A Low Bit Rate Audio Coder based on Segmental Sinusoidal Model

Florentinus Budi Setiawan
Electrical Engineering, Soegijapranata Catholic University, Semarang 50236,
email: fbudisetiawan@yahoo.com, f_budi_s@unika.ac.id.

Abstract

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 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 proposed audio signal coder consists of signal existention detector, period width detector, average amplitude counter, separation of the vibrating and the nonvibrating signal, determination of the period signal and parameters coding. One period signal is coded using sinusoidal model. 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.

Keywords : audio, coding, compression, peak, period, sinusoidal

1. Introduction

Information that transmited and received on the television broadcasting system are the moving picture and audio signal. 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. 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 coder developing is based on the hypotheses that the audio signal can be coded into the low bit rate and it can be reconstructed as the high quality synthetic signal. The proposed audio signal coder consists of signal existention detector, period width detector, average amplitude counter, separation of the vibrating and the nonvibrating signal, determination of the period signal and parameters coding. One period signal is coded using sinusoidal model. The decoder consists of the parameters detector, signal synthesizer, and the periodic signal generator. 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. The research results contribution for digital broadcasting is the method developing for decreasing the audio signal rate. So that the digital television communication channel usage can be saved.

2. Segmental Sinusiodal Model

Audio signal for digital television can be modelled by linear combination of sinusoidal component with time varying of amplitude, frequency and phase. 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. In this condition, unvoiced phase is random. Spectral envelope is defined with linear prediction coefficients. Audio signal can be represented by the following formula [9]

$$s(n) \approx \tilde{s}(n) = \sum_{k=1}^{K} A_k \cos(\varpi_k(n)n + \Phi_k(n) \quad (1)$$

$A_k(n)$ is representing amplitude, then $\varpi_k(n)$ is frequency and $\Phi_k(n)$ is representing the phase at the *k*-th of sinusoidal components. Signal in this model can be represent as the k sinusoidal of signal. Then the length of signal is infinite. If this signal will be quantized into sinusoidal components, *k* is infinite the large amount of sinusoidal components can be reduced with showing the significant components. The more components of sinusoidal signal are showed the higher quality of audio signal.

Audio signal maybe decomposed into modulated components[3]. Audio signal is also modeled into amplitude and frequency modulation[4]. System analysis and synthesis modeling based on overlap-add sinusoidal model combination for synthesis and audio quality enhancement is proposed by George[5]. 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 [13]:

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

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

$$s_{vp}(n) = \frac{v(i) + p(i+1)}{2} - \sum_{i=1}^{k} \frac{p(i+1) - v(i)}{2} \cos\left(\frac{(n - n_{v(i)})\pi}{n_{p(i+1)} - n_{v(i)}} \right) \quad (3)$$

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(3) 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.

In space, audio signal 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. 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.

Based on fluctuation of the audio signal, it could be analyzed by segmental peak to peak with sinusoidal model approach. On sinusoidal transformation method, signal was modelled 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. Processing of the signal analysis 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).

The quantization on the audio signal done is 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. Codebook of the period length 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

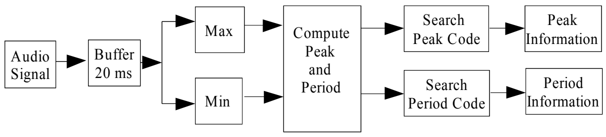

Fig 1. Signal analysis

Audio signal is fetched every 20 ms (960 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. 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.Encoder

Audio signal encoder at 48 kbps is designed in several block and algorithm. Detail of the encoder is shown in fig. 2. 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 [6-10]. 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 next block is codebook index searching based on periods and peaks. All of the coded parameter are sent to the decoder with rate 48 kbps, depends on the kind of the audio signal.

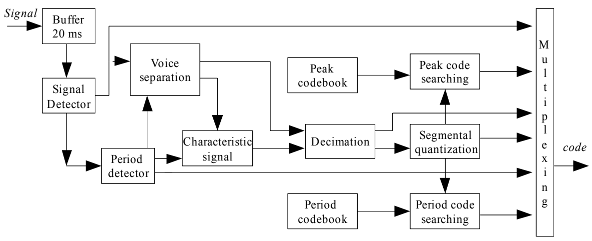

Figure 2. Encoder

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.

Periodicity of input signal is the useful parameter in encoding process. Based on the pitch period 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- pattern results 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[11-12].

One pitch 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 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.

Parameters have to send to the receiver are peaks, periods, pitch, 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. Codebook Structure

Input signal is fetched from microphone every 20 ms due to obtain more than probability of one pitch length of signal. Quantization of the period length on 20 ms of signal is divided into six sub-frames. A subframe is divided into blocks. Each of blocks contains combination of the five or four value of period length quantization. For each blocks will be coded into codebooks with vary in theirs index.

The period length quantization is coded into codebooks that vary in length due to theirs probability number of *kp* in a 20 ms of audio signal. The five or four blocks of signal that quantized in period length has coded into codebooks with length of 11 bits until five bits. The beginning of block is allocated the more number of bits because of its high of probability to come, especially for voiced signal. The following table shows the codebooks arrangement. The number of information is varied between one until 40 bin (number of period information). So, we cannot divide each bin with five or four to code this information. Combination between four and five sample is used to accommodate each number of period information. The combinations are two blocks of five-sample and four blocks of four-sample. It needs two five-codebook code vector and four four-codebook code vector.

The first step of the codebook generation is collecting the great number of period information. The period information would be applied as training vector code candidate. Then the next step is training this vector code by using the K-means algorithm, so that we would be obtained the codebook. Five-sample codebook is obtained by using the five-sample period information, and for sample-codebook is obtained by using the four-sample period information. Some codebooks will use a large number of memory place, in order to reduce coded data signal rate. The wider codebook has 2048 different codes, when the smallest has 32 ones. Each of codes consists of five period length quantization to include five dimension code-vector. The number of memory that is allocated to the highest codebook is 10240 places for integer. Each of the integers needs 8 bit in memory, so that memory allocation for 11 bit codebook is 81920 bits.

Table 1. Bit allocation for probability bin appear

Bin	Block	configuration	CB	bit
40	8	5555 5555	7	56
39	8	5555 5554	7	56
35	7	5555 555	7	49
34	7	5555 554	7	49
30	6	555 555	8	54
….				
….				
5	2	5	11	22
4	1	4	11	11
3	1	3	11	11
2	1	2	11	11
1	1	1	11	11

5. Experimental Results

There is a audio signal segment that modelled by sinusoidal approach for each interval between maximum peak to minimum peak and vice versa. This model would result a train of periods which change for every segment with changed amplitude. If amplitude was kept constant on normalization number, the result was the signal in frequency modulated. Period and peak value for every segment could be used as information which sent to decoder in order to save the transmission channel. In the decoder, signal would be reconstructed to obtain the original signal estimation by sinusoidal model.

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.

6. Conclusion

The periodic part of the auido signal can be decomposed by infinite sinusoidal signal with combination of amplitude, frequency and phase. Quantization is a methods to code or compress a audio signal. The useful parameters are peaks and period between consecutive peaks. Signal parameters are quantized into vector quantization form, so that codebook has to be generated in order to indexing of the vector quantization. Based on previous explanation and experimental results, we conclude that can be realized a codebook for period length quantization value in order to reduce the number of data to be sent. The more codebook index numbered, the higher performance of the 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*,* 198S.

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

[4] TF. Quatieri, TE. Hanna, GC. O'Leary, *AM-FM Separation Using Auditory-Motivated Filters,* IEEE TSAP, vol 5 hal. 465-480, 1997

[5] EB. George, MJT. Smith, *Audio Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model,* IEEE TSAP, vol 5, hal 389-406, 1997.

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

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