SUPERVISOR APPROVAL

IDENTIFICATION OF EXPRESSIVE SPEECH VIDEO SEGMENT USING ACOUSTIC FEATURES

By

NUR AMANINI SYAHIRAH BINTI ALIM

2014291566

This thesis was prepared under the supervision of the project supervisor, Puan Haslizatul Fairuz binti Mohamed Hanum. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science (Hons.) Multimedia Computing.

Approved by

.....

Puan Haslizatul Fairuz binti Mohamed Hanum

JULY 31, 2017

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty, I was able to finish this research. It will be difficult to me to complete this project without His blessings and permission.

Firstly, my special thanks goes to my supervisor, Puan Haslizatul Fairuz binti Mohamed Hanum for her support and as guidance throughout the course of this project as well as providing me such valuable advice and help to conduct this project. All the crucial guidance and motivation given by other lecturers including my instructor, Dr Marina binti Ismail are really appreciated where give me enough strength in completing this research.

Last but not least, I would like to give my gratitude to my family, my father, Encik Alim bin Hasim, my mother my sisters, and also my friends for giving me all the support and help that I need in making this project. Thank you for all the support and inspiring me in such mean that could not be described in words.

ABSTRACT

A sound retrieval method enables users to easily obtain their preferred sound. When we communicate, we exchange the expressive and related messages. This project reviews about identification of expressive speech video segment using acoustic features. Specifically, the segmented expressive speech retrieves the expressive speech and non-expressive speech from the video. From the sermon video that we have choose, the expression of motivator looks like similar from the beginning until the end. The audience cannot focus on what the motivator is talk about because there is no interesting part based on the motivator's expression. This project applies manual video segmentation to differentiate expressive speech and non-expressive speech. Then, this project extracted the audio features from segmented expressive and non-expressive speech such as pitch and intensity by using Pratt tools. Then, we used Random Forest Classifier technique in Spyder (IDE) using Python language to get the accuracy which is 43% and used the prediction method to classify the expressive speech and non-expressive speech as the intended results. The training audio features was trained to get the performance accuracy. The correctness of the project has been showed from the evaluation. The project compared the predicted and manually segmented data to get the percentage of matches using pitch, the percentage of match is 80% while using the intensity is 75%. The correctness of the results has been verified to improve the identification of expressive speech video segment automatically.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This chapter provides the background of this project. It contains the story about the project background which retrieves the speech from the video. It also gives detail about the problem statement, objectives, scopes, significant and conclusion of the project.

1.1 Project Background

This project identifies the expressive segment video using acoustic features. Speech retrieval is content-based retrieval of speech documents, i.e. audio recordings containing spoken text (Schauble & Glavitsch, 1994). Refer to MacLeod (2008) acoustic features is a voice of speaker can show as much meaning as the words themselves. Expressive speech is to expose (that is to say, to manifest) mental feels of the speaker such as joy, approbation (Vanderveken, 1990).

According to Sano, Shibata and Yagi (2009) various visual hints from the image have a great potentiality for helping persons to understand the content. To extract the representative images, a new idea of using the roles of shots and production rules were used. The sound signal is one of the main medium of interaction and it can be processed to detect the speaker, speech or even emotion. The sound signals do have some features that represent the emotional state of the speaker. The paper shows the problem of emotion classification for human speech (Davletcharova, Sugathan, Abraham & James, 2015).

As stated by Cole, Mahrt and Hualde (2014) from the vocal sound, the acoustic cues that signal prosody also serves as hints to the linguistic context of the prosodically noticeable word and the utterance to which it belongs. Listeners