

Overview of Arabic TTS and related Technologies

Mada Center

Nowadays, electronic devices are more able to output information in the form of sounds, which makes them more able to interact with humans. As Talking devices gain a prominent place in our daily lives and especially for persons with disabilities, it is increasingly important that these technologies have a speech capability similar to humans. The perceptual quality of the outcome of Text to Speech (TTS) software affects how well a person accepts these systems. For this reason, researchers on the domain of TTS are striving to make artificial speech more natural. In the market there exist several number of TTS systems that varies on quality and technology.

As an assistive technology, text-to-speech (TTS) software is designed to support people who have difficulties reading written text. Common reading disabilities can include blindness, dyslexia or any visual impairment, learning disability or other physical condition that impedes the ability to read. However, other persons can benefit from TTS technology, such as autistic children, attention deficit hyperactivity disorder (ADHD) or persons with intellectual disability.

“Festival” is one of the most common TTS frameworks that uses Hidden Markov Model based speech synthesis technology. It offers several tools and resources for making text to speech synthesis software. The framework includes examples of various modules. It allows making full text to speech applications through a set APIs: Scheme command interpreter, C++ library, Java Packages, and an Emacs interface. “Festival” is multi-lingual, and currently it supports English and Spanish. Other groups release new languages for the system. Many Open-source Arabic TTS are created using “Festival” and are available on “GitHub” for free.

“Sakhr TTS” is the industry leader in synthesizing a natural, human-sounding Arabic voice. Sakhr provides software for Arabic Text-to-Speech (TTS) and Automatic Speech Recognition (ASR). The TTS converts Arabic text into a natural synthetic voice. In developing its software, Sakhr is leveraging 28 years of research and development in Natural Arabic Language Processing (NLP). This research is considered critical to

overcome the Arabic text-to-speech challenges, such as the lack of accents and punctuation marks.

There are additional commercial engines such as “Amazon Polly”, “Google Tacotron”, and “IBM Watson Text to Speech”. Amazon Polly is a service that synthesizes speech from text, allowing developers to create talking applications, and build entirely new categories of speech-enabled software and products. Polly’s TTS engine uses advanced deep learning technologies to synthesize natural human-sounding speech. With dozens of lifelike voices across a broad set of languages. The standard voice of Polly supports Arabic language. However, Arabic is not yet included on the new Neural Text-to-Speech (NTTS) voices, which provides advanced improvements in speech quality through a new machine learning approach.

One of the top-rated TTS software is “Acapela”. Acapela is a High-Quality Text to Speech software. Acapela provides a large set of voices that cover 30 different languages including Arabic. Selected few voices can be purchased in special emotive versions that include multiple variations for different moods or perspectives. It also offers children’s voices. Acapella provides a large development kits for software developer covering mobile applications as well as desktop applications and cloud services.