ACOUSTIC EVENT DETECTION IN POLYPHONIC ENVIRONMENTS USING ARTIFICIAL NEURAL NETWORKS

Jayawardhana Pathiranage Manesh Mihiranga

(188016R)

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

May 2021

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/ dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters thesis/ dissertation under my supervision.

Name of the Supervisor: Dr. Sulochana Sooriyaarachchi

Signature of the Supervisor: Date:

ACKNOWLEDGEMENTS

It would never be possible to finish my dissertation without the encouragement, support, and supervision of various personalities, including my supervisor, family, friends and colleagues. At the end of this thesis, I would like to thank all those people who made this achievable and memorable experience for me.

First and foremost, I would like to thank my supervisor Dr. Sulochana Sooriyaarachchi for the continuous support and guidance I received in every aspect while completing this research.

I would also like to thank my progress review committee, Dr. Charith Chitraranjan and Dr. Ranga Rodrigo for their valuable insights and guidance. Their advice helped me to improve the state of my research work.

This work was funded by a Senate Research Committee (SRC) Grant of the University of Moratuwa. Further, I would like to thank for the committee for the financial support.

Finally, I would like to express my sincere gratitude to all of my friends and colleagues for motivating me all the time to do my best not limited to academic work. I thank my parents who have given me a fortunate life and always believing, trusting and supporting me.

ABSTRACT

Our environment is a mixture of hundreds of sounds that are emitted by different sound sources. These sounds are overlapped in both time and frequency domains in an unstructured manner composing a polyphonic environment. Identification of acoustic events in a polyphonic environment has become an emerging topic with many applications such as surveillance, context-aware computing, automatic audio indexing, health care monitoring and bioacoustics monitoring.

Polyphonic acoustic event detection is a challenging task aimed at detecting the presence of multiple sound events that are overlapped at a particular time instance and labeling. It requires a large amount of training data with a complex machine learning architecture thus making it a highly resource-consuming task. Hence, the accuracy of this research area is still not at a satisfactory level.

This study presents a neural networks-based classifier architecture with data augmentation and post-processing methods to improve accuracy. Two neural network architectures as a multi-label and combined single label are implemented and compared in the study. Previous literature reveals that Mel frequency cepstral coefficients and log Mel-band energies are the widely used features in the state of the art research in the area. Different data augmentation methods were used to ensure that the neural networks are trained for even the slight variations of the environmental sounds. A novel binarization method based on the signal energy is proposed to calculate the threshold value for binarizing the source presence predictions. Finally, the median filter based post processing was implemented to smoothen the detection results. The experimental results show that the proposed binarizing method improved the detection accuracy and recorded a maximum of 62.5% combined with the data augmentation and post-processing.

Keywords: Polyphonic Acoustic Event Detection, Dynamic Threshold Binarization, Deep Neural Networks

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACF	Autocorrelation Function
AED	Acoustic Event Detection
ANN	Artificial Neural Networks
BER	Band Energy Ratio
CNN	Convolutional Neural Network
CSL	Combined Single Label
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transformation
DNN	Deep Neural Network
DTB	Dynamic Threshold Binarization
DWTC	Discrete Wavelet Transform Coefficients
FFT	Fast Fourier Transformation
FTB	Fixed Threshold Binarization
GMM	Gaussian Mixture Models
GPU	Graphics Processing Unit
HMM	Hidden Markov Models
k-NN	k-Nearest Neighbor
LPC	Linear Prediction Coefficients
LPCC	Linear Prediction Cepstral Coefficients
MFCC	Mel Frequency Cepstral Coefficients
ML	Multi Label
NMF	Non-negative Matrix Factorization
PC	Personal Computer
PLP	Perceptual Linear Prediction
RMSE	Root Mean Square Energy
RNN	Recurrent Neural Network
SED	Sound Event Detection

SOM	Self Organizing Maps
STE	Short-Time Energy
SVM	Support Vector Machine
ZCR	Zero-Crossing Rate

TABLE OF CONTENTS