CHAPTER 1

INTRODUCTION

1.1 Introduction

In the era of artificial intelligence and machine learning, digital signal processing (DSP) system including audio engineering that have been widely used for large data manipulation. In general, DSP in audio engineering involves manipulation of signals originated from analogue data using digital signal devices. The system can be defined as digital representation involving the use of digital processors to perform wide variety of operations such as extracting, analysing, modifying data from captured signals. Figure 1.1 shows the basic process of digital signal processing in sound. Among other usable functions of DSP which they can easily be modified, reusable system performance, reliable of memory and logic DSP hardware, and allowing sophisticated application on complex devices (Gaydecki, 2004).

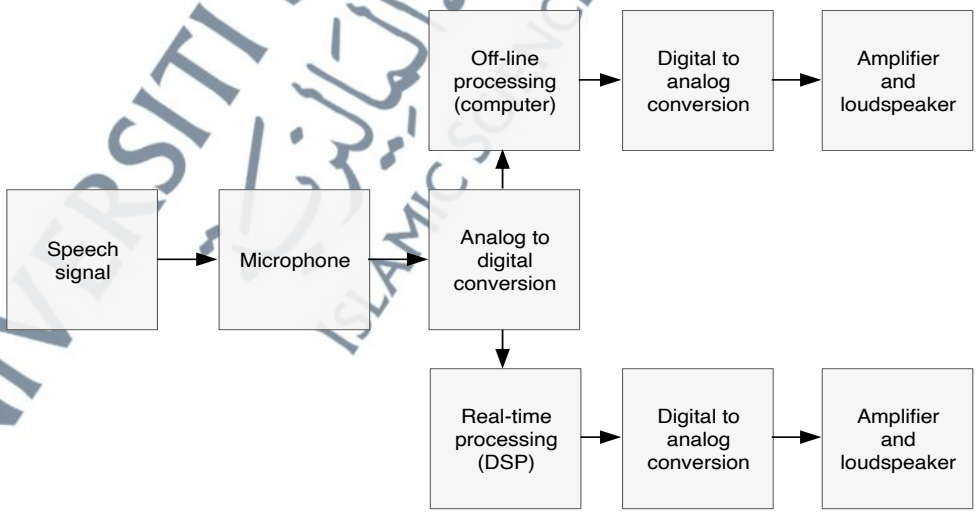

Figure 1.1 The foundations of digital signal processing in sound (Gaydecki, 2004)

Automatic speech recognition (ASR) is one of the speech processing systems which uses DSP tool to achieve its goals. The system concerned with interactions between human voice and machines such as computers and software as tools to process and analyse large amount of data, with the understanding of the voice content. ASR is the process of allowing a computer identify the words that a person speaks(spoken words) and convert it into text format by means of an algorithm implemented as a computer program for further computation (Yu, 2014). ASR is applied to many types of application such as voiced-based applications (Ni, 2015), language learning and translation of spoken sentences (Doremalen, 2016). Indeed, communicating fluently with machines and household appliance is intended to be dominant human machine interface in the near future, since this technology may eliminate the traditional handwriting problems, therefore increase the people and machines productivity at the same time. However, ASR still faces many challenges before it can employed by everyone everywhere because the machine capabilities developed are still quite primitive compared to their human counterparts. The system is widely used in various type of major languages around the globe, including not only English, Chinese, France and even Arabic too, due to common misinterpretation from the original meaning (Abushariah, 2012).

In learning the Quranic recitation, some areas such as *tajweed, haraqat* and *tabaqat* need to be learned appropriately to avoid misinterpretation or give different meaning of the original content (Ramli et al., 2018). One basic example of misinterpretation is stopping or starting at an incorrect place so that the meaning is spoilt, like stopping at ‘Laa ilaaha’ (i.e., there is nothing worthy of worship), without completing ‘illallaah’ (except Allah). This can be avoided by learning *tajweed*. Among

the proofs that the scholars bring to show the obligation of *tajweed* and its being an established Sunnah is that Almighty Allah Says in the Quran (what mean): "…And recite the Quran with measured recitation." (Quran 73:4)

One of the area is to learn maqamat. The study of maqamat is the study of *at-taghanni bi'l Quran*, or the study of beautifying the voice in recitation of the Quran. It is originated from the technique of Arabic maqam which imposed improvisation that defines the pitches, patterns, and development of a piece of music, and which is unique to Arabian art music (Albakri, 2020). Among other methods that has gained significance interest among researchers of Arabic language recently is automatic speech recognition system (Basem, 2017). Speech recognition for Arabic language is still at its early stages whereas western languages like English and Asian languages like Chinese and Mandarin are relatively well developed (Ambushariah et al., 2012).

Many researchers have addressed various challenges that most speakers faced when dealing with Arabic language in the Quran, due to its differences in recital technique (Shoaib et al., 2009). Like these varied melodies, the verses of the Quran vary widely in term of topics and event to generate different feelings to the listener. These audio features contained in the Quranic Maqamat Recitation (QMR) that need to be extracted and analyzed for the purpose of understanding the unique characteristics of the signal components . These properties contained uniquely in formants which define as a concentration of acoustic energy around a particular frequency that corresponds to a resonance in the vocal tract. Among other methods to extract strong features that characterise the complex nature of melodious audio speech signals is cepstral analysis (Giacobello et al., 2012). In recent Arabic recognition system studies, cepstral analysis

is commonly used technique by taking the DFT of log magnitude spectrum of a signal (Shahriar, 2021).

The overall process of profiling the acoustic profiles based on maqamat features, which consist of three stages. First stage is the data collection of the 32 samples of pre-recorded QMR audio files for each of seven (7) types of maqamat. That would give the total of 242 audio samples with average duration of 20-30s. Then it followed by extracting the maqamat features using existing and the proposed and enhanced mel-cepstral analysis based on warping function. The audio features will be validate by testing with simple recognition system model before performing the final stage which is the semantic analysis for an insight QMR knowledge base construction based on pre-designed ontology. Further areas of potential expansion, new ideas, and new areas of research for supporting Quranic learning for the Muslim community are also explored and identified.

1.2 Problem Statement

Many researchers have addressed various challenges speakers face when dealing with the Arabic language in the Quran, due to differences between that which is written and how it should be recited. The linguistic properties of Arabic are totally differed to those of other languages, the differences in Arabic include unique phonemes, phonetic features, and complicated morphological structures. Arabic is characterized by the presence of emphatic and pharyngeal phonemes, also known as *makhraj*. According to recent research, the phonetic features can be applied in the Quranic phonetic rules that the Quran's phonetic rules are connected by which the human tongue was created to function, and on the phonological side, the phonological rules can be applied in

Quranic phonological rules partially at some cases due to the complex nature of Quranic phonological rules and sound structures (Mohamed et al., 2020). Those properties and characteristics are considered unique and huge challenges for an Arabic recognition system. Moreover, when focusing on the problem of reciting the verses in the Quran, many weaknesses made by the reciters have been identified via scientific studies so the Quran needs to be properly learnt in order to avoid this problem (M. Ismail et al., 2011). The audio features contained in QMR are considered as a complex speech signal that need to be extracted and analyzed for the purpose of understanding the characteristics of the signal components. These properties contained uniquely in formants which define as a concentration of acoustic energy around a particular frequency that corresponds to a resonance in the vocal tract. A better understanding of the acoustic model and the differences between a correctly and wrongly pronounced recitation are considered key for efficient selection of mispronunciation detection method.

1.3 Objectives

The objectives of this thesis are to establish a new maqamat profiles based on acoustical elements contained in selected surah of Quranic verses using audio profiling system. Specifically, this research is to fulfil the following objectives:

1. QMR profiling method

   To extract and characterizes acoustical profiles from pre-selected QMR audio files using audio feature extraction method with warping function for audio profiling system.

2. Ontological acoustic model for Quranic profiling technique

   To define an acoustic model for Quranic profiling based on ontological extraction audio features which facilitate the development of a well-structured database system that support studio environment for audio engineering.

1.4 Motivations

Speech recognition has been widely studied for various type of languages, not only for English, but also for other languages including the Arabic audio recitations of the Holy Quran (Muhammad et al., 2010). Many works have been done to perform keyword search from Holy Quran. The main problem in all these works is that they have been widely developed based on task-driven which does not provide us semantic search based on acoustical elements with semantic measurements. Automatic analysis of audio content is a key aspect of information retrieval systems (Foote,1997, Wold et al., 1996) that deal with multimodal files. Specifically, for the purposes of recognition and classification, audio content-based analysis research has focussed on specific tasks which are the sound detection, classification, data retrieval. In the case of processing the complex and melodious speech, the Quranic verses is recited with acoustical properties contained in QMR that need to be extract and analyse for better understanding.

1.5 Research Scope

The scope of this thesis covers the speech processing system focusing on extracting significant features based on the acoustical elements contained in the Quranic recitation audio files. Three methods of feature extraction will be presented in analysing these audio signals, which are existing cepstral coefficient analysis, conventional and enhanced mel-cepstral analysis based on frequency warping function. The cepstral coefficients of the maqamat speech signal are used as parameters to extract significant features of the maqamat content. The final stage is to find the signal similarities using cross correlation method. The datasets consist of audio recording of Quranic recitation, which was recited by professional recitals also known as *qari.* There are various types of QMR exist for the reciters to recite the Quranic verses based on different origins and region. However, in this study only seven mostly used maqamat are selected which are Maqamat Bayyati, Hijaz Rast, Jiharkah, Nahawand, Soba, and Sikah. These maqamat have been selected since they were normally contributing to highest amplitude of the audio signal and hence lead to significant output of extracted audio features. The maqamat recordings are based on selected 32 Surah in Juzzu' Amma with Tabaqat Jawab. All audio samples are recorded in a recording studio in Faculty of Leadership and Management, Universiti Sains Islam Malaysia, Nilai for better audio quality.

1.6 Contributions

Currently, the difficulties for Quranic reciters to performs recitation without any standard guidelines which based on the Quranic content or theme have imposed motivations among researchers to study the characteristics of the sound profiles. This

thesis has three (3) major contributions that beneficial for the scientist in future. The contributions are mainly:

1. An enhanced audio feature extraction technique that combines cepstral analysis and direct warping function to optimise the extraction method for audio features and signal classification with the ability to characterized the sound profiles based on audio features.
2. An audio ontology that deduce the relationship between the emotions and the selected maqam based on extracted audio features.
3. Preliminary QMR acoustic profiles for ontological Quranic semantic audio search based on extracted audio features as a contribution to raising initiatives in profiling the acoustical elements of QMR for further audio processing analysis.

1.7 Structure of thesis

This thesis is organised into seven chapters and this section of Chapter 1 provides an overview and route map of the thesis. It includes the problem statements, objectives, scope and benefits and basic understanding of the whole structure of this thesis.

Chapter 2 presents the literature and background of this research by describing the overview of Quranic Maqamat, its types, characteristics and sound element contained therein. The first part consists of the overview of the existing cepstral analysis and warping function with some related works especially regarding on feature extraction and spectral descriptors. Later part discussed the overview and basic design

principles of semantic audio analysis related to ontological audio features contained in QMR.

Chapter 3 presents the overall framework and research methodology throughout the thesis. It includes data collection, audio feature extraction techniques based on ontological features and spectral descriptors as the first stage of profiling system and construction of knowledge base for QMR for final part.

Chapter 4 describes the first stage of the extraction method using spectral analysis and warping function. The proposed scheme with enhancement of an alternative frequency warping strategy by placing a direct warping of the spectrum during the filterbank process. The overall performance is evaluated for the optimum spectral descriptors and spectral envelope of the conventional mel-frequency and enhanced method. The final output of the signal envelope pattern is also discussed.

Chapter 5 present the final stage of the profiling system which covers ASR based on the extraction method in Chapter 4 by simple method of signal classification and recognition system. Here the knowledge base is constructed based on the ontological features deduce the significant audio features.

Chapter 6 describes further analysis on the selected audio recording using cross validation techniques. The performance after applying the profiling system using three validation techniques is also presented in this chapter. Then the results will be evaluated based on the most accurate classifiers. The overall performance after applying the profiling system is also presented in this chapter.

Chapter 7 concludes the thesis by presenting the thesis summary, contributions, and certain limitations that can be addressed to improve the efficiency of the audio feature extraction and signal matching for the profiling system in future works.

1.8 Summary

This preliminary chapter presents the introductions of the thesis including the objectives, motivations and benefits that contributes to the work done throughout this thesis. In the era of artificial intelligence and machine learning, DSP system among typical method that have been widely used for large data manipulation such as speech processing. Speech processing is one of the recent processing systems which uses artificial intelligence tools such as DSP. The system concerned with interactions between human voice and machines such as computers and software as tools to process and analyse large amount of data, with the understanding of the voice content. DSP are used to detect and characterized the complexity of such processing especially Arabic language. In recent Arabic recognition system studies, cepstral analysis is commonly used technique by taking the DFT of log magnitude spectrum of a signal. This has motivated the work done in this thesis by the possible finding on the significant sound profiles extracted from the audio signal.

The aim of this research is to develop a maqamat profiling system that enable to extract and characterizes acoustics profiles from the pre-recorded QMR and determine any correlation between acoustics properties and the Quranic rhetoric elements contained in QMR. This study will initiate an understanding on the characteristics of the maqamat content and hopefully contribute to raising initiatives to perform further quality checking of the recitations perform by the reciters. Further works is required to develop the overall integrated software application to be used by the reciters and Arabic speakers.