UNIVERSITI TEKNOLOGI MARA

IDENTIFICATION OF QURAN RECITATION SEGMENT FROM SPEECH VIDEO RECORDING

LILIANA BINTI NULKASIM @ MOHD KASSIM

BACHELOR OF COMPUTER SCIENCE (Hons.)
MULTIMEDIA COMPUTING

JULY 2017

ACKNOWLEDGEMENT

In The name of Allah, The Most Gracious and The Most Merciful. First and foremost, all praises to Allah The Almighty for giving me the strengths, guidance and patience in completing this thesis. I would like to express my sincere thanks to my dedicated supervisor, Mrs Haslizatul Fairuz Mohamed Hanum for giving me the opportunity to be her Dissertation students. I am deeply appreciated for her attentive support, dedication and encouragement throughout the course of my studies. Thank you very much for the professionalism support, guidance and her time at all stages of my research work in finishing this Dissertation Project.

I would like to convey my gratitude to my CSP650 lecturer, Dr Marina Binti Ismail whose assists, suggests, and encourages me in finishing this project report. The same goes to all of other lecturers who support me until this stage. I would like also to thanks for those who had been involved throughout the completion of this project either directly or indirectly, mostly to my friends who had always supported me.

Finally, thanks to my family who had given me moral support and advice throughout the completion of the project. Last but not least, once again I want to thank you to everyone for helping me from the starting until the end and I really appreciate it.

ABSTRACT

Identifying Quran recitation segment from speech video recording has become one of an active research themes in speech processing and in application based on Quran education. Therefore, a more efficient method for video segment identification within long speech video recording that will consuming time is urgently needed. This project develops a system to identify Quran recitation segment from speech video recording. This project applied manual video segmentation to differentiate between Quran and speech video content. This project selected 10 segmented video for Quran recitation and speech from one long speech video recording and extract the features using Praat tool. More specifically, two feature sets which are pitch and intensity are proposed to differentiate between Quran recitation and speech segment characteristics. A random forest classifier algorithm is employed in Spyder IDE using python language as a machine learning language for predict the type of an audio. The performance of the accuracy of the system will be trained and evaluated by the extracted audio features that will be compared with the segmented video which have been segmented manually. A classification accuracy of this project were 57% for pitch and 85% for intensity with the performance of 85% and 95% match accordingly. Therefore, by the accuracy of the result given has been proved that this project able to enhance the identification segment of Quran recitation.

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

1.1 Project background

Currently, identification of speech has achieved a high level of performance due to the enhancement of the algorithms and methods that have been used. Numerous field with different languages can be applied by this technology. Speech identification can be refer as a capability of a machine or program in verbal language to recognize and change words and phrases to a machine-readable format. It's already used in conversation of speech into text, medical and authorized career transcription tool and used on live television's subtitle.

However, to build speech identification systems, a speech database of an audio is required. This can be done by labelled and segmented the audio. The segmentation of the audio or video can be done by differentiating between the speech, music, Quran recitation segment, advertisement, or news segments from a speech video recording. Abdolali and Sameti (2012) comments that segmentation of an audio or video is the step of changing the identification of an argument for separating an input audio stream into each section. This step is necessary for advance one-dimensional and multi-dimensional signals of data processing (Prochazka and et al., 2008).

Nowadays there are few techniques that can be used to identify the segments of video such as using video key-frame and/or acoustic features extracted from the video. Initial observation of Malay sermon videos there is no difference between speech and Quran recitation frames. Therefore this project was proposed an enhancement of the identification of Quran recitation segment from speech video recording by evaluating the pitch and intensity features.