SWIFTLET SOUND IDENTIFICATION USING VECTOR QUANTIZATION AND GAUSSIAN MIXTURE MODEL

SITI NURZALIKHA ZAINI BT HUSNI ZAINI

Master of Engineering (Electronic)

UNIVERSITI MALAYSIA PAHANG
SUPERVISOR’S DECLARATION

We hereby declare that we have checked this thesis and in our opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Engineering (Electronic).

_______________________________
(Supervisor’s Signature)
Full Name : DR MOHD ZAMRI BIN IBRAHIM
Position : SENIOR LECTURER
Date : 7 AUGUST 2018

_______________________________
(Co-supervisor’s Signature)
Full Name : PROF MADA YA DR SAIFUL NIZAM BIN TAJUDDIN
Position : SENIOR LECTURER
Date : 7 AUGUST 2018
STUDENT’S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

__________________________________________
(Student’s Signature)

Full Name : SITI NURZALIKHA ZAINI BT HUSNI ZAINI
ID Number : MEL12004
Date : 7 AUGUST 2018
SWIFTLET SOUND IDENTIFICATION USING VECTOR QUANTIZATION AND
GAUSSIAN MIXTURE MODEL

SITI NURZALIKHA ZAINI BT HUSNI ZAINI

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Master of Engineering (Electronic)

Faculty of Electrical & Electronics Engineering
UNIVERSITI MALAYSIA PAHANG

AUGUST 2018
ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious, the Most Merciful.

All praise is due to Allah, the Creator and Sustainer of this whole universe, the Most Beneficent and the Most Merciful, for His guidance and blessing and granting me knowledge, patience and perseverance to accomplish this research successfully.

Firstly, I would like to acknowledge supportive supervisor Dr. Mohd Zamri Bin Ibrahim for his supervision and guidance to complete this thesis. Without patience and guidance from him, this research could not have been successfully completed. Special thanks to my co-supervisor Assoc. Prof. Dr. Saiful Nizam Bin Tajuddin and ex-supervisor Dr. Sunardi for their valuable motivation, moral support, encouragement and effort throughout this research.

Last but not least, most profound gratitude and respect to my beloved husband Mohd Kamarul Ridzuan bin Mohamad Himi, my son Muhammad Sayf Erfhan, my daughter Mia Sophea Emani my beloved mother Siti Hajar Binti Talib for their understanding and encouragement during my study. I would like to express my sincere appreciation to all my postgraduate's friends, who have been the ultimate source of my motivation to work hard in the completion of this research work and my entire study. Therefore, I proudly dedicate this work to both of them, may Allah SWT bless both of them.
CHAPTER 1 INTRODUCTION
1.1 Background of Study
1.2 Problem Statement
1.3 Research Objectives
1.4 Scope of research
1.5 Thesis contribution
1.6 Thesis Overview

CHAPTER 2 LITERATURE REVIEW
2.1 Introduction
2.2 Swiftlet
2.2.1 Swiftlet Attraction Using Sound
CHAPTER 3 METHODOLOGY

3.1 Introduction

3.2 Methodology

3.3 Feature Extraction

3.3.1 Mel Frequency Cepstral Coefficient (MFCC)

3.3.2 Linear Prediction Cepstral Coefficient (LPCC)

3.4 Alternative Techniques

3.4.1 Minimum Distance Classifier (MDC)

3.5 Propose models

3.5.1 Vector Quantization (VQ)

3.5.2 Gaussian Mixture Model (GMM)

3.6 Chapter Summary

CHAPTER 4 RESULTS AND DISCUSSION

4.1 Introduction

4.2 Data Collection

4.3 Experiment Setup

4.4 Result and Discussion
4.4.1 Result using Minimum Distance Classifier (MDC) 44
4.4.2 Result using Vector Quantization (VQ) 47
4.4.3 Result using Gaussian Mixture Model (GMM) 53
4.4.4 Comparison the best performance of VQ and GMM with MDC 58
4.4.5 Feature Improvement by Dynamic Feature Impact 61

4.5 Chapter Summary 64

CHAPTER 5 CONCLUSION 66

5.1 Introduction 66
5.2 Conclusion 66
5.3 Future Work 67

REFERENCES 69

APPENDIX A RESULT OF CONFUSION MATRIX 75

APPENDIX B LIST OF PUBLICATIONS 78
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Total samples of swiftlet sound</td>
<td>40</td>
</tr>
<tr>
<td>4.2</td>
<td>Confusion matrix using MDC with LPCC features</td>
<td>44</td>
</tr>
<tr>
<td>4.3</td>
<td>Confusion matrix using MDC with MFCC features</td>
<td>44</td>
</tr>
<tr>
<td>4.4</td>
<td>Confusion matrix using VQ (8-codebook) with LPCC features</td>
<td>47</td>
</tr>
<tr>
<td>4.5</td>
<td>Confusion matrix using VQ (8-codebook) with MFCC features</td>
<td>47</td>
</tr>
<tr>
<td>4.6</td>
<td>Confusion matrix using VQ (16-codebook) with LPCC features</td>
<td>48</td>
</tr>
<tr>
<td>4.7</td>
<td>Confusion matrix using VQ (16-codebook) with MFCC features</td>
<td>48</td>
</tr>
<tr>
<td>4.8</td>
<td>Confusion matrix using VQ (32-codebook) with LPCC features</td>
<td>49</td>
</tr>
<tr>
<td>4.9</td>
<td>Confusion matrix using VQ (32-codebook) with MFCC features</td>
<td>49</td>
</tr>
<tr>
<td>4.10</td>
<td>Confusion matrix using VQ (64-codebook) with LPCC features</td>
<td>50</td>
</tr>
<tr>
<td>4.11</td>
<td>Confusion matrix using VQ (64-codebook) with MFCC features</td>
<td>50</td>
</tr>
<tr>
<td>4.12</td>
<td>Confusion matrix using GMM (1-mixture) with LPCC features</td>
<td>54</td>
</tr>
<tr>
<td>4.13</td>
<td>Confusion matrix using GMM (1-mixture) with MFCC features</td>
<td>54</td>
</tr>
<tr>
<td>4.14</td>
<td>Confusion matrix using GMM (2-mixture) with LPCC features</td>
<td>55</td>
</tr>
<tr>
<td>4.15</td>
<td>Confusion matrix using GMM (2-mixture) with MFCC features</td>
<td>55</td>
</tr>
<tr>
<td>A.1</td>
<td>Confusion matrix using MDC with LPCC_D and LPCC_DA</td>
<td>75</td>
</tr>
<tr>
<td>A.2</td>
<td>Confusion matrix using MDC with MFCC_D and MFCC_DA</td>
<td>75</td>
</tr>
<tr>
<td>A.3</td>
<td>Confusion matrix using VQ (64-codebook) with LPCC_D</td>
<td>75</td>
</tr>
<tr>
<td>A.4</td>
<td>Confusion matrix using VQ (64-codebook) with LPCC_DA</td>
<td>76</td>
</tr>
<tr>
<td>A.5</td>
<td>Confusion matrix using VQ (64-codebook) with MFCC_D</td>
<td>76</td>
</tr>
<tr>
<td>A.6</td>
<td>Confusion matrix using VQ (64-codebook) with MFCC_DA</td>
<td>76</td>
</tr>
<tr>
<td>A.7</td>
<td>Confusion matrix using GMM (2-mixture) with LPCC_D</td>
<td>76</td>
</tr>
<tr>
<td>A.8</td>
<td>Confusion matrix using GMM (2-mixture) with LPCC_DA</td>
<td>77</td>
</tr>
<tr>
<td>A.9</td>
<td>Confusion matrix using GMM (2-mixture) with MFCC_D</td>
<td>77</td>
</tr>
<tr>
<td>A.10</td>
<td>Confusion matrix using GMM (2-mixture) with MFCC_DA</td>
<td>77</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 2.1  White-nest Swiftlets (*Aerodramus fuciphagus*)  7
Figure 2.2  Black-nest Swiftlets (*Aerodramus Maximus*)  8
Figure 2.3  Concept of Minimum Distance Classifier  14
Figure 2.4  Codewords in 2-dimensional space.  16
Figure 3.1  Basic step of speaker processing  21
Figure 3.2  Research Methodology  22
Figure 3.3  Mel frequency warping function  24
Figure 3.4  An example of mel frequency filterbank  25
Figure 3.5  Block diagram of the computation steps of MFCC  25
Figure 3.6  Block diagram of the computation steps of LPCC  31
Figure 3.7  Minimum Distance Classifier Model  34
Figure 3.8  Vector Quantization Model  36
Figure 3.9  Gaussian Mixture Model  37
Figure 4.1  Baby Spectrogram  40
Figure 4.2  Adult Spectrogram  41
Figure 4.3  Colony Spectrogram  41
Figure 4.4  A block diagram of training and testing swiftlet sound identification  42
Figure 4.5  The performance of the system using MDC as classifier with two different feature extraction method, LPCC, and MFCC  45
Figure 4.6  The performance of the system using MDC as classifier with two different feature extraction method, LPCC, and MFCC  46
Figure 4.7  The performance of total accuracy using VQ with 8, 16, 32 and 64 codebook  51
Figure 4.8  The performance of the system with different size of VQ codebook size using two types of feature extraction technique, LPCC, and MFCC  52
Figure 4.9  The performance GMM with 1-mixture and 2-mixture with LPCC and MFCC  56
Figure 4.10  The performance of the system with different number of mixtures of GMM using two types of feature extraction technique, LPCC, and MFCC  57
Figure 4.11  The performance using MDC, VQ (64-codebook) and GMM (2-mixture)  58
Figure 4.12  Details accuracy for each sound for MDC, VQ (64-codebook) and GMM (2-mixture)  60
Figure 4.13  Dynamic Features Impact Performance  62
Figure 4.14  Comparison GMM by 2-mixture with additional Delta-Acceleration (DA) features qualifier and Original (0)  63
Figure 4.15  Summary of chapter 4  65
LIST OF SYMBOLS

\( f \)  Frequency
\( f_s \)  Sampling frequency
\( H_{km} \)  Mel filter bank
\( H_z \)  Transfer function
\( M \)  Overlap window size
\( M_f \)  Number of filter bank
\( N \)  Window size
\( P \)  Magnitude Spectrum
\( S \)  Audio signal
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>Compact Disk</td>
</tr>
<tr>
<td>D</td>
<td>Delta</td>
</tr>
<tr>
<td>DA</td>
<td>Delta-Acceleration</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete Cosine Transform</td>
</tr>
<tr>
<td>DTW</td>
<td>Dynamic Time Wrapping</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-Maximization</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian Mixture Model</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>LPC</td>
<td>Linear Predictive Coding</td>
</tr>
<tr>
<td>LPCC</td>
<td>Linear Predictive Cepstral Coefficient</td>
</tr>
<tr>
<td>MDC</td>
<td>Minimum Distance Classifier</td>
</tr>
<tr>
<td>MDC</td>
<td>Minimum Distance Classifier</td>
</tr>
<tr>
<td>MFCC</td>
<td>Mel Frequency Cepstral Coefficient</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>VQ</td>
<td>Vector Quantization</td>
</tr>
<tr>
<td>VQLBG</td>
<td>Vector Quantization Linde-Buzo and Gray</td>
</tr>
</tbody>
</table>