The influence of sampling frequency on tone recognition of musical instruments

Linggo Sumarno*[1], Kuntoro Adi[2]
[1]Department of Electrical Engineering, Sanata Dharma University
Kampus III, Paingan, Maguwoharjo, Depok, Sleman, Yogyakarta, Indonesia 55282
[2]Department of Informatics Engineering, Sanata Dharma University
Kampus III, Paingan, Maguwoharjo, Depok, Sleman, Yogyakarta, Indonesia 55282
*Corresponding author, email: lingsum@usd.ac.id

Abstract

Sampling frequency of musical instruments tone recognition generally follows the Shannon sampling theorem. This paper explores the influence of sampling frequency that does not follow the Shannon sampling theorem, in the tone recognition system using segment averaging for feature extraction and template matching for classification. The musical instruments we used were bellyra, flute, and pianica, where each of them represented a musical instrument that had one, a few, and many significant local peaks in the Discrete Fourier Transform (DFT) domain. Based on our experiments, until the sampling frequency is as low as 312 Hz, recognition rate performance of bellyra and flute tones were influenced a little since it reduced in the range of 5%. However, recognition rate performance of pianica tones was not influenced by that sampling frequency. Therefore, if that kind of reduced recognition rate could be accepted, the sampling frequency as low as 312 Hz could be used for tone recognition of musical instruments.

Keywords: sampling frequency, Shannon sampling theorem, tone recognition

1. Introduction

Nowadays, the development of technology leads to digital technology. This digital technology requires the conversion of data from analog type to discrete type. In order to convert data from analog to discrete, the existence of a sampler is required. The major parameter in a sampler is sampling frequency. Generally in the field of digital signal processing, the sampling frequency used is the frequency that follows the Shannon sampling theorem. Basically, this theorem needs to be followed in order that analog signal can be perfectly recovered from its sampled version [1]. In addition, in the field of pattern recognition, the Shannon sampling theorem is generally followed, such as: in researches relating to tone recognition using the time domain approaches [2-5], the transformation domain approaches that are based on fundamental frequencies [6-11], and the transformation domain approaches that are not based on fundamental frequencies [12-18].

If it is analyzed further, in researches relating to musical instruments tone recognition, signal processing is not carried out on processes that contribute to producing the output signal. In this case, signal processing will stop before feature extraction process. There were no previous researches, relating to musical instruments tone recognition made use of sampling frequencies that did not follow the Shannon sampling theorem. Thus, a research of musical instruments tone recognition that makes use of sampling frequencies that do not follow the Shannon sampling theorem, is still wide open. As a note, the advantage of using sampling frequencies that do not follow the Shannon sampling theorem is smaller data for processing.

This paper will discuss the influence of sampling frequency that does not follow the Shannon sampling theorem, on musical instruments tone recognition. The tone recognition system uses a transformation domain approach that does not use fundamental frequencies [18]. The musical instruments use bellyra, flute, and pianica. They are chosen to represent tones that have one, a few, and many significant local peaks in the Discrete Fourier Transform (DFT) domain as shown in Figure 1. In more detail, it will be explored if there is a lowest sampling frequency that does not follow the Shannon sampling theorem, which can be used for tone recognition system.
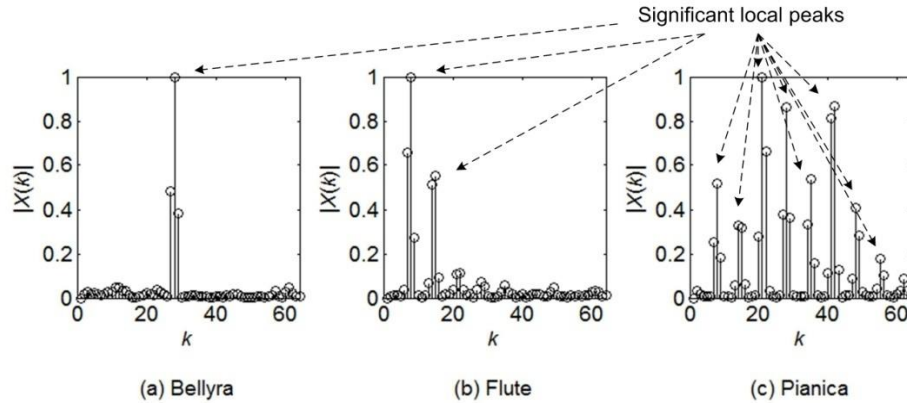
Figure 1. Representation of tone C in the normalized DFT domain X(k) for bellyra, flute, and pianica, by using a sampling frequency 5000 Hz and a DFT 128 points. As a note, only the left half of the normalized DFT domain is shown.

2. Research Method
2.1. Overall system development

A tone recognition system shown in Figure 2 had been developed in order to explore the influences of sampling frequency. The input is an isolated signal tone in wav format. The output is a text that indicates the recognized tone. The input and the function of each block in Figure 2 are described in details in Subsections 2.1.1-2.1.11.
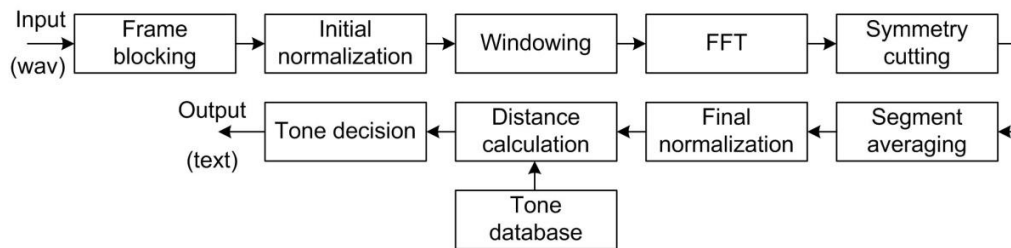

Figure 2. Block diagram of the developed tone recognition system

2.1.1. Input

The input of the developed tone recognition system was an isolated tone signal from a musical instrument (bellyra, flute, or pianica) in wav format. There were eight tones (C, D, E, F, G, A, B, and C) of each musical instrument. The tones were obtained by using various sampling frequencies: 5000 Hz, 2500 Hz, 1250 Hz, 625 Hz, 312 Hz, and 156 Hz. The highest sampling frequency of 5000 Hz was chosen because it met the following the Shannon sampling theorem [1]:

$$f_s \geq 2f_{max} \tag{1}$$

where f_{max} is the highest frequency component of the analog signal to be sampled, and f_s is the sampling frequency. Based on the evaluation of the signal spectrum, the highest significant frequency components for C' on bellyra, flute, and pianica were 2109 Hz, 1602 Hz, and 2100 Hz, respectively. The recording duration of 2 seconds was chosen since based on the evaluation of signal amplitude, it was sufficient to get more than half of the signal that was already in the steady state condition.

The three musical instruments we used for generating the tone signals above were, Isuzu ZBL-27 bellyra, Yamaha YFL-221 flute, and Yamaha P-37D pianica, as shown in Figure 3. In order that the generated tone signals could be processed by the computer, they were captured by using a Samson Meteor USB microphone.

(a) Bellyra   (b) Flute   (c) Pianica

Figure 3. Bellyra, flute, and pianica, which were used for this research

2.1.2. Frame Blocking

Frame blocking is the process of taking a short signal from a long signal [19]. This process is needed in order to reduce the number of signal data from the input. By reducing the number of signal data, it could reduce computing time. In this research, a short signal was taken from the beginning area of the tone signal that had been in the steady state condition. Based on the evaluation of the tone signals, 200 milliseconds after the silent part of the signal, the steady state condition had been reached. In this research, frame blocking length was evaluated with values of 16, 32, 64, and 128 points.

2.1.3. Initial Normalization

Initial normalization is a process for setting a maximum absolute value to a value of one. Initial normalization is required since the signals from the frame blocking process have variations in maximum absolute value. Initial normalization is carried out using the following (2):

$$x_{out} = \frac{x_{in}}{max(|x_{in}|)} \tag{2}$$

where signal vectors x_{in} and x_{out} are the input and output of normalization process respectively.

2.1.4. Windowing

Windowing is the process of reducing discontinuities in the areas of signal edges. These areas appear as a result of signal cutting in the frame blocking process. Discontinuity reduction will reduce the appearance of harmonic signals, after the normalized signals is transformed using FFT (Fast Fourier Transform). Hamming window [20] is a type of window that can be used for windowing purposes. This type of window is commonly used in digital signal processing [21]. In this research, the window width was the same as the frame blocking length.

2.1.5. FFT

FFT is a process to transform a discrete signal from windowing process, from the time domain to the DFT domain. Basically FFT is an efficient method to perform the DFT calculation. In this research, DFT calculation was performed using a radix-2 FFT. This type of FFT is widely used in the field of digital signal processing [21]. As well as the above windowing, the FFT length in this research had the same length as the frame blocking length.

2.1.6. Symmetry Cutting

Symmetry cutting is the process of cutting half part of the FFT result. This cutting is necessary since between the left and right half of the FFT result shows a symmetry property.

Therefore, when only half part of the left or right of the FFT result, it will be sufficient. In this research, only the left half part of the FFT result was used.

2.1.7. Segment Averaging

Segment averaging is a process to shorten the signal length from symmetry cutting. In this case, the shortened signal still shows similarity to the basic shape of the long signal. Segment averaging [18] used in this research using the principles of segment averaging inspired by Setiawan [22]. In this research, the segment length in segment averaging was evaluated with values 1, 2, 4, ..., and $2^{\log_2(N/2)}$ points, where N is the length of frame blocking.

2.1.8. Final Normalization

As with initial normalization, final normalization is also the process of setting the maximum absolute value to a value of one. Final normalization is necessary because the results of segment averaging have variations in the maximum absolute value. As with initial normalization, final normalization also carried out using (2). As a note, the final normalization result is called feature extraction of the input signal.

2.1.9. Distance calculation

Distance calculation is a comparison process between an input signal feature extraction and a number of tone signal feature extractions stored in the tone database. Distance calculation is an indication of pattern recognition using a template matching method [23, 24]. The Euclidean distance function can be used for this kind of distance calculation. This distance function is commonly used in the field of pattern recognition [25].

2.1.10. Tone decision

Tone decision is the process of deciding an output tone that corresponds to the input signal. The first step of tone decision is finding a minimum value from a number of distance calculation results. These are the results of distance calculation between an input signal feature extraction (after final normalization process) and a number of tone feature extractions in the tone database. The next step is to decide an output tone. A tone associated with one of the signal feature extraction in the tone database that has a minimum distance, will be decided as the output tone.

2.1.11. Tone database

The tone database shown in Figure 2 is generated using the tone feature extraction shown in Figure 4. In this research, for each musical instrument, we took 10 samples for each tone (C, D, E, F, G, A, B, and C'). It was assumed that by taking 10 samples, all variations for each signal tone have been obtained. The results of these 10 samples were 10 feature extraction results for each tone. Furthermore, the 10 feature extraction results were averaged as follows:

$$R_T = \frac{1}{10}\sum_{i=1}^{10} Z_i \qquad (3)$$

where vector $\{Z_i \mid 1 \leq i \leq 10\}$ is the 10 results of feature extraction, and vector $\{R_T \mid T = C, D, E, F, G, A, B, \text{ and } C'\}$ is a vector of eight tones stored in a tone database. As a note, a tone database is generated from a value of sampling frequency, a value of frame blocking length, and a value of segment length.
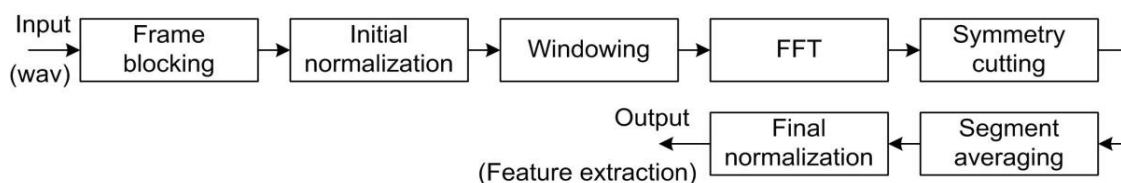

Figure 4. Block diagram of tone feature extraction

2.2. Test Tones and Recognition Rate

Test tones were used to examine the developed tone recognition system. The developed tone recognition system was tested using 160 tones for each musical instrument, each sampling frequency, each frame blocking length, and each segment length. Tones come from eight tones (C, D, E, F, G, A, B, and C), where each of these tones was recorded 20 times. In order to measure the performance of the developed tone recognition system, we used the recognition rate formula as follows.

$$Recognition\ rate = \frac{Number\ of\ recognized\ tones}{Number\ of\ test\ tones} \times 100\% \qquad (4)$$

3. Results and Analysis

The developed tone recognition system as shown in Figure 2, was used to test the effect of sampling frequency of the tone recognition system. The results of these tests are shown in Tables 1, 2 and 3, each of which is the result of pianica, flute, and bellyra, respectively. As a note, based on Figure 1, bellyra, flute, and pianica tones have one, a few, and many significant local peaks in the DFT domain, respectively.

Tables 1 and 2 indicate that from a sampling rate 5000 Hz down to 312 Hz, there is a little influence on the recognition rate, since it is reduced in the range of 5%. However, Table 3 indicates that from a sampling rate 5000 Hz down to 312 Hz, there is no influence on the recognition rate. As a note, the highest frequency components of bellyra, flute, and pianica are 2109 Hz, 1602 Hz, and 2100 Hz, respectively. Therefore, based on these highest frequency components, (1), and also Tables 1, 2 and 3, the sampling frequency of 5000 Hz follows the Shannon sampling theorem. Starting from a sampling frequency of 2500 Hz and below, the sampling frequency deviates from the Shannon sampling theorem.

Table 1. Test Results of Bellyra Musical Instrument, in Various Combinations of Sampling Frequency, Frame Blocking Length, and Segment Length.
Results Shown: Recognition Rate (%)

Sampling frequency (Hz)	Frame blocking length (points)	Segment length (points)						
		1	2	4	8	16	32	64
5000	16	98.75	91.88	44.38	12.50	–	–	–
	32	99.38	98.75	78.13	42.50	12.50	–	–
	64	100	100	98.75	80.63	37.50	12.50	–
	128	100	100	100	81.88	71.88	37.50	12.50
2500	16	98.13	86.88	56.25	12.50	–	–	–
	32	98.75	98.75	68.75	52.50	12.50	–	–
	64	98.75	98.75	92.50	72.50	40.00	12.50	–
	128	98.75	98.75	98.75	90.63	80.63	34.38	12.50
1250	16	90.63	89.38	66.25	12.50	–	–	–
	32	92.50	91.88	90.00	68.25	12.50	–	–
	64	96.88	94.38	92.50	91.25	65.00	12.50	–
	128	98.75	97.50	96.25	96.25	90.00	52.50	12.50
625	16	82.50	73.13	65.00	12.50	–	–	–
	32	96.88	91.88	85.00	60.00	12.50	–	–
	64	98.75	96.25	95.63	86.25	59.38	12.50	–
	128	98.75	98.75	98.75	97.50	70.63	39.38	12.50
312	16	93.13	75.00	46.25	12.50	–	–	–
	32	97.50	96.25	94.38	48.75	12.50	–	–
	64	98.75	98.75	96.25	90.00	55.00	12.50	–
	128	98.75	98.75	98.50	96.25	87.50	47.50	12.50
156	16	82.50	78.13	51.25	12.50	–	–	–
	32	91.25	90.00	68.75	48.75	12.50	–	–
	64	91.88	91.25	83.75	62.50	46.88	12.50	–
	128	93.25	91.25	86.88	80.00	53.75	36.25	12.50

Table 2. Test Results of Flute Musical Instrument, in Various Combinations of
Sampling Frequency,Frame Blocking Length, and Segment Length.
Results Shown: Recognition Rate (%)

Sampling frequency (Hz)	Frame blocking length (points)	Segment length (points)						
		1	2	4	8	16	32	64
5000	16	85.00	83.75	55.00	12.50	–	–	–
	32	97.50	96.88	94.38	46.25	12.50	–	–
	64	100	100	97.50	91.25	43.75	12.50	–
	128	100	100	100	95..63	83.13	36.88	12.50
2500	16	94.37	93.13	49.37	12.50	–	–	–
	32	97.50	97.50	80.00	45.63	12.50	–	–
	64	100	100	98.75	79.38	31.25	12.50	–
	128	100	100	100	96.88	73.75	23.75	12.50
1250	16	97.50	88.75	26.88	12.50	–	–	–
	32	100	100	95.63	45.00	12.50	–	–
	64	100	100	100	93.75	38.13	12.50	–
	128	100	100	100	95.00	85.00	48.75	12.50
625	16	86.88	64.38	23.75	12.50	–	–	–
	32	95.63	96.88	66.25	30.00	12.50	–	–
	64	99.38	97.50	95.00	61.88	36.25	12.50	–
	128	100	100	100	90.63	61.25	45.63	12.50
312	16	93.75	85.00	63.13	12.50	–	–	–
	32	96.88	95.63	85.63	63.13	12.50	–	–
	64	98.75	97.50	87.50	72.50	58.75	12.50	–
	128	98.75	98.75	98.75	87.50	74.38	70.63	12.50
156	16	90.00	88.13	38.75	12.50	–	–	–
	32	91.25	90.00	81.25	47.50	12.50	–	–
	64	93.75	92.50	88.75	78.13	50.63	12.50	–
	128	93.75	93.75	88.125	85.00	67.50	50.00	12.50

Table 3. Test Results of Pianica Musical Instrument, in Various Combinations of Sampling
Frequency, Frame Blocking Length, and Segment Length.
Results Shown: Recognition Rate (%)

Sampling frequency (Hz)	Frame blocking length (points)	Segment length (points)						
		1	2	4	8	16	32	64
5000	16	98.13	92.50	52.50	12.50	–	–	–
	32	100	99.38	93.13	49.37	12.50	–	–
	64	100	100	100	91.88	61.25	12.50	–
	128	100	100	100	100	94.38	61.25	12.50
2500	16	75.00	39.37	22.50	12.50	–	–	–
	32	100	100	78.13	36.88	12.50	–	–
	64	100	100	100	93.75	46.25	12.50	–
	128	100	100	100	100	94.38	37.50	12.50
1250	16	93.75	81.88	40.00	12.50	–	–	–
	32	100	100	97.50	63.13	12.50	–	–
	64	100	100	100	99.38	72.50	12.50	–
	128	100	100	100	100	98.75	80.63	12.50
625	16	92.50	64.38	34.38	12.50	–	–	–
	32	100	100	90.63	50.00	12.50	–	–
	64	100	100	100	98.13	53.13	12.50	–
	128	100	100	100	100	95.63	56.88	12.50
312	16	74.38	70.00	45.00	12.50	–	–	–
	32	100	100	76.88	49.38	12.50	–	–
	64	100	100	100	97.50	50.00	12.50	–
	128	100	100	100	100	96.25	50.00	12.50
156	16	88.13	61.88	41.25	12.50	–	–	–
	32	100	100	71.88	34.38	12.50	–	–
	64	100	100	100	89.38	51.88	12.50	–
	128	100	100	100	99.38	95.00	56.25	12.50

From the point of view of signal reconstruction, in general, when the sampling frequency decreases, the aliasing area (which is a high frequency region) increases. Therefore, the effect obtained from decreasing the sampling frequency is low pass filtering by decreasing the cutoff frequency. By the same reasoning, from the point of view of signal sampling, if the sampling

frequency decreases, low pass filtering (with a decrease in the cutoff frequency) will also appear. The effect is that the number of significant frequency components that can be obtained will be fewer. Therefore, if the number of significant frequency components (which in this case is represented by a number of significant local peaks in the DFT domain) of the signal becomes fewer, the recognition process will be more sensitive, when the frequency of sampling decreases. This is caused by the number of significant frequency components that are used to distinguish a tone with other tones becoming fewer. This incident will decrease the recognition rate.

From the point of view of data size, based on Tables 1 and 2, a decrease in the sampling frequency of 93.76% (from 5000 Hz down to 312 Hz) will reduce data size by 93.76%. However, a decrease in this data size will only cause a decrease in the recognition rate of less than 5%. This indicates that the tone recognition system is not sensitive to sampling frequencies that do not follow the Shannon sampling theorem.

4. Conclusion

The conducted research aims to explore if there was a lowest sampling frequency that did not follow the Shannon sampling theorem that could be used for the tone recognition system. For this purpose, we used the tone recognition system using segment averaging for feature extraction and template matching for classification.

Based on our experiments, until the sampling frequency is as low as 312 Hz, if the tone recognition system was used to recognize the tones that have one and a few significant local peaks in the DFT domain (such as bellyra and flute tones), the sampling frequency has a little influence on the recognition rate, since it reduced the recognition rate in the range of 5%. However, if the tone recognition was used to recognize the tones that have many significant local peaks in the DFT domain (such as pianica tones), that sampling frequency has no influence on the recognition rate. If the reduced recognition rate (in the range of 5%) could be accepted, the sampling frequency as low as 312 Hz could be used for the musical instruments tone recognition. In other words, if that kind of reduced recognition rate could be accepted, the sampling frequency does not need to follow the Shannon sampling theorem when recording musical instrument tones.

For further research, the exploration of sampling rates that do not meet the Shannon sampling theorem can be conducted for other tone recognition systems. In this case, the tone recognition systems could use a different approach of feature extraction (not segment averaging), and a different approach of classification (not template matching).

References

[1]   Tan L, Jiang J. Digital Signal Processing: Fundamentals and Applications 2nd Ed. *Elsevier Inc.* Oxford. 2013: 15-56.
[2]   Cheveigné A, Kawahara H. YIN, A Fundamental Frequency Estimator for Speech and Music. *The Journal of the Acoustical Society of America*. 2002: 111-117.
[3]   McLeod P, Wyvill G. *A Smarter Way to Find Pitch*. Proceedings of the International Computer Music Conference (ICMC'05). Barcelona. 2005.
[4]   Tadokoro Y, Yamaguchi M. *Pitch Estimation of Polyphony Based On Controlling Delays of Comb Filters for Transcription*. Proceedings of 11th DSP Workshop. Taos Ski Valley, New Mexico. 2004: 371-375.
[5]   Tadokoro Y, Saito T, Suga Y, Natsui M. *Pitch Estimation for Musical Sound Including Percussion Sound Using Comb Filters and Autocorrelation Function*. Proceedings of the 8th WSEAS International Conference on Acoustics & Music: Theory & Applications. Vancouver. 2007: 13-17.
[6]   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.

[7]   Brown JC, Puckette MS. A High Resolution Fundamental Frequency Determination based on Phase Changes of the Fourier Transform. *The Journal of the Acoustical Society of America*. 1993; 94(2): 662-667.

[8]   Mitre A, Marcelo Q, Régis F. *Accurate and Efficient Fundamental Frequency Determination from Precise Partial Estimates*. Proceedings of the 4th AES Brazil Conference. Brazil. 2006: 113-118.

[9]   Pertusa A, Inesta JM. *Multiple Fundamental Frequency Estimation using Gaussian Smoothness*. Proceedings of IEEE International Conference on Audio, Speech, and Signal Processing (ICASSP 2008). Las Vegas. 2008: 105–108.

[10]   Yeh C, Robel A, Rodet X. Multiple Fundamental Frequency Estimation and Polyphony Inference of Polyphonic Music Signals. *IEEE Transactions on Audio, Speech, and Language Processing*. 2010; 18(6):1116–1126.

[11]   Duan Z, Pardo B, Zhang C. Multiple Fundamental Frequency Estimation by Modeling Spectral Peaks and Non-peak Regions. *IEEE Transactions on Audio, Speech, and Language Processing*. 2010; 18 (8):2121–2133.

[12]   Smaragdis P, Brown JC. *Non-negative Matrix Factorization for Polyphonic Music Transcription*. Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2003). New Palz, New York. 2003: 177–180.

[13]   Bertin N, Badeau R, Richard G. *Blind Signal Decompositions for Automatic Transcription of Polyphonic Music: NMF and K-SVD on the Benchmark*. Proceedings of IEEE International Conferenece on Acoustics, Speech and Signal Processing (ICASSP 2007). Honolulu, Hawaii. 2007: I-65 – I-68.

[14]   Woojay J, Changxue M, Yan C. *An Efficient Signal-matching Approach to Melody Indexing and Search using Continuous Pitch Contours and Wavelets*. Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009). Kobe. 2009: 681–686.

[15]   Tjahyanto A, Suprapto YK, Wulandari DP, Spectral-based Features Ranking for Gamelan Instruments Identification using Filter Techniques. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2013; 11(1): 95-106.

[16]   O'Hanlon K, Nagano H, Keriven N, Plumbley MD. Non-negative group sparsity with subspace note modelling for polyphonic transcription. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*. 2016; 24(3): 530–542.

[17]   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*. 2016; 11(9): 5644-5649.

[18]   Sumarno L, Iswanjono. Feature Extraction of Musical Instrument Tones using FFT and Segment Averaging. *TELKOMNIKA Telecommunicatin Computing Electronics and Control*. 2017; 15(3): 1280-1289.

[19]   Meseguer NA. Speech Analysis for Automatic Speech Recognition. MSc Thesis. NTNU. Trondheim; 2009: 4-25.

[20]   Harris FJ. *On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform*. Proceedings of the IEEE. 1978; 66(1): 51-83.

[21]   Jenkins WK. Fourier Methods for Signal Analysis and Processing. In: Madisetti VK. Editor. *The Digital Signal Processing Handbook 2nd Ed: Digital Signal Processing Fundamentals*. CRC Press. Boca Raton. 2010: 1-1 - 1-29.

[22]   Setiawan YR. Numbers Speech Recognition using Fast Fourier Transform and Cosine Similarity (in Indonesian). Undergraduate Thesis. Sanata Dharma University. Yogyakarta; 2015: 66-70.

[23]   Jain AK, Duin, RPW, Mao J. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(1): 4-37.

[24]   Theodoridis S, Koutroumbas K. Pattern Recognition. 4th Ed. Elsevier Inc. San Diego California. 2009: 481-519.

[25]   Wilson DR, Martinez TR. Improved Heterogeneous Distance Function. *Journal of Artificial Intelligence Research*. 1997; 6: 1-34.