signal have to be held. One pitch period of signal is quantized using segmental sinusoidal model. The formant information for each pitch period is kept to obtain the variation information changing for 20 ms. The next block is codebook index searching based on periods, peaks, and formants. All of the coded parameter are sent to the decoder with rate 48 kbps, depends on the kind of the audio signal.

The audio input signal exist are detected by using existing signal detector. The signals are buffered with length of 20 ms. The detector identify the input audio signals whether there are signals exist or there are no signals exist. If there are no signals exist, they are called as silence. A sign is transmitted into decoder to inform this condition, so that the decoder is not process the signal during 20 ms. But if there are signals exist, the encoding process is continued with pitch detecting process.

Pitch is the useful parameter in encoding process. Based on the pitch value, we would identify the signals whether voiced or unvoiced. The pitch period is detected by using autocorrelation process. The first step, the buffered signal is detected on its peak. Based on the peak value, it can be found the threshold for the centre clip process. The threshold is half of the peak value over entire signal in the buffer. The audio signals on the buffer are clipped, so that we would reduce computation complexity. The clipped signals are processed in autocorrelation computation. The autocorrelation process would result two kind patterns. There are peak-valley-peak pattern and peak-valley pattern. The pitch value is detected based on the distance between peaks of the peak-valley-peak pattern. The peak-valley pattern indicates that the signals are unvoiced.

The voiced and the unvoiced signals are classified by using the pitch detection process results. The autocorrelation results pattern are used as reference to identify the signals whether voiced or unvoiced. If the pattern is peak-valley-peak and if the distance between peaks is longer than 2.5 ms but less than 20 ms, it means that the signals is voiced. Then, if the pattern is peak-valley or peak-valley-peak with distance between peaks is longer than 20 ms, it means that the signal is unvoiced. The voiced and unvoiced signals would process in the different methods. The unvoiced signals would be process without referring the pitch period, then the voiced signals would process based on the pitch period.

The voiced signals are fetched on the one pitch period that representing entire voiced signals on the buffer. The one pitch period signals are called as characteristic signal in waveform interpolative signal terminology. The length of the characteristics signals is referred as the pitch period. The characteristic signal is quantized on its peaks and periods by using segmental sinusoidal model. For the unvoiced signal, the decimation process is implemented to obtain the smaller size of signal. Then the peak and period quantization is applied. Based on the segmental sinusoidal model, peaks and periods information is extracted. The processed signal would be generated by using the peaks and periods quantization.

The peak information size is reduced by applying look-up table. The look-up table is also called as codebook. The codebook is trained by using the peak information code-vector. Large amount of the peak information code-vectors are trained with LBG algorithm to obtain the peaks codebook. The index number of the peak codebook is varied from 6 to 10 to obtain the optimum process. The period information size is also reduced by applying look-up table. The index number of the peak codebook is also varied from 6 to 10 to obtain the optimum process. The period accuracy has to maintain to obtain the good receiver perception on the decoder side.

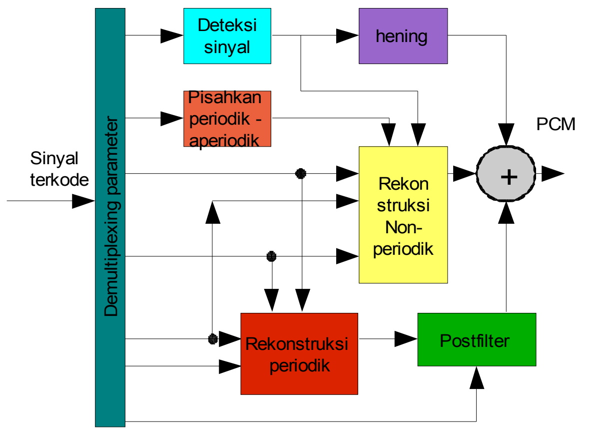

Fig 2. Decoder

Parameters have to send to the receiver are peaks, periods, pitch, formants, segment, and decimation. The peaks information needs 360 bits, the period information needs 360 bits, pitch needs 42 bits, and the segment information needs 36 bits. Total coded signal bits resulted for one frame (20 ms) is 960 bits. Thus, the coded audio signals data rate is 48 kbps.

4. Experimental Results

Part of signal between minimum to maximum was approached by negative cosine at half period. When the other part of signal between maximum to minimum would be approach by a half period of cosine signal. The number of sinusoidal signals which resulted would vary respect to the number of peaks on the signal frame. The more peak would decrease compression factor.

Result of signal reconstruction by sinusoidal approach seems smoother than the original that had arbitrary form between one peak to the consecutive peak. Nevertheless, roughness of original signal means containing high frequency component. Therefore, the spectral power is reduced on the high frequency component. Unfortunately, increasing of the number of peak would decrease the dynamic range of period changing variation for each segment. Thus, compression ratio would be increased to compensate decreasing of compression ratio caused by the number of peaks.

The proposed audio coder is simulated in personal computer using *coder 48kbps* program that developed by using C++ program. The coder is applied on digital signal processor starter kit TMS320VC5416 from *Spectrum Digital*, using *Code Composer Studio ver.3.1*. Based on the experimental results, hearing perception of the reconstructed signal is fairly good. The coder complexity is less than 10 MIPS. It needs less than 16 kB for encoder and 3 kB for decoder.

5. Conclusion

Reconstruction process is realized based on periods and peaks information. The experimental results show that this is a correlation between the size of codebook and performance of reconstructed signal.

Acknowledgment

This paper is written based on my research on Audio Signal Coder, funding by Indonesian Ministry of Research and Technology on Applied Research Incentive Program for 2010 Scheme.

References

[1] RJ.McAulay, TF Quatiery, Audio, *Analysis/ Synthesis Based on a SinusoidalRepresentation*, IEEE TASSP, vol. ASSP-34, no. 4, 1986

[2] TF Quatiery, RJ.McAulay, *Audio Transforma tions Based on a Sinusoidal Representation,* IEEE TASSP**, vol**. ASSP-34, no. 6*,* 198SL. Rabiner, BH. Juang, *Fundamentals of Audio Recognition*, Prentice Hall international, 1993, New Jersey.

[3] FB. Setiawan, *Reconstruction of Quantized Audio Signal by Peak to Peak Segment with Sinusoidal Model*, International Joint Conference TSSA WSSA, Bandung, 2006

[4] FB. Setiawan, *A Codebook for Period Information of Segmental Sinusoidal Model*, ICEEI, Bandung, 2007

[5] S. Furui, *Digital Audio Processing, Synthesis, and Recognition,* Marcel Dekker Incorporation, 1989 New York.

[6] JR. Deller, JG. Proakis, JHL. Hansen, *Discrete-Time Processing of Audio*, Macmillan Publishing Company, 1993, New York.

[7] BS. Atal, V. Cuperman, A. Gersho, *Advances in Audio Coding*, Kluwer Academic Publishers, 1991, Massachusetts.

[8] BS. Atal, V. Cuperman, A. Gersho, *Audio and Audio Coding for Wirelles and Network Applications*, Kluwer Academic Publishers, 1993, Massachusetts.

[9] C. Rowden, *Audio Proccesing*, McGraw-Hill Book Co., 1992, London.

[10] AM. Kondoz, *Digital Audio : Coding for Low Bit Rate Communications Systems*, John Wiley & Sons Ltd, 1995,West Sussex.

[11] A. Rao, R. Kumaresan, *On Decomposing Audio into Modulated Components,* IEEE TSAP vol 8, hal 240-254, 2000