Low Bit-Rate Parametric Coding

Djadjat Sudaradjat¹, Andi Rosano²

¹,² Universitas Bina Sarana Informatika
Kramat 98 Jakarta, Indonesia

e-mail: drjadjat.dsj@bsi.ac.id

Abstract - With a low bit rate will be obtained savings in the use of bandwidth transmission channels and memory. In coding the human voice signal, an artificial voice-producing model is made to obtain a low bit rate, which is known as the parametric coding method. With the parametric coding method, the human voice signal can produce a lower bit rate than the waveform coding method. For example, the waveform coding method will be limited by a minimum sampling frequency which according to the Shannon-Nyquist theorem states that so that no information is lost when sampling the signal, the sampling speed must be at least twice the bandwidth of the signal. Thus, the lowest bit rate waveform encoding that can be achieved is by the Delta Modulation system which is 8 Kbit/sec. While the signal parametric coding can achieve a lower bit rate of 2.4 Kbit/sec by the LPC (Linear Predictive Coding) system.

Keywords: LPC, Speech Processing, Speech Compression

INTRODUCTION

The progress of human aids in the form of digital computers which is very extraordinary at this time makes all types of information in analog form must be converted first into digital form so that it can be processed by a digital computer. As in the human voice as a communication tool in the form of analog signals, it can be represented into two types of methods of converting analog form to binary digital form, namely waveform coding and encoding analog signal parameters (Singh, 2015), as shown in Figure 1. Waveform coding is a classic coding method that was discovered since the progress of digital computers has not been as fast as it is today, so that its manufacture still uses analog electrical circuits such as the PCM (Pulse Code Modulation), ADPCM (Adaptive PCM) sistem, and Delta Modulation (Sudaradjat, Teknik Adaptive Pada Modulasi Delta, 2019). However, after the rapid development of digital computers supported by the development of software technology, the analog signal coding system can also be simulated using programming languages.
For example, the human voice signal generated by the human oral cavity can be imitated so as to produce a human voice like the original, namely by imitating the parameters that form human speech, which consists of vocal cords, cavities that change sound vibrations, voice volume, and the type of voice generated. as shown by figure 2 (Djadja Sudaradjat, 2020).

\[ S(z) = E(z).G \cdot \frac{1}{A(z)} \]  
\[ \text{(synthesis model)} \]  
\[ \text{with:} \]  
\[ A(z) = \sum_{i=0}^{p} a(i)z^{-i} \]

\[ (a(i) = \text{coef.filter, } a(0) = 1) \]  
\[ Z^{-1} \]

\[ S(z) \Leftrightarrow s(nT) = s(t) \]  
\[ t = nT \]  

The discrete time signal \( s(nT) \) is the inverse of the z-transformation of \( S(z) \). For the normalized discrete time period \( T=1 \), then:

\[ S(z) \Leftrightarrow s(n) \]

With the same relationship as the signal \( s(n) \) above,

\[ E(z) \Leftrightarrow e(n) \]

is a model of vocal cord output. If the type of sound vibrates, \( e(n) \) will be in the form of a pulse function. If the type is non-vibrating, \( e(n) \) will be a random function. The \( e(n) \) signal after amplified \( G \) is passed into the vibration converter cavity model in the form of an all-pole \( 1/A(z) \) filter to produce an \( s(n) \) signal. From equation (1), the analysis model can be obtained in the form of an equation:

\[ E(z) = 1/G \cdot S(z) \cdot A(z) \]  
\[ \text{(analysis model)} \]  

where \( S(z) \) is the z-transformation of the voice signal \( s(n) \) as the inverse filter input (filter all-zero) \( A(z) \), the output \( E(z) \) will carry information on the parameters of the input voice signal.

**RESEARCH METHODOLOGY**

The process flow diagram for the LPC vocoder sistem simulation is shown in Figure 3. At first, the source voice signal is recorded, then a digital data file is created with the name RECORD.EXE and called the original speech data file. This original speech data can be seen in the time domain named PLOT-DATA, and in the frequency domain named NONPARAMETRIC SPECTRUM. Furthermore, the speech data series is processed with the CODER program to produce parameters that have been encoded into digital form, namely filter coefficients, pitch period, V/UV, and Gain. The coding process is carried out on each frame of the input signal, so that a code for the parameters of the filter coefficient, pitch period, V/UV, and Gain is generated for each frame. The DECODER program is used to reconstruct the speech data again based on the parameters stored in the LPC parameter encoding data file, and the reconstruction results are recorded back in a speech reconstruction data file. To get back the analog output form, PLAYBACK.EXE software is used which will convert the digital data in the speech reconstruction data file into analog data. After
going through the LPF filter and speaker, the reconstructed speech sound can be heard by our ears.

![Diagram](image)

Figure 3. Simulation Process Flowchart.

To test the sound quality, the MOS (Mean Opinion Score) technique is used, which is to calculate the average of the number of speakers, listeners, and other conditions used in the study as shown in table 1.

Table 1. Sound Quality Test Conditions

| Objective test method | Mean Opinion Score (MOS) |
|-----------------------|-------------------------|
| Score                 |                          |
| 5 Good                | Perceptible but not annoying |
| 4 Fair                | Slightly annoying        |
| 3 Poor                | Annoying                |
| 2 Bad                 | Very annoying           |

Table 2. Calculation of bit-rate with a frame length of 30 seconds and a sampling frequency of 8 Khz

| Parameter | Jml bit tiap frame |
|-----------|--------------------|
| k_4       | 6                  |
| k_2       | 4                  |
| k_4 sampai k_4 | 2x3, 2x2, 2x9 |
| k_5 sampai k_10 | 6x2, 6x5, 6x8 |
| Pitch     | 7                  |
| V/UV      | 1                  |
| Gain      | 5                  |
| Sinkronisasi | 1                  |
| Jumlah total bit | 42, 72, 102 |
| Bit-rate (Kbit/det) | 1,4, 2,4, 3,4 |

Table 3. MOS for sentences ekor-ikan-biru

| N    |      |      |
|------|------|------|
|      | 1.4  | 2.4  |
|      | Kbit/det | Kbit/det | Kbit/det |
| 1    | 2    | 2    | 2 |
| 2    | 2    | 2    | 3 |
| 3    | 2    | 3    | 3 |
| 4    | 3    | 3    | 3 |
| 5    | 2    | 2    | 2 |
| 6    | 2    | 3    | 3 |
| 7    | 2    | 2    | 3 |
| 8    | 3    | 2    | 3 |
| 9    | 2    | 2    | 2 |
| 10   | 2    | 3    | 3 |
| 11   | 2    | 2    | 2 |
| 12   | 3    | 2    | 3 |
| 13   | 3    | 3    | 3 |
| 14   | 3    | 4    | 4 |
| 15   | 3    | 4    | 4 |
| 16   | 2    | 4    | 4 |
| 17   | 2    | 4    | 3 |
| 18   | 2    | 4    | 3 |
| 19   | 3    | 4    | 4 |
| 20   | 3    | 4    | 4 |
| 21   | 3    | 4    | 4 |
| 22   | 2    | 3    | 3 |
| 23   | 3    | 4    | 3 |
| 24   | 3    | 4    | 3 |
| 25   | 3    | 2    | 3 |
| 26   | 2    | 3    | 3 |
| 27   | 2    | 3    | 3 |
| 28   | 2    | 3    | 3 |
| 29   | 2    | 2    | 2 |
| 30   | 2    | 3    | 3 |
| 31   | 3    | 3    | 3 |

RESULTS AND DISCUSSION

To test the sound quality on a vocoder system with a certain bit-rate, it is necessary to plan in advance the design of the coding system and the synthetic filter used. In this research, a coding system with bit-rate is designed as shown in table 2, with the design of the synthetic filter used is a filter with a lattice structure. The results of the test for the sound of the ekor-ikan-biru sentence are shown in table 3, the sound of the aneka-sate-ikan sentence in table 4, the sound of the petugas-sekolah sentence in table 5, and the sound of the biru-api sentence in table 6. In general, the results of the quality test The sound system in tables 3, 4, 5, and 6 can be said that the voice signal coding system with the parametric encoding method has an intelligible sound quality and the minimum bit-rate limit that can be used is around 2.4 Kbit/second. In contrast to the waveform coding system, which has almost the same sound quality but with a higher bit-rate, which is 8 Kbit/sec.
Source: (Sudaradjat, Pemrosesan Sinyal Suara Dengan Metoda LPC, 1993)

### Table 4. MOS for sentences aneka-sate-ikan

| N | 1,4 Kbit/det | 2,4 Kbit/det | 3,4 Kbit/det |
|---|--------------|--------------|--------------|
| 1 | 2            | 2            | 3            |
| 2 | 2            | 2            | 3            |
| 3 | 2            | 2            | 3            |
| 4 | 2            | 3            | 3            |
| 5 | 2            | 2            | 2            |
| 6 | 2            | 3            | 3            |
| 7 | 2            | 3            | 3            |
| 8 | 2            | 3            | 3            |
| 9 | 2            | 2            | 3            |
| 10| 2            | 2            | 3            |
| 11| 3            | 2            | 2            |
| 12| 3            | 2            | 3            |
| 13| 3            | 3            | 3            |
| 14| 3            | 4            | 3            |
| 15| 3            | 4            | 4            |
| 16| 2            | 4            | 4            |
| 17| 2            | 4            | 4            |
| 18| 2            | 4            | 3            |
| 19| 2            | 4            | 3            |
| 20| 3            | 4            | 3            |
| 21| 3            | 3            | 3            |
| 22| 3            | 3            | 3            |
| 23| 3            | 4            | 3            |
| 24| 3            | 4            | 4            |
| 25| 3            | 3            | 3            |
| 26| 2            | 3            | 3            |
| 27| 2            | 3            | 3            |
| 28| 2            | 3            | 4            |
| 29| 2            | 3            | 2            |
| 30| 2            | 3            | 3            |
| 31| 3            | 3            | 3            |
| 32| 2            | 3            | 3            |

| \( \bar{X} \) | 2.37 | 2.97 | 3.00 |

Source: (Sudaradjat, Pemrosesan Sinyal Suara Dengan Metoda LPC, 1993)

### Table 5. MOS for sentences petugas-sekolah

| N | 1,4 Kbit/det | 2,4 Kbit/det | 3,4 Kbit/det |
|---|--------------|--------------|--------------|
| 1 | 2            | 2            | 3            |
| 2 | 2            | 2            | 3            |
| 3 | 2            | 2            | 3            |
| 4 | 2            | 2            | 3            |
| 5 | 2            | 2            | 2            |
| 6 | 2            | 3            | 3            |
| 7 | 2            | 3            | 3            |
| 8 | 3            | 3            | 3            |

| \( \bar{X} \) | 2.41 | 3.00 | 3.03 |

### Table 6. MOS for sentences biru-api

| N | 1,4 Kbit/det | 2,4 Kbit/det | 3,4 Kbit/det |
|---|--------------|--------------|--------------|
| 1 | 2            | 2            | 3            |
| 2 | 2            | 3            | 3            |
| 3 | 3            | 3            | 3            |
| 4 | 3            | 3            | 3            |
| 5 | 2            | 2            | 2            |
| 6 | 2            | 2            | 2            |
| 7 | 2            | 2            | 2            |
| 8 | 3            | 3            | 3            |
| 9 | 2            | 2            | 2            |
| 10| 2            | 3            | 3            |
| 11| 2            | 2            | 2            |
| 12| 3            | 2            | 3            |
| 13| 3            | 3            | 3            |
| 14| 3            | 4            | 3            |
| 15| 3            | 4            | 4            |
| 16| 2            | 4            | 4            |
| 17| 2            | 4            | 3            |
| 18| 2            | 3            | 3            |
| 19| 3            | 3            | 3            |
| 20| 3            | 4            | 3            |
| 21| 3            | 4            | 4            |
| 22| 2            | 3            | 3            |
| 23| 3            | 4            | 3            |
| 24| 3            | 4            | 3            |
| 25| 3            | 3            | 3            |
| 26| 2            | 3            | 3            |
| 27| 2            | 3            | 3            |
| 28| 2            | 3            | 4            |
| 29| 2            | 3            | 2            |
| 30| 2            | 3            | 3            |
| 31| 3            | 3            | 3            |
| 32| 2            | 3            | 3            |

| \( \bar{X} \) | 2.41 | 3.00 | 3.00 |

Source: (Sudaradjat, Pemrosesan Sinyal Suara Dengan Metoda LPC, 1993)
CONCLUSION

Parametric encoding method on human voice signal can produce lower bit-rate (2.4 Kbit/sec) than waveform encoding method (8 Kbit/sec). With digital technology, it is possible to create an imitation model of a human speech sound-producing system. In order to obtain a lower bite-rate without compromising sound quality, it can be further investigated by using mock models that allow for use such as the selection of adaptive frame length, the type of digital filter used, or a more precise pitch algorithm.

REFERENCES

Adam J. Chong, 1. M. (2020). Effects of consonantal constrictions on voice quality. The Journal of the Acoustical Society of America 148, EL65 (2020); doi: 10.1121/10.0001585, 65-71.

Djadjat Sudaradjat, S. A. (2020). Aplikasi Pengolahan Sinyal Suara pada Teknologi Kecerdasan Buatan. INSANtek – Jurnal Inovasi dan Sains Teknik Elektro, 88-95.

Gibson, J. (2019). Mutual Information, the Linear Prediction Model, and CELP Voice Coders. Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2019 doi:10.20944/preprints201904.0184.v1, 1-14.

Gibson, J. D. (2016). Speech Compression. Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93118, USA.: Information 2016, 7, 32; doi:10.3390/info7020032.

Hamid, O. K. (2017). Speech Sound Coding Using Linear Predictive. Journal of Information, Communication, and Intelligence Systems (JICIS) Volume 3, Issue 1, May 2017, 13-17.

Hernando Castaneda Mari, M. J. (2018). Theory, Algorithms, Implementation and Practice of Power Density Signal by Autocorrelation Modeling. https://www.researchgate.net/publication/325019798, 1-20.

Jurafsky. (2018). Speech and Language Processing. Third Edition Draft.