TELKOMNIKA, Vol.15, No.3, September 2017, pp. 1280~1289
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v15i3.3381

Feature Extraction of Musical Instrument Tones using FFT and Segment Averaging

Linggo Sumarno*, Iswanjono
Electrical Engineering, Sanata Dharma University, Paingan, Maguwoharjo, Yogyakarta, Indonesia 55281
*Corresponding author, email: lingsum@usd.ac.id

Abstract

A feature extraction for musical instrument tones that based on a transform domain approach was proposed in this paper. The aim of the proposed feature extraction was to get the lower feature extraction coefficients. In general, the proposed feature extraction was carried out as follow. Firstly, the input signal was transformed using FFT (Fast Fourier Transform). Secondly, the left half of the transformed signal was divided into a number of segments. Finally, the averaging results of that segments, was the feature extraction of the input signal. Based on the test results, the proposed feature extraction was highly efficient for the tones, which have many significant local peaks in the Fourier transform domain, because it only required at least four feature extraction coefficients, in order to represent every tone.

Keywords: *feature extraction, musical instrument tones, FFT, segment averaging*

1. Introduction

From the human perception point of view, musical instruments have two aural characteristics, namely scale and timbre [1]. The scale corresponds with high and low tones of a musical instrument. The timbre corresponds with the type of a musical instrument. Based on the timbre, when it is viewed from the Fourier transform domain, the tone of a musical instrument can be divided into two groups. The first one is the tones that have multiple significant local peaks (also called polyphonic), and the second one is the tones that have single significant local peaks (also called monophonic). Figure 1 shows an example of such tones.

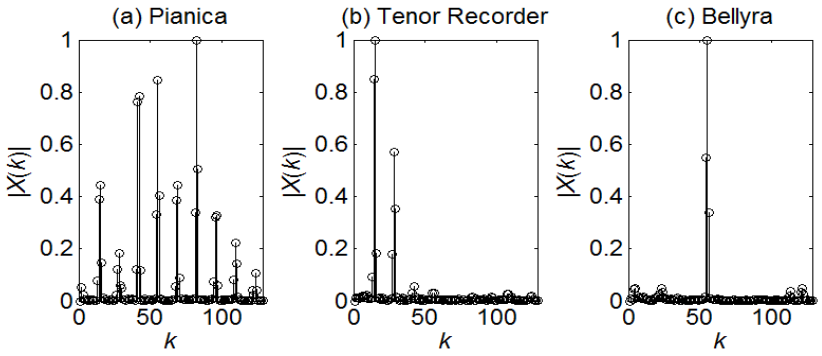

Figure 1. Representation of tone C in the left half of the normalized Fourier transform domain, when using FFT 256 points

As shown in Figure 1, pianica and tenor recorder tones are the examples of the tones that have multiple significant local peaks. The pianica tone has many (i.e. nine) significant local peaks, while the tenor recorder tone has several (i.e. two) significant local peaks. The bellyra tone is an example of the tone that has a single significant local peak.

In the case of tone recognition by humans, for the trained humans, the recognition is done by taking into account the characteristic of the tone that is heard through the ear and then

Received April 14, 2017; Revised July 20, 2017; Accepted August 5, 2017

match them with the characteristic of the tones which had heard in the past [2]. Today, by using computers, it can be developed a tone recognition system, which is based on the characteristic of the tone. In this case that characteristic is called as feature extraction.

An approach to get the feature extraction is to use a transform domain. FFT and DCT (Discrete Cosine Transform) are two examples of the transform methods that can be used. In the transform domain, there are two ways to get the feature extraction. The first one is based on fundamental signals [3-6] and the second one is not based on fundamental signals [7-11].

There was a problem in the field of feature extraction, which using the transform domain approach that was not based on fundamental signals. It still showing a relatively large number of the feature extraction coefficients. For example, based on the references, the feature extraction for musical instrument sounds gave 34 coefficients [7], 32 coefficients [8], 12 coefficients [9], 9 coefficients [10], and 8 coefficients [11]. Therefore, there is still an opportunity to further reduce the number of the feature extraction coefficients.

This paper proposes a feature extraction for musical instrument tones, with the approach of the Fourier transform domain, which is not based on fundamental signals. The proposed feature extraction has the ability to further reduce the number of feature extraction coefficients that described above. In this paper, it will be discussed also the performance of the proposed feature extraction, if it deals with the signals that have multiple and single significant local peaks in the Fourier transform domain.

2. Research Methodology

2.1. Overall System Development

In order to explore the proposed feature extraction in this research, it had been developed a tone recognition system, which is shown in Figure 2. The input is an isolated signal tone in wav format, while the output is a text that indicates a recognized tone. Here is an explanation of the blocks shown in Figure 2.

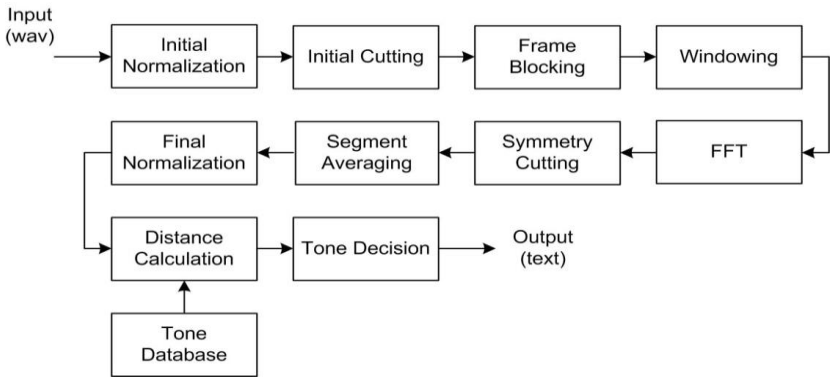

Figure 2. Overall system block diagram

2.1.1. Input

The input of the recognition system is a signal tone of a musical instrument (pianica, tenor recorder, or bellyra) in an isolated wav format. From each musical instrument there are eight tones (C, D, E, F, G, A, B, and C') used. That signal tones were obtained by recording the sounds that coming out from the musical instrument, by using a sampling frequency of 5000 Hz. That magnitude of the sampling frequency was chosen since it had met the Shannon sampling theorem [12] as follow:

$$f_S \geq 2f_{\max} \quad (1)$$

where $f_{\max}$ is the highest frequency of the sound, which is coming out from a musical instrument, f_S is the sampling frequency. The magnitude of the sampling frequency (5000 Hz)

has already exceeded the highest tone C' on pianica, tenor recorder, and bellyra. Based on the test spectrum, the highest tones C' on pianica, tenor recorder, and bellyra, were 1584 Hz, 547 Hz, and 2097 Hz respectively. Based on the evaluation, recording duration of 1.5 seconds was sufficient, in order to obtain the middle area data which was already in the steady state condition. This area was required for the purpose of frame blocking.

In order to generate the tone sounds, three musical instruments were used. They were pianica Yamaha P-37D, tenor recorder Yamaha YRT-304B II, and bellyra Isuzu ZBL-27 (see Figure 3). In order to capture the tone sounds, a USB microphone AKG Perception 120 was used.

(a) Pianica (b)Tenor recorder (c) Bellyra

Figure 3. Pianica, tenor recorder, and bellyra, which were used for the research

2.1.2. Initial Normalization

Initial normalization is a process of setting the maximum value to one, from a series of signal data, which comes from recording the sound that coming out from a musical instrument. By setting the maximum value to one, it will eliminate the differences in the maximum values on a number of the series of signal data.

2.1.3. Initial Cutting

Initial cutting is a process of cutting the silence area followed by cutting the transition area of the tone signal. Silence area is an area that has no tone information, whereas the transition area is the area where the tone signal is not in a steady state condition yet. In this research, based on the graphical observation of the signals, the cutting of silence area could be carried out by using an amplitude threshold |0.5|, while the cutting of the transition region could be carried out by cutting for 0.1 seconds at the beginning area of the tone signal.

2.1.4. Frame Blocking

Frame blocking is a process of taking a frame signal from a long series of signal [13]. The purpose of frame blocking is to reduce the number of data signals to be processed. The effect of this reduction is a reduction in computing time. In this research, a frame signal was taken from the beginning area of the tone signal that had already in the steady state condition. In this research, the length of a frame signal was equal to the length of the FFT that used in the next process.

2.1.5. Windowing

Windowing is the process of reducing discontinuities at the edges of the signal [13]. This reduction is necessary in order to reduce the appearance of harmonic signals after the signal is transformed using FFT. In this research, windowing used a Hamming window. Hamming window $w(n)$ with a width of N points defined [14].

$$w(n) = 0.54 - 0.46\cos\left(2\pi \frac{n}{N-1}\right), \quad 0 \le n \le N-1 \tag{2}$$

Hamming window is a window that is widely used in the field of signal processing. In this research, the length of the windows was the same with the length of the frame blocking.

2.1.6. FFT

FFT is a process of changing the signal from the time domain to the Fourier transform domain. Basically FFT is an efficient method for calculating the DFT (Discrete Fourier Transform). DFT is a form of Fourier transforms that used for discrete signals. Mathematically, DFT is formulated as follows.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi}{N}kn}, \qquad k = 0, \cdots, N-1 \tag{3}$$

where $x(n)$ = input series, $X(k)$ = output series, and N = number of samples. In this research FFT radix-2 [15] was used. This FFT is widely used in the field of signal processing. In this research, the length of the FFT was evaluated with values of 16, 32, 64, 128, and 256 points.

2.1.7. Symmetry Cutting

Symmetry cutting is cutting the half of the right side of the FFT result. This cutting is necessary, because basically FFT will give the symmetry result. Therefore, if only the half side (right or left) is used, it is sufficient.

2.1.8. Segment Averaging

Segment averaging is a process for reducing the size of the data signal. Basically, the result of segment averaging can represent the basic shape of the data signal. This research made use of the segment averaging that had been inspired by Setiawan [16]. The algorithm of the segment averaging is shown as follows.

1. Let a series $\{X(k) \mid 0 \le k \le M-1\}$, where $M = 2^p$ for $p \ge 0$.
2. Calculate the segment length L, where $L = 2^q$ for $0 \le q \le p$.
3. Cut uniformly the series $X(k)$ by using segment length L. So, it will give a number of S segments.

$$S = \frac{M}{L} , \tag{4}$$

and also a series $\{F(r) \mid 1 \le r \le L\}$ in every segment.

4. Calculate the mean value of every segment $Z(v)$ as follows.

$$Z(v) = \frac{1}{L} \sum_{r=1}^{L} F_V(r), \quad 1 \le v \le S . \tag{5}$$

In this research, the segment lengths in the segment averaging were evaluated with the values 1, 2, 4, ..., and $2^{\log_2(N/2)}$ points, where N is the length of the FFT used.

2.1.9. Final Normalization

Final normalization is a process of setting the maximum value to one, from a series of signal data, which comes from segment averaging. As described above, by setting the maximum value to one, it will eliminate the differences in the maximum values on a number of the series of signal data.

2.1.10. Distance Calculation

Distance calculation is a process for comparing the feature extraction of an input signal and the feature extraction of a number of the tone signals, which is stored in a tone database. Distance calculation is an indication of a pattern recognition method known as template matching [17]. Template matching is basically trying to find the degree of similarity between two vectors in the feature space. In order to find the degree of similarity, it can be used a distance function. If using a distance function, there is a general guideline, the degree of similarity of the two vectors will be increased away from zero (getting more similar) if the value of the distance function decreases toward zero. On the contrary, the degree of similarity of the two vectors will be decreased toward zero (getting not similar) if the value of the distance function is increased away from zero.

Euclidean distance is a distance function that can be used to find the degree of similarity between two vectors in the feature space. Euclidean distance is a distance that is commonly used in pattern recognition [18]. Euclidean distance is defined as follows.

$$E(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \quad (6)$$

where **x** and **y** are two vectors of the same length, and *m* is the length of the vector x and y. In a pattern recognition system which using template matching method, one of the vector (x or y) is a vector that will be determined its class pattern, whereas the other vector is a vector that stored in a pattern class.

2.1.11. Tone Decision

Tone decision is a process for determining the output tone of the input signal, which is in the form of an isolated tone signal in wav format. Tone decision is carried out by finding a minimum value from a number of distance calculation results. These results come from a number of distance calculation, between the feature extraction of an input signal and the feature extraction of a number of the tone signals, which is stored in the tone database. A tone, which has a minimum distance, will be determined as the output tone.

2.2. Feature Extraction and Tone Database

The above distance calculation process requires a tone database. This database is generated by using the tone feature extraction shown in Figure 4. From Figure 4, the input is an isolated tone signal in wav format. The output is the feature extraction of the input. The output is a vector of length *S*, where *S* value is obtained from the Equation (4).

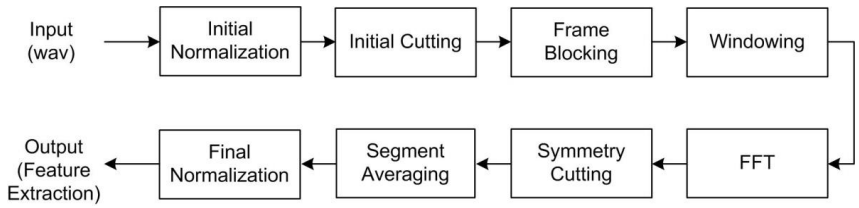

Figure 4. Tone feature extraction block diagram

Based on Figure 4, the process of segment averaging gets a signal from the symmetry cutting process. In this cutting symmetry process, the signal from the FFT process is divided into two equal lengths. If the signal from the FFT process has length *N*, then the Equation (4) can be rewritten to be as follows:

$$S = \frac{N/2}{L} \quad (7)$$

where *S* is the length of the feature extraction vector, *N* is the FFT length, and *L* is the segment length in the segment averaging process.

By using the process illustrated in Figure 4, for each instrument (pianica, recorder tenor, or bellyra), it is generated the feature extraction from a number of *K* samples for each tone (C, D, E, F, G, A, B, or C'). Next, for each tone of a musical instrument, it is calculated the average value of the following:

$$\mathbf{R}_T = \frac{\mathbf{Z}_1 + \mathbf{Z}_2 + \cdots + \mathbf{Z}_K}{K} \tag{8}$$

where $\mathbf{Z}_1$, Z_2, … $\mathbf{Z}_K$ are the feature extraction vectors of a tone of a musical instrument, which have a number of *K* samples. In this research it was evaluated the value *K* with the values 1, 5, 10, and 15. Therefore, for this evaluation, it was necessary to provide a number of 120 tones for each musical instrument, or number of 360 tones for all three instruments (pianica, tenor recorder, and bellyra). The addition operation $\mathbf{Z}_1 + Z_2 + \ldots + \mathbf{Z}_K$ is the addition operation of vector elements. Vector $\{ \mathbf{R}_T | T$ = C, D, E, F, G, A, B, or C'$\}$ is called as the reference vector.

In the generation of a tone database, every musical instrument has a tone database, which consists of eight reference vectors, namely $\mathbf{R}_C$, $\mathbf{R}_D$. $\mathbf{R}_E$. $\mathbf{R}_F$. $\mathbf{R}_G$, $\mathbf{R}_A$, $\mathbf{R}_B$, and $\mathbf{R}_{C'}$. Thus, for the three instruments (pianica, tenor recorder, and bellyra), there are three tone databases. On the tone recognition system in this research, the tone database that is used depends on the tone input, namely, whether the tone comes from pianica, tenor recorder, or bellyra.

2.3. Test Tones

Test tones are used to evaluate the performance of the developed tone recognition system. In this research, for each instrument (pianica, tenor recorder, or bellyra), a number of 10 samples for each tone (C, D, E, F, G, A, B, and C ') were used. Thus, there were a number of 80 test tones for each musical instrument, or a number 240 test tones for all three musical instruments (pianica, tenor recorder, and bellyra).

2.4. Recognition Rate

The recognition rate is the performance a recognition system when given a number particular inputs. The recognition rate is calculated using the following equation.

$$\text{Recognition rate} = \frac{\text{Number of recognized tones}}{\text{Number of test tones}} \times 100\% \tag{9}$$

In this research, as described above, a number of 80 test tones were used for each musical instrument (pianica, tenor recorder, or bellyra).

3. Results and Analysis

3.1. Test Results

Test results of the tone recognition system for a variety of the FFT length, the segment length, and the number of feature extraction vectors, are shown in Table 1, 2, and 3. As the first note, the number of feature extraction vectors refers to the Equation (8). As the second note, the segment length refers to the segment length used for the segment averaging (see Equation (4) and (5)).

3.2. The Smallest Number of feature Extraction Coefficients

The main objective of this research is to find the smallest number of the feature extraction coefficients, which can represent a tone. Therefore, by using Table 1, 2, and 3, it can be explored the combination of the values of the FFT length, the segment length, and the number of feature extraction vectors that can give the smallest number of the feature extraction coefficients, at the highest recognition rate (in this case 100%). The exploration results of the combination of these values, along with the number of the feature extraction coefficients are shown in Table 4. As a note, in order to calculate the number of the feature extraction coefficients, the equation (7) was used.

As shown in Table 4, the features of a pianica tone can be represented by at least four feature extraction coefficients, while a tone of tenor recorder or bellyra can be represented by at least 16 feature extraction coefficients. Therefore, in general, it can be said that, for the tones of musical instrument (e.g. pianica) which have many significant local peaks in the Fourier transform domain, the feature extraction which using FFT and segment averaging is highly efficient. The reason is that it can give at least four feature extraction coefficients.

Table 1. Test results of pianica musical instrument, in various combinations of the FFT length, the segment length, and the number of feature extraction vectors
Results shown: Recognition rate (%)

FFT length (points)	Number of feature extraction vectors	Segment length (points) 1	2	4	8	16	32	64	128
16	1	86.25	71.25	45.00	12.50	-	-	-	-
	5	86.25	75.00	50.00	12.50	-	-	-	-
	10	87.25	76.25	50.00	12.50	-	-	-	-
	15	87.50	78.75	52.50	12.50	-	-	-	-
32	1	100	95.00	78.75	46.25	12.50	-	-	-
	5	100	97.50	86.25	52.50	12.50	-	-	-
	10	100	100	92.50	53.75	12.50	-	-	-
	15	100	100	95.00	56.25	12.50	-	-	-
64	1	100	100	100	81.25	57.50	12.50	-	-
	5	100	100	100	81.25	58.75	12.50	-	-
	10	100	100	100	85.00	58.75	12.50	-	-
	15	100	100	100	91.25	61.25	12.50	-	-
128	1	100	100	100	98.75	81.25	52.50	12.50	-
	5	100	100	100	98.75	88.75	55.00	12.50	-
	10	100	100	100	98.75	92.50	55.00	12.50	-
	15	100	100	100	98.75	98.75	60.00	12.50	-
256	1	100	100	100	100	100	98.75	62.50	12.50
	5	100	100	100	100	100	**100**	65.00	12.50
	10	100	100	100	100	100	100	65.00	12.50
	15	100	100	100	100	100	100	67.50	12.50

Table 2. Test results of tenor recorder musical instrument, in various combinations of the FFT length, the segment length, and the number of feature extraction vectors
Results shown: Recognition rate (%)

FFT length (points)	Number of feature extractions vectors	Segment length (points) 1	2	4	8	16	32	64	128
16	1	57.50	43.75	11.25	12.50	-	-	-	-
	5	60.00	48.75	16.25	12.50	-	-	-	-
	10	63.75	53.75	17.50	12.50	-	-	-	-
	15	63.75	58.75	27.50	12.50	-	-	-	-
32	1	98.75	86.25	62.50	30.00	12.50	-	-	-
	5	98.75	88.75	62.50	35.50	12.50	-	-	-
	10	98.75	88.75	71.25	37.00	12.50	-	-	-
	15	98.75	90.00	77.00	43.75	12.50	-	-	-
64	1	100	98.75	73.75	58.75	37.50	12.50	-	-
	5	100	**100**	85.00	56.25	42.50	12.50	-	-
	10	100	100	87.50	62.50	42.50	12.50	-	-
	15	100	100	87.50	71.25	43.75	12.50	-	-
128	1	100	100	97.50	68.75	46.25	31.25	12.50	-
	5	100	100	100	81.25	47.50	36.25	12.50	-
	10	100	100	100	85.00	58.75	38.75	12.50	-
	15	100	100	100	85.00	60.00	45.00	12.50	-
256	1	100	100	100	96.25	71.25	48.75	31.25	12.50
	5	100	100	100	98.75	81.25	53.75	33.75	12.50
	10	100	100	100	100	82.50	61.25	36.25	12.50
	15	100	100	100	100	78.75	65.00	40.00	12.50

Table 3. Test results of bellyra musical instrument, in various combinations of the FFT length, the segment length, and the number of feature extractions vectors. Results shown: Recognition rate (%).

FFT length (points)	Number of feature extraction vectors	Segment length (points)							
		1	2	4	8	16	32	64	128
16	1	57.50	40.00	25.00	12.50	-	-	-	-
	5	81.25	60.00	27.50	12.50	-	-	-	-
	10	85.00	62.50	35.00	12.50	-	-	-	-
	15	88.75	65.00	36.25	12.50	-	-	-	-
32	1	73.75	71.25	42.50	23.75	12.50	-	-	-
	5	95.00	91.25	53.75	27.50	12.50	-	-	-
	10	97.50	95.00	70.00	32.50	12.50	-	-	-
	15	96.25	95.00	72.50	35.00	12.50	-	-	-
64	1	100	93.75	83.75	65.00	28.75	12.50	-	-
	5	100	97.50	91.25	76.25	32.50	12.50	-	-
	10	100	97.50	92.50	83.75	33.75	12.50	-	-
	15	100	98.75	92.50	80.00	36.25	12.50	-	-
128	1	100	98.75	93.75	83.75	77.50	30.00	12.50	-
	5	100	100	97.50	95.00	83.75	33.75	12.50	-
	10	100	100	98.75	97.50	92.50	36.25	12.50	-
	15	100	100	98.75	97.50	90.00	42.50	12.50	-
256	1	100	100	100	92.50	85.00	73.75	26.25	12.50
	5	100	100	100	**100**	96.25	83.75	38.75	12.50
	10	100	100	100	100	98.75	87.50	38.75	12.50
	15	100	100	100	100	98.75	87.50	42.50	12.50

Table 4. The combination of the values of FFT length, segment length, number of feature extraction vectors, which resulted in the smallest number of feature extraction coefficients

Parameter	Musical Instrument		
	Pianica	Tenor Recorder	Bellyra
FFT length (points)	256	64	256
Segment length (points)	32	2	8
Number of feature extraction vectors	5	5	5
Number of feature extraction coefficients	**4**	16	16

For the tone of a musical instrument that has only one significant local peak in the Fourier transform domain (e.g. bellyra), or the tone of a musical instrument that has a few significant local peaks in the Fourier transform domain (e.g. tenor recorder), the feature extraction is relatively inefficient. The reason is that it can give at least 16 coefficients feature extraction to represent each tone. Figure 5 and Table 5 shows the reason of this case. As a note to Figure 5, the results of feature extraction are described by linear interpolation line, for easy visual comparison.

As shown in Figure 5, the feature extractions of tone C and D on tenor recorder and also bellyra are lookalike (because they look almost overlap), when compared with pianica. In order to be more exact, the feature extraction coefficients of tone C and D, on tenor recorder and also bellyra, as shown in Table 5, has Euclidean distances that are much smaller, when compared to pianica. It means, the feature extraction coefficients from the adjacent tones on tenor recorder and also bellyra, are more difficult to be distinguished each other, compared with pianica. Furthermore, in order that adjacent tones of tenor recorder and also bellyra can be distinguished easier, there is a need in increasing the number of feature extraction coefficients. Therefore, tenor recorder and also bellyra require a greater number of feature extraction coefficients, when compared with pianica.

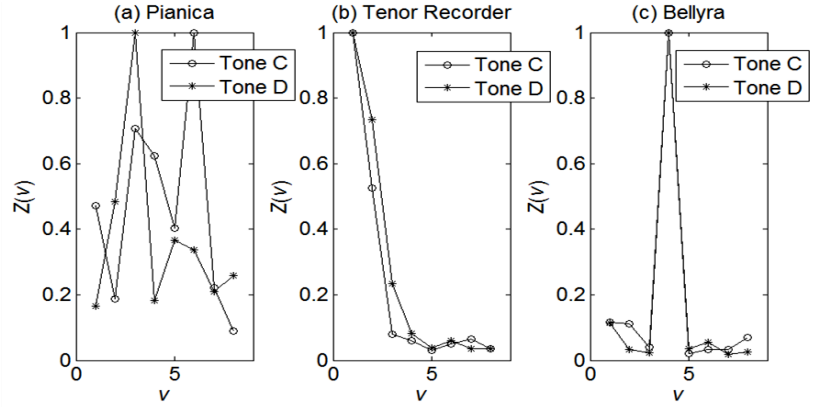

Figure 5. The examples of feature extraction differences between pianica, tenor recorder, and pianica, when using an FFT 256 points and a segment length 16 points

Table 5. Euclidean distances for the feature extractions of tone C and D, for pianica, tenor recorder, and bellyra, in Figure 5

Musical Instrument	Euclidean distance of tone C and D
Pianica	0.966
Tenor Recorder	0.262
Bellyra	0.097

3.3. Comparison with other Feature Extractions

As has been described above, the feature extraction that used FFT and segment averaging is highly efficient for use in the musical instrument tones that have many significant local peaks in the Fourier transform domain. Table 6 compares the performance of several feature extraction for pianica tone recognition. As shown in Table 6, the feature extraction used in this research, when compared with the other feature extraction that have been published previously [7-11], is highly efficient for use in the musical instrument tones that have many significant local peaks in the Fourier transform domain. The reason is that it can give the number of feature extraction coefficients that are significantly much smaller.

Table 6. The comparison of the number of coefficients of the smallest feature extraction results, which resulted in a 100% recognition rate, of several feature extraction for pianica tone recognition

Feature extraction	Number of feature extraction coefficients
Spectral Features [7]	34
DCT and windowing coefficients [8]	32
FFT and windowing coefficients [9]	12
Musical Surface [10]	9
DCT and segment averaging [11]	8
FFT and segment averaging (this research)	**4**

4. Conclusions and Future Work

Based on the things that have been described above, it can be concluded the following:

1. A feature extraction with FFT and segment averaging is highly efficient for musical instrument tones, which have many significant local peaks in the Fourier transform domain. The reason is that it only required at least four feature extraction coefficients, in order to represent every tone.

2. A feature extraction with FFT and segment averaging is not efficient for musical instrument tones, which have one or several significant local peaks in the Fourier transform domain. The reason is that it required at least 16 feature extraction coefficients, in order to represent every tone.

Here are some suggestions for future work:

1. Further studies of the feature extractions that are not based on a fundamental signal, which can produce a number of feature extraction coefficients less than 16 coefficients, for the tone of musical instruments that have one or several significant local peaks in the Fourier transform domain.
2. Further studies about the performance of feature extraction, which uses FFT and segment averaging, when using the tone recognition methods other than the template matching.

Acknowledgements

This work has been supported by The Institution of Research and Community Service of Sanata Dharma University, Yogyakarta.

References

[1] Forster C. Musical Mathematics: On the Art of Science and Acoustic Instruments. California: Chronicle Books LLC. 2010. 7-8.
[2] McAdams S. Recognition of Auditory Sound Sources and Events. Thinking in Sound: The Cognitive Psychology of Human Audition. Oxford: Oxford University Press. 1993: 146-198.
[3] Noll 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. Brooklyn, New York. 1970; 19: 779-797.
[4] Mitre A, Queiroz M, Faria R. *Accurate and Efficient Fundamental Frequency Determination from Precise Partial Estimates*. Proceedings of the 4th AES Brazil Conference. 2006: 113-118.
[5] Izzudin A, Santoso TB, Dutono T. Recognition of Guitar Single Tones using Digital Signal Processing. *EEPIS Journal Online System*. 2005; 10(1).
[6] Gaffar I, Hidayatno A, Zahra AA. An Application to Convert Single Instrument Tones become a Chord using Pitch Class Profile Method. *Transient*. 2012; 1(3): 121-127.
[7] Tjahyanto A, Suprapto YK, Wulandari DP, Spectral-based Features Ranking for Gamelan Instruments Identification using Filter Techniques. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(1): 95-106.
[8] Sumarno L. Recognition of Pianica Tones using Gaussian Window, DCT, and Cosine Distance. *Research Journal*. 2013; 17(1): 8-15.
[9] Sumarno L. Recognition of Pianica Tones using Blackman Window and Fast Fourier Transform Feature Extraction (in Indonesian). *Media Teknika*. 2014; 9(2): 84-93.
[10] Mutiara AB, Refianti R, Mukarromah NRA, Musical Genre Classification Using Support Vector Machine and Audio Features. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2016; 14(3): 1024-1034.
[11] Sumarno L. On The Performace of Segment Averaging of Discrete Cosine Transform Coefficients on Musical Instruments Tone Recognition. *ARPN Journal of Engineering and Applied Sciences*. 2016; 11(9): 5644-5649.
[12] Tan L, Jiang J. Digital Signal Processing Fundamentals and Applications. Second Edition. Oxford: Elsevier Inc. 2013: 15-56.
[13] Meseguer NA. Speech Analysis for Automatic Speech Recognition. MSc Thesis. Trondheim: NTNU; 2009.
[14] Harris FJ. *On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform*. Proceedings of the IEEE. 1978; 66(1): 51-83.
[15] Proakis JG, Manolakis DG. Digital Signal Processing: Principles, Algorithm, and Applications. Fourth Edition. New Jersey: Prentice Hall Inc. 2007: 511-562.
[16] Setiawan YR. Numbers Speech Recognition using Fast Fourier Transform and Cosine Similarity. Undergraduate Thesis. Yogyakarta: Sanata Dharma University. 2015.
[17] Jain AK, Duin RPW, Mao J. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(1): 4-37.
[18] Wilson DR, Martinez TR. Improved Heterogeneous Distance Function. *Journal of Artificial Intelligence Research*. 1997; 6: 1-34.