INTRODUCTION

Automatic Speech Recognition (ASR) is a key technology for a variety of industrial and IT applications; and it extends the reach of IT across people as well as applications. Automatic Speech Recognition (ASR) is gaining a growing role in a variety of applications, such as hands-free operation and control (as in cars and airplanes), automatic query answering, telephone communication with information systems, automatic dictation (speech-to-text transcription), government information systems, etc. In fact, speech communication with computers, PCs, and household appliances is envisioned to be the dominant human-machine interface in the near future. In spite of the tangible success of this technology for the English language and other languages, there are still many issues for Arabic language that need to be addressed by

DOI: 10.4018/jitr.2009062905
researchers to catch up with the progress of the ASR technology in the other languages.

One of the key components of the modern large-vocabulary speech recognition systems is the pronunciation or phonetic dictionary. This dictionary serves as an intermediary between the Acoustic Model and the Language Model in speech recognition systems. It contains a subset of the words available in the language and the pronunciation of each word in terms of the phonemes or the allophones available in the acoustic model.

For instance, the CMU dictionary for North American English contains over 125,000 words and their transcriptions (CMU, 2008). The format of this dictionary is particularly useful for speech recognition and synthesis, as it has mappings from words to their pronunciations in the given phoneme set. The current phoneme set contains 39 English phonemes, for which the vowels may also carry lexical stress. Because of the large number of pronunciation exceptions in English, this dictionary was essentially built manually by experts over many years.

On the other hand, pronunciation of Arabic text follows specific rules when the text is fully diacritized. Many of these pronunciation rules can be found in Elshafei (1991), and Alghamdi et al. (2004).

The statistical approach for speech recognition (Huang et al., 2001; Jelinek, 1998; Rabiner & Juang, 1993) has virtually dominated Automatic Speech Recognition (ASR) research over the last few decades, leading to a number of successes (Lee, 1988; Soltan et al., 2007; Stallard et al., 2008; Young, 1997; Zhou et al., 2003). The statistical approach is dominated by the powerful statistical technique called Hidden Markov Model (HMM) (Rabiner 1989). The HMM-based ASR technique allowed to build many successful applications that depend on large vocabulary speaker-independent continuous speech recognition.

The HMM-based technique essentially consists of recognizing speech by estimating the likelihood of each phoneme at contiguous, small frames of the speech signal (Huang et al., 2001; Rabiner & Juang, 1993). Words in the target vocabulary are modeled into a sequence of phonemes, and then a search procedure is used to find, amongst the words in the vocabulary list, the phoneme sequence that best matches the sequence of phonemes of the spoken word.

Two notable successes in the academic community in developing high performance large vocabulary speaker independent speech recognition systems are the HMM tools, known as the HTK tool kit, developed at Cambridge University, (HTK, 2007), and the Sphinx system developed at Carnegie Mellon University (Huang et al., 1993; Lamere et al., 2003; Noamany et al., 2007; Placeway et al., 1997).

Development of an Arabic speech recognition is a multi-discipline effort, which requires integration of Arabic phonetic (Alghamdi, 2000; Alghamdi et al., 2004), Arabic speech processing techniques (Elshafei, 1991; Elshafei et al., 2002), and Natural languages processing (Elshafei et al., 2006).

Development of Arabic speech recognition systems has recently been addressed by a number of researchers (Elshafei. et al., 2008; Hiyassat, 2007; Noamany et al., 2007; Soltan et al., 2007). Saroti et al. (2007) used Sphinx tools for Arabic speech recognition. They demonstrated the use of the tools for recognition of isolated Arabic digits. The data was recorded from 6 speakers. They achieved digits recognition accuracy of 86.66%. Hiyassat (2007), in his Ph.D. thesis, developed a tool to generate Arabic pronunciation dictionaries. The generated dictionaries are based on a small MSA speech corpus consisting of digits or command and control vocabulary.

A workshop was held in 2002 at John Hopkins University (Kirchhoff et al., 2003) to define and address the challenges in developing a speech recognition system using Egyptian dialectic Arabic for telephone conversations. They proposed to use Romanization method for transcription of the speech corpus.

Billa et al. (2002) addressed the problems of indexing of Arabic news broadcast, and discussed a number of research issues for Arabic speech recognition.
Implementation Failure of an Integrated Software Package: A Case Study from the Far East
Suprateek Sarker and Saonee Sarker (2000). Organizational Achievement and Failure in Information Technology Management (pp. 249-262).
www.igi-global.com/chapter/implementation-failure-integrated-software-package/27864?camid=4v1a

DataNaut Incorporated
www.igi-global.com/article/datanaut-incorporated/44494?camid=4v1a