



Generation of a phonetic transcription for modern standard Arabic: A knowledge-based model[☆]

Allan Ramsay^a, Iman Alsharhan^{a,*}, Hanady Ahmed^b

^a School of Computer Science, University of Manchester, United Kingdom

^b Arabic Department, Qatar University, Qatar

Received 1 February 2013; received in revised form 29 January 2014; accepted 27 February 2014

Available online 15 March 2014

Abstract

This paper outlines a comprehensive system for automatically generating a phonetic transcription of a given Arabic text which closely matches the pronunciation of the speakers. The presented system is based on a set of (language-dependent) pronunciation rules that works on converting fully diacriticised Arabic text into the actual sounds, along with a lexicon for exceptional words. This is a two-phase process: one-to-one grapheme to phoneme conversion and then phoneme-to-allophone conversion using a set of “phonological rules”. Phonological rules operate on the phonemes and convert them to the actual sounds considering the neighbouring phones or the containing syllable or word. This system is developed for the purpose of delivering a robust Automatic Arabic Speech Recognition (AASR) system which is able to handle speech variation resulting from the mismatch between the text and the pronunciation. We anticipate that it could also be used for producing natural sounding speech from an Arabic text-to-speech (ATTS) system as well, but we have not extensively tested it in this application.

© 2014 Elsevier Ltd. All rights reserved.

Keywords: Phonological rules; Phonetic transcription; Speech processing; Modern standard Arabic; Sound-spelling correspondences

1. Introduction

Arabic is the most common Semitic language in terms of the number of native speakers (Lewis, 2009). There are two main forms of Arabic used at the present time: Modern Standard Arabic (MSA) is used in formal writing and speech (e.g. education, newspapers, broadcast news, etc.) and is considered to be the official language in all Arabic speaking countries. In contrast, people generally speak in their own dialects in daily communication. These dialects are neither taught in schools nor even have any organised written form. Arabic dialects differ substantially from MSA in terms of phonology, morphology, vocabulary, and syntax.

With the increasing role of the computer in today’s society, there is an ever growing desire to develop robust ASR systems to communicate with computers in a more natural way. In ASR systems, the lexicon is the main source for finding the pronunciation of the words. Providing a comprehensive pronunciation lexicon for Arabic requires a huge

[☆] This paper has been recommended for acceptance by Saraclar Murat.

* Corresponding author. Tel.: +44 7544251750.

E-mail addresses: ramsay@cs.man.ac.uk (A. Ramsay), alsharhan@cs.man.ac.uk (I. Alsharhan), hanady@qu.edu.qa (H. Ahmed).

amount of memory and lots of effort to cover the enormous number of word forms generated due to the richness of Arabic morphology. Furthermore, the “correct” pronunciation of the word which can be found in the lexicon is not always the real pronunciation, especially in continuous speech and with languages. Deriving a phonetic transcription from a given Arabic text is made worse by the problem of the boundary effects on pronunciation and the influence of the regional dialect when speaking in MSA.¹ This variation may result in a mismatch between the speech and the text used in modelling the ASR system. Mismatch between the text and the pronunciation is one of the main issues encountered by ASR researchers, as they confirm that the pronunciation variation has the greatest influence in the system’s performance when compared with the other factors (McAllaster et al., 1998; Fosler-Lussier et al., 1999). Therefore, modelling pronunciation variation during recognition in order to predict the phonetic realisation of the input speech is essential to enhance recognition performance (Slobada and Waibel, 1996; Saraclar et al., 2000).

The current study aims to develop a comprehensive system for grapheme-to-allophone conversion of MSA. The motivation behind developing this system is to map the written text onto phonetic content that can be used in designing the acoustic model of the Arabic ASR system and therefore to enhance the quality of the recogniser and to reduce the Out-Of-Vocabulary (OOV) words.

This paper is divided into four sections. Section 1 gives a brief overview of the Arabic phonetic system, followed by laying out the theoretical dimensions of the pronunciation modelling approaches and reviewing the related works in the literature. Section 2 describes the design of the developed grapheme-to-allophone system. The developed system is assessed in Section 3 of this paper. Finally, the last section reflects on what has been discussed throughout the paper and gives suggestions for future work.

1.1. Arabic orthography and pronunciation

The Arabic alphabet consists of 29 letters, 26 of which are consonants. The remaining three letters represent the long vowels (the phonemes /i:/, /a:/, /u:/). Arabic is written in script from right to left. Each letter in the alphabet can appear in up to four different shapes, depending on its location in the word. For example the letter ‘h’ *هـ* can appear in the following shapes (*هـ* *هـ* *هـ* *هـ*).

The Arabic writing system has available nine optional orthographic symbols (diacritics) that normally appear above or between the letters (three short vowels *َ* *ِ* *ُ*, dagger Alif *ا*, three nunation diacritics *ً* *ٌ* *ٍ*, gemination diacritic *ّ*, and sukun *◌*). These diacritics are extremely helpful for the readability and understanding of the text. Diacritics, however, are mainly restricted to religious texts and language learning textbooks, and are absent in most other texts.

The correspondence between the spellings and the pronunciation of the lexical items, when all the diacritics are included, is deterministic, unlike for example English or French. However, transcribing Arabic texts does not just involve converting the graphemes into phonemes but also converting them to phones that represent the actual sound of the language. This task is done partly by analysing the boundary effects on pronunciation. The issue here is that the local context can have considerable impact on what a phoneme sounds like. This can have substantial consequences for speech recognisers, since it means that the phonetic transcription which is required for any speech recogniser needs to be sensitive to the local context.

Table 1 illustrates how the acoustic realisation of the word varies depending on its position in the context. The table provides the result of generating the phonetic transcription of the preposition ‘*min*’² *من* “from” when occurring in different phonetic context. This process is called nasal assimilation and will be discussed fully later.

1.2. Pronunciation modelling approaches

Modelling the pronunciation variation is a crucial step in any ASR system. Such models can help to improve the performance by shrinking the mismatch between the speech and the text used in training the acoustic model. There are two main methods used in the literature in modelling the pronunciation variation (Amdal and Fosler-Lussier, 2003; Wester and Fosler-Lussier, 2000):

¹ We will refer below to the systematic variations in pronunciation that arise from the influence of regional dialects as ‘accents’.

² Throughout the paper we used Buckwalter transliteration scheme, italicised and single-quoted, in the examples to present Arabic script, and IPA when transcribing actual speech sounds (between two slashes for the phonemic representation or two square brackets for the phonetic representation). We use the ASCII-friendly SAMPA presentation of the IPA in rules

Download English Version:

<https://daneshyari.com/en/article/10368493>

Download Persian Version:

<https://daneshyari.com/article/10368493>

[Daneshyari.com](https://daneshyari.com)