# Research Efforts on Arabic TTS in the State of Qatar

Mada Center

Text-to-Speech (TTS) technology uses automated speech synthesis to produce speech. TTS technology employs a speech synthesizer that converts symbolic linguistic representations into sound in conjunction with another solution (usually software) that parses raw text input and assigns its phonetic transcriptions by marking and dividing the text input into relevant words, sentences, and punctuations.

Over the past two decades, TTS has become a key area of interest due to its potential usage across various application areas like assistive technology and educational software consisting of multimedia output and relevant Interactive solutions. Likewise, the quality of TTS has significantly improved over time by sounding more like natural human voices. Following are the different aspects that measure the quality of a TTS output:

- **Naturalness:**the degree of the speech generated to be as close to a human-sounding speech in terms of its timing structure, pronunciation, and rendering of emotions.
- **Intelligibility:** the quality of the audio generated, or the degree of each word being produced in a sentence**.**
- **Preference**: a better liking by end-users of a particular TTS over other available alternatives; preference and naturalness are influenced by TTS system, signal quality, and voice, in isolation and in combination.
- **Comprehensibility**: the degree of the speech output being interpretable

Many advances have been made in Text-to-Speech Engine (TTS) over the past decade. TTS has played a major role in developing technologies for the blind and visually impaired, as it allows to read text from a screen display. Most research on TTS has been done in languages such as English and French, while many other languages, such as Arabic, have not been substantially worked on until the recent decade. The field of Arabic Text-to-Speech can still be considered to be in its early stages of development compared to other Latin languages.

**Text-to-Speech Synthesizer Components**

TTS Synthesizer comprises of two major components which are the Natural Language Processing (NLP) Engine and the Digital Signal Processing (DSP).

The natural language derived from the interaction between computers and humans is called **Natural Language Processing** (NLP) which is a branch of Artificial Intelligence. (NLP) reads, decipher, and interprets human languages which are commonly achieved through machine learning. There are four major components of (NLP) namely; Text Processing Module, Text Analyzer, Pronunciation Module, and a Prosody Generator.

**Figure 1: Major components of NLP**

The **Digital Signal Processing (DSP)** is the component of the TTS synthesizer that converts the list of phonetic transcription and their prosodic information into digital audio through mathematical models, algorithms, and computational methods to deliver a natural-sounding speech. The algorithm for generating the digital audio will vary based on the requirements, complexity, and technology used. DSP ultimately transforms symbolic information processed from NLP into speech.

Arabic Text-to-Speech Challenges

In addition to common challenges faced in the process of developing TTS solutions, Arabic TTS development poses additional significant challenges which are as follows:

· Diacritization

Arabic is a diacritized language with a complex diacritization system. Written Arabic text often omits the detailed diacritic properties of characters leading to the unavailability of key information about its accurate pronunciation to be performed by the TTS. The absence of diacritization is a source of confusion for computational systems that adds ambiguity to both text analysis and sound generation. Each character in an Arabic word must be assigned with diacritics that give the information about its accurate pronunciation. Additionally, the correct pronunciation of a word is not often obvious from its spelling and there will exist many words with multiple pronunciations based on the linguistic context.

· Dialects

Arabic is spoken in more than 23 countries by more than 300 million people worldwide. The large geo-demographic spread of Arabic speakers means that the language is

spoken by a socio-culturally diverse range of population with various dialects. The varieties in dialects impose a problem for speech synthesis as it would have to vary the speech output based on the pronunciations of the concerned dialect. Every dialect will have a relatively limited number of users from specific regions where the dialect is practiced. In addition to dialects, Arabic TTS systems may generate speech output in Modern Standard Arabic (MSA). However, MSA is understood primarily by individuals with relatively higher levels of literacy limiting the number of users who would use TTS supporting MSA language. Thus, the development of Arabic TTS will involve the creation of TTS systems that support multiple dialects and MSA all of which have a limited customer base on their own.

## Qatari Research Efforts

### ArabicProsody TTS – Intonation and stress generator for Arabic text-to-speech

A team from Qatar University contributed towards developing the **ArabicProsody TTS** – **Intonation and stress generator for Arabic text-to-speech system.** This research project involved the use of fine-grained linguistic analysis to help produce spoken output which was both intelligible and natural sounding from didacticized Arabic text.

A team from Qatar University developed a Natural Language Process (NLP) engine- which included a prosody generator- for converting Arabic text into a phonetically transcribed and prosodically labeled text. The prosodic generator is the final module of the NLP engine that derives the pitch information automatically. The (Multi-Band Resynthesis OverLap Add) MBROLA system was used which is a diphone-base synthesizer to produce the signal waves.

This TTS system is designed for Modern Standard Arabic (MSA). This system generates speech from unrestricted Arabic text by carrying on the following tasks:

- Assigning diacritics to the written text.
- Automatic phonetic transcription.
- Syntactic analysis in order to assign a global intonation contour.
- Calculating and producing the local pitch contours for sentences.
- Connecting the NLP engine with MBROLA synthesizer

## References

Rizk, Y. Mohanna. “Arabic Text to Speech Synthesizer: Arabic Letter to Sound Rules”. In International Review on Computers and Software (I.RE.CO.S.), January 2011

H. Mansour. “ArabicProsody TTS – Intonation and stress generator for Arabic text-to-speech” In Intonational Variation in Arabic (IVA09), UK, September 2009

 