turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Linggo Sumarno

Assignment title: Periksa similarity

Submission title: Chord Recognition using Segment Averaging Feature Extract...

File name: Recognition_using_Segment_Averaging_Feature_Extraction_w...

File size: 1.5M

Page count: 12

Word count: 5,344

Character count: 25,322

Submission date: 30-Jun-2022 09:47AM (UTC+0700)

Submission ID: 1864832408

Chord Recognition using Segment Averaging Feature Extraction with Simplified Harmonic Product Spectrum and Logarithmic Scaling

Linggo Sumarno

Electrical Engineering Study Program, Sanata Dharma University, Yogyakarta, Indonesia
lingsum@usd.ac.id

Abstract: This paper proposes a feature extraction method for a chord recognition, which gives a fewer number of feature extraction coefficients than the previous works. The method of the proposed feature extraction is segment averaging with SHPS (Simplified Harmonic Product Spectrum) and logarithmic scaling. The chords used in developing the proposed feature extraction were guitar chords. In a more detail, the method of the proposed feature extraction basically is as follows. Firstly, the input signal is transformed using FFT (Fast Fourier Transform). Secondly, the left portion of the transformed signal is then processed in succession using SHPS, logarithmic scaling, and segment averaging. The output of segment averaging is the result of the proposed feature extraction. Based on the test results, the proposed feature extraction is quite efficient for use in chord recognition, since it requires only at least eight coefficients to represent each chord.

Keywords: chord recognition, feature extraction, segment averaging, Simplified Harmonic Product Spectrum, logarithmic scaling

1. Introduction

A chord is two or more tones that played at the same time. A chord is named based on the origin of the tones being played. For example, if the tones are coming from the first, third, and fifth notes on the major scale, then the chord is included in the major chord. There are many variations of the chords that depend on the scale and sequence of notes on the scale [1].

The chords are usually used to accompany the song. The existence of these chords will make the song more alive. In order to recognize the chords being played, someone with a good musical ability can recognize it easily. However, it will be difficult to do by someone who does not have a good musical ability.

Today the computer can be programmed to imitate someone with good musical ability, in order to recognize chords. Based on the previous works, a chroma-based chord recognition is a popular approach. A feature extraction based on the chroma feature is the Pitch Class Profile (PCP). Basically PCP introduced by Fujishima [2] produces a 12-dimensional vector for each chord. In that vector, each vector dimension represents the power of each pitch. The PCP of Fujishima is quite popular today, as it is still used in recent works [3] [4].

In addition to the original PCP feature extraction above, there are also a number of derivatives of PCP feature extraction, such as Improved Chromagram [5], Improved PCP [6], and CRP (Chroma DCT-Reduced log Pitch) [7]. A number of derivatives of PCP feature extraction, also produce 12-dimensional vectors for each chord.

Hartquist [1] has shown that, the signal from the guitar chord has many significant local peaks in the discrete Fourier transform domain. In addition, Sumarno [8] and Sumarno [9] also have shown that, the signal from a pianica tone also has many significant local peaks, in the discrete cosine transform domain and discrete Fourier transforms domain, respectively. In order to perform feature extraction on a signal which has many significant local peaks, Sumarno [8] and Sumarno [9] have used the segment averaging feature extraction which can produce eight and four feature extraction coefficients for each tone, respectively.

This paper proposes a feature extraction method, which has a fewer number of feature extraction coefficients than the previous works. Specifically, this paper proposes a variant of the segment averaging feature extraction, that previously used in the tone recognition [8] [9], to be

753

Chord Recognition using Segment Averaging Feature Extraction with

by Sumarno Linggo

Chord Recognition using Segment Averaging Feature Extraction with Simplified Harmonic Product Spectrum and Logarithmic Scaling

Linggo Sumarno

Electrical Engineering Study Program, Sanata Dharma University, Yogyakarta, Indonesia
lingsum@usd.ac.id

Abstract: This paper proposes a feature extraction method for a chord recognition, which gives a fewer number of feature extraction coefficients than the previous works. The method of the proposed feature extraction is segment averaging with SHPS (Simplified Harmonic Product Spectrum) and logarithmic scaling. The chords used in developing the proposed feature extraction were guitar chords. In a more detail, the method of the proposed feature extraction basically is as follows. Firstly, the input signal is transformed using FFT (Fast Fourier Transform). Secondly, the left portion of the transformed signal is then processed in succession using SHPS, logarithmic scaling, and segment averaging. The output of segment averaging is the result of the proposed feature extraction. Based on the test results, the proposed feature extraction is quite efficient for use in chord recognition, since it requires only at least eight coefficients to represent each chord.

Keywords: chord recognition, feature extraction, segment averaging, Simplified Harmonic Product Spectrum, logarithmic scaling

1. Introduction
A chord is two or more tones that played at the same time. A chord is named based on the origin of the tones being played. For example, if the tones are coming from the first, third, and fifth notes on the major scale, then the chord is included in the major chord. There are many variations of the chords that depend on the scale and sequence of notes on the scale [1]. The chords are usually used to accompany the song. The existence of these chords will make the song more alive. In order to recognize the chords being played, someone with a good musical ability can recognize it easily. However, it will be difficult to do by someone who does not have a good musical ability.

Today the computer can be programmed to imitate someone with good musical ability, in order to recognize chords. Based on the previous works, a chroma-based chord recognition is a popular approach. A feature extraction based on the chroma feature is the Pitch Class Profile (PCP). Basically PCP introduced by Fujishima [2] produces a 12-dimensional vector for each chord. In that vector, each vector dimension represents the power of each pitch. The PCP of Fujishima is quite popular today, as it is still used in recent works [3] [4].

In addition to the original PCP feature extraction above, there are also a number of derivatives of PCP feature extraction, such as Improved Chromagram [5], Improved PCP [6], and CRP (Chroma DCT-Reduced log Pitch) [7]. A number of derivatives of PCP feature extraction, also produce 12-dimensional vectors for each chord.

Hartquist [1] has shown that, the signal from the guitar chord has many significant local peaks in the discrete Fourier transform domain. In addition, Sumarno [8] and Sumarno [9] also have shown that, the signal from a pianica tone also has many significant local peaks, in the discrete cosine transform domain and discrete Fourier transforms domain, respectively. In order to perform feature extraction on a signal which has many significant local peaks, Sumarno [8] and Sumarno [9] have used the segment averaging feature extraction which can produce eight and four feature extraction coefficients for each tone, respectively.

This paper proposes a feature extraction method, which has a fewer number of feature extraction coefficients than the previous works. Specifically, this paper proposes a variant of the segment averaging feature extraction, that previously used in the tone recognition [8] [9], to be

suitable for the chord recognition, especially the guitar chords.

2. Methodology

A. *System development*

The development of the chord recognition system is shown in Figure 1. The input is a chord signal in wav format, whereas the output is a text that indicates a recognized chord. Here is the explanation of the blocks shown in Figure 1.
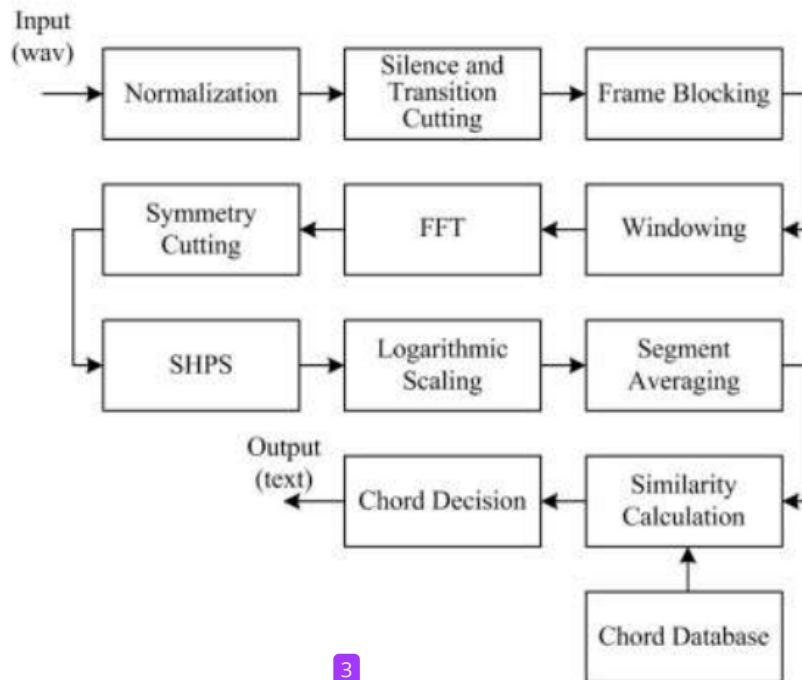
Figure 1. The block diagram of the developed chord recognition system.

A.1. *The input*

The input of the chord recognition system is a recording of guitar chords in isolated wav format. The guitar is an acoustic-electric guitar Yamaha CPX 500-II (see Figure 2). The chords are the major chords C, D, E, F, G, A, and B. Furthermore, in this paper, when the chords are mentioned, they are the major chords.

The input of the chord recognition system were obtained by recording the above chords with a sampling frequency of 5000 Hz. The magnitude of sampling frequency was chosen by considering Shannon's sampling theorem and previous works using segment averaging feature extraction [8] [9]. Based on the Shannon's sampling theorem, the 5000 Hz of sampling frequency has already exceeded twice the highest fundamental frequency 392 Hz (tone G4) from chord G. Based on the evaluation results, the duration of the recording for 2 seconds was sufficient to obtain the steady state region for frame blocking purposes.

Figure 2. The guitar used in this work.

A.2. Normalization

Normalization is a process of setting the maximum value of the data of the recording signal to be 1 or -1. This normalization aims to eliminate the differences in the maximum values of a number of recording data signal sequence.

A.3. Silence and transition cutting

Silence and transition cutting is a process of cutting the silence and transition regions that are on the left side of the recording signal. Based on the observations, the silence region can be removed by applying an amplitude threshold value of |0.5|. It starts from the far left region of the recording signal, if the signal amplitude is less than |0.5| then the signal is cut off. After cutting the silence region, based on the observations also, the transition region can be removed by cutting as long as 200 milliseconds of the left portion of the signal.

A.4. Frame blocking

Frame blocking is a process of getting a frame of data signal from a long data signal [10]. The purpose of using frame blocking is to reduce the number of data signal to be processed. In this work, the width of the frame blocking length is the same as the length of FFT in the FFT process.

A.5. Windowing

Windowing is a process of reducing the discontinuities that appear at the edges of the signal [10]. This reduction is necessary to reduce the emergence of harmonic signals that appear after the FFT process. In this work, the window used in windowing is the Hamming window [11]. This window is a window that is widely used in the field of digital signal processing [12]. In this work, the width of the window is the same as the length of FFT in the FFT process.

A.6. FFT (Fast Fourier Transform)

FFT is a process for transforming signals from the time domain to the Fourier transform domain. This work used FFT radix-2. This kind of FFT is widely used in the field of signal processing [12]. The length of the FFT in this process is 2^n ($n = 0, 1, 2, \ldots$), since this work used FFT radix-2.

A.7. Symmetry cutting

Symmetry cutting is a process for cutting the half portion the FFT result. This cutting is necessary because between the left and the right half portion of the FFT result shows a symmetry property. Therefore, if using only the left or right half portion of the FFT result, it is sufficient. In this work, the left half portion of the FFT is used.

A.8. SHPS (Simplified Harmonic Product Spectrum)

SHPS is a process for eliminating harmonic signals. This paper proposes this SHPS that derived from HPS (Harmonic Product Spectrum) introduced by Noll [13]. Based on the observations, this SHPS could clarify the difference between a chord with the other one. By algorithm, this SHPS is as follows.

SHPS (Simplified Harmonic Product Spectrum) algorithm

1. Consider a sequence $x(k)$={$x(0), x(1), \ldots, x(N\text{-}1)$} with $N = 2^p$ and $p \geq 0$.
2. Do downsampling for $x(k)$ in order to get
   $x_d(k)$={$x(0), x(2), \ldots, x(N\text{-}2)$}
3. Do zero padding for $x_d(k)$ in order to get
   $x_z(k) = \{x(0), x(2), \ldots, x(N\text{-}2), z(0), z(1), \ldots, z((N/2)\text{-}1)\}$

755

where

$z(0) = z(1) = \dots = z((N/2)\text{-}1) = 0$

4. Do element multiplication $x(k)$ and $x_z(k)$ in order to get

$x_m(k) = x(k) . x_z(k)$ (1)

5. The result of SHPS $y(k)$ is the left half of $x_m(k)$

$y(k) = \{ x_m(0), x_m(1), \dots, x_m((N/2)\text{-}1)\}$

A.9. Logarithmic scaling

Logarithmic scaling is a process for reducing the difference in the peak values in a signal data sequence. The effect of this reduction is the increasing of the number of significant local peaks. Based on the previous works [8] [9], feature extraction using segment averaging, showed its superiority for a data signal sequence which has many significant local peaks. Mathematically, the logarithmic scaling is formulated below.

$\mathbf{y}_{out} = \log(\alpha \mathbf{y}_{in} + 1)$ (2)

where y_{in} and y_{out} are input data vector and output data vector, and α is a logarithmic scale factor. The addition of '1' to the above formula is to avoid the infinite logarithmic results when there is a zero y_{in} vector element.

A.10. Segment averaging

Segment averaging as inspired from Setiawan [14], is a process for reducing the signal data sequence. Basically, the result of this reduction still shows the basic form of the original signal data sequence. By algorithm, this segment averaging is as follows.

Segment averaging algorithm

1. Consider a sequence $y(k) = \{y(0), y(1), \dots, y(N\text{-}1)\}$ where $N = 2^p$ and $p \geq 0$.

2. Determine the segment length L with $L = 2^q$ for $0 \leq q \leq p$.

3. Divide the sequence $y(k)$ by using the segment length L. Thus a number of M segments will be generated as follows

$M = \dfrac{N}{L}$ (3)

and also the sequence $g(u) = \{g(1), g(2), \dots, g(L)\}$ in each segment.

4. Calculate the average value in each segment $h(v)$ as follows

$h(v) = \dfrac{1}{L}\sum_{u=1}^{L} g_v(u), \quad 1 \leq v \leq M$ (4)

A.11. Similarity calculation

Similarity calculation is basically a process for comparing the feature extraction of the input signal with a number of feature extraction of chord signals stored in a chord database. This similarity indicates that the chord recognition system in this work using a template matching method [15] [16]. This work used cosine similarity since it is a popular similarity measure [17]. Cosine similarity is formulated as follows.

$S(\mathbf{x}, \mathbf{y}) = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$ (5)

where x and y are two equal length vectors, and n is the length of the vectors x and y. In a pattern recognition system using a template matching method, one of the vectors (x or y) is a vector to be searched for its pattern class, while the other vector is a vector stored in a pattern class database.

A.12. Chord decision

Chord decision is a process to determine an output chord corresponding to the input signal. The first step of determining the output chord is to find a maximum value from some similarity values. These values are the result of the similarity calculations between the feature extraction of the input signal with a number of the chord feature extractions present in the chord database. The second step is to determine an output chord. A chord associated with one of the chord feature extractions present in the chord database, which has the highest similarity, will be determined as the output chord.

B. Chord database

Chord database shown in Figure 1 is generated using the proposed chord feature extraction shown in Figure 3. Based on Figure 3, the input is a chord sample in the wav format, while the output is the feature extraction of the chord sample.
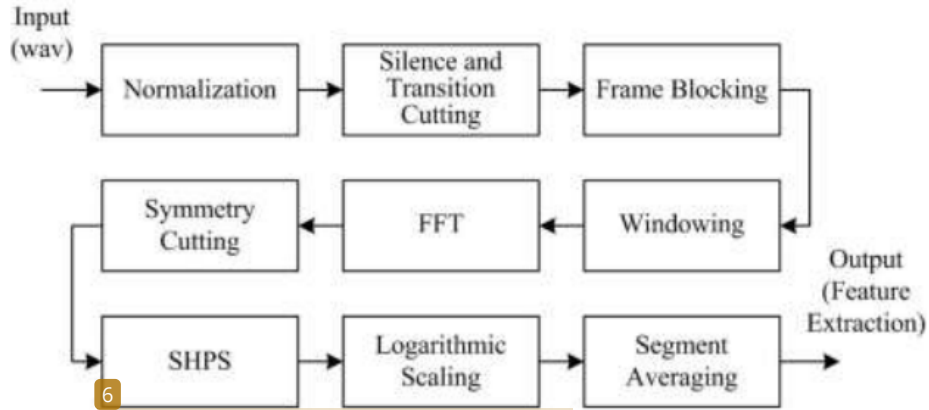

Figure 3. The block diagram of the proposed chord feature extraction.

In this work, 10 samples were taken for each of the chords C, D, E, F, G, A, and B. The assumptions used for the collection of 10 samples were, with a total of 10 samples, all the signal variations of each chord were obtained. The results of the 10 samples were 10 feature extraction results for each chord. Furthermore, the 10 feature extraction results are averaged as follows.

$$\mathbf{Z}_T = \frac{1}{10}\sum_{i=1}^{10} T_i \tag{6}$$

where the vector $\{T_i \mid 1 \leq i \leq 10\}$ are 10 feature extraction results, and the vectors $\{Z_T \mid T = C, D, E, F, G, A, \text{and } B\}$ are the seven chord vectors stored in a chord database. Note that a single chord database is generated from a single frame blocking value, a single segment length value, and a single logarithmic scale factor.

C. Test chords

A number of test chords are used to test the developed chord recognition system. For each frame blocking length, each segment length, and each logarithmic scale factor used, the developed recognition system was tested using 140 recorded chords. These chords were came from the seven chords C, D, E, F, G, A, and B, where each chord had been recorded 20 times.

D. Recognition rate

The recognition rate is used to measure the performance of the developed chord recognition system. Mathematically the recognition rate is formulated below.

$$\text{Recognition rate} = \frac{\text{Number of recognized chords}}{\text{Number of test chords}} \times 100\% \tag{7}$$

As described above, this recognition rate is calculated for each frame blocking length, each

segment length, and each the logarithmic scale factor used. The number of test chords in this work was 140 recorded chords.

3. Results and discussion
A. Test results

The developed chord recognition system (see Figure 1), was tested under two conditions, namely without SHPS and with SHPS. The test results are shown in Table 1 and Table 2 respectively. In each of these tables, the tests were performed for various frame blocking length, segment length, and logarithmic scale factor (α).

Table 1. Results of chord recognition test without using SHPS.
Results shown: Recognition rate (%).

Frame blocking length (points)	Segment length (points)							
	2	4	8	16	32	64	128	256
(a) $\alpha = 1$								
64	78.57	60.71	57.86	23.57	14.29	–	–	–
128	100	94.43	77.86	65.71	26.43	14.29	–	–
256	100	100	98.57	75.00	64.29	32.86	14.29	–
512	100	100	100	100	77.14	69.29	37.86	14.29
1024	100	100	100	100	97.86	80.71	65.00	32.14
(b) $\alpha = 50$								
64	84.29	65.00	59.29	25.71	14.29	–	–	–
128	100	94.29	82.14	71.43	25.71	14.29	–	–
256	100	100	100	83.57	65.00	35.71	14.29	–
512	100	100	100	100	82.14	70.00	33.57	14.29
1024	100	100	100	100	97.86	83.57	69.29	37.14
(c) $\alpha = 100$								
64	85.00	65.71	59.29	28.57	14.29	–	–	–
128	100	95.71	79.29	71.43	33.57	14.29	–	–
256	100	100	100	83.57	65.71	35.71	14.29	–
512	100	100	100	100	82.14	66.42	31.43	14.29
1024	100	100	100	100	97.86	83.57	67.86	36.43
(d) $\alpha = 500$								
64	85.71	62.86	57.14	29.29	14.29	–	–	–
128	100	93.57	76.43	69.29	30.00	14.29	–	–
256	100	100	100	82.14	66.43	32.86	14.29	–
512	100	100	100	100	82.14	63.57	28.57	14.29
1024	100	100	100	100	95.71	82.86	67.86	37.14
(e) $\alpha = 1000$								
64	85.00	63.57	57.86	30.00	14.29	–	–	–
128	100	93.57	77.14	68.57	30.00	14.29	–	–
256	100	100	100	79.29	65.71	31.43	14.29	–
512	100	100	100	100	82.86	63.57	29.29	14.29
1024	100	100	100	100	96.43	82.14	70.00	37.14
(f) $\alpha = 5000$								
64	83.57	60.00	57.86	30.71	14.29	–	–	–
128	100	90.71	75.71	67.14	31.43	14.29	–	–
256	100	100	100	79.29	65.71	30.71	14.29	–
512	100	100	100	100	81.43	61.43	28.57	14.29
1024	100	100	100	100	95.71	82.86	69.29	35.00

Table 2. Results of chord recognition test using SHPS.
Results shown: Recognition rate (%).

Frame blocking length (points)	Segment length (points)							
	2	4	8	16	32	64	128	256
(a) α = 1								
64	67.86	65.71	37.14	14.29	–	–	–	–
128	96.43	92.86	57.85	42.85	14.29	–	–	–
256	97.14	93.57	88.57	48.57	45.71	14.29	–	–
512	100	100	98.57	92.14	58.57	41.42	14.29	–
1024	100	100	100	100	96.43	57.14	34.29	14.29
(b) α = 50								
64	64.29	52.86	44.29	14.29	–	–	–	–
128	96.43	93.57	63.57	44.29	14.29	–	–	–
256	100	100	100	68.57	42.14	14.29	–	–
512	100	100	100	95.71	61.43	39.29	14.29	–
1024	100	100	100	100	91.43	64.28	45.00	14.29
(c) α = 100								
64	62.86	50.71	43.57	14.29	–	–	–	–
128	96.43	92.86	65.00	42.14	14.29	–	–	–
256	100	100	100	68.57	37.14	14.29	–	–
512	100	100	100	93.57	64.29	36.42	14.29	–
1024	100	100	100	100	87.86	63.57	45.71	14.29
(d) α = 500								
64	59.29	53.57	39.29	14.29	–	–	–	–
128	96.42	90.00	69.29	42.86	14.29	–	–	–
256	100	100	100	70.00	37.86	14.29	–	–
512	100	100	100	92.14	65.71	39.29	14.29	–
1024	100	100	100	100	80.71	66.43	46.43	14.29
(e) α = 1000								
64	60.71	55.71	37.86	14.29	–	–	–	–
128	96.43	87.86	65.71	45.00	14.29	–	–	–
256	100	100	100	70.71	40.00	14.29	–	–
512	100	100	100	92.86	64.29	40.00	14.29	–
1024	100	100	100	100	80.71	66.43	42.86	14.29
(f) α = 5000								
64	60.00	57.86	33.57	14.29	–	–	–	–
128	94.42	85.00	65.00	42.14	14.29	–	–	–
256	100	100	98.57	67.86	37.14	14.29	–	–
512	100	100	100	92.86	63.57	41.43	14.29	–
1024	100	100	100	100	80.71	66.43	43.57	14.29

B. Discussion

Viewed from the use of SHPS point of view, Table 1, Table 2, and Table 3 show that for the same number of feature extraction coefficients, the recognition rates in Table 2 are generally higher than in Table 1. This case can be explained by using Figure 4. As shown in Figure 4, the use of SHPS causes two things to be observed. The first thing is the decrease of cosine similarity value between C and D chords. While the second thing is the increase of variance average value of C and D chords.

759

Table 3. The number of feature extraction coefficients associated with Table 1 and Table 2. Results shown: The number of feature extraction coefficients (coefficients).

Frame blocking length (points)	Segment length (points)							
	2	4	8	16	32	64	128	256
(a) Associated with Table 1								
64	16	8	4	2	1	–	–	–
128	32	16	8	4	2	1	–	–
256	64	32	16	8	4	2	1	–
512	128	64	32	16	8	4	2	1
1024	256	128	64	32	16	8	4	2
(b) Associated with Table 2								
64	8	4	2	1	–	–	–	–
128	16	8	4	2	1	–	–	–
256	32	16	8	4	2	1	–	–
512	64	32	16	8	4	2	1	–
1024	32	64	32	16	8	4	2	1
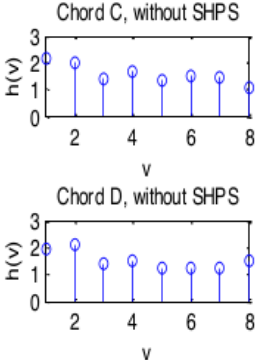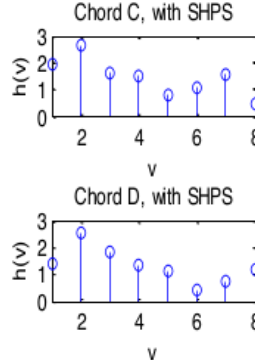
SHPS usage	Without SHPS	With SHPS
Feature extraction result		
Cosine similarity between C and D chords	0.994	0.963
Variance average of C and D chords	0.133	0.538
Notes:		
Frame blocking length (points)	256	256
Segment length (points)	16	8
Logarithmic scale factor (α)	500	500

Figure 4. The example of SHPS influences on feature extraction, cosine similarity, and variance average, for C and D chords.

The decrease of cosine similarity value above indicates that the feature extraction of C and D chords becomes more dissimilar. This means that the feature extraction of C and D chords becomes more distinct. In general, if one feature extraction with the other feature extraction becomes more distinct, it means the discrimination level of feature extraction is increasing. Furthermore, in general the increasing of this discrimination level will finally increase the recognition rate.

The increase in variance average value above indicates a greater variation in the values of feature extraction result. The existence of this greater variation (see the feature extraction of C

and D chords in Figure 4), will cause the feature extraction of C and D chords becomes more distinct. As discussed above, in general this one will finally increase the recognition rate.

In general, based on the above discussion, the use of SHPS will cause two things that can be observed: the decrease in cosine similarity value and the increase in variance average value. As discussed above, in general, they will finally increase the recognition rate.

Viewed from the frame blocking length point of view, Table 2 shows, if the frame blocking length is getting longer, it will further increase the recognition rate. The cause of this incident can be described as follows. If the frame blocking length is getting longer, it will increase the number of the feature extraction coefficients. This increasing number of the feature extraction coefficients will make one feature extraction and other feature extraction becomes more distinct. As discussed above, in general, this one will finally increase the recognition rate.

Viewed from the segment length point of view, Table 2 shows, if the segment length is getting longer, it will further decrease the recognition rate. The cause of this incident can be described as follows. If the segment length is getting longer, it will decrease the number of feature extraction coefficients. This decreasing number of the feature extraction coefficients will make one feature extraction and other feature extraction becomes more indistinct. If one feature extraction with the other feature extraction becomes more indistinct, it means the discrimination level of the feature extraction is decreasing. Furthermore, if the discrimination level of the feature extraction is decreasing, it will finally decrease the recognition rate.

Viewed from the logarithmic scale factor point of view, Table 2 shows two cases. The first case is, if the logarithmic scale factor is increased up to 50, then the recognition rate will increase. The second case is, if the logarithmic scale factor is increased above 1000, then the recognition rate will decrease.
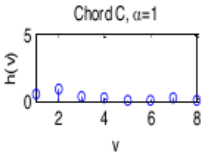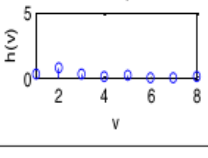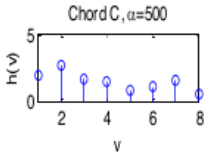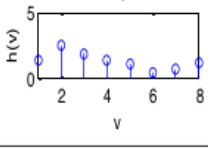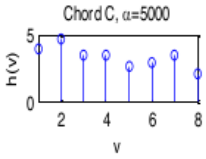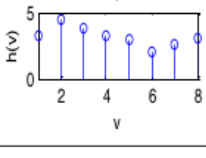
Logarithmic scale factor (α)	1	500	5000
Feature extraction result			
Cosine similarity between C and D chords	0.947	0.963	0.980
Variance average of C and D chords	0.069	0.538	0.585
Notes:			
Frame blocking length (points)	256	256	256
Segment length (points)	8	8	8

Figure 5. The example of logarithmic scale factor (α) influences on feature extraction, cosine similarity, and variance average, for C and D chords.

In the first case, as shown in Figure 5, the increase of the logarithmic scale factor, causing the increasing of cosine similarity value and also the increasing of variance average value. The increasing of cosine similarity value indicates that the feature extraction of C and D chords becomes more similar. This means that the feature extraction of C and D chords becomes more indistinct. In general, if one feature extraction with the other feature extraction becomes more indistinct, it means the discrimination level of the feature extraction is decreasing. Furthermore, if the discrimination level of the feature extraction is decreasing, it will finally decrease the recognition rate.

The increasing of variance average value in the first case above indicates a greater variation in the value of the feature extraction of C and D chords. This greater variation in the value will cause the feature extraction of C and D chords becomes more distinct. As discussed above, in general this one will finally increase the recognition rate. Furthermore, since the final result of the first case is the increase of the recognition rate, so in this first case, it can be said that the increase of variance average value has more dominant influence, so that it can increase the recognition rate.

In the second case, as shown in Figure 5, the increase of the logarithmic scale factor also causes the increase of cosine similarity value and also the increase of variance average value. As discussed above, the increase of cosine similarity value, will finally decrease the recognition rate. In addition, as discussed above also, the increase of variance average value, will finally increase the recognition rate. Since the final result of the second case is the decrease of the recognition rate, so in this second case, it can be said that the increase of cosine similarity value has a more dominant influence, so that it can decrease the recognition rate.

In general, based on the above discussion, the use of the logarithmic scale factor will lead to the increase in cosine similarity value and the increase in variance average value. When it is observed in a more detail, in the first case, the variance average value increases significantly, while in the second case it increases insignificantly. Therefore, it can be said that in the first case, the significant increase of variance average value has a more dominant influence in increasing the recognition rate, than the increase of cosine similarity value. In contrast, in the second case, the increase of cosine similarity value has a more dominant influence in reducing the recognition rate, than the insignificant increase of variance average value.

C. Best feature extraction result

Based on Table 2, in order to obtain the highest recognition rate, up to 100%, it is required the use of frame blocking length 256 points, segment length 8 points, and logarithmic scale factor 50-1000. By using that frame blocking length and segment length, it will be obtained a chord feature extraction that has eight coefficients.

D. Comparison with the other feature extractions

Table 4 compares the performance of some feature extractions for chord recognition. As seen in Table 4, the proposed feature extraction in this work is quite efficient. This is due to the proposed feature extraction can give a number of feature extraction coefficients that is only two-third of the previous works of feature extractions.

Table 4. The performance comparison of some feature extraction for chord recognition.

Feature Extraction	Number of Feature Extraction Coefficients	Recognition Rate (%)	Test Chords
Improved Chromagram [5]	12	94	1440 test chords from 360 recorded guitar chords
Improved PCP [6]	12	95.83	192 test chords from 192 generated guitar chords
CRP (Chroma DCT-Reduced log Pitch) Enhanced PCP [7]	12	99.96	4608 test chords from 576 generated guitar chords
Segment averaging with SHPS and logarithmic scaling (this work)	8	100	140 test chords from 140 recorded guitar chords

4. Conclusion and future work

In this paper, it has been presented a chord recognition system using a proposed feature extraction called segment averaging with SHPS and logarithmic scaling. The best result of the chord recognition system using this proposed feature extraction, can be obtained by using frame blocking length 256 points, segment length 8 points, and logarithmic scale factor 50-1000. By using that frame blocking length and segment length, it will be obtained a chord feature extraction which has eight coefficients.

The chord recognition system in this work using a template matching method. In the future work, it can be studied the performance of the proposed feature extraction, if the chord recognition system using other than template matching method.

5. Acknowledgement

6. References

[1] Hartquist J., *Real Time Musical Analysis of Polyphonic Guitar Audio*, Master Thesis, California Polytechnic State University, 2012.

[2] Fujishima T., "Realtime chord recognition of musical sound: a system using Common Lisp Music", *Proceeding of the International Computer Music Conference (ICMC)*, Beijing, pp. 464–467, 1999.

[3] Muludi K., Aristoteles, Loupatty A.F.S, "Chord Identification Using Pitch Class Profile Method with Fast Fourier Transform Feature Extraction", *International Journal of Computer Science Issues*, Vol. 11, Issue 3, No. 1, pp. 139-144, 2014.

[4] Gaonkar P., Varma S., Nikhare R., "A Survey on Content-Based Audio Retrieval Using Chord Progression", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4, No. 1, pp. 629-636, 2016.

[5] Stark A.M., Plumbey M.D., "Real-Time Chord Recognition for Live Performance", *Proceedings of The International Computer Music Conference, ICMC 09*, pp. 85-88, 2009.

[6] Ma K., *Automatic Chord Recognition*, Department of Computer Sciences, University of Wisconsin-Madison, http://pages.cs.wisc.edu/~kma/projects.html, (May 2016).

[7] Rajparkur P., Girardeau B., Migimatsu T., "A Supervised Approach to Musical Chord Recognition", *Stanford Undergraduate Research Journal*, Vol. 15, pp. 36-40, 2015.

[8] Sumarno L., "On The Performance of Segment Averaging of Discrete Cosine Transform Coefficients on Musical Instruments Tone Recognition", *ARPN Journal of Engineering and Applied Sciences*, Vol. 11, No. 9, pp. 5644-5649, 2016.

[9] Sumarno L., Iswanjono, "Feature Extraction of Musical Instrument Tones using FFT and Segment Averaging", *Telkomnika*, Vol. 15, No. 3, pp. 1280-1289, 2017.

[10] Meseguer N.A., *Speech Analysis for Automatic Speech Recognition*, MSc Thesis, Norwegian University of Science and Technology (NTNU), Trondheim, 2009.

[11] Harris F.J. "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform". *Proceedings of the IEEE*, Vol. 66., No. 1, pp. 51-83, 1978.

[12] Jenkins W.K., "Fourier Methods for Signal Analysis and Processing", *The Digital Signal Processing Handbook 2nd Ed: Digital Signal Processing Fundamentals*. Madisetti V. K. (ed), CRC Press, Boca Raton, pp. 1-1 - 1-29, 2010.

[13] Noll A.M., "Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum and a Maximum Likelihood Estimate", *Proceedings of the Symposium on Computer Processing in Communications*, Vol. 19, Polytechnic Press, Brooklyn, New York, pp. 779-797, 1970.

[14] Setiawan Y.R., *Numbers Speech Recognition using Fast Fourier Transform and Cosine Similarity (in Indonesian)*, Undergraduate Thesis, Sanata Dharma University, Yogyakarta, pp. 66-70, 2015.

[15] Jain A.K., Duin R.P.W., Mao J., "Statistical Pattern Recognition: A Review", *IEEE Transactions and Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, pp. 4-37, 2000.

[16] Theodoridis S., Koutroumbas K., *Pattern Recognition*, 4th Edition, Elsevier Inc., San Diego, California, pp. 481-519, 2009.

[17] Zhu S., Wu J., Xiong H., Xia G., "Scaling up top-K similarity search", *Data and Knowledge Engineering*, Vo. 70, pp. 60-83, 2011.

Linggo Sumarno received doctorate degree from Electrical Engineering Study Program, Gadjah Mada University, Indonesia, in 2010. He is currently a lecturer in Electrical Engineering Study Program, Sanata Dharma University, Indonesia. His research interests include signal processing, image processing, and pattern recognition

Chord Recognition using Segment Averaging Feature Extraction with

8	Internet Source	<1 %

9	Jesus Guerrero-Turrubiates, Sergio Ledesma, Sheila Gonzalez-Reyna, Gabriel Avina-Cervantes. "Guitar Chords Classification Using Uncertainty Measurements of Frequency Bins", Mathematical Problems in Engineering, 2015 Publication	<1 %

10	www.easychordz.com Internet Source	<1 %

11	www.nextlevelguitar.com Internet Source	<1 %

12	www.hindawi.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		