International Journal on Islamic Applications in Computer Science And Technology, Vol. 7, Issue 4, December 2019, 01-10

Study to Build a Holy Quran Text-To-Speech System

Nadjla Bettayeb [1], Mhania Guerti [2]

[1,2] Signal and Communications Laboratory , Electronics Department,
Ecole Nationale Polytechnique, Algiers, Algeria
[1]nedjla.bettayeb@g.enp.edu.dz , [2] mhania.guerti@enp.edu.dz

Abstract

Building a Text-To-Speech (TTS) system needs the study and the preparation of several modules. In this paper we present some of these parts to build a speech synthesis system for the Quran recitation. Firstly, we will describe the database building process, because of its importance, especially in concatenative speech synthesis methods. Secondly, a Grapheme to Phoneme Transcription (GPT) module will be presented. This latter was developed by elaborating a set of Tajweed and Arabic reading rules. They were also organized in a new way to gain some computational time. Those developed modules are evaluated by verifying the correct transcription and the database's comprehensiveness, which gives satisfactory results.

Keywords: Database, Grapheme to Phoneme Transcription, Speech Synthesis, Holy Quran.

1. Introduction

Speech synthesis systems (TTS) are widely used in our actual life. Applying this technique to the Holy Quran recitation may be advantageous by using less memory space (instead of storing all surahs and verses) and give the user a freedom to choose the part he/she wants to hear. One of the currently used speech synthesis methods are the corpus-based, in which it is based on the concatenation of natural sound segments. Thought, for a good result they require a big database, rich with units with different prosodic and linguistic contexts (Dutoit, 2008).

The transcription is the first module in any TTS system that refers to the grapheme to phoneme conversion. It is also called phonetization as it is a transformation of a written text into a representation of its phonetic pronunciation (Elshafei, Al-Muhtaseb, & Al-Ghamdi, 2002). In previous researches, we find three main transcription methods (El-Imam, 2004; Laurent, Meignier, & Deléglise, 2014):

- The dictionary-based method that needs rich dictionary of the language words with their phonetic transcription;
- The rule-based method that transcribes the text by applying well-defined and regular rules;
- The data driven-based method, that applies the latest data mining techniques and algorithms.

The choice of the suitable method depends on the language and the complexity of its spelling and pronunciation system. When Standard Arabic (SA) is written with all its diacritics marks (as the Quran case), it can be considered as a language with a regular spelling system and has better grapheme to phoneme direct conversion compared to others like English. Therefore, a rule-based transcription is the best method with a dictionary of exception for some special words (Bellegdi & Muhtaseb, 2015; El-Imam, 2004; Imedjdouben & Houacine, 2013).

In this work we describe the process to build a database for a Quran speech synthesis system, because of the lack of specialized one. We also discuss the GPT module by the rule-based method as it seems to be the best one for Arabic language (Quranic text). This transcription process differs from the existing works (Bellegdi & Muhtaseb, 2015) in the separation between the SA rules and the Tajweed rules application. Adopting two successive modules was to facilitate the implementation and the verification of the second process (the Tajweed rule verification). The rules structure is quite similar to the works of (Bellegdi & Muhtaseb, 2015; El-Imam, 2004; Saidane, Zrighi & Ben Ahmed 2004; Sawalhaa, Brierley, Atwell, & Dickins 2017) but we grouped and applied them depending on the case of the grapheme (its position, its type, etc.) to speed up the entire process.

2. Work methodology

In this paper our work methodology consists of the following steps:

2.1 Database building

To build our database we used some Holy Quran verses' recordings that were downloaded from a website with many reciters and recitation styles. Among the existing records, we chose the ones recited by Dr Aymen Roshdy Sweed (Sweed, 2014), because of its normal and monotonous recitation style (with no melody) as they are originally designed to teach the Tajweed rules. With an objective to gather all the existing phonemes' [1] combinations in at least two different contexts, we opted to 772 whole or a part of verses. The sentences were from 1 to 10 words length and with a sampling rate of 44.1 kHz in most cases. This verses selection resulted in 45 min from 39 h of the total Quran recording.

After the corpus preparation, the chosen sounds were segmented into smaller units, to extract the audio files of each segment for the database building. Using Praat [2] software, the segmentation step was performed manually, in which we based on our listening and analysing the audiogram and the spectrogram of each verse sound. However, before passing to the final segments, the phonemes boundaries were first defined to facilitate the final segmentation (Figure.1).

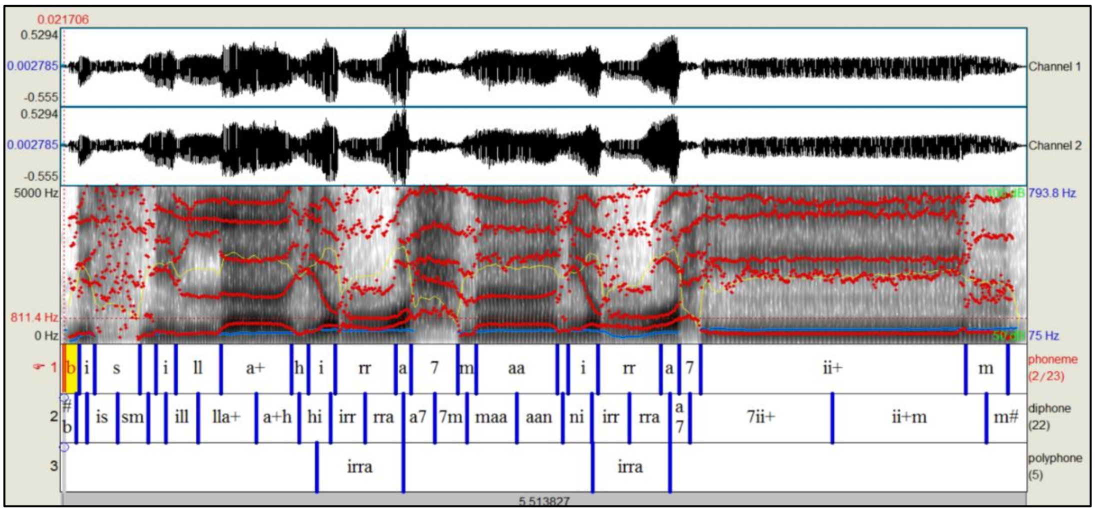

Figure 1. Segmentation of the sentence [bismillahirraћmaanirraћiim].

[1] the basic sound unit
[2] http://www.fon.hum.uva.nl/praat/

The main unit used in the database is the diphone that extends from the stable part of a phoneme to the stable part of the adjacent one, including the transition between them. The choice of the diphone was done to avoid the maximum of spectral discontinuities that can occur with the concatenation of smaller units, and not to overload the database with a bigger sound segment. Beside the diphone we adopted the polyphone unit which is a combination of 3 to 4 phonemes. It was specially used to phonemes that present short duration like the Hamza [ʔ], or difficult to segment like the case of semi-vowels when they merge with the adjacent vowels.

After the segmentation that results of 11070 units, the extracted sound files had been saved in the database with a specific name as shown in Figure. 2. After that, each sound unit was analysed using the software speech analyser [3]. This analysis was performed to extract new files that include some of the units' acoustic features: Energy values, Fundamental frequency (F_0) and other formant values. Those files were also saved in the database with the same name as the sound files. Then the 12 Mel Frequency Cepstral Coefficients (MFCC) of each sound segment were calculated using a MATLAB file exchange "HTK MFCC MATLAB" [4] and saved after that in the database.

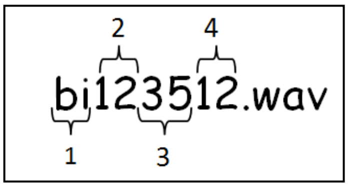

Figure 2. Notation of the extracted audio files.

In which:

1: the phonemes composing the unit ([b] and [i]), which are written in our own proposed code (Table 1);
2: a code for the unit positions (the word in the sentence and the unit in the word);
3 and 4: codes represent the Left and Right Contexts of the unit respectively.

Table 1. International Phonetic Alphabet (IPA) and proposed phonetic code for the arabic characters.

Arabic character	IPA	Proposed code	Arabic character	IPA	Proposed code	Arabic character	IPA	Proposed code
أ	[ʔ]	[!]	ش	[ʃ]	[c]	م	[m]	[m]
ب	[b]	[b]	ص	[ʂ]	[$]	ن	[n]	[n]
ت	[t]	[t]	ض	[ɖ]	[£]	ه	[h]	[h]
ث	[θ]	[8]	ط	[ʈ]	[6]	و	[w]	[w]
ج	[ʤ]	[5]	ظ	[ʐ]	[%]	ي	[j]	[j]
ح	[ħ]	[7]	ش	[ʃ]	[c]	َ	[a]	[a]
خ	[x]	[x]	ع	[ʕ]	[3]	ـَا	[aa]	[aa]
د	[d]	[d]	غ	[ɣ]	[g]	ُ	[u]	[u]
ذ	[ə]	[4]	ف	[f]	[f]	ـُو	[uu]	[uu]
ر	[r]	[r]	ق	[q]	[q]	ِ	[i]	[i]
ز	[z]	[z]	ك	[k]	[k]	ـِي	[ii]	[ii]
س	[s]	[s]	ل	[l]	[l]			

[3] https://software.sil.org/speech-analyzer/
[4] https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab

Special codes for the Tajweed rules	
[aa +]	For the 4 counts Madd
[aaa]	For the 6 counts Madd
[~]	Joinder the "n" in the case of Ikhfaa (hiding) ([nc]=>[n~c])
[-]	Stands between two assimilated letters ([n-l])
[_]	Represents the Qalqalah (the silent phoneme vibration) ([bq]=>[b_bq])
[']	Added to the [m] that was changed from [n] ([n b]->[m'b])) in the replacing (Iqlab) case

2.2 Transcription

To build the GPT module for the Holy Quran using the rule-based method we need to implement all the SA and the Tajweed rules (Abd-Elhamid, 2009) in addition to an exception dictionary for the special words. Figure 3 presents the block diagram of steps followed to transcribe a fully vocalized (contains all the necessary diacritics marks) Quranic text that are described below.

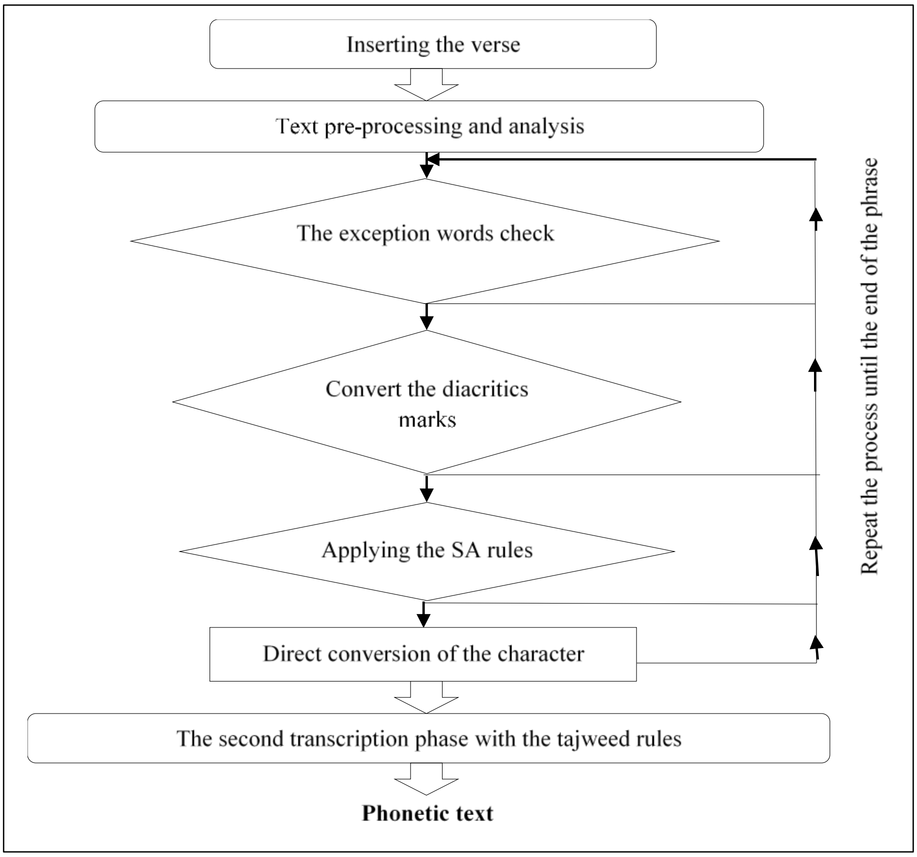

Figure 3. Block diagrame of the Holy Quran phonetic transcription module.

2.2.1 Text pre-processing

This phase consists of interpreting or removing the special characters and defining the beginning and end of the text to be synthesised.

The input text of the designed module is a simple Quranic verse (free of symbols, numbers, abbreviation, etc.) that will be entered using the keyboard. Therefore, this text doesn’t need much pre-processing. So, in this step we just dealt with some little keyboard symbols (full stop and comma) to indicate the end of the verse to be synthesised.

2.2.2 The alphabetic text analysis and the first transcription phase

In this step the pre-processed text is transformed into a phonetic written, in which we used a personal phonetic code to facilitate the use of this phonetic writing (Table 1). This code base on the International Phonetic Alphabet (IPA) nomination with replacement of complex symbols by keyboard characters. To realize this conversion the following step must be verified and implemented in order:

2.2.2.1 Exception words

When the Quranic text is entered into MATLAB workspace, it is automatically interpreted as a numeric code (windows-1256). Then it is analysed letter by letter and word by word, to find if it contains any special word (words that does not obey to any of the Arabic reading rules and have special pronunciation). Comparing the word code with codes in the exception dictionary, the transcription is performed by uploading from the database the corresponding one to each word. This dictionary was built based on the used word in (Bellegdi & Muhtaseb, 2015; Imedjdouben & Houacine, 2013; Ramsay, Alsharhan, & Ahmed, 2014) and after analysing the Quran words and their pronunciation. Table 2 shows a part of this dictionary.

Table 2. Phonetic transcription of some exception words in the dictionary.

The special word	Phonetic transcription
هَؤُلاَء	[haa!ulaa!]
أولَئِك	[!uulaa!ik]
هَذَا	[haa4aa]

2.2.2.2 Diacritics marks

After dealing with the exception words, the diacritics marks must be converted into phonetic characters : the Fatha, Damma and Kasra presented as vowels (َ-> [a], ُ-> [u], ِ-> [i]); the nunation as vowel + [n] (ً-> [an], ٌ-> [un], ٍ-> [in]) ; the gemination mark into double Consonant (e.g: نّ -> [nn]) and the lengthening (Madd) as a double vowel (ـَا -> [aa], ـُو -> [uu], ـِي -> [ii], آ -> [!aa]), and finally symbolize all the Hamzah kinds with one phonetic character (ئ , ؤ , أ ,ء -> [!]).

2.2.2.3 Applying the main SA rules

While converting the diacritics marks, the text analysis module continues to transcribe the whole verse by applying the essential Arabic reading rules. This step consists of replacing the alphabetic character by none, one or two phonetics characters, depending on its position (in the word or the sentence), its left or right neighbour characters as follows.

a) *The definition " ال " [al] rule*: when the character “ ا ” (Alif) comes in the beginning of the sentence and followed by a silent "ل" [l], the “ ا ” is transcribed as [!a]; and if the character " ل " is followed by one of the solar letter **ت، ث، د، ذ، ر، ز، س، ش، ص، ض،ط، ظ، ل، {ن}**, it will be removed and the third character (the solar letter) will be doubled. In some cases, the third letter already had the gemination mark, so we just need to check its presence. Otherwise if the third letter is a lunar letter {**أ، ب، ج، ح، خ، ع، غ، ف، ق، ك، م، هـ، و، ي**}, the "**ل**" is kept silent (Czerepinski, 2000; El-Imam, 2004). When the character " ا " does not come in the beginning of the sentence, it will be removed without changing the transcription rules for the rest of the word.

E.g: (and the sun) “ **وَ الشَمسِ** ” -> [waccamsi] ; (The Shocker) “ **الْقَارِعَةُ** ” -> [!alqaariratu].

b) Hamzat Wasl rules: The Hamzat wasl is a silent Alif “ ا ” in the beginning of the word. This character is removed when it comes in the middle of the sentence and has different reading ways when we start the sentence with it (El-Imam, 2004; Bellegdi & Muhtaseb, 2015). Its reading rules depends on the word type (noun or verb)(Abd-elhamid, 2009), in which we simplified them into rules works with an exception words dictionary and does not need the word type as follows:

If the third letter is followed by the marks (َ , ِ) or the word is one of the exception list {' 'ابْنٌ' 'ابْنَت' 'امْرُؤٌ' 'امْرَأَت' 'امْرَأَة' 'اثْنَتَ' 'اثْنَ' 'اسْم'} the Alif is transcribed as [!i], otherwise if the third letter is followed by the mark (ُ) the Alif will be transcribes as [!u] except for the words {' امْضُو ' 'اقْضُو' 'امْشُو' 'ابْنُو' 'اثْنُو'} where it will be transcribed as [!i].
E.g : (read) " اقْرَأ " -> [!iqra!] ; (go on) " امْضُو " -> [!im£uu].

c)The Alif Maqsura rule: It is kind of a silent Alif " ى " that comes in the end of certain words (El-Imam, 2004; Bellegdi & Muhtaseb, 2015), in the Quran it has two cases :

- if it is preceded by the Fatha mark (َ) it disappear and the vowel [a] becomes longue vowel [aa]. E.g: (upon) " عَلَى " ->[3alaa]
- if it was precedes by the Tanween mark (ً) it disappear without any other change. Eg : (a guide) " هُدًى " -> [hudan].

2.2.2.4 The direct conversion: After checking all the previous cases, the rest of characters are transcribed by direct conversion to the phonetics writing (Table 1). In addition, the Sukune mark " ْ " is removed, and finally the silent phoneme [#] is added to the beginning and end of the transcribed text.

2.2.3 The Tajweed rules

With the same process as the main rules, the Tajweed rules are checked and applied in a specific order to cover the variants that incorporate with the recitation of Quran, by a second transcription of the previous phonetic text. To organize this process and reduce the execution time, these rules' application was divided into cases depending on type and position of the character. This is because, there are rules work in the end of the word, others applied to a specific consonant type (like the assimilation in the [n] case), etc.

2.2.3.1 Rules in the end of the phrase

These kinds of rules are applied when certain phoneme type comes in the end of the phrase like:

- When a vowel is the last phoneme in the phrase, it is removed. While if it is preceded by "ة" [t], both characters are removed and replaced by the phoneme [h]. E.g:" الْقَارِعَةُ " ->[!alqaari3atu#] -> [! a l q aa r i 3 a h #];
- The Tanween ([in] and [un]) is removed too, but we made an exception word list that end with a real [in] or [un] to differentiate between them and the words ending with Tanween. E.g: (Working hard and exhausted) " عَامِلَةٌ نَاصِبَةٌ " -> [3aamilatun naa$ibatun#] -> [# 3 aa m i l a t u n-n aa $ i b a h #].

2.2.3.2 The double phoneme Rules

When a double phoneme occurs, two cases must be checked:

- If it is one of the phonemes {q, 6, b, 5, d}, the Qalqalah rule must be applies by replacing the found phoneme by its specified symbol (Table 1). E.g: (pouring abundantly) " ثَجَّاجًا " -> [8a55aa5an#] -> [8 a 5_55_5 aa 5_5 aa #].

- If it is a vowel, the Madd rules are called, and replace the longue vowel characters by its right character.
  E.g: (O you) " يَا أَيُّهَا " -> [jaa !ajjuhaa] -> [j aa+ ! a jj u h aa].

2.2.3.3 The general cases' rules

It regroups the rest of the Tajweed rules (Bellegdi & Muhtaseb, 2015; Czerepinski, 2000) like:

- The assimilation (Idgham) rule: when the phoneme [n] followed by one of the phonemes {j, r, m, w, l, n} in another word, a new representation takes the place of these two phonemes.
  E.g: (Because he sees himself) " أَنْ رَآهُ " -> [!anra!aahu] -> [! a n-r a ! aa h];
- The hiding (Ikhfaa): when the [n] followed with one of the phonemes {$, 4, 8, k, 5, c, q, s, d, 6, z, f, t, £, %}, we change the [n] symbol by [n~].
  E.g: (And who were) "وَمَن كَانَ " -> [wamankaana] -> [w a m a n~ k aa n a].

While checking the previous cases and applying the Tajweed rules, the space between words is also removed because of the speech continuous feature. The output of this module is a list of characters that their concatenation gives the desired phonetic transcription.

2.3 The Abbreviated letters verses

Beside the normal type of sentences in the Quran, we find an exceptional type of verses called the " the Abbreviated letters ". the later consist of a combination of one to five characters from fourteen Arabic letters (just consonants) and appear in the beginning of 29 Surahs (E.g. " كهيعص" [khjʕʂ] , " ألم " [ʔlm], etc). Those verses are recited by concatenating the pronunciation of each composing letter out of context, and they have no linguistic meaning. E.g." ألم " => [ʔalif laam miim].

The database part of these verses consists of word segments for each letter pronunciation, assigned with its position, preceding and following letter codes. Their transcription is done using an exception dictionary, in which it assigns to each letter its correct recitation.
Eg : , " ألم " => [!aliflaaammiiim].

2.4 The evaluation

Because of the lack of standard method to evaluate the transcription system, it seems that the comparison between the automatic transcription result and a manual one, is the best way (El-Imam, 2004; Imedjdouben & Houacine, 2013; Ramsay et al., 2014). In our tests, we included 20 different verses that were selected in a way to cover the maximum of transcription cases. Then as a fully vocalized text, the verses were entered to the phonetization program and a comparison between the output result and a manual transcribed version was done. Table 3 presents some of the test verses and their manual transcription, while Figure 3 shows the results of the first transcription phase, and the final automatic transcription of the chosen examples. Even with the few test samples compared to the other phonetization systems (El-Imam, 2004; Imedjdouben & Houacine, 2013; Ramsay et al., 2014), the results in Figure 4 seem very good with a 100% correct transcription of phrases contain all the implemented rules, in addition to a mean run-time of 0.07s for the verse (using an intel (R) Core (TM) i7-4600U CPU @ 2.10 GHz 2.70 GHz).

Table 3. Manuel transcription of some test examples.

The example number	Quranic verse	Manual transcription
1	يَجْعَلُونَ أَصَابِعَهُم فِي آذَانِهِمْ مِنَ الصَّوَاعِقِ حَذَرَ الْمَوْتِ'	[#ja5_53aluuna!a$aabi3ahum fii+!aa4aanihim mina $$awaa3iqi7a4aralmaw+t#]
2	'وَمِنْ شَرِّ النَّفَّاثَاتِ فِي العُقَدِ'	[#wamin~ carri nnaffaa8aatifil3uqadd#]
3	'كَلَّا لَيُنبَذَنَّ فِي الْحُطَمَةِ'	[#Kallaalajum'ba4anna fil7u6amah#]
4	أولَئِكَ عَلَى هُدًى مِنْ رَبِّهِمْ وَ أولَئِكَ هُمُ ' الْمُفْلِحُونَ'	[#!uulaa!ika3alaahudan-min-rabbihimwa!uulaa!ika humu lmufli7uu+n#]
5	"فَلْيَعْبُدُوا رَبَّ هَذَا الْبَيْتِ"	[#falja3buduurabbaha4aa lbaj+t#]
6	'أَنْ رَآهُ اسْتَغْنَى'	[# !an-ra!aahustagnaa #]
7	'فَأُمُّهُ هَاوِيَةٌ'	[#fa!ummuhuuhaawijah#]

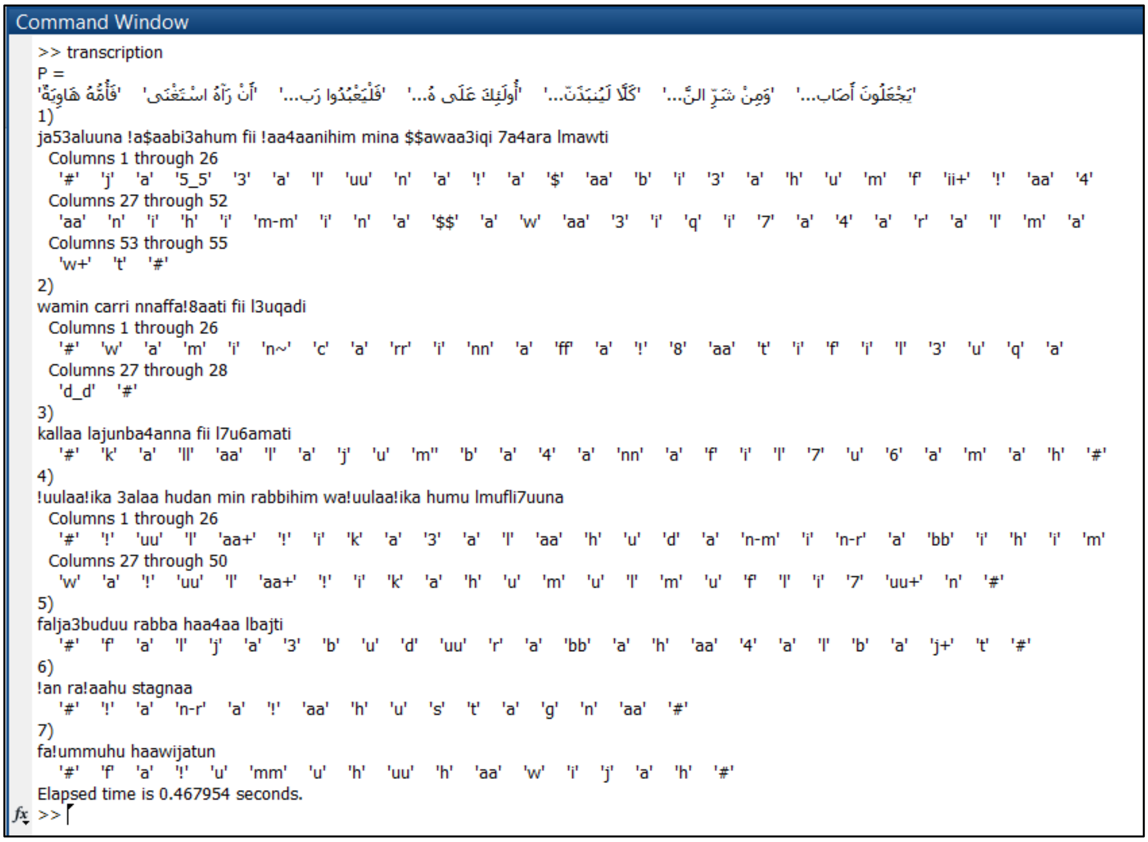

Figure 4. First transcription part and the full automatic transcription of some test examples.

Beside the transcription's evaluation, we also assessed the database comprehensiveness. By a simple concatenative synthesis test, we generated around 45 verses and checked if there is any missing sound part in them. Those verses were chosen with different length (from 17 to 70 phonemes) to cover the maximum of phonemes combinations. During this test we did not get any missing part in the synthesised verses, which is a good result.

In this evaluation we did not focus on the resulted speech quality, as it depends on other factors like how the units are selected or the way to concatenate them. However, we encountered some bad quality sound units, especially when they are concatenated with good ones. Those bad units were because of the original recordings' environment, in which we were obliged to take them because of their monotonous style. Fortunately, this can be reduced by some speech smoothing techniques.

3. Conclusion

Aiming to developpe a speech synthesis system for the Quran recitation, this paper discussed two main preparation's steps to build such system. It explains the database building, from the corpus preparation and the sound units' segmentation and annotation. It also presented the building steps of a transcription module. With the rule-based method for the transcription, the system developing steps were discussed and some of the implemented rules were presented and explained. The evaluation of the transcription module gives an encouraging result to integrate it in a TTS system and the database shows its units' coverage.

References

Abd-elhamid, S. (2009). تيسير الرحمان في تجويد القرآن [Tayseer Al - Rahman in the recitation of the Quran]: dar Altakoa, Egypt.

Bellegdi, S., & Muhtaseb, H. (2015). Automatic Rule Based Phonetic Transcription and Syllabification for Quranic Text. International Journal on Islamic Applications in Computer Science And Technology, 3(4).

Czerepinski, K. C. (2000). Tajweed Rules Of The Qu'ran : part 1 (Vol. 1): Dar Al-Khair for Pub .& Dist.

Dutoit, T. (2008). Corpus-Based Speech Synthesis. In J. Benesty, M. M. Sondhi & Y. A. Huang (Eds.), Springer Handbook of Speech Processing (pp. 437-456). Berlin, Heidelberg: Springer Berlin Heidelberg.

El-Imam, Y. A. (2004). Phonetization of Arabic: rules and algorithms. Computer Speech & Language, 18(4), 339-373.

Elshafei, M., Al-Muhtaseb, H., & Al-Ghamdi, M. (2002). Techniques for high quality Arabic speech synthesis. Information Sciences, 140(3–4), 255-267.

Imedjdouben, F., & Houacine, A. (2013). Automatic Phonetization of Arabic Text. In A. Amine, A. M. Otmane & L. Bellatreche (Eds.), Modeling Approaches and Algorithms for Advanced Computer Applications (pp. 85-94). Cham: Springer International Publishing.

Laurent, A., Meignier, S., & Deléglise, P. (2014). Improving recognition of proper nouns in ASR through generating and filtering phonetic transcriptions. Computer Speech & Language, 28(4), 979-996.

Sawalhaa, M., Brierley, C., Eric Atwell, James Dickins. (2017). Text analytics and transcription technology for quranic arabic. International Journal on Islamic Applications in Computer Science And Technology, 5(2),45-51.

Ramsay, A., Alsharhan, I., & Ahmed, H. (2014). Generation of a phonetic transcription for modern standard Arabic: A knowledge-based model. Computer Speech & Language, 28(4), 959-978.

Sweed, A. R. (2014). Islamway : the quran teacher Retrieved 30/04/2018, from https://ar.beta.islamway.net/collection/11899/%D8%A7%D9%84%D9%85%D8%B5%D8%AD%D9%81-%D8%A7%D9%84%D9%85%D8%B9%D9%84%D9%85

Saidane, T., Zrigui, M., Ben Ahmed, M., (2004). La transcription orthographique phonetique de la langue arabe [Graphem to phonem transcription of Arabic language]. Paper presented at the RÉCITAL, Fès, Morocco.

Abstract in Arabic

دراسة لإنشاء نظام قراءة النصوص للقرآن الكريم

نجلاء بالطيب [1]، مهنية قرتي[2]

[1،2] مخبر الإشارة والإتصالات، دائرة الإلكترونيك،

المدرسة الوطنية المتعددة التقنيات، الجزائر العاصمة، الجزائر

[1] nedjla.bettayeb@g.enp.edu.dz, [2] mhania.guerti@enp.edu.dz

الخلاصة: يحتاج بناء نظام لقراءة النصوص إلى دراسة وإعداد عدة وحدات. خلال هذا العمل نعرض مرحلتين لإنشاء جهاز التركيب الآلي للكلام لتلاوة القرآن الكريم. أولاً، سنقدم وصفا لعملية بناء قاعدة البيانات، لأهميتها، خاصة في أساليب تركيب الكلام بالتسلسل. ثانيا، سيتم عرض وحدة التحويل من الحرف الى الصوت .(GPT) تم تطوير هذا الجزء عن طريق اعداد مجموعة من قواعد قراءة اللغة العربية وقواعد التجويد. كما انه تم تنظيمها بطريقة جديدة لتقليص الوقت المستغرق للعملية. تم تقييم هذه الأجزاء المقدمة من خلال التحقق من التحويل الصحيح وشمولية قاعدة البيانات، مما أعطى نتائج مرضية.

الكلمات الجوهرية. قاعدة بيانات، التحويل من الحرف الى الصوت، تركيب الكلام، القرآن الكريم