



Sudan University of Science and Technology  
College of Graduate Studies

**SOUND RECOGNITION FOR SELECTED VERSES OF  
THE HOLY QURAN**

التعرف الصوتي على آيات مختارة من القرآن الكريم

BY  
ASHRAF MOHAMMED ABDALLAH ELAMIN

SUPERVISOR  
Dr. ISMAIL EL-AZHARY

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF  
PHILOSOPHY IN COMPUTER SCIENCE

MAY 2012

Dedication:

To my mother.

## **Abstract**

In this study of sound recognition for selected verses of the Holy Quran, five verses selected from the Holy Quran. This done as a small number of verses contains all the Arabic phonemes. One hundred recitations of each verse were prepared in wave files. Verses had been recited by famous certified readers of the Holy Quran. A file created from each recite by extracting only the first word of the verse. A wave file of noise is set by the researcher's voice, for testing purpose. All wave files recorded at Sampling rate =22.050 kHz, PCM signed 16 bit mono.

Three types of coefficients are extracted from each wave file to represent features of speech. They are Mel Frequency Cepstrum Coefficients MFCC, Power Spectral Density PSD and Reflection Coefficients RC.

Also three techniques of speech recognition are used. They are Hidden Markov Models, Dynamic time warping and Artificial Neural Networks.

Test were done at two levels. The first stage was applied at the full verse level. All the three techniques mentioned were used. HMMs and ANNs trained by the first 30 samples and test done by all the 100 sample of each verse.

The second stage was applied in the same way, but the used samples were of the first word of the verse. HMMs technique only was selected to be used in recognition. That due to high recognition rates scored at the first stage. This stage repeated in the same way , but used the first 50 samples in training instead of the first 30 samples.

Mainly, HMMs scored high rates of recognition to coefficients used(MFCC, PSD and RC). Low recognition rates with high confusion scored by ANNs and DTW at

verse level. For all coefficients used, high scores of recognition rates with low confusion rate concentrated in HMMs with MFCC. MFCC scored higher recognition rates than PSD and RC.

## الخلاصة

في هذه الدراسة للتعرف الحاسوبي الصوتي على آيات من القرآن الكريم، تم اختيار خمس آيات من القرآن الكريم وهو أقل عدد من الايات يمكن تحتوي على جميع الوحدات الصوتية العربية ويكون زمن التلاوة أقصر زمن ممكن ليكون حجم البيانات أقل وتركيز الخصائص الصوتية المطلوب أكبر.

أعدت مائة ملف صوت لكل آية بأصوات قراء مشهورين ومجازين. اعد من كل منها ملف صوت اخر يحتوي فقط على الكلمة الاولى . اعد ملف بصوت الباحث يحتوي ضجيج، للتحكم أو معرفة مدى قرب النتائج منه وذلك بوضعة كنموذج للمقارنة.ملفات الصوت مسجلة بمعدل ٢٢,٠٥٠ كيلوهرتز ١٦ بت مونو.

استخرجت ثلاثة انواع من المعاملات لتمثل الخصائص الصوتية من هذه الملفات. والمعاملات هي معاملات السبسترم وكثافة الطاقة ومعاملات الانعكاس. واستخدمت ثلاثة تقنيات للتعرف على الكلام هي نماذج ماركوف المخفية والبرمجة الديناميكية والشبكات العصبية الاصطناعية.

أجريت الاختبارات على مرحلتين. المرحلة الأولى على مستوى الآية ، باستخدام التقنيات الثلاث المذكورة للتعرف على الكلام. نماذج ماركوف المخفية والشبكات العصبية الاصطناعية تم تدريبها على العينات ال ٣٠ الأولى من كل آية، والاختبار استخدمت فيه جميع العينات ال ١٠٠ من كل آية.

المرحلة الثانية بنفس الطريقة للمرحلة الأولى ، ولكن على مستوى الكلمة (الكلمة الأولى من الآية). استخدمت فقط نماذج ماركوف المخفية في الاختبارات. نظرا لانها سجلت ارتفاع في معدلات التعرف على الكلام في المرحلة الأولى. اعيدت المرحلة الأخيرة بنفس الطريقة، واستخدم فيها أول ٥٠ عينة في التدريب بدلا من ٣٠ عينة.

بخلاصة النتائج وجد الاتي:

- نماذج ماركوف المخفية (HMMS) سجلت أكبر معدل تعرف مع معاملات MFCC وتميزت عموما بالاستقرار والاتزان في ادائها.
- الشبكات العصبية و تقنية DTW سجلت أقل معدلات تعرف ما عد في الاية الاولى مع MFCC والاية الرابعة مع PSD وتميزت بالتباين في ادائها لذلك لم تستخدم في المرحلة التالية في البحث.
- في نماذج ماركوف عموما سجلت معاملات MFCC معدلات تعرف أعلى من PSD و RC .
- بعض الكلمات سجلت معدلات تعرف عليها بصورة اكبر من غيرها.

## **Acknowledgements**

It is here where my deeply thanks go to the various people who, during several stages of this thesis, provided me with useful and helpful assistance. My wife who, made efforts to create environment as suitable as possible. I would like to commend the interest, encouragement and efforts done by Dr. Ismael Elazhary. Thaks also go to Professors from the Islamic Institute for Translation; Attoum and Ahmed Khalid for Review and the language.

## Contents

Dedication .....	II
Abstract .....	III
Acknowledgements .....	VI
List of figures .....	XI
List of tables .....	XIII
1. Introduction .....	1
1.2 Objectives .....	3
1.3 Goal of the study .....	3
1.4 Study limits .....	3
1.5 Methods .....	4
1.6 Review of previous studies .....	5
1.7 Thesis organization .....	10
2 Adaptation of speech recognition technologies to the Holy Quran .....	11
2.1 Introduction .....	12
2.2 Steps of signal processing .....	13
2.2.1 Pre-emphasis .....	13
2.2.2 Voice Activation Detection .....	14
2.2.3 Signals framing .....	15
2.2.4 windowing with hamming window .....	15
2.3 features extraction .....	16
2.3.1 MFCC calculation steps .....	16
2.3.1.1 Fast Fourier Transformation FFT .....	17
2.3.1.2 Using a bank of 20 filters .....	17

2.3.1.3 Natural logarithm calculation .....	18
2.3.1.4 Discrete cosine transformation calculation .....	18
2.3.1.5 Derivative calculation .....	19
2.3.1.6 Mean and variance .....	19
2.3.2 Power Spectral Density( PSD) .....	20
2.3.2.1 PSD Mean and variance .....	21
2.3.3 Reflection coefficients (RC).....	21
2.3.3.1 The RC mean and variance .....	22
2.4 Recognition techniques used .....	22
2.4.1 Hidden Markov Models (HMM) .....	22
2.4.2 Dynamic Time Wrapping .....	23
2.4.3 Artificial Neural Networks .....	23
2.5 The template program SRHQ.....	24
3. Testing analysis and results .....	27
3.1 introduction .....	28
3.2 Hidden Marcove Models.....	28
3.2.1 Mel Frequency Cepstrum Coefficients (MFCC) and HMMs	28
3.2.2 Power Spectral Density (PSD) and HMMs .....	29
3.2.3 Reflection Coefficients (RC) and HMMs .....	30
3.2.4 HMMs recognition rates summary.....	31
3.3 Dynamic Time Warping( DTW).....	32
3.3.1 Mel Frequency Cepstrum Coefficients mean (MFCC mean) and DTW .....	32
3.3.2 Mel Frequency Cepstrum Coefficients variance (MFCC var) and DTW .....	33



3.3.3 Power Spectral Density mean (PSD mean) and DTW ....	34
3.3.4 Power Spectral Density Variance (PSD var) and DTW ....	35
3.3.5 Reflection Coefficients mean (RC mean) and DTW .....	36
3.3.6 Reflection Coefficients variance (RC var) and DTW.....	37
3.4 Artificial Neural Networks (ANNs) .....	38
3.4.1 Mel Frequency Cepstrum Coefficients (MFCC) with ANN	38
3.4.2 Power Spectral Density (PSD) with ANN .....	39
3.4.3 Reflection Coefficients (RC) with ANN.....	40
3.5 Summary of DTW and ANNs recognition rates .....	41
3.6 Testing one word analysis and results.....	42
3.7 Testing one word analysis and results using 50 sample in training	45
4. Discussions .....	49
4.1 Introduction .....	50
4.2 Verse level .....	50
4.2.1 Comparisons using MFCC and HMMs .....	50
4.2.2 Comparisons using PSD and HMMs on the Verse level .....	50
4.2.3 Comparisons using RC and HMMs on the Verse level .....	51
4.2.4 Observations of using MFCC mean and variance and DTW ..	51
4.2.5 Using PSD mean and variance and DTW .....	52
4.2.6 Using RC mean and variance and DTW on the Verse level	52
4.2.7 Using MFCC and ANN on the Verse level.....	52
4.2.8 Using PSD and ANN on the Verse level	53
4.2.9 Using RC and ANN on the Verse level	53
4.3 Word level	53
4.3.1 Comparisons using MFCC and HMMs	53

4.3.2 Comparisons using PSD and HMMs on the Word level ....	54
4.3.3 Comparisons using RC and HMMs .....	55
4.4 Summary.....	55
4.4.1 Word level using 30% of samples for training the models	56
4.4.2 Word level using 50% of samples for training models ...	56
4.5 Previous results for (Hafss ).....	56
4.6 Experimental results for (E-Hafiz ).....	57
5. Conclusions and future recommendations.....	58
5.1 Conclusion.....	59
5.2 Future recommendations.....	59
References .....	61
List of appendices tables .....	64
Appendix (A) Names of the recitors .....	68
Appendix (B) Tables contain results of testing utterance of verses .....	83
Appendix (C) Results of testing utterance of words using HMMs training F30	179
Appendix (D) Results of testing utterance of words using HMMs training F50	209

## List of figures

1.1	Features extraction and model initialization	3
2.1	The wave form of recite No: 1 of verse (1) by Ahmed Ali Alajami.	12
2.2	The wave form of the pre-emphasized signal.	13
2.3	Undesired ends removed from the sides of the wave form.	13
2.4	Frame No.1, Length = 660 samples (30 milliseconds) begins at sample1 and ends at 660.	14
2.5	hamming window, length=660.	14
2.6	Frame No:1 windowed with Hamming window	15
2.7	Frame No:1 transformed to the frequency domain the FFT is used.	16
2.8	Frame1 filtered with 20 filters in a filter bank.	16
2.9	Natural logarithm calculated for the filtered data	17
2.10	The discrete cosine transform applied.	17
2.11	MFCC 12 coefficients extracted from frame No. 1	18
2.12	Periodogram Power Spectral Density estimation .	19
2.13	The 12 max power frequencies.	19
2.14	The 12 lags autocorrelation coefficients.	20
2.15	Reflection coefficients (RC).	20
2.16	SRHQ program GUI	23
2.17	SRHQ Dynamic Time Warping function (DWT).	24
3.1	distribution of recognition and confusion, data of Table 3.1	27
3.2	PSD distribution of recognition and confusion, data of Table3.2.	28
3.3	RC distribution of recognition and confusion, data of Table 3.3.	29
3.4	Summary recognition rates using HMMs.	30

3.5	MFCC mean distribution of recognition and confusion.	31
3.6	MFCC var distribution of recognition and confusion.	32
3.7	PSD mean distribution of recognition and confusion.	33
3.8	PSD mean distribution of recognition and confusion.	34
3.9	RC mean distribution of recognition and confusion.	35
3.10	RC var distribution of recognition and confusion.	36
3.11	MFCC distribution of recognition and confusion.	37
3.12	MFCC distribution of recognition and confusion.	38
3.13	MFCC distribution of recognition and confusion.	39
3.14	Summary of recognition rates using DTW, ANNs.	40
3.15	MFCC distribution of recognition and confusion.	41
3.16	PSD distribution of recognition and confusion.	41
3.17	RC distribution of recognition and confusion.	42
3.18	summary recognition rates using HMMs.	42
3.19	MFCC distribution of recognition and confusion.	43
3.20	PSD distribution of recognition and confusion.	44
3.21	RC distribution of recognition and confusion.	44
3.22	summary recognition rates using HMMs.	45

### **List of tables:**

2.1	Verses recite wave files used.	12
2.2	12 MFCC extracted from frame No.1.	18
2.2.1	Mean and variance of the 12 MFCCs of Table 2.2.	18
2.3	The first 12 max power frequencies selected of PSD .	19
2.4	The 12 lags Autocorrelation coefficients.	20
2.5	Reflection coefficients (RC).	20
2.6	Verses, model names, files used for training and verses in Arabic.	21
2.7	models initialized from three types of features extracted from five verses wave files and a wave file of human noise.	21
3.1	MFCC distribution of recognition and confusion.	27
3.2	PSD distribution of recognition and confusion using HMMs.	28
3.3	RC distribution of recognition and confusion using HMMs	29
3.4	summary of recognition rates using HMMs .	30
3.5	MFCC mean distribution of recognition and confusion using DTW.	31
3.6	MFCC var distribution of recognition and confusion using DTW.	32
3.7	PSD mean distribution of recognition and confusion using DTW.	33
3.8	PSD var distribution of recognition and confusion using DTW.	34
3.9	RC mean distribution of recognition and confusion using DTW.	35
3.10	RC var distribution of recognition and confusion using DTW.	36
3.11	MFCC distribution of recognition and confusion using ANN.	37
3.12	PSD distribution of recognition and confusion using ANN.	38

3.13	RC distribution of recognition and confusion using ANN.	39
3.14	Summary of recognition rates using DTW, ANNs.	40
3.15	MFCC distribution of recognition and confusion.	41
3.16	PSD distribution of recognition and confusion.	41
3.17	RC distribution of recognition and confusion.	42
3.18	Summary recognition rates using HMMs from the three previous Tables.	42
3.19	MFCC distribution of recognition and confusion.	43
3.20	PSD distribution of recognition and confusion.	43
3.21	RC distribution of recognition and confusion.	44
3.22	Summary recognition rates using HMMs from the three previous Tables.	45