

Chapter XVII

Mobile Speech Recognition

Dirk Schnelle

Technische Universität Darmstadt, Germany

ABSTRACT

This chapter gives an overview of the main architectures for enabling speech recognition on embedded devices. Starting with a short overview of speech recognition, an overview of the main challenges for the use on embedded devices is given. Each of the architectures has its own characteristic problems and features. This chapter gives a solid basis for the selection of an architecture that is most appropriate for the current business case in enterprise applications.

OVERVIEW

Voice-based interaction is a common requirement for ubiquitous computing (UC). However, the idea of having speech recognition on wearable devices is not simply copying the recognizer to such a device and running it. The limitations of the device, especially computational power and memory, pose strong limitations that cannot be handled by desktop size speech recognizers. This chapter gives a brief overview of the different architectures employed to support speech recognition on wearable devices. A background in speech recognition technology is helpful in order to understand them better, but is not required. At some points you will be provided with pointers to the literature to achieve a better understand-

ing. A detailed understanding of the available architectures is needed to select the appropriate architecture for the enterprise, if it wants to support audio-based applications for mobile workers. The selection process has to consider the available resources, such as servers, wireless network, the software that has already been bought in order to save the investment, and to be able to justify the decision to invest more money in required infrastructure.

Most of the figures use UML 2.0 as a means of communicating architectural descriptions. The diagrams are easy to read, even if the reader is not familiar with this modeling language. The UML specification can be obtained from the Object Management Group (OMG, 2006).

Figure 1. Speech recognition

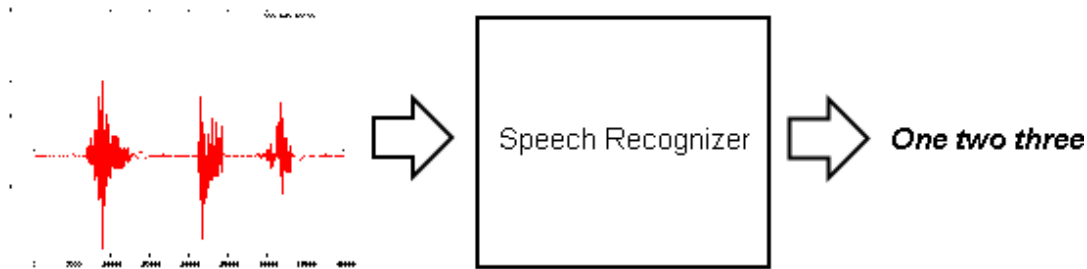
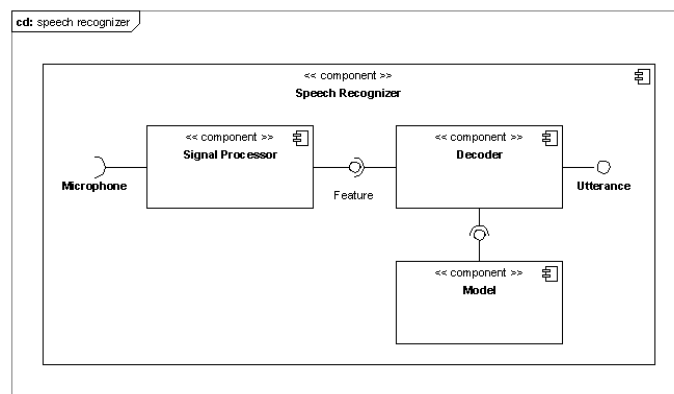


Figure 2. General architecture of a speech recognizer



A speech recognizer has the task of transcribing spoken language into a text (see Figure 1). The input is the *speech signal*, the human voice that is recorded, for example, with a microphone. The textual output, in this case “one two three,” is called an *utterance*.

The architecture of a speech recognizer has not changed over the past decades. It is illustrated in Figure 2 based on Jelinek (2001).

It comprises the main components of recognizers as they are used today, regardless of the technology used. They are available as pure software solutions or implemented in hardware

to gain speed. In the following sections we focus only on the main components involved. Some recognizers may use additional components or components that are slightly different. However, the architectures presented show the main functionalities of each of them and discuss the main challenges that have to be faced when applied to mobile devices.

The *signal processor* generates real valued vectors σ_i from a speech signal, obtained from a microphone. They are also called feature vectors. Currently, most speech recognizers use at least 13 features in each vector. We will have a closer look

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/mobile-speech-recognition/21777

Related Content

Research on Simulation Algorithm of Series Hybrid Electric Vehicle Energy and Intelligent Control

Tao Gao (2017). *International Journal of Advanced Pervasive and Ubiquitous Computing* (pp. 33-77).
www.irma-international.org/article/research-on-simulation-algorithm-of-series-hybrid-electric-vehicle-energy-and-intelligent-control/189227

Research on Root Locus Correction Algorithm in Automatic Control System

Weifang Zhai and Juan Feng (2020). *International Journal of Security and Privacy in Pervasive Computing* (pp. 30-45).
www.irma-international.org/article/research-on-root-locus-correction-algorithm-in-automatic-control-system/259350

u-City: The Next Paradigm of Urban Development

Jong-Sung Hwang (2010). *Ubiquitous and Pervasive Computing: Concepts, Methodologies, Tools, and Applications* (pp. 1601-1612).
www.irma-international.org/chapter/city-next-paradigm-urban-development/37870

Student Learning in an Online Environment: Differences in Study Approaches

Rodney Arambewela (2010). *Strategic Pervasive Computing Applications: Emerging Trends* (pp. 224-234).
www.irma-international.org/chapter/student-learning-online-environment/41592

Scrambling Keypad for Secure Pin Entry to Defeat Shoulder Surfing and Inference Attacks

Samuel Selassie Yakohene, Winfred Yaokumah and Ernest Barfo Boadi Gyebi (2021). *International Journal of Security and Privacy in Pervasive Computing* (pp. 12-33).
www.irma-international.org/article/scrambling-keypad-for-secure-pin-entry-to-defeat-shoulder-surfing-and-inference-attacks/282085