

## **Interview with Gabi Ilan, Command Speech**

### ***Speech recognition and speech synthesis done on the device***

*Gabriel (Gabi) Ilan, CEO, **Command Speech Ltd.**, was interviewed by Bill Meisel in mid-December. Gabi is a co-founder of Command Speech, an Israeli company formed in 2008. In 1990, he co-founded ART - Advanced Recognition Technologies, a pioneer in software algorithms for voice and handwriting recognition for the cellular and PDA markets, and served as its president for 11 years. ART was acquired by ScanSoft (now **Nuance**) in 2004. Gabi has degrees of M.Sc. and B.Sc., both in Electrical Engineering, from the Technion.*

#### **Please briefly describe the focus of Command Speech products.**

Command Speech speech recognition and text-to-speech products are aimed at voice command applications in mobile devices. Our main goal is to keep high recognition rate in noisy and reverberating places such as in vehicles, while maintaining compact code size and data storage. Recognition is done in realtime locally on the device, with no need to connect to a remote server, guaranteeing reliability and absolute privacy in all situations.

#### **What types of noise can Command Speech handle in its speech recognition?**

Our speech recognition works well in moving cars and in crowded and noisy and crowded places such as shopping centers, train stations, and conferences. The software was designed from the ground up for this environment.

#### **Is the noise tolerance part of the speech recognition processing, or is there a separate noise cancellation algorithm? Are multiple microphones required?**

The noise tolerance is integral with the recognition process. A single, low cost microphone as in most handsets is sufficient.

#### **How complex can the speech recognition task be and yet maintain accuracy?**

State-of-the-art speech recognition algorithms such as ours are extremely complex. However, if these algorithms are carefully designed, processors currently used in smartphones and in other mobile devices are powerful enough to carry out complex DSP algorithms in real-time without compromising accuracy. Our algorithms include many innovations that accelerate the recognition process, and all calculations are done in fixed point, to eliminate the need for a mathematical co-processor. Pipelining is used to push the response time down even further.

#### **How is the speech “grammar” specified?**

Grammar may be defined by the application. Command Speech provides a high performance basic recognition engine recognizing a single word or phrase out of a large pre-defined vocabulary. The vocabulary may include thousands of entries, such as all user contacts, all streets names in a city, or all the cities in a state, etc.

#### **What is different about your text-to-speech synthesis product?**

Command Speech's text-to-speech software module, originally developed to give the user a vocal feedback of the phrase recognized, is purely statistical, unlike the conventional text-to-speech design based on a “unit selection” approach. This statistical design requires significantly less data while lacking unpleasant audible discontinuities common in unit selection systems having equal footprint size. Being small and performing fast are two big challenges for software that normally requires lots of storage space and floating-point processing, and our software succeeds in overcoming both of these challenges.

#### **Any final comments?**

Command Speech's innovative R&D results in one of the most noise-robust, quickest, smallest, accurate and self-contained speech-recognition software engines for mobile devices and one of the most pleasant sounding text-to-speech modules currently available in the market.