“I hereby acknowledge that the scope and quality of this thesis is qualified for the award of the Bachelor of Electrical Engineering (Electronics)”

Signature : ______________________________

Name : DR. SUNARDI

Date : 20 JUNE 2012

SOUND IDENTIFICATION IN THE SWIFTLET HOUSE

AIDA SHAFINA BINTI SHOFI

This thesis is submitted as partial fulfillment of the requirements for the award of the Bachelor of Electrical Engineering (Electronics)

Faculty of Electrical & Electronics Engineering
Universiti Malaysia Pahang

JUNE, 2012

"All the trademark and copyrights use herein are property of their respective owner. References of information from other sources are quoted accordingly; otherwise the information presented in this report is solely work of the author."

Signature : ______________________________

Author : AIDA SHAFINA BINTI SHOFI

Date : 20 JUNE 2012

ACKNOWLEDGEMENT

First of all, Alhamdulillah to Allah the almighty for granting me good health and patience upon completing this project. I was very fortunate to have a number of wonderful and supportive people who have been giving me assistance and direction towards completing this project.

I would like to express my utmost gratitude to my project supervisor, Dr. Sunardi for his advice, aid and guidance in order to realize the development and implementation of this project. He has been kind enough to facilitate and let me explore my project without hindering creativity and freedom from me.

I am also thankful to my family and friends who have been giving their full support to me all these while. Their encouragement, support, care and comfort have been equally vital to me towards the completion of this project.

ABSTRACT

Swiftlets farming industry has the potential to grow into a multi-million ringgit industry due to the market demanding on using edible birds' nests as base materials for producing natural and organic health supplement products. To encourage the Swiftlets making more nests, their cave-like house should have a great sound identity to make them stay convenient and permanently in that building as if they are in caves. Therefore, it is very important to discover the vocalization of the Swiftlet chirps and mating sounds before being played in the cave-like building. The characteristics of recorded sound will be analyzed using the Sound Analysis Pro (SAP 2011) software. The software will display a desired waveform and the sound will be analyzed based on several featured such as Wiener entropy, pitches, frequency modulation and spectral continuity. The final product is the identification of desired sound that may attract more Swiftlets to make nests in the cave-like building. This will encourage more Swiftlets to mate, play and stay permanent in the areas that have a desired quality of sound as they interest.

ABSTRAK

Industri penternakan burung Walit mempunyai potensi untuk menjana pendapatan berbillion ringgit selaras dengan permintaan sarang burung dipasaran bagi menghasilkan produk suplemen kesihatan yang organik dan asli. Bagi menggalakkan burung Walit menghasilkan lebih banyak sarang, rumah burung Walit yang seakan-akan gua itu haruslah mempunyai identiti bunyi yang bagus untuk menggalakkan Walit selesa dan menjadikan rumah burung tersebut sebagai habitat mereka seolah-olah seperti tinggal di dalam gua. Oleh itu, mengkaji vokal kicauan dan pengawanan Walit adalah sangat penting sebelum bunyi itu dimainkan di rumah Walit. Karakter bunyi yang dirakamkan dianalisis menggunakan aplikasi "Sound Analysis Pro 2011" (SAP 2011). Aplikasi tersebut akan memaparkan gelombang dan bunyi dianalisa berdasarkan beberapa ciri seperti entropi Wiener, nada, modulasi frekuensi dan kesinambungan spektrum. Produk akhir adalah penentuan bunyi yang dikehendaki dimana bunyi tersebut boleh menarik lebih banyak Walit untuk membuat sarang di dalam bangunan yang seakan-akan gua. Ini akan menggalakkan lagi lebih banyak Walit untuk mengawan, bermain dan tinggal lebih lama di kawasan yang mempunyai kualiti bunyi yang mereka inginkan.

CONTENT

CHAPTER	CONTENT	PAGE
	ACKNOWLEDGEMENT	iii
	ABSTRACT	iv
	ABSTRAK	v
	CONTENT	vi
	LIST OF TABLES	ix
	LIST OF FIGURES	x
1	**INTRODUCTION**	1
	1.1 Background of Study	1
	1.2 Overview of Swiftlets	2
	1.3 Problem Statement	4
	1.4 Objectives	5
	1.5 Project Scope	5
	1.6 Project Outline	6
2	**LITERATURE REVIEW**	**8**
	2.1 Echolocation	8
	2.2 Review of Previous Research	9
	2.2.1 Characterization on animal behavior through the use of audio and video signal processing	10
	2.2.2 A Procedure for an Automated Measurement of Song Similarity	11
	2.2.3 Deafening-Induced Vocal Deterioration in Adult Songbirds Is Reversed by Disrupting	

		a BasalGanglia-Forebrain Circuit	12
	2.2.4	Dual Pre-Motor Contribution to Songbird Syllable Variation	13
	2.2.5	Spatial pattern of Song Element Sharing and its Implications for Song Learning in the Chowchilla,Orthonyxspaldingii	14
3		**METHODOLOGY**	16
	3.1	Overview	16
	3.2	Partitioning of Sound Motif	18
	3.3	Integration of Song Measures	19
	3.4	Features Used to Characterize and Compare Songs	21
		3.4.1 Wiener Entropy	22
		3.4.2 Spectral Continuity	24
		3.4.3 Pitch	27
		3.4.4 Frequency modulation	29
	3.5	Similarity Measurements	30
4		**RESULT AND DISCUSSIONS**	32
	4.1	Introduction	32
	4.2	Reducing Scoring Ambiguity	32
	4.3	Integration of the Song Features	37
	4.4	The Internal Sound	37
	4.5	The External Sound	39
	4.6	Comparison of Sound Features	41
	4.7	Symmetric Comparison of Sound Features	43
		4.7.1 Symmetric Comparison of Internal sound	44
		4.7.2 Symmetric Comparison of External Sound	47
	4.8	Discussion	53

5	**CONCLUSION AND RECOMMENDATIONS**		**54**
	5.1	Introduction	54
	5.2	Limitations	54
	5.3	Suggestion for Future Experiments	55
	5.4	Conclusion	56
	REFERENCES		57

LIST OF TABLES

TABLE NO.	TITLE	PAGE
4.1(a)	Feature Scores for External Sound 1	33
4.1(b)	Similarity Score of Sound 1 and Sound 2 at 32-ms	34
4.1(c)	Similarity Score of Sound 1 and Sound 2 at 550-ms	35
4.1(d)	Similarity Score of Sound 1 and Sound 2 at 1030-ms	36
4.2	Feature Scores for Internal Sound 1	38
4.3	Feature Scores for External Sound 1	39
4.4	Feature Scores for External Sound 2	40
4.5	Similarity Score for Internal Sound 1.1 and Sound 1. 2	44
4.6	Similarity Score for Internal Sound 1.1 and Sound 1. 3	45
4.7	Similarity Score for Internal Sound 1.2 and Sound 1. 3	46
4.8	Similarity Score for External Sound 10.1 and Sound 10. 2	47
4.9	Similarity Score for External Sound 10.1 and Sound 10. 3	48
4.10	Similarity Score for External Sound 10.2 and Sound 10. 3	49
4.11	Similarity Score for External Sound 5.1 and Sound 5. 2	50
4.12	Similarity Score for External Sound 5.1 and Sound 5.3	51
4.13	Similarity Score for External Sound 5.2 and Sound 5. 3	52

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	Swiftlet on its Saliva	3
1.2	Modern Cave-like Nest Farm	3
1.3	Nest Ready to be Sold	4
2.1	Echo click	9
3.1	Flow of Procedures for Sound Analysis	17
3.2	Splitting Sound Using Audacity 1.3.14-beta (Unicode)	18
3.3	Sound Conversion Using dBpoweramp	19
3.4	Measurement of Sound Features	20
3.5	Similarity Score between Two Sound	21
3.6	Feature Expression	22
3.7	Level of Wiener Entropy	23
3.8	Range of Entropy	23
3.9	Computation of Spectral Derivatives and Frequency Contours	25
3.10	Features of Song Chunks	27
3.11	Sample of Goodness of Pitch	28
3.12	Level of Frequency Modulation	29
3.13	Steps of Overall Similarity	31
4.1(a)	Similarity matrix of External sound at 32-ms	34
4.1(b)	Similarity matrix of External sound at 550-ms	35
4.1(c)	Similarity matrix of External sound at 1030-ms	36
4.2	Graph of measurement of features for Internal sound 1	41
4.3	Graph of measurement of features for External sound 1	42
4.4	Graph of measurement of features for External sound 2	43
4.5	Similarity matrix of Internal sound 1.1 and 1.2	44

4.6	Similarity matrix of Internal sound 1.1 and 1.3	45
4.7	Similarity matrix of Internal sound 1.2 and 1.3	46
4.8	Similarity matrix of External sound 10.1 and 10.2	47
4.9	Similarity matrix of External sound 10.1 and 10.3	48
4.10	Similarity matrix of External sound 10.2 and 10.3	49
4.11	Similarity matrix of External sound 5.1 and 5.2	50
4.12	Similarity matrix of External sound 5.1 and 5.3	51
4.13	Similarity matrix of External sound 5.2 and 5.3	52

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Vocalizations of Swiftlets are particularly interesting, because Swiftlets are almost unique among birds by the ability of some species to echolocate. Their ability is to use a simple yet effective form of echolocation as a navigator in the darkness through the chasm and shafts of the caves where they roost at night and breed. Songs and calls play a very important role in the lives of Swiftlets as they do not have a strong sense of smell that they can rely only on vision and sound. The sounds that Swiftlets use for echolocation are rather stereotyped clicks, which are usually double click, but sometimes also of a single click design.

Due to the market demanding on using edible birds' nests, Swiftlets had become a farming industry. Nowadays, the Swiftlets' farmers will record the sounds from the original habitats of in caves and play it through speakers which had been installed in the new Swiftlets' building to make the surrounding exactly as in the cave of the Swiftlets' original habitats. The Swiftlets sounds are different for the Internal and External of the farm. To maintain the number of nests produced, the Swiftlets farm itself have to be convenient and the most important thing is to have a good mating and chirps sound.

1.2 Overview of Swiftlets

Swiftlets are small Swifts as can be seen in Figure 1.1. People often confused with Swallows as because they are having resemble appearance and obtain a similar food niche of aerial insects. Swiflets are insectivorous and first-class flyers with high velocity and maneuverability. By having prominent flying abilities, this enables them to catch their prey in the air or from leaves while fluttering through tree canopies. Swiftlets occur at all altitudes, from lowland to high in the mountains. In group, they occupy a rather wide range of feeding habitat, covering dense forests, open cultivated land, barren mountain ranges, and even towns, which provide buildings as modern cave-like nest farms. One of the notable features of Swiftlets is that they roost and nest in caves or cave-like structures as shown in Figure 1.2.

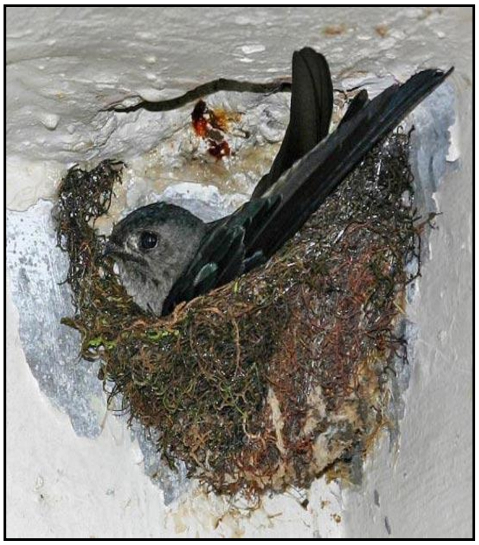


Figure 1.1 Swiftlet on its saliva-Cemented nest in cave-like building.

Figure 1.2 Building as Modern cave-like nest farm

Swiftlets are probably best known for the edible nests that are produced by several species. In early seventh century, their nests were famous among the Chinese, who attribute healing powers to the nests. This is the reason why many Chinese make tinctures and soups with Swiftlet nests as one of the ingredients. The edible part of the nests consists of the hardened Swiftlet saliva that is used to glue twigs or feathers together and to the wall of a cave as shown in Figure 1.1.

The most valuable nests are the nests that completely consist of saliva. Present day prices are up to a few thousand US dollars for one kilo of good quality, cleaned nests, which are ready for consumption as shown in Figure 1.3. Thus, Swiftlet farming has become a lucrative business. Swiftlet colonies are therefore carefully nursed and protected against nest thieves. The nests are harvested twice during a breeding season, where once just before the first eggs are laid and once after the chicks have fledged.

Figure 1.3 Box of nests ready to be sold for consumption

1.3 Problem Statements

Vocal abilities may have played an important role in the evolution of songbirds. The sound recorded from the original habitats of the Swiftlets does not undergo any further analysis. The farmers will simply played the recorded sound through speakers to attract more Swiftlets enter and stay permanently in the building. The characteristics of Internal and External sounds of the Swiftlets are very important to be identified as both sounds may be different in features and spectrums.

The echo click features will influence the number of Swiftlets entering the building and also the quality and quantity of nests produced. However, as yet less attention has been paid to inter specific comparison of the structure of Internal and External of echo clicks. Hence, further analysis on the structure of spectrum produced by both Internal and External sounds make it easier for the farmers to distinguish between both echo clicks. Because vocalizations appear to play such an important role in Swiftlets farm, it is very crucial to identify the behavior and features of the sound in the Swiftlets house.

1.4 Objectives

The objectives of this project are:

i. To analyze the Swiftlets' vocalizations and features from different location in the Swiftlet house.
ii. To analyze the similarity scores of feature characteristics and compare the results between Internal and External song motif of Swiftlets.

1.5 Project Scope

There are several scopes that need to be accomplished upon doing this project. The details are as follows:

i. The sounds will be analyzed from different location of the Swiflets habitats which is Internal and External sounds.

ii. The Internal sound is being recorded in the area of the cave where the Swiftlets stayed permanently and make nests.

iii. The External sound is at the location at the entrance of the cave, where the route that Swiftlets usually used to enter and exit.

iv. Each sample of sounds will be measured based on 4 features comprising Wiener entropy, spectral derivatives, pitches and frequency modulation.

v. The characteristics of song motif will be determined based on the scores of the features.

vi. The similarity of features between several samples will be measured for both Internal and External sounds.

1.6 Project Outline

This thesis comprise 5 chapters altogether. Chapter 1 is on introduction, Chapter 2 is literature review, Chapter 3 contains methodology, Chapter 4 is on result and discussion followed by last chapter which is Chapter 5 is conclusion.

In Chapter 1, here will be discussed about the introduction of the thesis. An overview of the Swiftlets phylogeny will be outlined including problem statement, the projects' objective and also the project scope.

For Chapter 2, there will be discussions on the literature review of previous thesis, journals and experiments that relates with my project. The reviews are about birds' behavior, and the methods that they used to analyze sound.

Chapter 3 represents the research methodology of this project. The step to run this project from the beginning till end will be explained here in detail. This will also include the procedures and processes involved for the software development of the entire project.

Chapter 4 consists of the experimental results and its respective analysis. The result obtained will be discussed and explained here with the aid of figures and tables. The comparison of the song features as well as the similarity measurement also will be stated here.

In Chapter 5, here would be the summarization of this thesis and suggestions that can be done to upgrade the analysis method for further research in the future.

CHAPTER 2

LITERATURE REVIEW

2.1 Echolocation

One interesting feature of Swiftlets is that many species utilize a sonar-like system to navigate in the dark area of caves that they inhabit. Through this way of navigation, called echolocation, the bird produces a click-like sound, of which the echo provides information about the bird's speed and position relative to an object. Other mammals group such as bats, whales and dolphins are capable to echolocate. Above all, Swiftlets are virtually unique among birds in terms of the ability to echolocate. The frequencies of echolocation sounds that are used by Swiftlets are much lower than those of bats until it is not suitable to detect small objects. Bats use echolocation to locate their small insect prey. Echolocation acuity of Swiftlets was examined by several researchers [2],[3],[4], who found the smallest detectable objects to range between 6.3 mm and 10 mm. Thus, with prey items that are much smaller than the smallest detectable objects, Swiftlets can only use echolocation as a navigator.

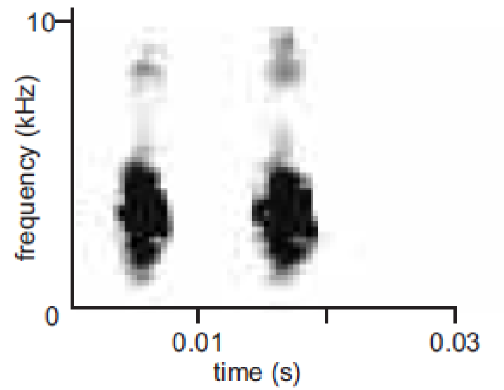


Figure 2.1 Echo click of *Aerodramusvulcanorum*, typical for echolocating Swiftlets. The click is of double click design, with a very short interval.

The echolocation sound of Swiftlets is a short click-like sound with a sharp on and offset as shown in Figure 2.1.

2.2 Review of Previous Research

Findings and the development of animal sound features and characteristics would be discussed under this section.

2.2.1 Characterization on animal behavior through the use of audio and video signal processing

In this paper, authors used two examples of video and audio by analyzing long series of behavioral data to quantify the characterization of animal behavior. To obtain complete characterization, a long time series are ideal. In previous time, people used visual inspection of spectrograms (time-frequency plots) for characterization of birdsong [6].

Even this method was insightful and yielded many important findings, it does not an entirely objective technique and does not obtain accurate scale to high-throughput experiments. We circumvent these problems by using time and frequency derivatives of the spectrogram, which can be used to accurately detect events (e.g., peaks) in the time frequency plane. Moreover, multitaper (MT) spectral methods are being used to calculate spectrograms, which produce superior estimates to the traditional spectrogram [7], [8].

The obvious features of the motion of an oscillating membrane are the period and the regularity (entropy) of the oscillation. They are reflected in sound as pitch and Wiener entropy, respectively. Pitch is defined as the median difference between consecutive harmonics. While Wiener entropy is defined as the ratio of the geometric mean to the arithmetic mean of the spectrum S(*f*) or can be simplified as the spectrogram within a time window.

$$W = \log \left(\frac{\exp[\int df Log(S(f))]}{\int df S(f)} \right)$$

As the songs progress with time, it is necessary to measure how pitch changes with time by frequency modulation. Frequency modulation is defined as the angle of the maximum directional derivative of the spectrogram. Another metric to measure the progress of sound in time is spectral continuity. Spectral continuity is the continuity of frequency contours across time windows. Thus, each time window of the spectrogram can be represented by feature vector of the four components.

2.2.2 A Procedure for an Automated Measurement of Song Similarity

This paper discuss about measuring the sound similarities between the song of a tutor and that of its pupil. A fully automated procedure was presented to measures parametrically the similarity between songs. The performance has been tested on a large database of zebra finch songs. This analysis provides a superior sound spectrogram that is then reduced to a set of simple acoustic features. Based on these features, the procedure detects similar sections between songs automatically.

In advance, the procedure can be used to examine others such as : (1) imitation accuracy across acoustic features; (2) song development; (3) the effect of brain lesions of specific song features; and (4) variability across different renditions of a song or a call produced by the same individual, across individuals and across populations.

2.2.3 Deafening-Induced Vocal Deterioration in Adult Songbirds Is Reversed by Disrupting a Basal Ganglia-Forebrain Circuit

In detail, this paper showed the output of an anterior forebrain pathway (AFP) through the basal ganglia directly contributes to the expression of deafening-induced vocal changes in adulthood. In birds, as in mammals, the contribution of basal ganglia-thalamic-cortical circuits to motor may change when feedback is absent or unexpected and includes both “active” and “permissive” roles.

The similarity between songs produced on different days was measured using the Sound Analysis Pro 2011 software using the asymmetric similarity analysis that calculates similarity based on measurement of pitch, frequency modulation (FM), amplitude modulation (AM), Wiener entropy (noisiness), and goodness of pitch (how periodic or harmonic the acoustic structure is).

Three separate measures of song similarity were obtained from Sound Analysis Pro 2011 and the details are as follows:

1) Percentage similarity (the percentage of the template (d0 motif) achieving a threshold level of similarity to the sample
2) Accuracy (the average local similarity score for those sections of the sample motif that are judged to be similar to the template)
3) Sequential match (the degree to which similar sounds in the sample and template correspond in temporal order)

In measurement and statistical analysis of deafening- and lesion-induced changes in syllable structure of availability, the features batch module in Sound Analysis Pro 2011 software were used to measure the acoustic character and consistency of identified song syllables. For each syllable, they measured for the syllable duration, mean pitch, mean FM, mean entropy, mean pitch goodness, and mean frequency. Also, to capture how much these features were modulated within each syllable, they measured the variation in mean pitch, FM, entropy, pitch goodness, and mean frequency.

2.2.4 Dual Pre-Motor Contribution to Songbird Syllable Variation

In this paper, the ability of basal-ganglia pathway has been tested to generate pre-motor vocal variation within the spectral and temporal dimensions of zebra finch song structure. The study is about the vocal control system of zebra finches using the pre-motor where the results shows that the pre-motor variation generated by the basal ganglia pathway may be sufficient to adjust vocal output toward highly acoustically dispersed targets of imitation.

It has been suggested that complete acquisition of the pronounced variation in syllable pitch that characterizes adult song will necessitate a gradual developmental interaction between the basal ganglia and vocal motor pathways. For the measurement of temporal variation in the production of individual syllables before and after LMAN ablation in adult birds, the temporal variation of song has been analyzed at three levels which is: syllable duration, intersyllable duration and motif duration. All motifs were analyzed using the Explore and Score module in the Sound Analysis software, which generates measures of the temporal onset and duration of each syllable.

The Sound Analysis Pro 2011 software was used to partition bouts into syllables and measure the spectral quality of syllables based on pitch, FM, entropy and pitch goodness.

2.2.5 Spatial pattern of Song Element Sharing and its Implications for Song Learning in the Chowchilla, Orthonyxspaldingii

Their study is about examining the spatial pattern of song similarity in the chowchilla, determining in what scale the song similarity changes, whether its pattern is indicative of the timing learning and dispersal strategy of the rain forest. They divided the songs into their syllables and created a catalogue of it. To calculate similarity index for each recorded group of songs, they used pairwise comparisons of syllable sharing. The songs were analyzed using Avisoft-SASlabPro[12].

Sound of chowcillas were divided in their syllables and stored electronically as a spectrogram for each separate syllable. Then, the spectrograms were visually sorted according to the similarity of syllable shape and frequency. There are more objective methods to classify song syllables, yet visual classification was deemed an appropriate method of subdividing each song because chowcilla songs have such visually distinct elements.

They determine the song similarity by using standard song sharing index which is as follows:

$$S = \frac{2N_S}{(R_A + R_B)}$$

N_s = number of all syllables shared in both groups' repertoires

R_A = syllable repertoire of group A

R_B = syllable repertoire of group B

The sharing index normally used to determine the number of song types shared between two birds, but in this case, it has been applied to share syllable types between two groups of birds. To test whether song similarity is correlated with distance, they used nonparametric Mantel test, which evaluates the null hypothesis of no relationship between two similarity matrices. The results show that levels of song similarity in the chowcilla are significantly higher among neighbours and decrease very rapidly and significantly 1km between groups within a site.

CHAPTER 3

METHODOLOGY

3.1 Overview

Assessment of vocal characteristics of Swiftlets requires a widely accepted way of describing and measuring several features to characterize and compare the songs. Through this chapter, the overall process that involved upon completing this project would be discussed, including the information about equation involved, as well as the procedures and processes involved in the software development and implementation of the project. The flow of the process can be summarized such in figure 3.1.

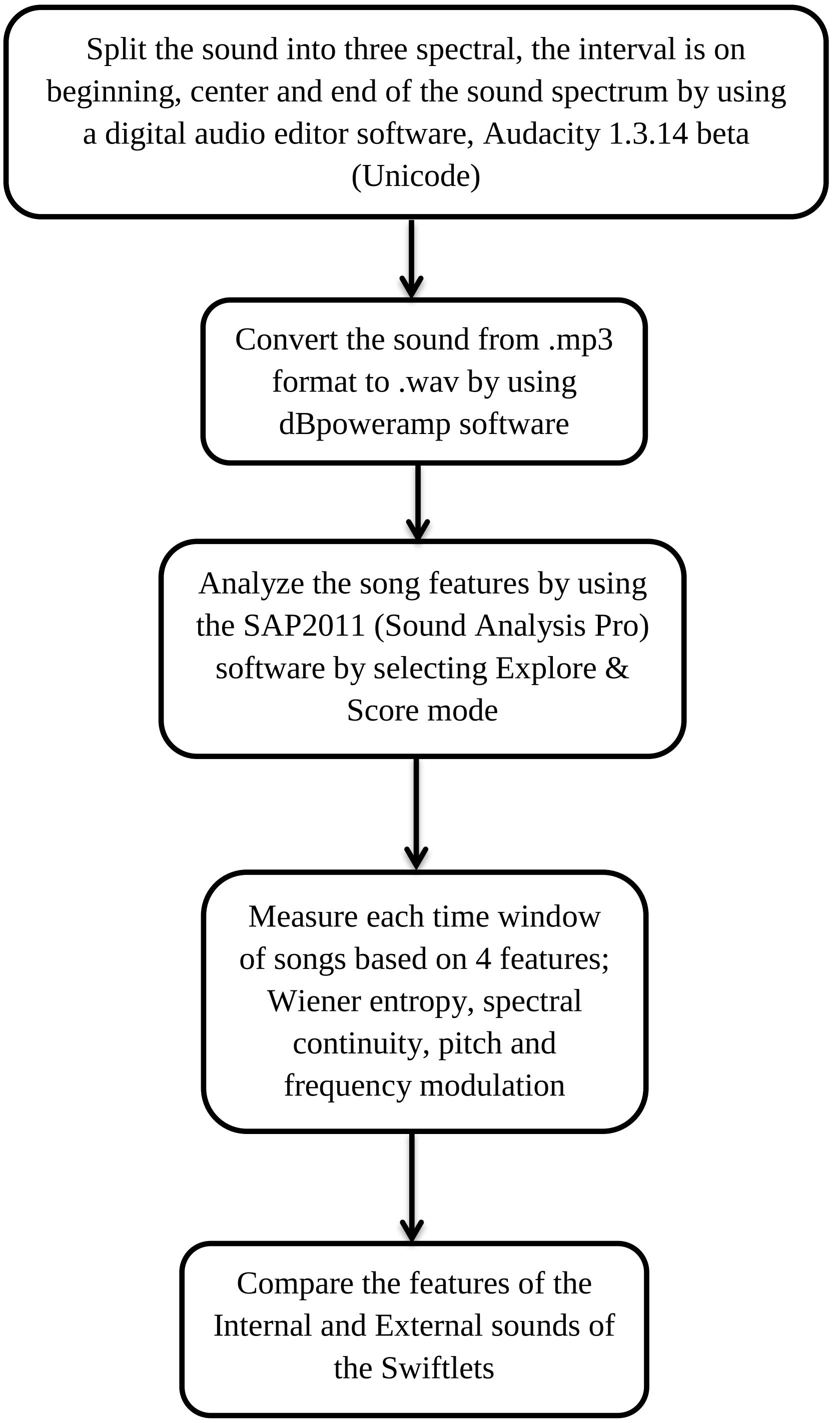


Figure 3.1 The flow of procedures for sound analysis

3.2 Partitioning of Sound Motif

The Swiftlets song motif consists of discrete notes that produce in chunks of variable sizes. The total length of the original sound is 70 minutes. For intervals that are sufficiently long, they will contain enough information to identify a unique song segment. In contrast, if the intervals are too long, it might be a problem too as the features that are at a smaller interval size may be rejected and that would reduce the power of analysis.

Hence, by taking a sample sound, the sound has been split into three different time interval, each by 10 second using the Audacity ® 1.3.14-beta (Unicode) software. For the purpose of analysis, 10 second from the beginning, middle and at the end of the interval of the sound has been taken for analyzing the vocal structure of Swiftlets sounds based on several features. The sound has been split as shown in Figure 3.2.

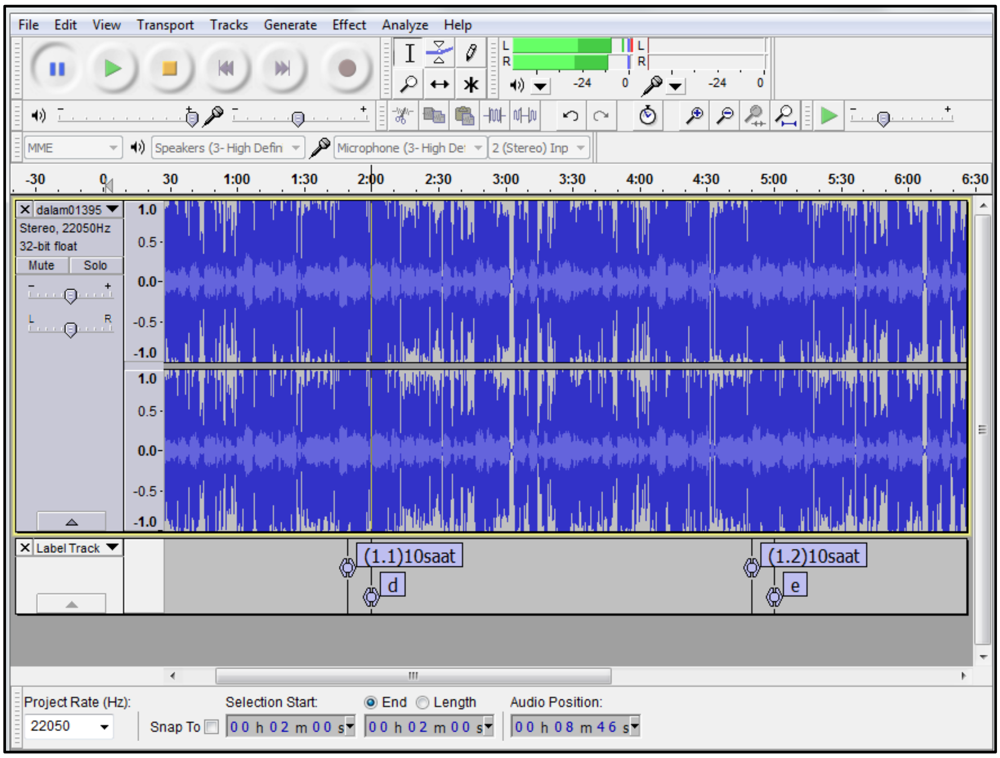


Figure 3.2 The sound is split using Audacity ® 1.3.14-beta (Unicode) software

After splitting the sound, it will be converted from .mp3 format into .wav format by using dBpoweramp Music Converter software. The sound conversion is shown in figure 3.3 below.

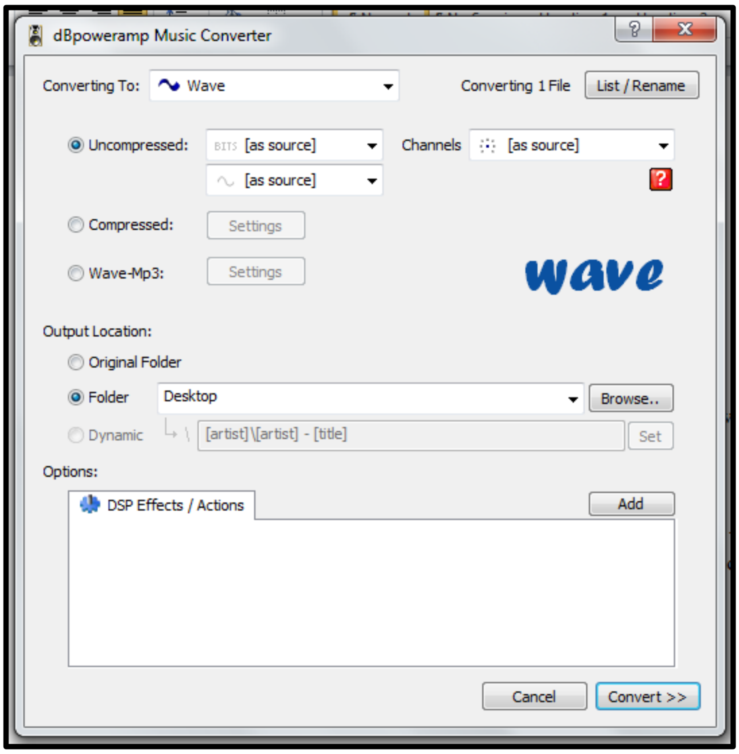


Figure 3.3 Sound format conversion using dBpoweramp software

3.3 Integration of Song Measures

Each time window of the Swiftlets sounds is represented by measurements of four features: Wiener entropy, spectral continuity, pitch and modulation. Each of these features has different units and different statistical distributions in the population of song studied. To arrive at an overall score of similarity, the units for each feature has been transformed to a common type of unit that could be added. For example, one can transform the units of pitch from Hertz to units of statistical distances. I scaled the measures based on their distribution for the several sample of internal and external songs which consist of different waveform. The software that has been used is Sound Analysis Pro 2011 and it is shown in figure 3.4 below.

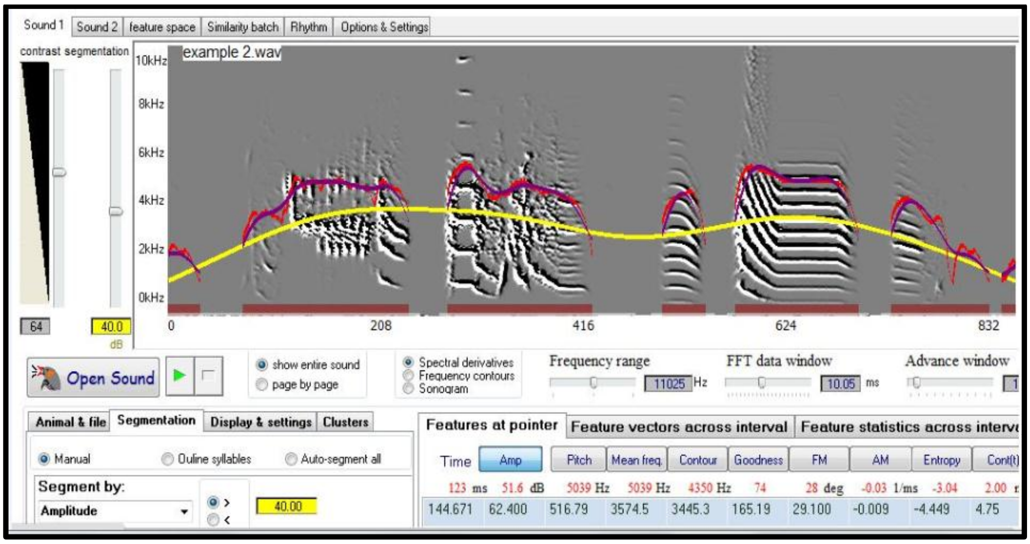


Figure 3.4 The measurement of sound features using Sound Analysis Pro 2011

The elements of the Swiftlets sounds may be repeated in each sample, as the waveform produce are almost similar. Therefore the tendency of the song features to be similar to each other is high. Using automatic procedure to recognize homologous notes can be very difficult. Therefore, I chose to examine each time window of song motif for similarity, by omitting silent intervals. This approach is allowed for the detection of discrete segments of similarity that typically emerge from the analysis. The similarity score between 2 sound intervals can be obtained as shown in figure 3.5 below.

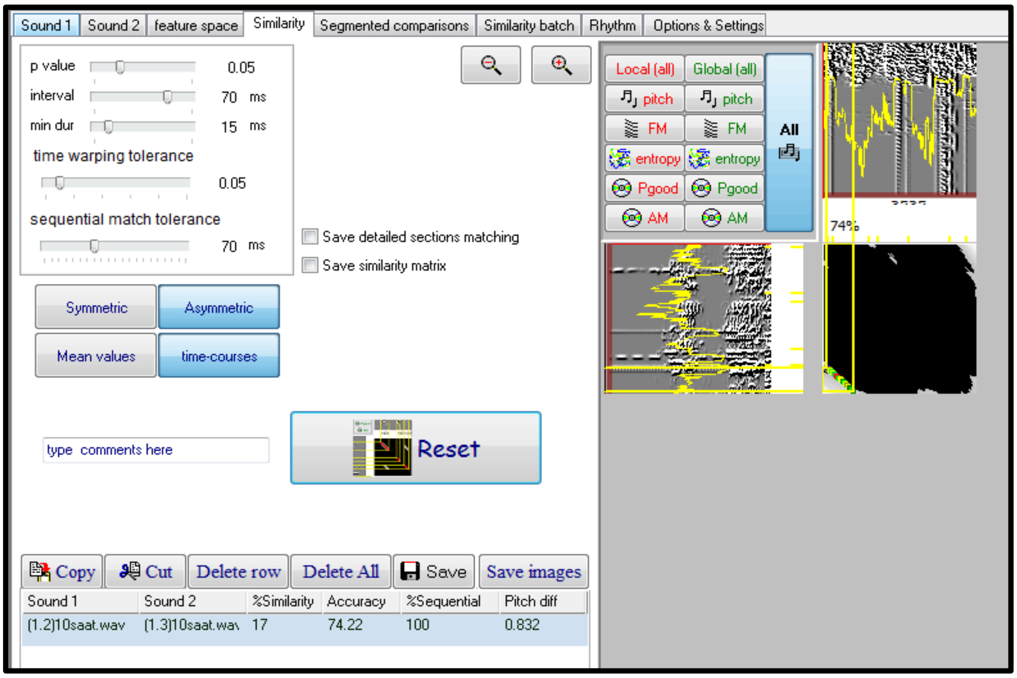


Figure 3.5 The measurement of similarity score between two sound interval

To simplify, here is the demonstration on how one measure, pitch, performs when comparing the sounds that show perfect similarity. First, I singled out a particular time window of 10 second of the internal sound and compared its measures to those of each other interval of 10 second in the similar sound sample. I repeated this procedure also to the external sound. The features then will be compared both between the internal and external sounds. The resulting matrix spans all possible combinations of pairs of internal and external sounds. The difference in pitch between each pair of window is encoded into a color scale. There is a marked tendency for the strongest similarity between pairs of window to show as a red diagonal line.

3.4 Features Used to Characterize and Compare Songs

There are four sound features that had been used to identify the characteristics of the Swiftlets song.

3.4.1 Wiener Entropy

Wiener entropy is a measure of the width and uniformity of the power spectrum. It is also suitable for measuring the randomness that can be applied to sounds as shown in figure 3.6 and figure 3.7.

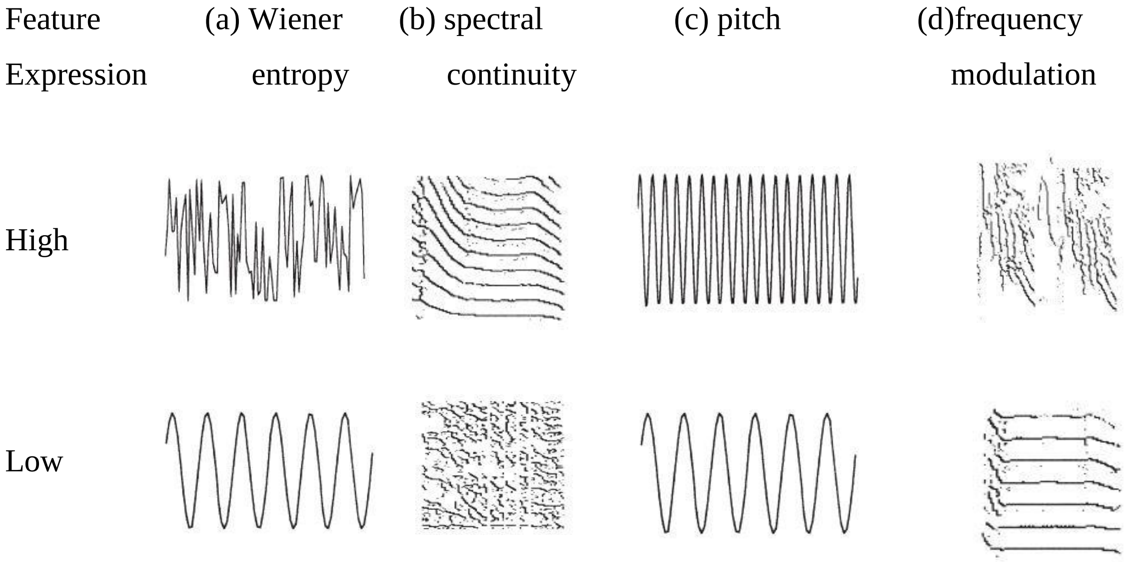


Figure 3.6 . (a) Wiener entropy (a measure of randomness) is high when the waveform is random and low when the waveform is of pure tone. (b) The spectral continuity value is high when the contours are long and low when the contours are short. (c) Pitch's is a measure of the period of the sound and its value is high when the period is short and low when the period is long. (d) Frequency modulation is a measure of the mean slope of frequency contours.

Noise is typically broadband with song energy smeared rather smoothly within the noise range, whereas animal sounds, even when multi-harmonic, are less uniform in their frequency structure as shown in figure 3.7 below.

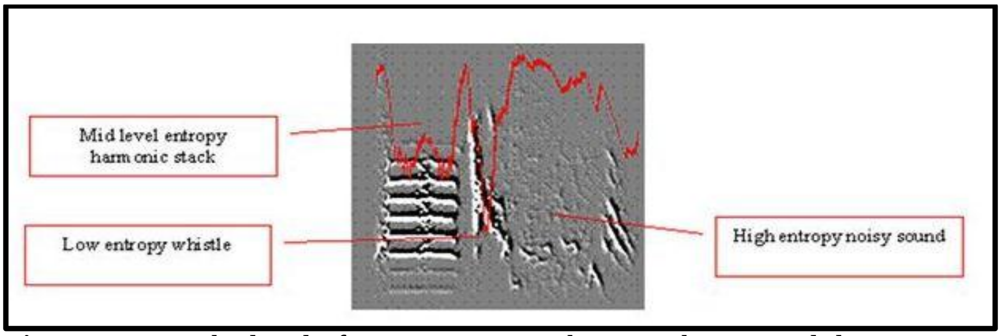


Figure 3.7 The level of Wiener Entropy base on the spectral derivatives.

Wiener entropy is a pure number which is unitless. On a scale of 0-1, the entropy value of white noise is 1 and complete order; for instance, a pure tone has an entropy value of 0. To expand the dynamic range, the Wiener Entropy is measured on a logarithmic scale, in the range of 0 to minus infinity (white noise: log1=0; complete order: log0=minus infinity). The Wiener Entropy of a multi-harmonic sound depends on the distribution of the power spectrum. For narrow power spectrum (the extreme of which is a pure tone), the Wiener Entropy approaches minus infinity; if broad, it approaches zero as illustrated in figure 3.8 below.

Figure 3.8 Range of entropy based on their spectral derivatives.

The amplitude of the sound does not affect its Wiener Entropy value. Yet, the entropy time curve of a song motif is negatively correlated with its amplitude time series. This is due to the noisy sounds that tend to have less energy than tonal sounds. A similar phenomenon has also been observed in human speech, where unvoiced phonemes have low amplitude. Wiener Entropy may also correlate with the dynamic state of the syringeal sound generator, which shifts between harmonic vibrations and chaotic states. Such transitions may be among the most primitive features of song production.

The formula used to calculate Wiener entropy is the ratios of geometric mean to arithmetic mean of the spectrum, which is as follows;

$$W = \log\left(\frac{\exp[\int df Log(S(f))]}{\int df S(f)}\right)$$

3.4.2 Spectral Continuity

Spectral continuity estimates the continuity of frequency contours across the time windows as illustrated in figure 3.6(b), figure 3.9(d), and figure 3.10(b). Frequency contours are mostly continuous in example1 shown in figure 3.9(d), but not in the more complex set of notes in example 2. Hence, it is clearly shown from figure 3.6(d) that the noisier a sound, the lower its spectral continuity score and the higher its Wiener entropy.

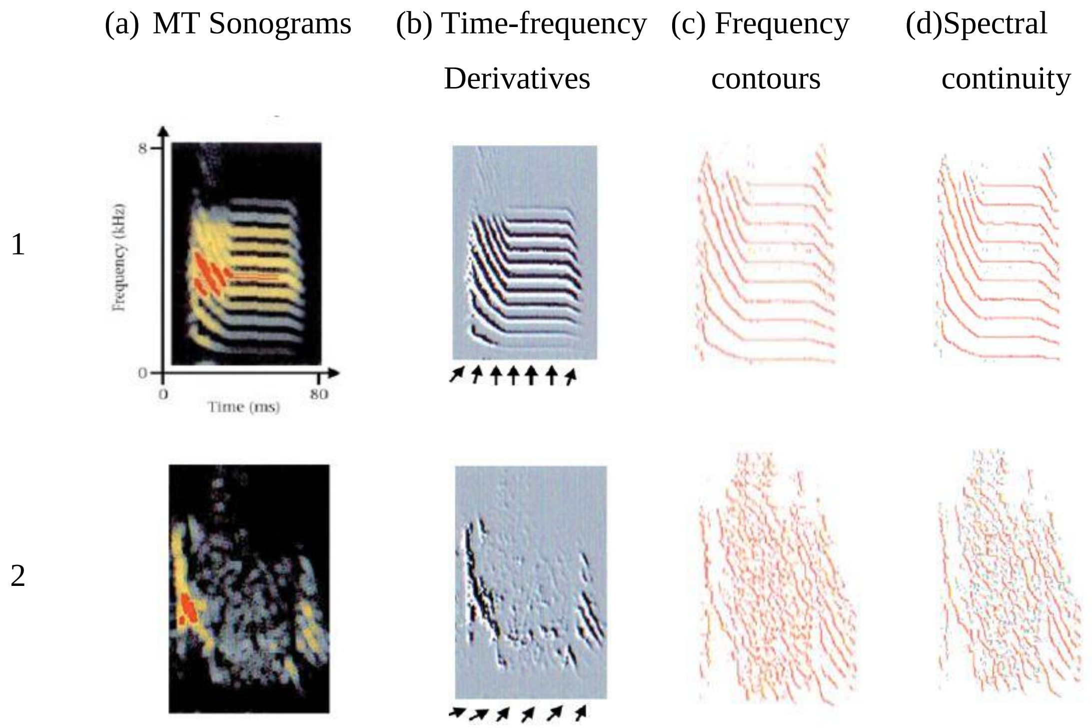


Figure 3.9 Computation of spectral derivatives and frequency contours of a note (example 2). (a) Multitaper sound spectrograph improves the definition of frequencies. This technique allows us to approximate the spectral derivatives as shown in (b), where the light areas represent an increase of power, while the dark area, the decrease of power. The arrows below the X-axis in (b) indicate the direction of the derivatives presented. The direction of the arrow chosen is to maximize the derivatives and hence the sharp transition between white and black at the middle of each frequency trace. This will allow the frequency peaks of modulated and unmodulated frequencies to be located accurately. (c) The red lines in (d) corresponds to continuous frequency contours and the grey lines indicate discontinuous contours.

Both Wiener entropy and spectral continuity measures are related to ‘noise’, where they are measured orthogonally to each other; Wiener entropy is measured on the Y-axis while spectral continuity is measured on X-axis. Although at their extremes, Wiener entropy and spectral continuity are correlated because there is a broad middle range in these two measures where one does not predict the other. For instance, adding more harmonics to a sound would not change spectral continuity but would increase the Wiener entropy value. There are two contour-derived features; which is continuity over time and continuity over frequency. Sounds are examined across a grid of time and frequency pixels. If a contour continues across five consecutive pixels, it crosses a section of approximately the resolution of the analysis, it is defined as continuous.

The detection threshold is weighted by the distance from the mean frequency (the gravity center of the frequencies) and by the width of the power spectrum. This can be simplified into an equation below:

$$\mathrm{T}(t_i, f_i) = T' \frac{abs\{(Wiener\_entropy(t)\}}{abs\{f_i - mean\_frequency_{(t_i)}\}}$$

T’ = user defined threshold

On a scale 0-1, continuity is 1 when all the frequency contours of a time window are continuous and 0 when none of the contours is non-continuous. Figure 3.10(b) presents the examples of the continuity measurement.

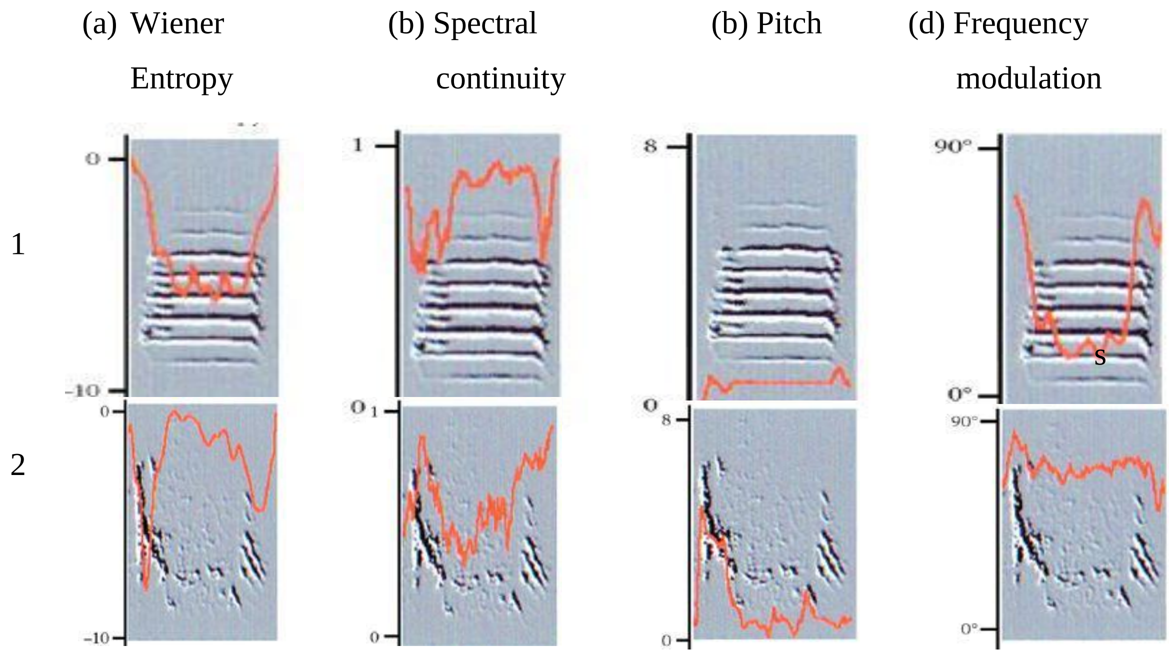


Figure 3.10 Values in red represented for each of the four features in two examples of song chunks. The grey traces correspond to the time frequency of the sounds represented. Each sound has its own unique combination of feature values.

Continuity over time is defined by for each time window t_i, all of the contours will be detected and the duration of each counter will be measured. The same goes to continuity over frequency. For each time window t_i, all of the contours will be detected and the frequency range of the contour will be measured. Yet, the mean frequency range across the contours is the continuity over frequency.

3.4.3 Pitch

The formal definition of pitch is the peak of the derivative-cepstrum calculated for harmonic pitch. Its units are comparable to amplitude and can be converted to dB by subtracting a baseline and converting to log scale. The goodness of pitch is an estimate of harmonic pitch periodicity. The high goodness of pitch can be used as a detector of harmonic stack, whether modulated or not. By referring to the figure 3.11 below, the red arrows indicate the harmonic sounds while the blue one is a clearly non-harmonic sound.

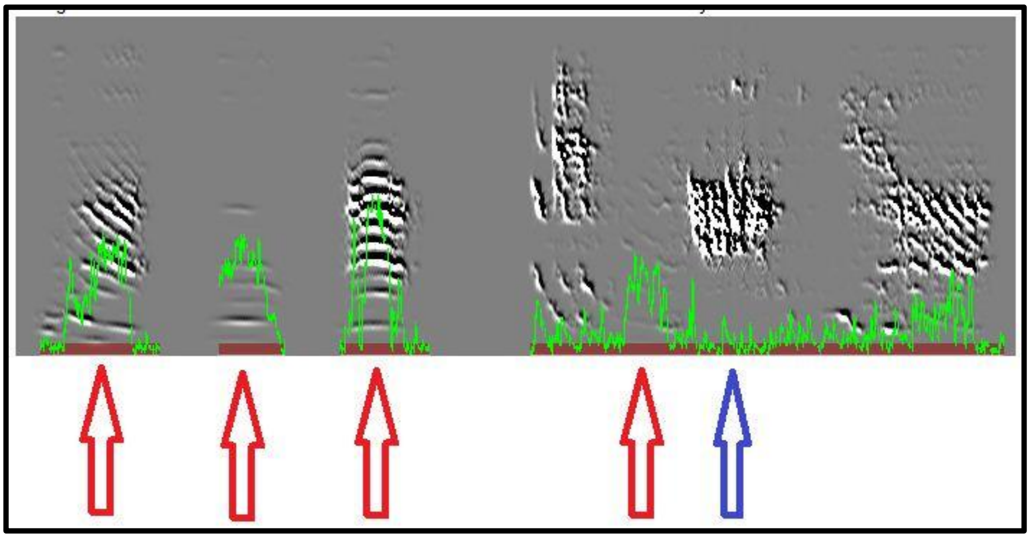


Figure 3.11 The red arrows show the sample of the goodness of pitch for the harmonic sounds while the blue arrow is the non-harmonic one.

Pitch is determined by the period of a sound. Its value is high when the period is short and low when the period is long. To measure pitch of a pure tone, the frequency of the tone is its pitch. In sounds contains many harmonics, pitch is the fundamental frequency, as defined by the separation between successive harmonics, and the median difference between consecutive frequency contours is the estimated harmonic pitch. This is a measure of how periodic the sound is. It only captures the 'goodness' of harmonic pitch, whereas both noisy sounds and pure tones give low values.

3.4.4 Frequency modulation

By referring to figure 3.10(d), frequency modulation also can be computed as spectral derivatives. It is defined as the angle of the directional derivatives as shown in figure 3.10(d). The formulae to compute frequency modulation is as follows;

$$FM = \tan^{-1}\left(\frac{Max_f \delta P_t^2 dt}{Max_f \delta P_t^2 df}\right)$$

FM is the angular component of squared time and frequency derivatives. Frequency modulation is estimated based on time frequency derivatives across frequencies. When the frequency derivatives are much higher than the time derivatives, FM is low and vice versa. Visually, FM is an estimate of the (absolute) slope of frequency traces in reference to the horizontal line. The results can be clearly seen in figure 3.12 below.

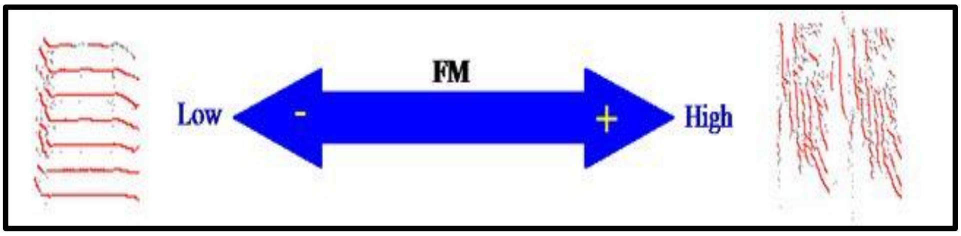


Figure 3.12 FM is low when the frequency derivatives are much higher than the time derivatives, while it is high when the frequency derivatives are much lower than the time derivatives

3.5 Similarity Measurements

Similarity measurements are necessary for comparing two complex sounds, e.g., a pair of songs or several pairs of complex syllables. The aim of analysis is to address three issues:

i. Assessing the likelihood that two sounds are related to each other.
ii. Quantifying the accuracy of the vocal match (assuming that sounds are related).
iii. Accounting for the temporal order (syntax) of sounds when scoring similarity.

Symmetric comparison is the best choice if we know in advance the sound units that should be compared and also when we have identified a cluster of syllables we might want to measure the similarity across sounds of this cluster. The following figure 3.13 shows the step of overall similarity captured by each section of the similarity process.

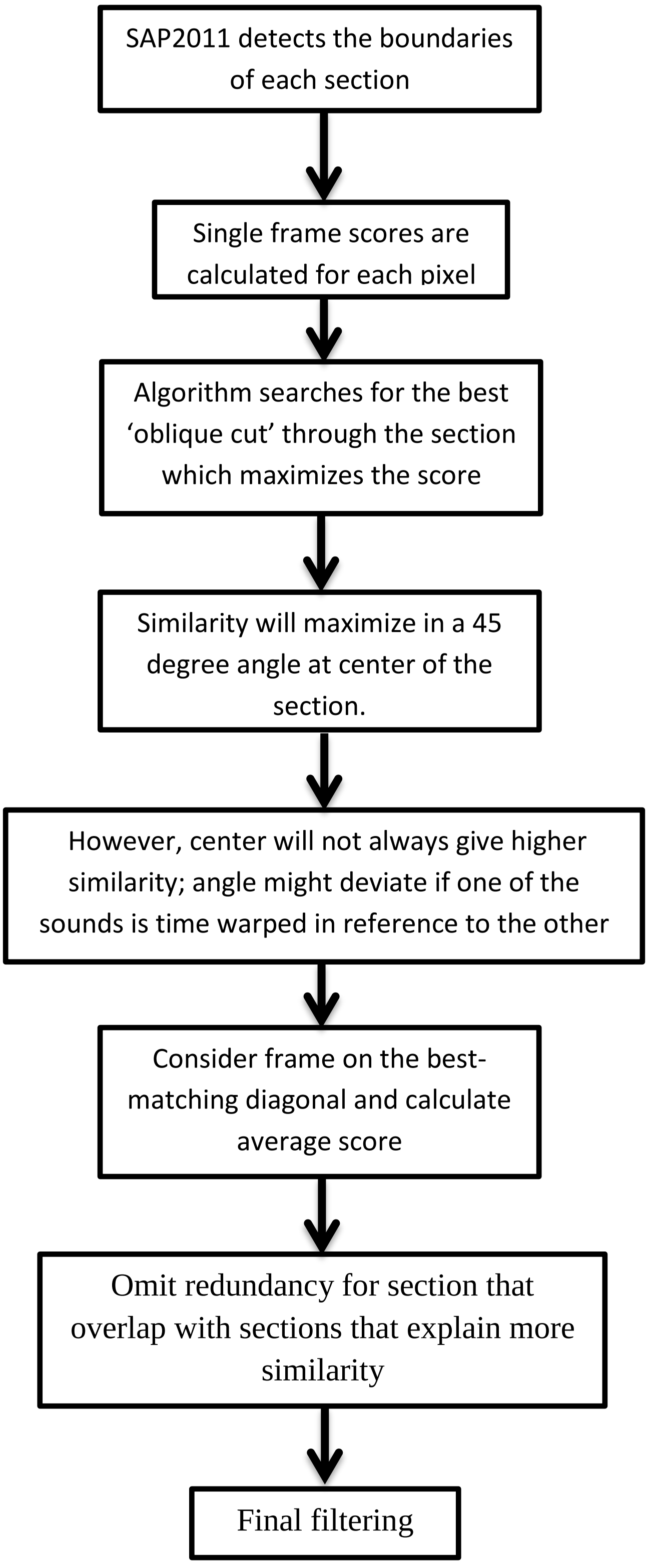


Figure 3.13 Steps of overall similarity captured by each section

CHAPTER 4

RESULT AND DISCUSSIONS

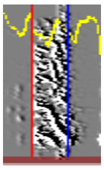
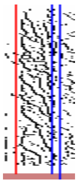
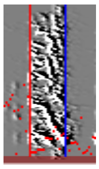
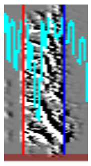
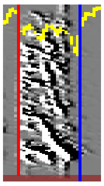
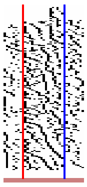
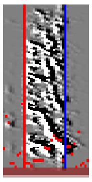
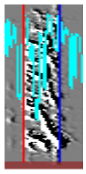
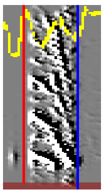
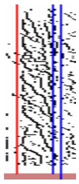
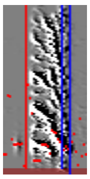
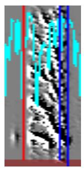
4.1 Introduction

In this chapter, there will be a detail description on the results obtained from the implementation of this project. This will be included the results of analysis and the output calculated by Sound Analysis Pro 2011 software. Discussions will be included too as an explanation to the results obtained. The result will be represented by figures and tables.

4.2 Reducing Scoring Ambiguity

Comparing two pair of sounds will obtained a percentage of similarity and accuracy. Scoring similarity between songs on the scale of a single window is hopeless, as is comparing pictures one pixel at a time. Therefore, the solution is to compare intervals of sound consists of several windows. If the intervals are sufficiently long, they will contain enough information to identify a unique song segment. Yet, if the intervals are too short, similarities that are real could not been traced at a smaller interval size. An example of short interval sound is illustrated in table 4.1(a) below.

Table 4.1(a) Scores of features for the sample of External sound 1

Sound name	Wiener entropy	Spectral continuity	Pitch	Frequency modulation
(5.1) 10sec1 Duration = 31.836 ms	 = -1.387	 = 0 ms	 = 1243.0 Hz	 = 61.250°
(5.2) 10sec1 Duration = 31.836 ms	 = - 1.433	 = 0 ms	 = 1354.8 Hz	 = 57.811°
(5.3) 10sec1 Duration = 31.836 ms	 = -1.568	 = 0 ms	 = 1706.6 Hz	 = 60.922°

Empirically, it have been discovered that comparisons using 500-ms to 1100-ms range of intervals were satisfactory compared to the smaller interval. Figure 4.1a-c and Table 4.1a-c illustrates the results.

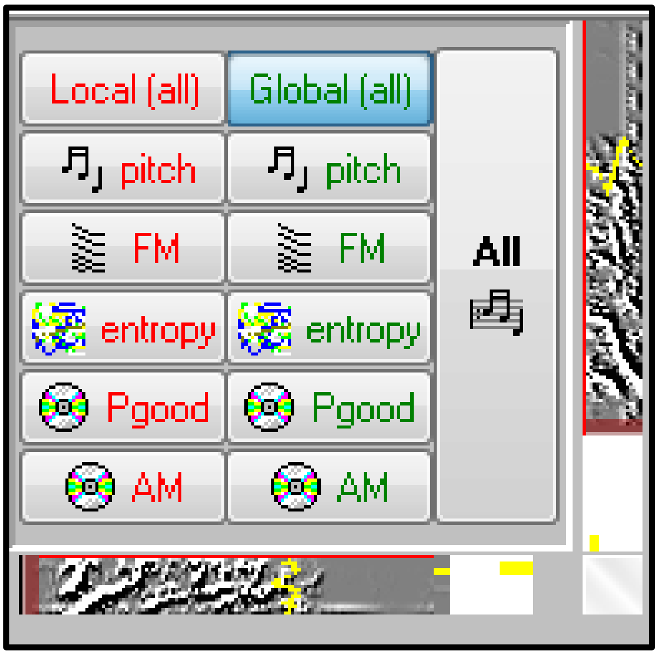


Figure 4.1(a) Similarity matrix between two External sounds at duration 32-ms

Table 4.1(b) Similarity score of sound comparison at duration 32-ms

Sound 1	Sound 2	%Similarity	Accuracy
(5.1) 10saat.wav	(5.3) 10saat.wav	0	-1

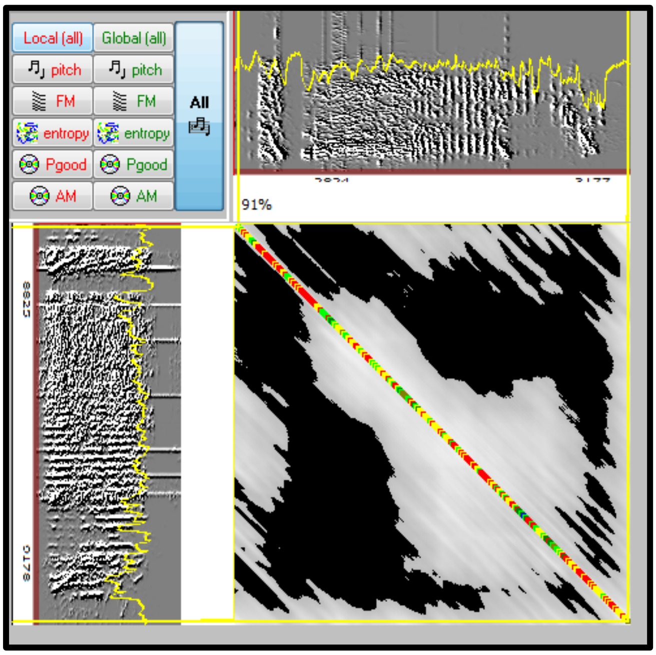


Figure 4.1(b) Similarity matrix between two External sounds at duration of 550-ms.

Table 4.1(c) Similarity score of sound comparison at duration 550-ms

Sound 1	Sound 2	%Similarity	Accuracy
(5.1) 10saat.wav	(5.2) 10saat.wav	98	91.08

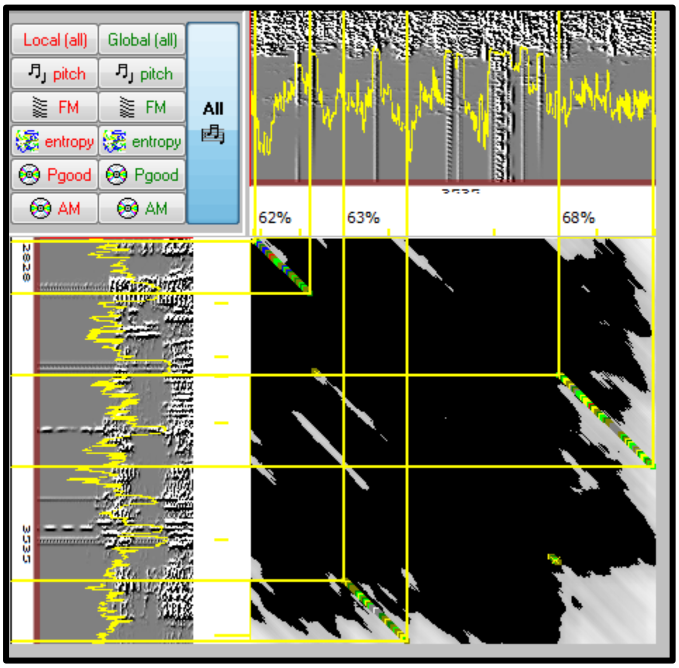


Figure 4.1(c) Similarity matrix between two External sounds at duration of 1030-ms.

Table 4.1(d) Similarity score of sound comparison at duration 1030-ms

Sound 1	Sound 2	%Similarity	Accuracy
(1.1) 10saat.wav	(1.2) 10saat.wav	51	67.94

By observing the similarity score, from Figure 4.1(a) and Table 4.1(a), at duration 32-ms, the score of %similarity and accuracy are extremely low, seems like cannot be detected. This is because the sound interval is not sufficient for measuring the sound features. In contrast with the duration of 32-ms, at sound duration 550-ms, as shown in Figure 4.1(b) and Table 4.1(b), the %similarity and accuracy obtained scores as well as the sample of song at duration 1030-ms in Figure 4.1(c) and Table 4.1(c). This can be simplified that the interval are sufficiently long, that contain enough information to identify a unique song segment.

4.3 Integration of the Song Features

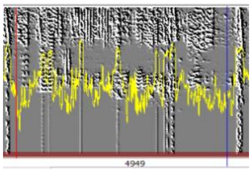
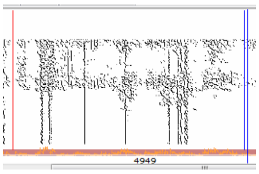
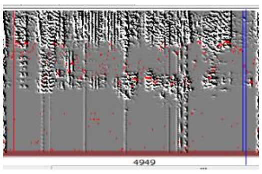
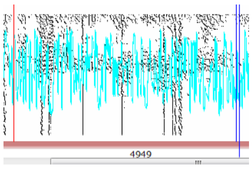
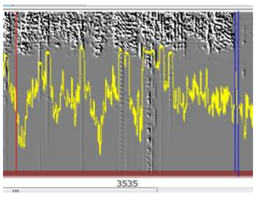
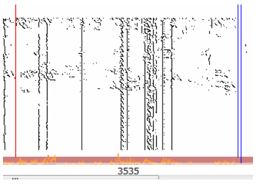
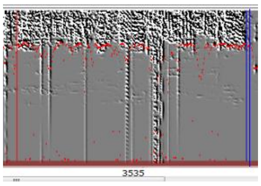
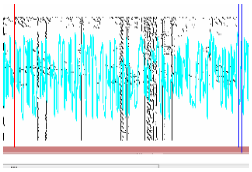
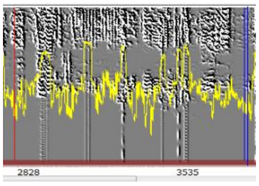
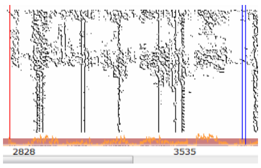
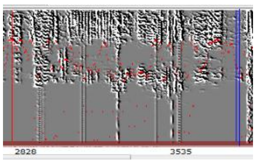
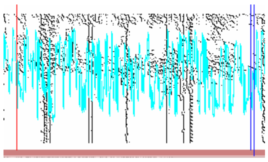
Each time window of the Swiftlets sound is represented by measurements of four features: Wiener entropy, spectral derivatives, pitch and frequency modulation. Each of these features has different units and different statistical distributions in the sound spectrum. To observe and analyze the data calculated by Sound Analysis Pro 2011, few experiments and observations has been carried out. The data obtained were recorded in table. The method of implementation has been described in the previous chapter. This following section will be detailed on the result obtained and together with the explanation of the result.

Different windows often share similar patterns of power spectrum. Therefore, I scaled the measures based on their distribution in a sample of Internal and External sounds. Even all four measures are taken into account, there a likely to be several windows in the sound itself show close similarity to each other as the pattern of waveform is almost similar at certain interval and it is like repeating. By taking a sample sound, the sound is split into three different time interval, each by 10 second. For the purpose of analysis, 10 second from the beginning, middle and at the end of the interval of the sound has been taken to analyze based on several features. These steps were applied to the Internal as well as the External sound.

4.4 The Internal Sound

The Internal sound has been recorded in the space of cave where the Swiftlet stayed permanently and make nest there. The result of song features are shown in table 4.2.

Table 4.2 Scores of features for the sample of Internal sound 1

Sound Name	Wiener entropy	Spectral continuity	Pitch	Frequency modulation
(1.1) 10 sec Duration = 1025.8 ms	 = -3.152	 = 3.4463 ms	 = 2959.2 Hz	 = 45.416°
(1.2) 10 sec Duration = 1029.6 ms	 = -2.707	 = 2.0234 ms	 = 3179.8 Hz	 = 45.640°
(1.3) 10 sec Duration = 1022.3 ms	 = -2.954	 = 4.2742 ms	 = 3132.4 Hz	 = 46.657°

4.5 The External Sound

The source of External sound has been recorded at the entrance of the cave, where the route that Swiflets usually used to enter and exit. Most of the time, this will be the place where they play around. The resultant measurement of sound features can be seen in table 4.3 and table 4.4.

Table 4.3 Scores of features for the sample of External sound 1

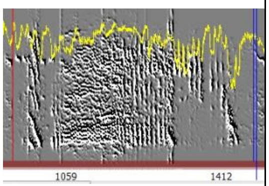
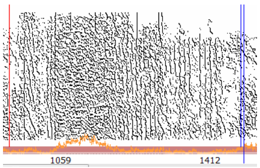
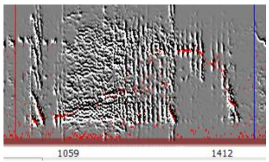
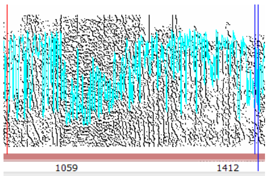
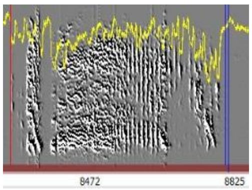
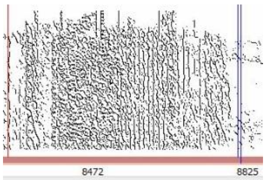
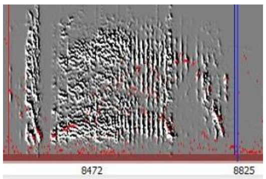
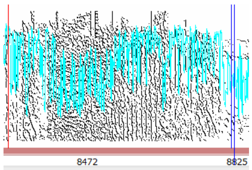
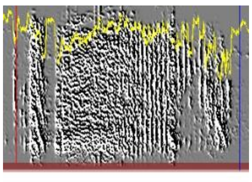
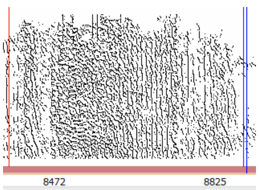
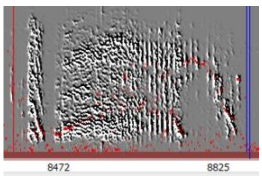
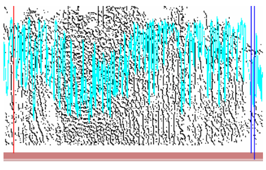
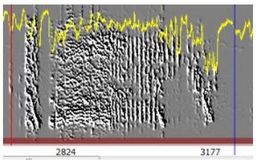
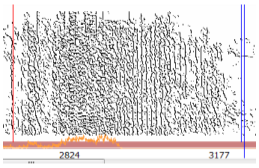
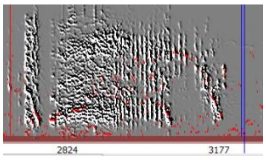
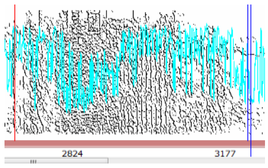
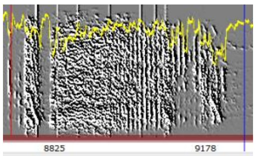
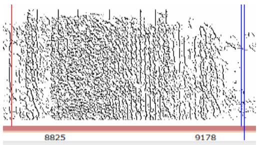
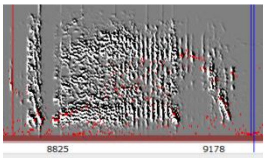
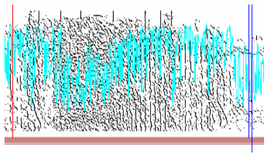
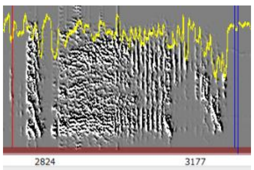
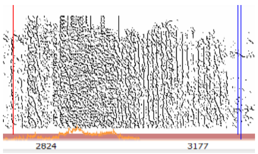
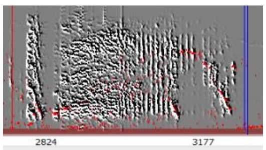
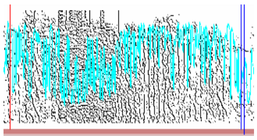
Sound name	Wiener entropy	Spectral continuity	Pitch	Frequency modulation
(10.1) 10sec1 Duration = 548.29 ms	 = -1.543	 = 5.6692 ms	 = 2440.8 Hz	 = 52.654°
(10.2) 10sec1 Duration = 535.30 ms	 = - 1.519	 = 0 ms	 = 2445.5 Hz	 = 55.278°
(10.3) 10sec1 Duration = 516.46 ms	 = -1.568	 = 0 ms	 = 2296.9 Hz	 = 54.085°

Table 4.4 Scores of features for the sample of External sound 2

Sound Name	Wiener entropy	Spectral continuity	Pitch	Frequency modulation
(5.1) 10sec1 Duration = 548.29 ms	 = -1.545	 = 3.4107 ms	 = 2314.2 Hz	 = 55.259°
(5.2) 10sec Duration = 539.45 ms	 = - 1.542	 = 0 ms	 = 2334.4 Hz	 = 54.939°
(5.3) 10sec Duration = 523.53 ms	 = -1.595	 = 3.7043 ms	 = 2524.1 Hz	 = 54.915°

4.6 Comparison of Sound Features

By referring to figure 4.2 below, for Internal sound, the score of Wiener entropy is approaching to - ∞. Thus, the waveform can be simplified as narrow which consist of pure tone. By looking at the spectral continuity score, the value is high. Therefore, the contours of the Internal sound is longer. By comparing to the pitch scores, the value is high. Consequently, the period of the spectrum in the Internal song can be simplified as short. By observing the pitch value of Internal sound, the range is higher than in External sound. Therefore, the period of song can be categorized as short.

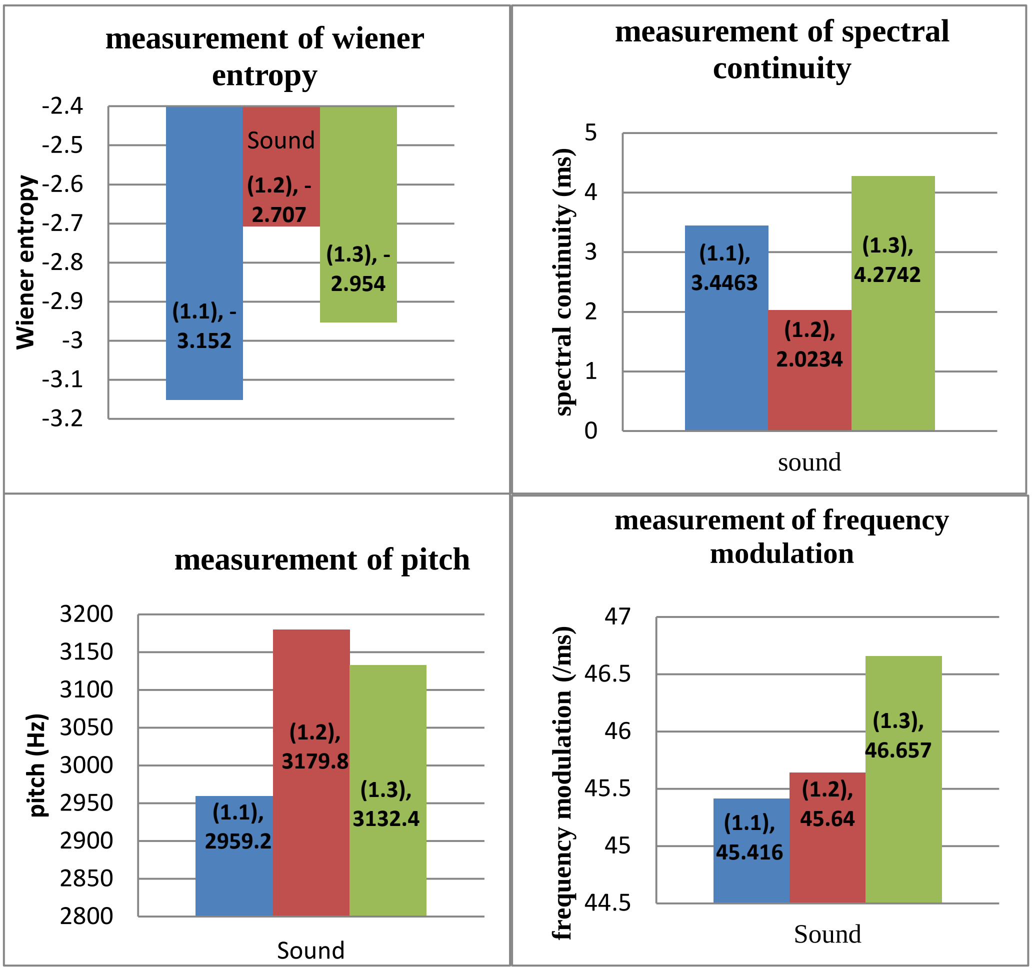


Figure 4.2 Graph of measurement of features for Internal sound 1

Based on figure 4.3 and figure 4.4, for both sample of the External sound, the value of Wiener entropy is nearly approaching to 0 values. Thus, the waveform is random due to the high value of entropy and consists of noisy sounds. As the value of entropy is near to 0 values, the spectrum can be categorized as broad. For the feature of spectral continuity, the score is low. As a result, the contours are not continuous and short. By observing the average scores of pitch, it is not much higher for the External sounds. This can be categorized to low pitch value, whereby the period is long.

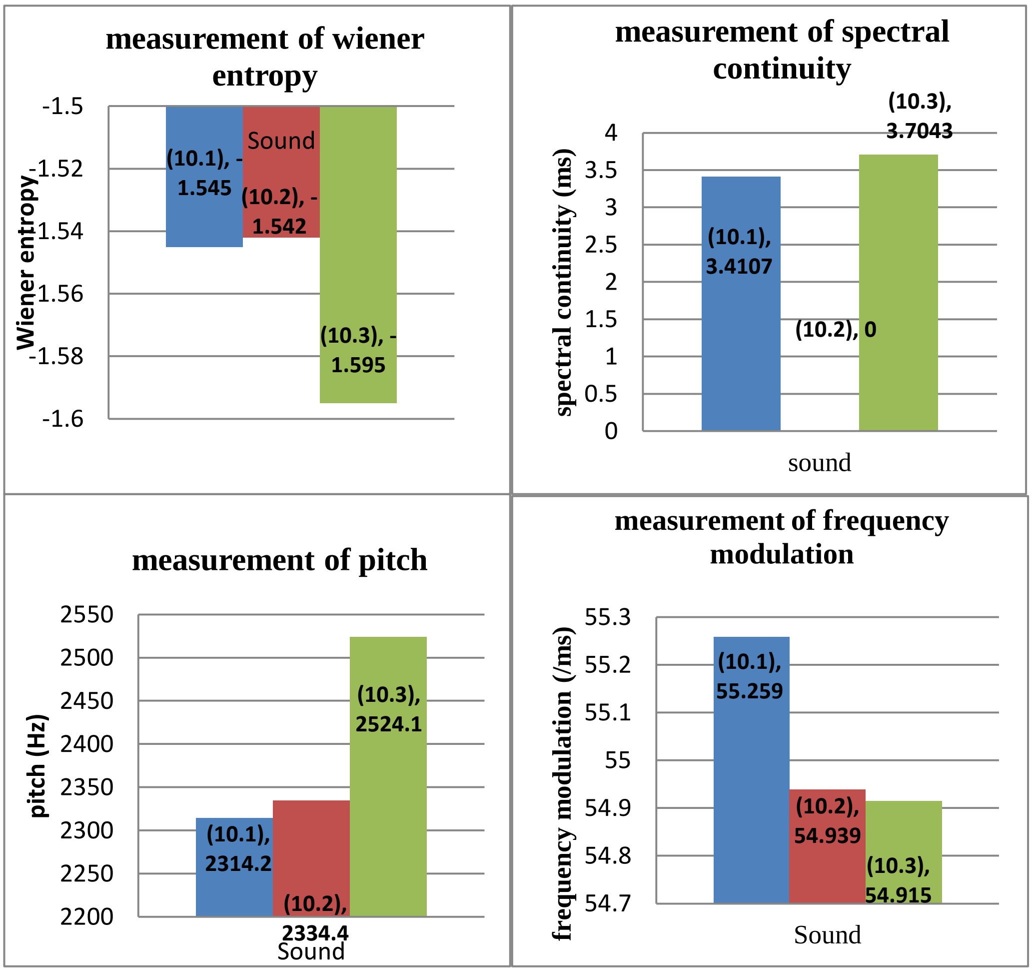


Figure 4.3 Graph of measurement of features for External sound 1

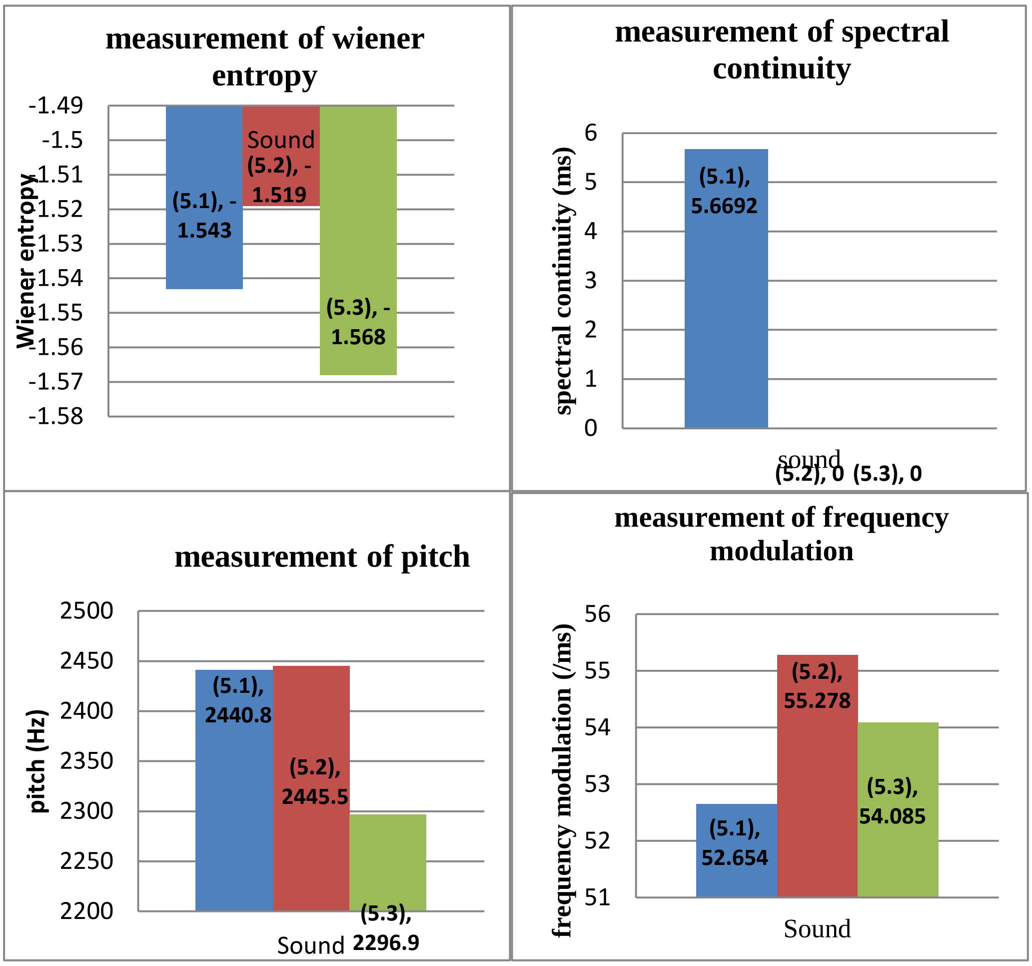


Figure 4.4 Graph of measurement of features for External sound 2

4.7 Symmetric Comparison of Sound Features

The resulting matrix spans all possible combinations of pairs of Internal and External sounds. The difference in features between each pair of window is encoded into a color scale. There is a marked tendency for the strongest similarity between pairs of window to show as a red diagonal line. Different windows often share similar patterns of power spectrum and it is not restricted to a unique pair of windows. Hence, checking similarity of sound features has been done on several windows.

4.7.1 Symmetric Comparison of Internal sound

For the sample of Internal sound, the comparison are done between 3 interval of sound in the sample of Internal sound 1. The results of matrix spans has been visualize as in figure 4.5, figure 4.6 and figure 4.7. The scores are recorded in table 4.5, table 4.6 and table 4.7.

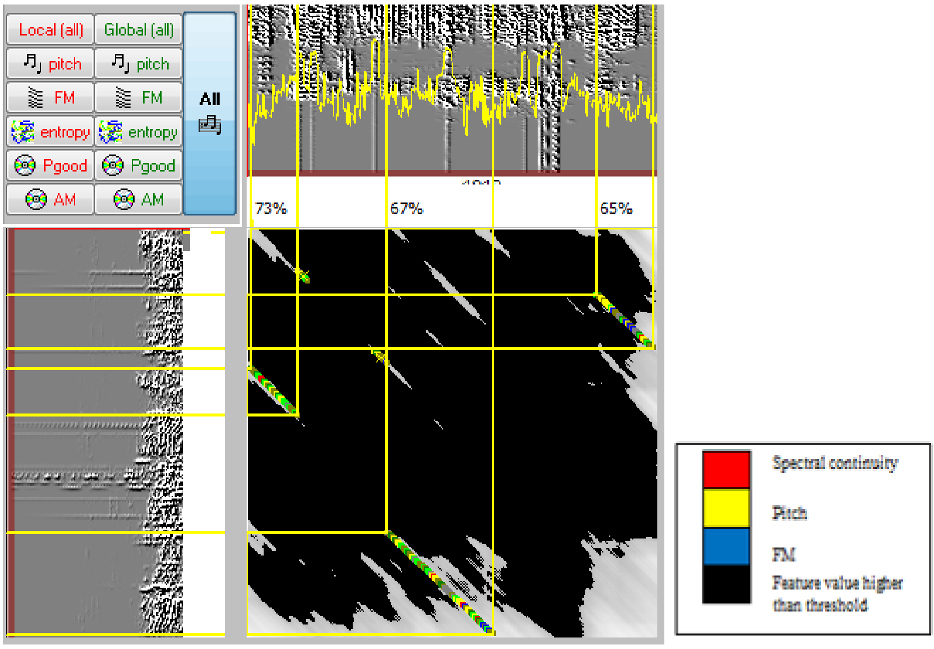


Figure 4.5 Similarity matrix for Internal sound between Song 1.1 and Song 1.2

Table 4.5 Scores of similarities for Internal sound between Song 1.1 and Song 1.2

Sound 1	Sound 2	%Similarity	Accuracy
(1.1) 10saat.wav	(1.2) 10saat.wav	51	67.94

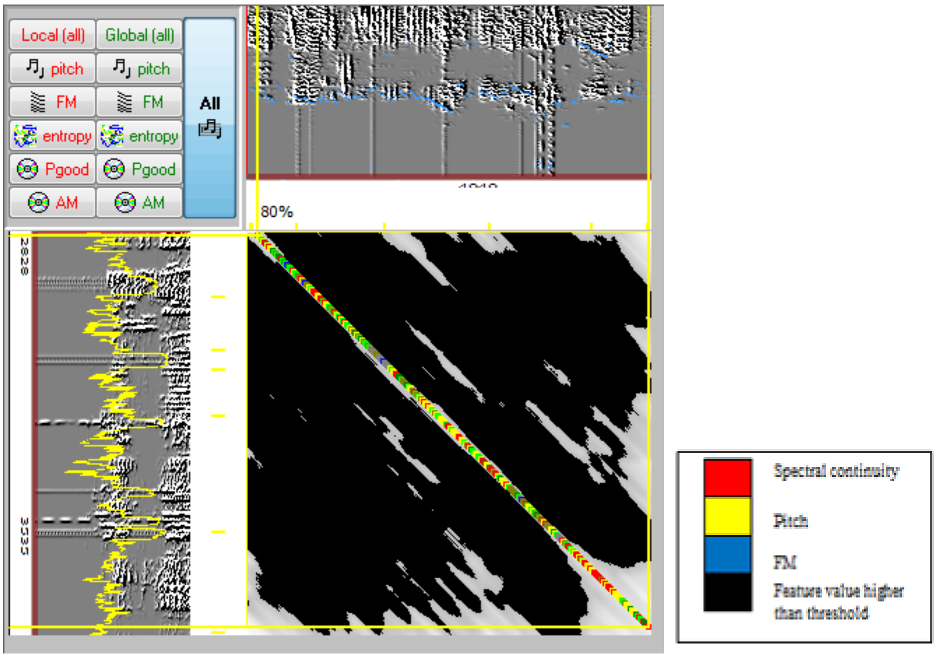


Figure 4.6 Similarity matrix for Internal sound between Song 1.1 and Song 1.3

Table 4.6 Scores of similarities for Internal sound between Song 1.1 and Song 1.3

Sound 1	Sound 2	%Similarity	Accuracy
(1.1) 10saat.wav	(1.3) 10saat.wav	96	80.8

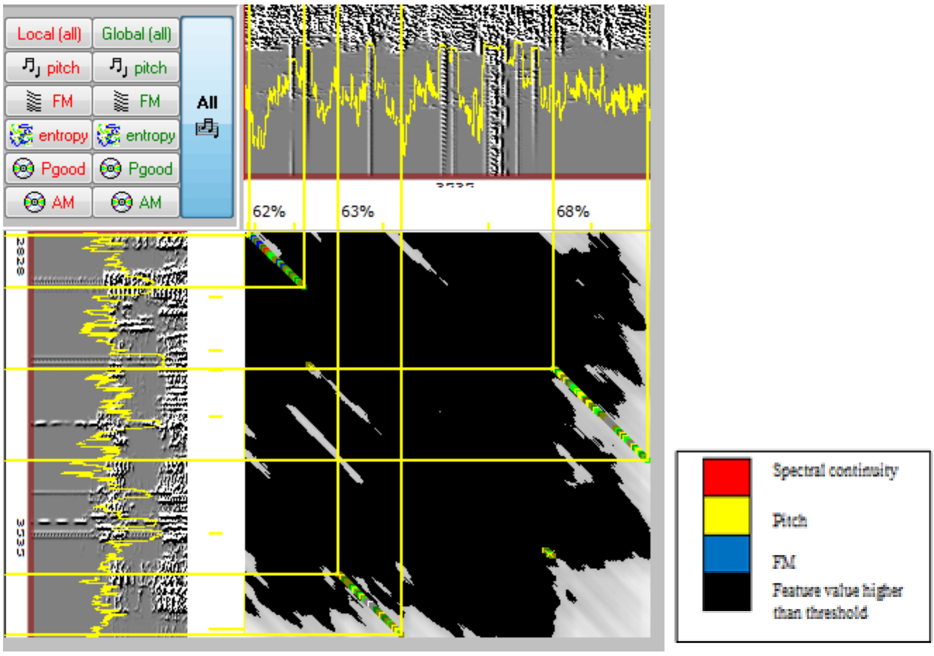


Figure 4.7 Similarity matrix for Internal sound between Song 1.2 and Song 1.3

Table 4.7 Scores of similarities for Internal sound between Song 1.2 and Song 1.3

Sound 1	Sound 2	%Similarity	Accuracy
(1.2) 10saat.wav	(1.3) 10saat.wav	52	65.52

The symmetric similarity of Internal sound has been measured in term of time course as it is good for detecting similarity between two sequences of features that shows similar curves. As a result, it gives narrower sections that capture the sequential in diagonal match. Overall, the %similarity is only half of the spectrum. By comparing the song motif at beginning, middle and end of sound notes, these indicate that the song notes of Internal sound is not repeating continuously.

4.7.2 Symmetric Comparison of External Sound

For External sound, the comparisons of song motif are also taken from three different intervals of sound notes. By this time, two sample of External sound has been taken into account for measuring the symmetric similarity of sound motifs. The results of matrix span have been visualized in figure 4.8 until figure 4.13. The similarity scores were recorded in table 4.8 until table 4.13.

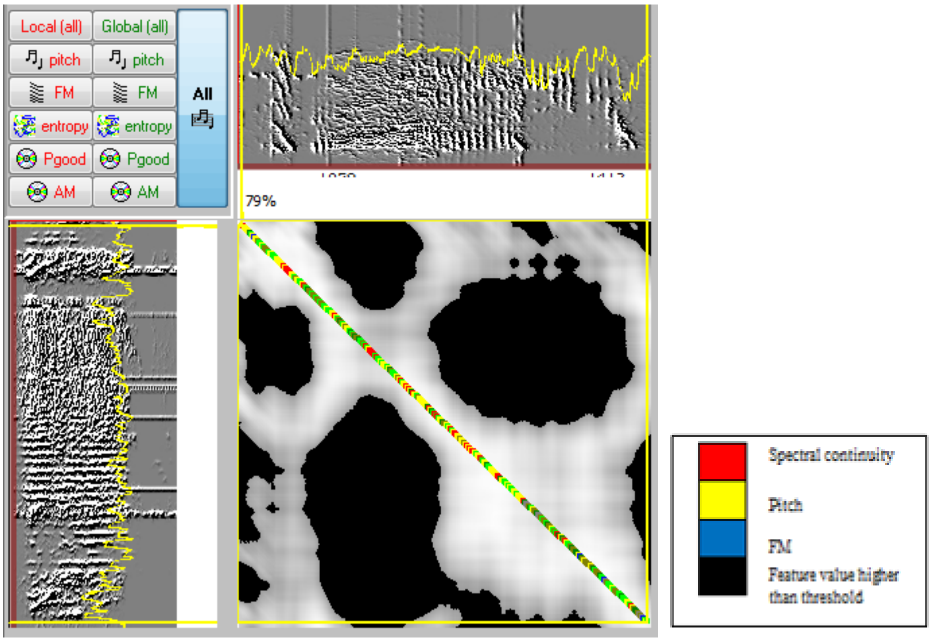


Figure 4.8 Similarity matrix for External sound between Song 10.1 and Song 10.2

Table 4.8 Scores of similarities for External sound between Song 10.1 and Song 10.2

Sound 1	Sound 2	%Similarity	Accuracy
(10.1) 10saat1.wav	(10.2) 10saat1.wav	97	79.62

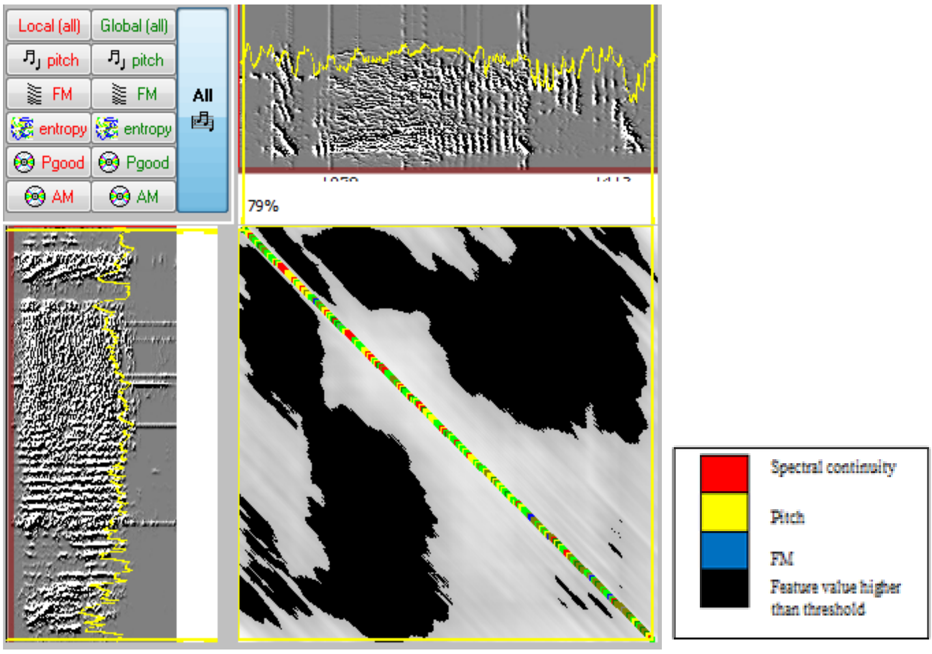


Figure 4.9 Similarity matrix for External sound between Song 10.1 and Song 10.3

Table 4.9 Scores of similarities for External sound between Song 10.1 and Song 10.3

Sound 1	Sound 2	%Similarity	Accuracy
(10.1) 10saat1.wav	(10.3) 10saat1.wav	97	79.38

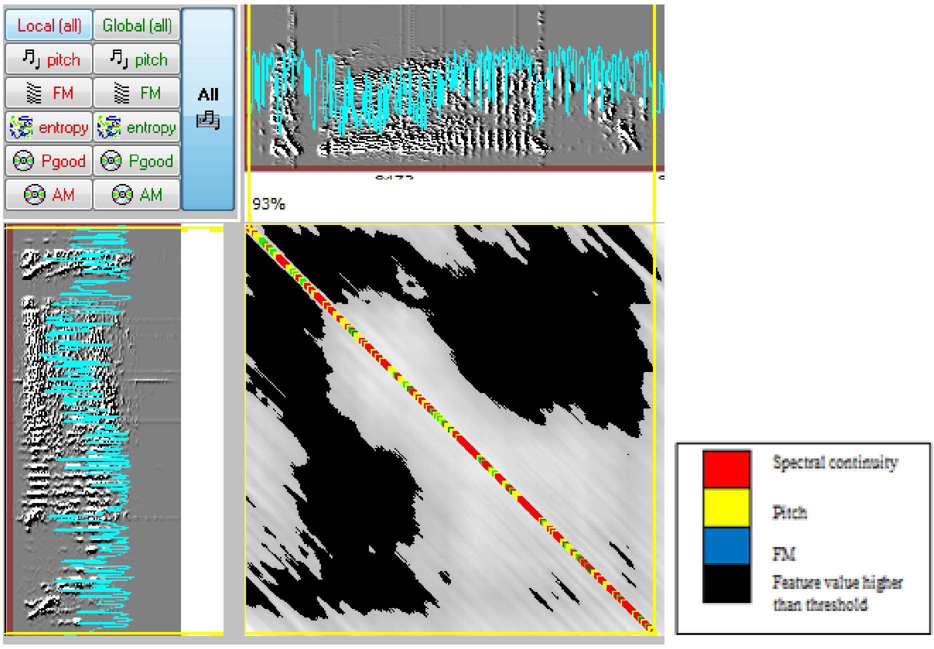


Figure 4.10 Similarity matrix for External sound between Song 10.2 and Song 10.3

Table 4.10 Scores of similarities for External sound between Song 10.2 and Song 10.3

Sound 1	Sound 2	%Similarity	Accuracy
(10.2) 10saat1.wav	(10.3) 10saat1.wav	96	93.48

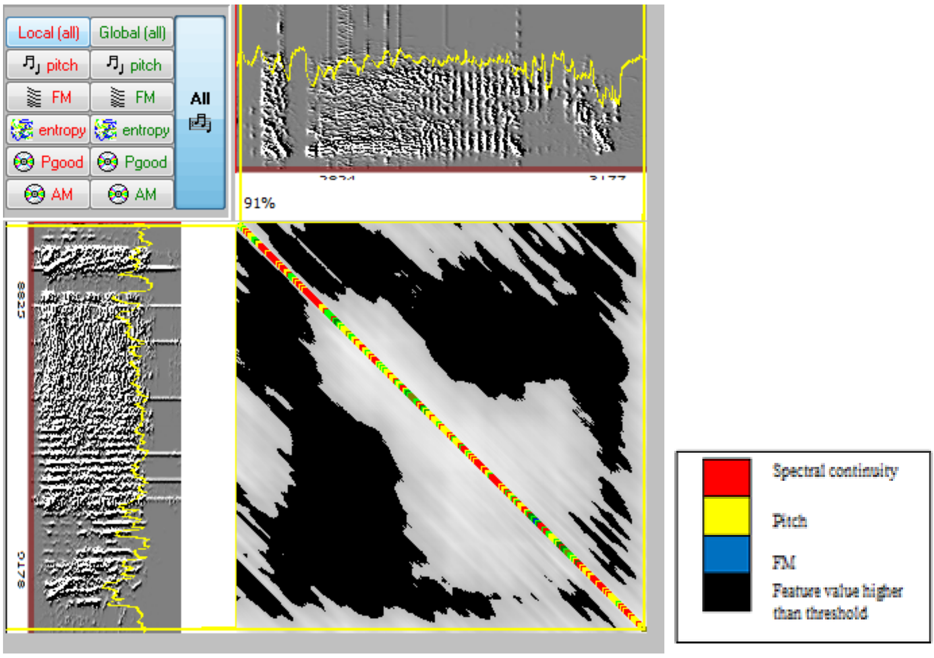


Figure 4.11 Similarity matrix for External sound between Song 5.1 and Song 5.2

Table 4.11 Scores of similarities for External sound between Song 5.1 and Song 5.2

Sound 1	Sound 2	%Similarity	Accuracy
(5.1) 10saat.wav	(5.2) 10saat.wav	98	91.08

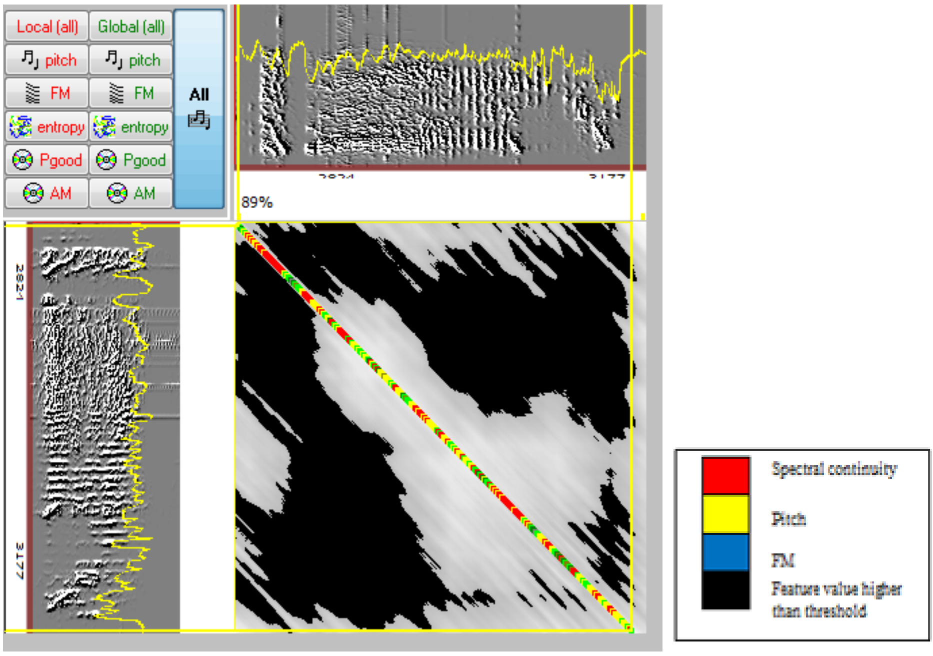


Figure 4.12 Similarity matrix for External sound between Song 5.1 and Song 5.3

Table 4.12 Scores of similarities for External sound between Song 5.1 and Song 5.3

Sound 1	Sound 2	%Similarity	Accuracy
(5.1) 10saat.wav	(5.3) 10saat.wav	95	89.93

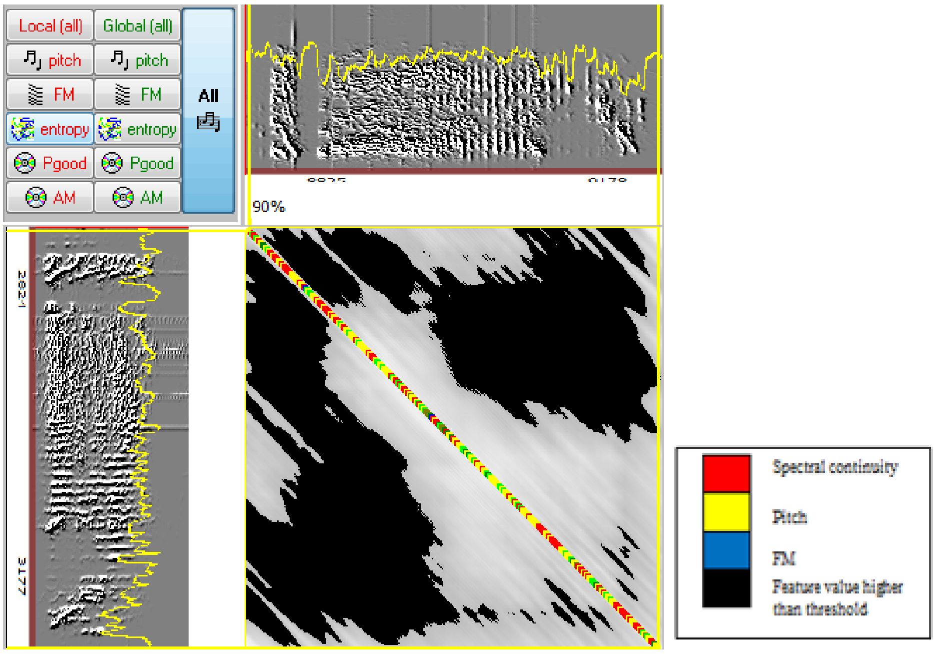


Figure 4.13 Similarity matrix for External sound between Song 5.2 and Song 5.3

Table 4.13 Scores of similarities for External sound between Song 5.2 and Song 5.3

Sound 1	Sound 2	%Similarity	Accuracy
(5.2) 10saat.wav	(5.3) 10saat.wav	97	90.92

In this similarity score, the first priority is to observe on the score of %similarity and accuracy. Overall, the scores between the selected song motifs of the External sound are all in the range above 90%. While doing a visual inspection on the waveform of the sound, the spectral derivatives of the song motif looks similar to each other. Due to the high scores of similarity that occurs by chance with the similarity of the spectral derivatives, this can be simplified that the song notes of External sound consists of several repeated song motifs. Due to the contrast scores of features for External and Internal sound, both sounds are totally different in features and spectrums.

4.8 Discussion

In practice, however, similar feature values are seldom restricted to a unique pair of windows. Comparing features between songs on the scale of a single window is hopeless, as is comparing pictures one pixel at a time. Therefore, the solution is to compare song notes consisting of several song motifs. For intervals that are sufficiently long, they will contain enough information to identify a unique song segment. In contrast, if the intervals are too short, it might be a problem as the similarities that are at a smaller interval size may be rejected and that would reduce the power of analysis.

The four features had been chosen to be taken into account because those features are thought to bear a close relation to the articulatory variables involved in sound production. The measurement of similarity between two sounds is in the form of explicit and reliable metric. This approach can be used for more rigorous and quantitative study of vocal learning, vocal imitation and vocal communication.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter consists of the limitations, contributions on project, and the recommendation for the purpose of future analysis. To be précised, the conclusion of this project will be stated here too.

5.2 Limitations

Upon completing this project, several problems have been encountered which had interrupted and slower down the working progress. The interruption are as follows:

a) The waveform of sound spectrums consist a lot of features. Selecting the interval for feature analysis has to be appropriately done to avoid selecting the noise
b) The sounds that has been split failed to be analyzed due to the limitations of the sound converter software
c) Exploring and understanding the application in the Sound Analysis Pro 2011 software is time consuming

5.3 Suggestion for Future Experiments

In this project, the sounds are only needed to be identified by measuring the features from the spectrum of the sound to determine the characteristics of the internal and external sounds. The analysis is only restricted to determining several features for each category of sound and checking the similarity of sound based on each category itself.

However, the analysis can be varies for further experiments in the future. This project can be extended by doing assessment of vocal imitation in the way of describing and measuring any similarities between the song of a tutor and that of the pupil of Swiftlets. In addition, a development of GUI can also be done for identifying the type of Swiflets sound; either it is internal, external or pulley sound. The GUI will aid to choose the most suitable sound to be played in the Swiftlets building.

5.4 Conclusion

The analysis comprise of two parts of sound which is the Internal and External sound. The analysis involved the measurement of several features and similarity scores within the interval of spectrum of the Internal and External sounds. This project succeeded to identify the scores of sound features which are different for both Internal and External sounds.

REFERENCES

[1] Henri Thomassen and HendrikAalbert (2005) Swift as soun, Design and evolution of the Echolocation in Swiftlets (Apodidae:Collocaliini). *Thesis Leiden University*

[2] Fenton, M.B. (1975) Acuity of echolocation in *Collocaliahirundinacea*(Aves:Apodidae), with comments on the distribution of echolocatingSwiftlets andMolossid bats. *Biotropica***7** (1): 1-7.

[3] Griffin, D.R. and Suthers, R.A. (1970) Sensitivity of echolocation in Cave Swiftlets.*Biological Bulletin* **139**: 495-501.

[4] Griffin, D.R. and Thompson, D. (1982) Echolocation by cave Swiftlets. *BehaviouralEcology and Sociobiology* **10**: 119-123.

[5] Haibin Wang, Andrews, P., Mitra, P.P., Saar, S., Tchernichovski, O., Golani, I., Benjamini, Y. (2007) Characterization of animal behavior through the use of audioand video signal processing. Multimedia, IEEE, vol. 14, issue. 4, p. 32-41.

[6] W.H. Thorpe, “The process of song-learning in the chaffinch as studied by means of the sound spectrograph,” *Nature*, vol. 173, 1954, p. 465.

[7] D. Thomson, “Quadratic-inverse spectrum estimates: applications to palaeoclimatology,” *Philos. Trans. R. Soc. London, Ser. A,* vol. 332, 1990, p.539–597.

[8] D. Thomson, “Non-stationary fluctuations in stationary time series,” *Proc. International Society of Optical Engineering*, vol. 2027, 1993, p. 236–244.

[9] Tchernichovski, O., Nottebohm, F., c, C.E., Pesarant, B. &Pratim, P, “A Procedure for an Automated Measurement of Song Similarity,”*Animal Behaviour*, vol. 59, issue 6, June 2000, p. 1167-1176.

[10] K.W. Nordeen and E.J Nordeen, “Deafening-Induced Vocal Deterioration in Adult Songbirds Is Reversed by Disrupting a Basal Ganglia-Forebrain Circuit,”Jurnal of Neuroscience, vol. 30, issue 21,May 2010, p. 7392-7400.

[11] John A. Thompson, Mark J. Basista, Wei Wu, Bertram R., & Johnson F., (2011) “Dual Pre-Motor Contribution to Songbird Syllable Variation,”Jurnal of Neuroscience, vol. 31, issue 1, p. 322-330.

[12] Anna H.Koetz, David A. Westcott & Bradley C.Congdon, (2007) “Spatial pattern of Song Element Sharing and its Implications for Song Learning in the Chowchilla, Orthonyxspaldingii,” Animal Behaviour, vol. 74, issue 4, p. 1019-1028.

[13] “Sound Analysis Pro 2011” http://soundanalysispro.com/manual-1/manual