2014 1st International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)

Period Information Deviation on the Segmental Sinusoidal Model

Florentinus Budi Setiawan
Electrical Engineering, Soegijapranata Catholic University,
Semarang 50234, Indonesia
fbudisetiawan@yahoo.com

Abstract— **Speech signal can be modeled by sinusoidal model. On the sinusoidal model, there are many kinds for representing the signal. One of model is Segmental Sinusoidal model. The segmental sinusoidal model is an approximation method based on sinusoidal model for speech signal, especially for periodic detection. The periodic signal can be decomposed by infinite sinusoidal signal with combination of amplitude, frequency and phase. After the signal is decomposed, parameter will be quantized. The proposed quantization method in this paper is sampling signal on big part between minimum and maximum part over observation block. Some parameters of speech signal are detected. The useful parameters are peaks and period between consecutive peaks. Period information obtained from this quantization tends to different than the original, In this paper, we show the experimental results that there are many differences between period information on encoder side with the decoder side. It caused by quantization error on period information and quantization error on the codebook design. Effect of differences is degradation of signal quality, especially on frequency signal accuracy. On this paper, deviation of the reconstructed signal from original signal will be evaluated. Deviation from the original signals means that some error occur on period quantization.**

Keywords— ***frequency, error, period, segmental, sinusoidal***

I. INTRODUCTION

Recent communication needs simple form coding method. Because the number of transmission channel for communication become limited due to increase of the communication channel usage. Limited channel capacity had endorsed all side more efficient. Speech signal data rate can be developed by using encoder to obtain the simpler information. The research aim is obtained the low rate speech signal encoder with high quality. But some handicaps must be

The research aim is obtaining error of period quantization for coding algorithm on the rate of 4 kbps with high perception quality, on Mean Opinion Score 3,5 - 4. The main research consists of encoder and decoder development at conversation frequency between 300 Hz and 3400 Hz. For The simulation is developed by using the Borland C++ software at 4.5 version. Software simulation consists of speech signal acqusition, speech signal coder, speech signal reconstruction, speech signal reproduction as results of the signal reconstruction and performance tests. Error of period quantization is including on the transcoder performance test.

Speech signal can be modelled by sinusoidal model. On the sinusoidal model, there are many variation for representing the signal. On of the model is Segmental Sinusoidal model. The segmental sinusoidal model is an approximation method based on sinusoidal model for speech signal, especially for periodic detection. The periodic signal can be decomposed by infinite sinusoidal signal with combination of amplitude, frequency and phase. Some parameters of speech signal are detected. The useful parameters are peaks and period between consecutive peaks. Period information obtained from this quantization tends to different than the original, In this paper, we show the experimental results that there are many differences between period information on encoder side with the decoder side. It caused by quantization error on period information and quantization error on the codebook design. Effect of differences is degradation of signal quality, especially on frequency signal accuracy. On this paper, deviation of the reconstructed signal from original signal will be evaluated. We need other method to reduce error. This method have to improve the period accuracy. By finding number of error, we can obtain the proper method for improvement.

II. SINUSOIDAL MODEL

A. *Sinusoidal Model*

There are many part of speech signal would be represented into k sinusoidal signals that vary in amplitude, frequency, and phase. Speech signal was reproduced by articulation of human organ. For modeling signal, voiced signal could be quantized into infinite sinusoidal components. Speech signal also would be decomposed into modulated components [1]. The speech signals could be represented into amplitude modulation and frequency modulation [2]. Speech signal modeling for speech signal synthesis and quality enhancement. The system based on overlap-add sinusoidal model combination is proposed by George [3].

Periodic signal sampled at sampling frequency Fs would generate N discrete signal $s(n)$ from n=0 until n=N-1. The discrete signal $s(n)$ could be represented into Fourier series. The n represents signal sampling, , a_0 represents 0-th Fourier series coefficient, a_k and b_k are k-th Fourier series coefficients , then k represents the number of sinusoidal components and $\omega = 2\pi/T$, where T is signal period, we would formulate the discrete signal as :

$$s(n) = a_0 + \sum_{k=1}^{\infty} a_k \cos k\omega + \sum_{k=1}^{\infty} b_k \sin k\omega \quad (1)$$

B. Segmental Sinusoidal Model

Speech signal could be represented into dynamics segments [4]. Based on the dynamic segments, speech signal would be represented into varied length depends on maximum peak positions and minimum peak positions. This method is more efficient than the recent sinusoidal model that needs ten or more Fourier coefficients. Ears would hear sound because of fluctuation of air pressure. This fluctuation contains peaks and valleys in certain time interval. In this paper, we mention peaks as maximum peaks and valleys as minimum peaks. Signal characteristics with maximum and minimum peaks could be used as a model to approximate the speech signal form. Part from a peak to the consecutive peak would be represented as a segment of sinusoidal signal [5-7]. Peak to peak pattern is important for determining signal periodicity level. Then, the signal periodicity would help he human hearing perception, especially for voiced signal that have the largest energy of the speech signal.

Peaks of the speech signal could be quantized into theirs distance and magnitude. Then, the quantized value could be used as information to reconstruct a speech signal approximation. Quantization is realized by detecting time distance between maximum peak and consecutive minimum peak. The next is detecting time distance between minimum peak and consecutive maximum peak. This process is repeated until the end of observed block.

Sinusoidal model can be developed to obtain less parameters, so that the signal data rate can be reduced. This model is called as segmental sinusoidal model. By using this model, there are two harmonic signal to estimate the original signal between two consecutive peaks, (maximum to minimum or minimum to maximum). Peak means minimum peak or maximum peak on the frame. Therefore, one segment means part of signal between maximum peak and consecutive minimum peak or part of signal between minimum peak and consecutive maximum peak.

Time distance between i-th maximum peak and consecutive minimum peak called as period information, and denoted by $p_d(i)$. Maximum peak or minimum peak called as peak information, and denoted by $p_k(i)$. Peak information is obtained by detecting the maximum peaks and minimum peaks over the frame observed. Period information is obtained by counting the time distance between the consecutive peaks. The proposed method is a process on time domain. On the extreme waveform coding, signal is fetched on its peaks [8]. On the one frame with length of N, there are M maximum peaks and L minimum peaks. On this frame, there are large number sinusoidal signal components. It can be written as :

$$s(n) = \sum_{k=0}^{K} a_k \cos(\omega_k(n)n + \phi_k(n)) \quad (2)$$

$$\text{for}: 0 \le k \le K\text{-}1,\ n = 0,1,2 \ldots N\text{-}1,\ K \le N$$

The first and second coefficients (k=0 and k=1) are used as components to reconstruct the estimated signal from one maximum peak until the consecutive minimum peak or from one minimum peak until the consecutive maximum peak. This is the equation for estimated signal

$$s(n) = a_0 + a_1 \cos(\omega_1(n) + \phi_1(n)) \quad (3)$$

The estimated signal over the frame is a train of the s_{pv} and s_{vp} for i=0 until i=I-1. based on the previous explanation, for $p_k(0) > p_k(1)$, the reconstructed signal using the segmental sinusoidal model can be written as :

$$s_r(i,n) = \frac{p_k(i) + p_k(i+1)}{2} + \frac{p_k(i) - p_k(i+1)}{2} \cos\left(\frac{\pi.(n - n_k(i))}{p_d(i)} + i.\pi\right) \quad (4)$$

for $i = 0,1,2 \ldots (I\text{-}1)$

If $p_k(0) < p_k(1)$, the reconstructed signal using the segmental sinusoidal model can be written as :

$$s_r(i,n) = \frac{p_k(i) + p_k(i+1)}{2} + \frac{p_k(i) - p_k(i+1)}{2} \cos\left(\frac{\pi.(n - n_k(i))}{p_d(i)} + (i+1).\pi\right) \quad (5)$$

for $i = 0,1,2 \ldots (I\text{-}1)$

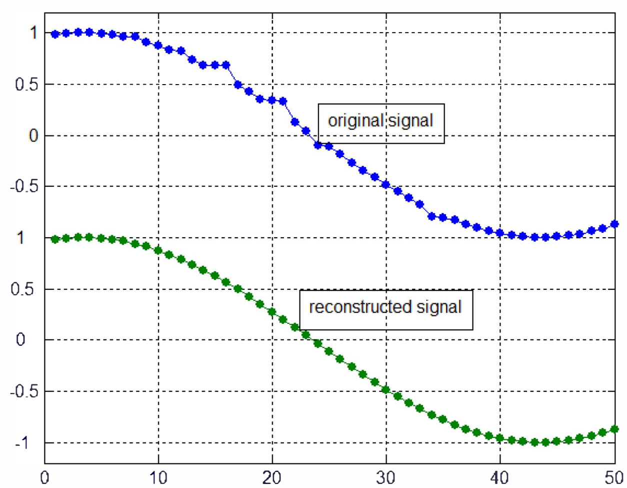

Figure 1. Sinusoidal Segmental Model

III. Period Detection

Location of the peaks of speech signal must be detected to obtain period information have to send to receiver. Simple following program listing, describe how to obtain the peak location. The first, we have to find the first peak. Based on the first peak location, we have to find the next peaks. Distance between peaks is determined as period of this signal.

Speech signal is quantized based on peaks value and distance between two consecutive peaks with segmental sinusoidal model. For reducing size of transmitted signal, 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.

```
for n = 1:m-2
  if sx(n) <= sx(n+1)
    tu = tu + 1;
    if sx(n+1) > sx(n+2)
      pd(j)=tu;              // estimated period
      pk(j)=sx(n+1);         // fetching peak
      j = j + 1;
      tu = 0;
    end
  else
    td = td + 1;
    if sx(n+1) < sx(n+2)
      pd(j)=td;              // estimated period
      pk(j)= sx(n+1);        // fetching peak
      j = j + 1;
      td = 0;
    end
  end
end
```

In period length quantization, codebook is design based on the quantization of period length value of the large number of speech 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. Speech signal is fetched every 30 ms (240 samples) before separated into voiced and unvoiced. This signal will be coded into segmental sinusoidal model with bit allocation is 120 bits for 30 ms in order to obtain the rate of 4000 bit per second. Speech signal can be coded into segmental sinusoidal model for every 120 samples. If the length of pitch period is more than 80 samples, it is applied the 2:1 decimation process. So the maximum number of maximum and minimum peaks is no more than 25. Decimation process also potentially shifted the period information into wrong number.

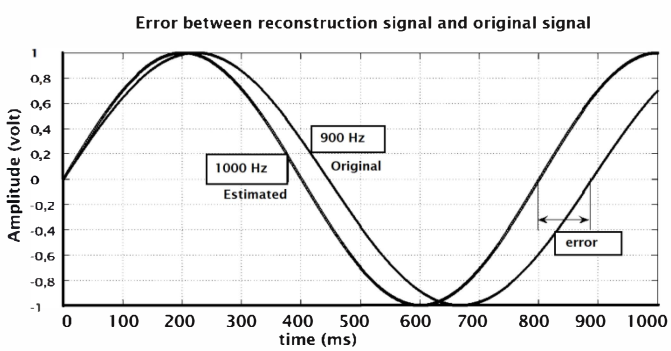

Figure 2. Error between reconstructed signal and original signal

Parameters that have to be sent to the receiver are peaks, periods, pitch, formants, segment, and decimation. In decoder, the coded signal is reconstructed to obtain the good quality of speech signal approximation. The encoder sends speech signal information as a multiplexed parameter. In the decoder, parameters are de-multiplexed to obtain the useful information for signal reconstruction. Parameters would be generated are peaks, periods information, formants information, pitch, number of segment, and decimation. The voiced signal is reconstructed by generating the characteristic signal along the 20 ms segment. The number of characteristic signal along this segment varies between 1.5 until 12, depends on the pitch period. The characteristic signal is generated by using the segmental sinusoidal model from peaks and periods information. The reconstructed signal, especially the voiced signal is passed into post-filter. The post-filter proposed is a train of four adaptive band-pass filters. The center frequency of each filter is changed every 20 ms segment. The center frequencies are detected by encoder, and then they are sent to decoder. Then the comb filter and compensation filter are applied to improve the hearing perception quality [9-11]. Comb filter also changing the estimated period is shifted. Perception quality may be decreased on its frequency accuracy.

Deviation of period information is defined by difference between original signal period and detected period signal on period quantization process.

$$d = T - Tpd \tag{6}$$

$$d\,\% = \frac{T-Tpd}{T}.100\% \tag{7}$$

By using this equation, we obtained deviation or error data of period reconstructed signal as shown on experimental result.

IV. EXPERIMENTAL RESULTS

The experimental result for calculating the period information error for sinusoidal segmental signal shows that the accuracy of period signal is more than 86%. Potential error will happen when the signal period not match with the sampling time. When the frequency is 1000 Hz on sampling frequency at 8000 Hz, distance between existing signal to previous signal with the same value (one period) can be exactly period information (8 point). When the frequency is modified by small number of Hertz, period information still 8 point.

Period information will accurate when the frequency is modified near period of 7 or 9. The period of 7 is represented as 1143 Hz and the period of 9 is represented as 889 Hz. Period information between them frequency tend to give error. On the following table show that period information tend to error when frequency move to the higher or lower than 1000 Hz.

TABLE 1. PERIOD ERROR OF RECONSTRUCTED SIGNAL

frequency	error(%)
880	13.636364
890	12.359551
900	11.111111
910	9.8901099
920	8.6956522
930	7.5268817
940	6.3829787
950	5.2631579
960	4.1666667
970	3.0927835
980	2.0408163
990	1.010101
1000	0
1010	0.990099
1020	1.9607843
1030	2.9126214
1040	3.8461538
1050	4.7619048
1060	5.6603774
1070	6.5420561
1080	7.4074074
1090	8.2568807
1100	9.0909091
1110	9.9099099
1120	10.714286
1130	11.504425
1140	12.280702
1150	13.043478

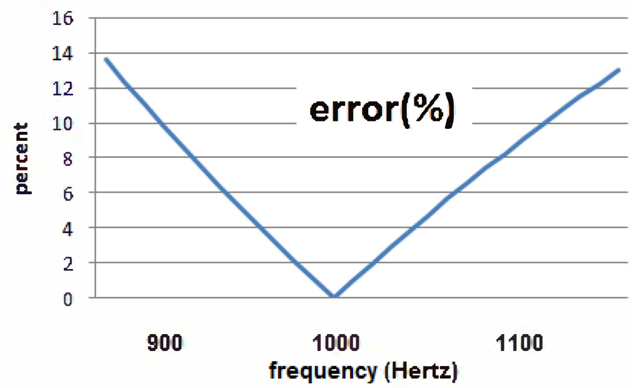

Figure 3. Period error percentage of reconstruction signal compared by the original period signal between 880 Hz and 1150 Hz

For signal with frequency sampling 8000 Hz tends to change its frequency until 13,63%. It can be amend hearing perception for this signal. When the sampling frequency decreased, potential error will be increased. For the lower frequency, potential error will be decreased.

Frequency restoration is needed to give the same perception between original signal on encoder side with the reconstructed signal on decoder side.

V. DISCUSSION

Frequency shifting will make different perception the human hearing system. The reconstructed signal will be heard as the different signal from the original. When the period information is compressed with code book, there will much error on decoder.

Error of period quantization will change on the lower value when the sampling frequency is greather than before. It will be happen because of decreasing the consecutive sampling distance. For the same signal frequency, error will be reduced into 50% for doubling of sampling frequency. If sampling frequency decreased into half than before, error of period quantization is increased until 200%.

VI. CONCLUSION

Segmental sinusoidal signal needs some parameter to obtain simplest form of encoded signal. The most important paramater are peak information and period information. Period information is represented frequency signal. Period information fetched from peak detector location will be used for reconstructed estimating signal. Error between peak location detection depends on the number of sample per second (frequency sampling).

REFERENCES

[1] TF Quatiery, RJ.McAulay, *Speech Transformations Based on a Sinusoidal Representation,* IEEE TASSP, vol. ASSP-34, no. 6, 1986

[2] TF. Quatieri, TE. Hanna, GC. O'Leary, *AM-FM Separation Using Auditory-Motivated Filters,* IEEE TSAP, vol 5, 1997, hal. 465-480.EB. George, MJT. Smith, Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model, IEEE TSAP, vol 5, 1997, hal 389-406.

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

[4] R. Boyer, dan Abed-Meraim, K, *AudioModeling Based on Delayed Sinusoids,* TSAP vol 12, No.2, 2004.

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

[6] FB. Setiawan, S. Soegijoko, Sugihartono, dan S. Tjondronegoro, *A Low Bit Rate Speech Coder using Segmental Sinusoidal Model for Disaster and Emergency Telemedicine*, Journal of eHealth Technology and Application, September 2008

[7] Gottesman, O. dan Gersho, A. (2000) : High Quality Enhanced Waveform Interpolative Coding at 2.8 kbps, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing,* 0-0.

[8] Ridkosil F., *Extreme Waveform Coding*, US Pattent

[9] Gottesman, O. dan Gersho, A. (2001) : Enhanced Waveform Interpolative Coding at Low Bit-Rate, *IEEE Transactions on Speech and Audio Processing*, 9, 1-13.

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

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