Towards A Minimal Phonetic Set for Quran Recitation

Husni A. Al-Muhtaseb[1], Sameh A. Bellegdi[2]

[1]Information and Computer Science Department, KFUPM, Dhahran 31261, Saudi Arabia
[2]Academic Leadership Center, Ministry of Education, Saudi Arabia
[1]muhtaseb@kfupm.edu.sa, [2]bellegdi@kfupm.edu.sa

Abstract

Speech is the most important interaction mechanism between human beings. Text-to-Speech synthesis problem has been addressed by many researchers in the literature for different languages. However, the Arabic language did not receive that much attention. This paper addresses a computational linguistic aspect in a phonetically transcribed syllabified Quranic text that is essential for developing speech synthesis prototype. The main objective of this work is to find a set of Quran verses (Ayat) that has the complete set of distinct syllables. An algorithm to find a reduced set for Quran verses that contains all Quran syllables is proposed. One of the motivations for this work is compressing the sound files of the Quran recitation. The current work proposes a technique to extract a reduced phonetic set of Quran recitation that can be used to develop Text-to-Speech system. It is found that out of 211,573 syllables – which Quran consists of – there are 2,642 distinct syllables that represents less than 1.25% of the Quranic syllables set. In addition, a reduced set of Quran verses that contains the whole set of distinct syllables is identified. The extracted set of verses represents around 16% of Quran verses.

***Keywords*:** Quranic text; phonetic set; Quran recitation; reduced Quranic phonetic set; text statistics; Quran recitation synthesis

1. Introduction

Text-to-Speech synthesis (TTS) research field received a lot of attention from the research community. However, only limited research work addressed the problem of Arabic speech synthesis. One of the main motivations for this work is to compress the size of the sound files of the Quran recitation. TTS has mainly two phases to transform text to speech, viz. text analysis and speech signal generation. Figure 1 shows four modules in TTS system: text analysis, phonetic analysis, prosodic analysis, and speech synthesis. Text analysis includes preprocessing steps to replace numbers and abbreviations by their corresponding words or phrases. Phonetic analysis – phonetic transcription or grapheme-to-phoneme conversion – transforms text to phonemes. Prosodic analysis module is responsible for attaching stress and intonation features. Speech synthesis module is responsible for generating speech (Dutoit, 1997). The work presented in this paper deals with the phonetic analysis phase.

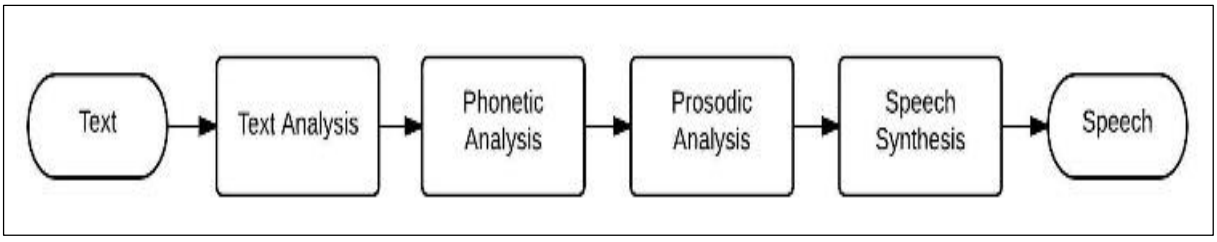

Figure 1. TTS Modules

Speech synthesis relies heavily on language and requires broad linguistic knowledge of the language (Dutoit, 1997). Computational linguistics were very much involved in speech synthesis systems. Concatenative TTS requires that basic units must contain all units used in the language to be synthesized. Statistical studies of a language can significantly reduce the amount of text needed to be recorded and obtain high quality synthesis result (Stan & Giurgiu, 2010).

All Muslims around the Globe including Arabs and non-Arabs read, listen, and respect the Holy Quran. Quranic text is a form of classical Arabic text. In fact, the Holy Quran represents a reference for the standard Arabic language (Harrag & Mohamadi, 2010). Quranic text is unique all over the world. It is fully vocalized (proper diacritics are placed on each letter of the text). The accuracy of Quranic text is very critical as it is a basis for correct recitation. Quranic text has many symbols and special characters to provide some hints about correct recitation.
In this paper, a reduced set of Quran verses (Ayat) that has the complete set of distinct syllables is identified. Syllables have been used by different researchers as speech production basic units (Cholin, Schiller, & Levelt, 2004). Similarly, this work uses syllables as basic units. An algorithm to find a reduced set for Quran verses that contains all Quran syllables is proposed.
The rest of this paper is organized as follows. Section 2 presents related work. The used phonetic transcription and syllabification technique is described in section 3. The obtained phonetic statistics are described in section 4. Conclusion and future work are presented in section 5.

2. RELATED WORK

There are limited research papers which reported the statistical analysis for non-Arabic languages (Samlowski, Möbius, & Wagner, 2011) (Stan & Giurgiu, 2010) (Stanescu, Buzo, Cucu, & Burileanu, 2012). Samlowski et al. compared syllable frequencies in different German language corpora (Samlowski, Möbius, & Wagner, 2011). They showed that syllable frequencies from written corpora can be used mostly to estimate their frequency in spoken language. Stan and Giurgiu presented text statistics for different linguistic units, as phonemes, syllables, and words (Stan & Giurgiu, 2010). They used News- Romanian (News-RO) of 4,500 newspaper articles as a text corpus for their study. Maximal Onset Principle was used for syllabification (Cox, n.d.). Stănescu et al. conducted statistical phonetic analysis based on a Romanian text corpus (Stanescu, Buzo, Cucu, & Burileanu, 2012). They addressed different types of units, as diphones and triphones. The authors reported that less than 50% of the diphones cover 99% of the text.

Al-Muhtaseb et al. presented minimal Arabic text that covers all of the Arabic alphabets in all positions (viz. standalone, initial, medial, and terminal) (Al-Muhtaseb, Mahmoud, & Qahwaji, 2009). In addition, they presented statistical analysis of Arabic corpora for estimating the number of occurrences of various letters shapes in large text corpora.

Many studies proposed different techniques and tools for the Holy Quran recitation (Ibrahim, Idris, Razak, & Rahman, 2013) (Abdou, et al., 2006) (Abdo, Kandil, El-Bialy, & Fawzy, 2010) (Ahmed, 2004). Abdou et al. developed a tool for teaching of the correct recitation of the holy Quran called HAFSS© (Abdou, et al., 2006). An automatic generation of pronunciation hypotheses was built as part of his system. In addition, a phoneme duration classification algorithm to detect recitation mistakes was proposed. Some researchers proposed different techniques to detect common pronunciation errors in Quran recitation (Ibrahim, Idris, Razak, & Rahman, 2013) (Abdo, Kandil, El-Bialy, & Fawzy, 2010). Elhadj et al. presented an independent recognizer for allophonic sounds of the classical Arabic based on Quranic

recitation (Elhadj, Alghamdi, & Alkanhal, 2013). The authors selected speech sounds extracted from recitations of a part of the Holy Quran of ten reciters. Speech signals were segmented and annotated manually into three levels: word level, phoneme level, and allophone level. They used Hidden Markov Models (HMM) with 3-emitting states and three continuous probability distributions. Ahmed studied the rules of Quran recitation as true acoustic phenomenon (Ahmed, 2004). He analyzed the voice of Sheikh AlHosary during reciting the Quran. The author analyzed Quran voices based on the modern scientific research and the rules of Tajeweed. The author studied the effect of Nasalization with Noon (ن) and Meem (م) with respect to letters' description and origination. He used 215 verses from the Holy Quran from different Chapters to study these two letters.

3. PHONETIC TRANSCRIPTION AND SYLLABIFICATION

In this work, we used our phonetic transcription and syllabification technique for Quranic text (Bellegdi & Al-Muhtaseb, 2015). General overview of the used prototype for phonetic transcription and syllabification is shown in Figure 2. The prototype can take any preprocessed Quranic text as an input. The authors used Ayah "verse" as the basic unit, i.e. they assumed that there is no stop within any verse. The prototype was constructed from three main modules, viz. preprocessing, syllabification, and phonetic transcription. In the preprocessing phase, a set of preprocessing tasks was used to remove or replace some special Quranic characters using their Unicode value. In addition, other tasks were used as standardizing Hamza letter forms – as "أ، ئ، ؤ" – and dealing with doubling a letter mark (Shaddah).

In the syllabification phase, the given text was divided into its syllables according to a set of rules. Given 'c', 'V', and 'W' as consonant, vowel, and long vowel letters, respectively; Arabic syllable patterns can have these forms: cV, cW, cVc, cWc, and cVcc. More information about these patterns could be found in (Elshafei, Al-Muhtaseb, & Al-Ghamdi, 2002). In addition, cWcc is another Arabic syllable form.

In the transcription phase, a set of rules was defined according to intonation (Tajweed) rules according to Hafss way. In addition, a set of special words along with their syllabified phonetic transcription was defined. Table 1 shows the list of exception words. If a given word is in the list of special words, its corresponding syllabified phonetic transcription was brought from the set. Otherwise, the rules were used to produce the transcription and syllabification of the given word. The defined rules were mainly for Nasal tone (Ghunnah) rules, Qalqala, Raa letter "R", long vowels, and extended long vowels (Madd). More information about phonetic transcription and syllabification could be found in (Bellegdi & Al-Muhtaseb, 2015).

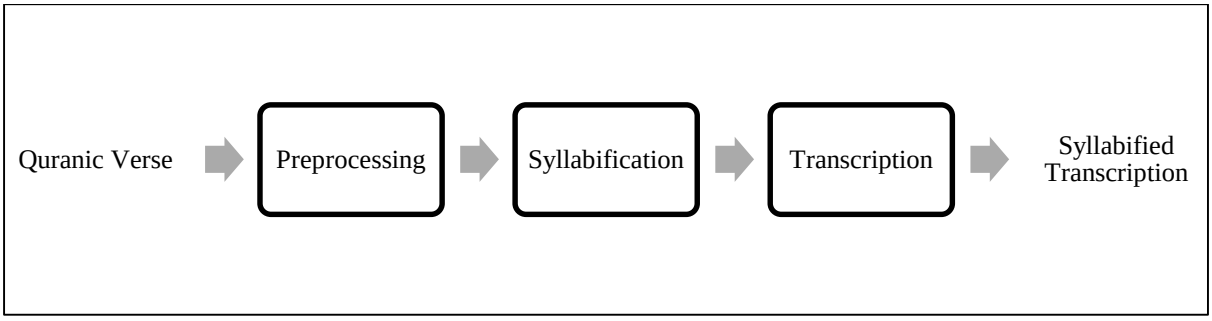

Figure 2. Proposed Prototype Architecture

4. PHONETIC STATISTICS

The main objective of this work is to find the set of Quran verses (Ayat) that has the complete set of distinct syllables. An algorithm to find a reduced set for Quran verses that contains all Quran syllables is proposed. The proposed algorithm is shown in Figure 3. In the figure, "Extracted Verses" list contains the extracted verses. If there are more than one verse containing a given unit, the verse is chosen depending on the number of syllables not already extracted. The chosen verse has the maximum number of those syllables.

1. Find the syllabified Quran phonetic transcription of a given Quranic text.
2. Calculate N-gram units using SRILM.
3. Order the units in increasing order according to their frequencies.
4. For all N-gram units (not in the “Extracted Verses”)
   - Find all verses with the given unit
   - Extract the verse that has maximum number of unprocessed units
   - Add the extracted verse to the “Extracted Verses”

Figure 3. Reduced Set of Quranic Verses Algorithm

Table 1: Special Words

الٓمٓصٓ	الٓمٓ	بِسۡمِ	ٱللَّهِ
كٓهيعٓصٓ	الٓمٓر	الٓر	مَجۡرٖىٰهَا
يٰسٓ	طٰسٓمٓ	طٰسٓ	طه
قٓ	عٓسٓقٓ	حٰمٓ	صٓ
			نٓ

To calculate N-gram statistics, SRI Language Modeling Toolkit (SRILM) was used (SRILM - The SRI Language Modeling Toolkit, 2013) (Stolcke, 2002). SRILM is an open-source toolkit for building and evaluating statistical language models. It consists of C++ class libraries that implements language models. In addition, it contains a set of executable programs built on top of these libraries to perform standard tasks such as training and testing language models. The toolkit does not perform text tokenization and treats everything between whitespaces as a word. According to (Stolcke, 2002), there is no built-in limit on the length of N-grams.

By using SRILM, it is found that, the total number of syllables in Quran is 211,573 syllables. The distinct ones are 2,642 syllables that represent less than 1.25% of the whole set. The complete transcribed Quranic text can be reproduced from this set. It is clear that any syllable can be either in one word or two words at maximum. The number of words that has all distinct syllables is at most 2×2,642 = 5,284 words by considering the worst case that every syllable is in two different words. Table 2 shows the most frequent syllables with their frequencies and examples.

Table 2: Most Frequent Syllables with Frequencies

Syllable	Word	Frequency %
وَ	هُوَ	4.10
لَ	قِيلَ	3.20
نَ	خَالِدِينَ	3.13
لَا	لَا	3.06
عَ	جَعَلَ	1.64
فَ	رَفَعَ	1.62
مَا	مَالَهُ	1.55
تَ	تَبَيَّنَ	1.50
لِ	ءَالِهَتَكُمْ	1.50
بِ	بِوَالِدَيْهِ	1.47
مَ	يَوْمَ	1.42
نَا	قُلْنَا	1.38
كَ	كَشَفْنَا	1.38
هُ	بَيْنَهُمْ	1.18
ءَ	ءَايَاتِهِ	1.17

By analyzing transcribed Quranic text in relation with the distinct syllables, it is found that out of 6,348 verses of Quran – with Basmalah which is a special verse that appears mostly at the beginning of every Quranic chapter – the obtained minimal phonetic set has only 1,001 verses. This phonetic set covers all syllables in the Quran. It represents approximately 16% of the Holy Quran. By obtaining recordings of those verses in the obtained set, the complete Holy Quran can be synthesized. This procedure will allow users to save space so that they can get complete Holy Quran in their cell phones and of many reciters. In addition, instead of recording the complete Holy Quran for a reciter, it is required to get his recitation of only the reduced set.

5. CONCLUSION AND FUTURE WORK

The current work proposes a technique to extract the reduced phonetic set of Quran recitation that can be utilized in building Text-to-Speech system for Quran recitation. A set of useful statistical results is obtained. It is found that the number of distinct syllables in the Quran is 2,642 syllables that build the complete set of Quran of 211,573 syllables. In addition, it is found that all Quranic text of 6,348 verses could be obtained from a set of Quranic verses of 1,001 verses only. This technique will allow users to save space and obtain new recitation from a reduced set of recordings. As a future work, the authors will minimize the suggested minimal set to include only words that contain acquired syllables instead of complete verses. We suggest utilizing the proposed technique in synthesizing the Holy Quran. The authors will seek the authentication and approval of specialized scholars and trusted bodies in sciences of the Holy Quran before announcing any system using this work to ensure the correctness of the synthesized recitation.

Acknowledgment

The authors would like to acknowledge KFUPM.

References

Abdo, M., Kandil, A., El-Bialy, A., & Fawzy, S. (2010). Automatic detection for some common

pronunciation mistakes applied to chosen Quran sounds. 5th Cairo International Biomedical Engineering Conference, (pp. 219–222).

Abdou, S., Hamid, S., Rashwan, M., Samir, A., Abd-elhamid, O., Shahin, M., & Nazih, W. (2006). Computer Aided Pronunciation Learning System Using Speech Recognition Techniques. INTERSPEECH, (pp. 849–852).

Ahmed, A. R. (2004). Quran Phonology: Quran reciting rules based on modern acoustics. Ain Shams University. Master's thesis, Ain Shams University.

Al-Muhtaseb, H. A., Mahmoud, S. A., & Qahwaji, R. S. (2009). A novel minimal Arabic script for preparing databases and benchmarks for Arabic text recognition research. 8th WSEAS International Conference on Signal Processing (SIP '09), (pp. 37-43). Istanbul, Turkey.

Bellegdi, S., & Al-Muhtaseb, H. (n.d.). Automatic Rule Based Phonetic Transcription and Syllabification for Quranic Text. unpublished.

Cholin, J., Schiller, N., & Levelt, W. (2004). The preparation of syllables in speech production. Journal of Memory and Language, 50(1), 47-61.

Cox, J. H. (n.d.). PHONETICS AND PHONOLOGY, The Syllable and Phonotactic Constraints. Retrieved from http://clas.mq.edu.au/speech/phonetics/phonology/syllable/syll_phonotactic.html

Dutoit, T. (1997). An introduction to text-to-speech synthesis. Dordrecht: Kluwer Academic Publishers.

Elhadj, Y., Alghamdi, M., & Alkanhal, M. (2013). Approach for Recognizing Allophonic Sounds of the Classical Arabic Based on Quran Recitations. In Theory and Practice of Natural Computing (pp. 57-67). Berlin Heidelberg: Springer.

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

Harrag, A., & Mohamadi, T. (2010). QSDAS: new Quranic speech database for Arabic speaker recognition. Arabian Journal for Science and Engineering, 35(2), 7-19.

Ibrahim, N., Idris, M., Razak, Z., & Rahman, N. A. (2013). Automated tajweed checking rules engine for Quranic learning. Multicultural Education & Technology Journal, 7(4), 275–287.

Samlowski, B., Möbius, B., & Wagner, P. (2011). Comparing syllable frequencies in corpora of written and spoken language. Interspeech, (pp. 637-640).

SRILM - The SRI Language Modeling Toolkit. (2013, Nov). Retrieved 6 24, 2014, from Speech Technology and Research (STAR) Laboratory: http://www.speech.sri.com/projects/srilm/

Stan, A., & Giurgiu, M. (2010). Romanian language statistics and resources for text-to-speech systems. 9th International Symposium on Electronics and Telecommunications (ISETC), (pp. 381-384).

Stanescu, M., Buzo, A., Cucu, H., & Burileanu, C. (2012). Statistical phonetic analysis of the Romanian language for speech recognition and synthesis tasks. ELMAR, (pp. 219-222).

Stolcke, A. (2002). SRILM-an extensible language modeling toolkit. INTERSPEECH.