Ultra-Low-Power Command Recognition for Ubiquitous Devices

Chris Rowen, CEO, BabbleLabs

Rapid advanced in deep learning over the past 5 years have permanently changed expectations for speech as a fundamental interface for many systems. The high compute demands of the early systems, have pushed most of the processing into the cloud, where cost, latency, energy, robustness and privacy are all compromised in favor of development flexibility. Speech, however, cannot become a universal interface for everyday devices until cost is measured in the pennies and power in a few milli-watts. Moreover, most of today's speech systems are tuned for the ideal environment of quiet offices and living rooms. Real-world devices must sustain high speech response accuracy even under the chaotic real-world conditions of traffic, crowds, wind, and interfering music.

In this talk I decipher the current speech recognition spectrum from key-word detection to full natural language dictation, and highlight the emerging role for noise-robust rich command recognition. I present a new training regimen, network structure and optimized 8b integer implementation that recognizes large suites of device commands, while running on small microcontrollers, and consuming single-digit mW. Moreover, this new system achieves effective accuracy almost one order of magnitude better than general-purpose cloud-based speech recognizers for the target command sets under high-noise conditions.