Chapter 8.18
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 understanding. 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).

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
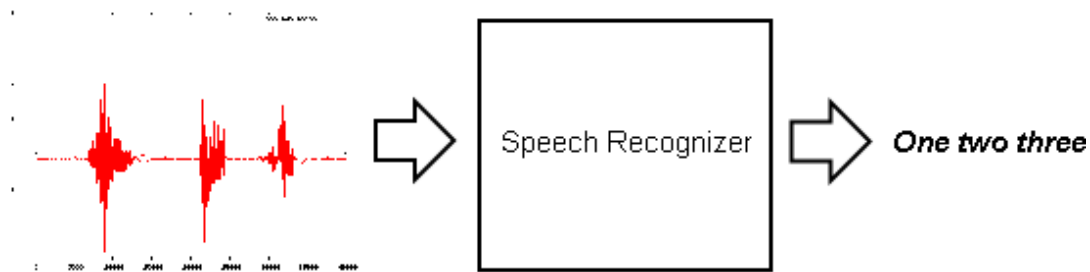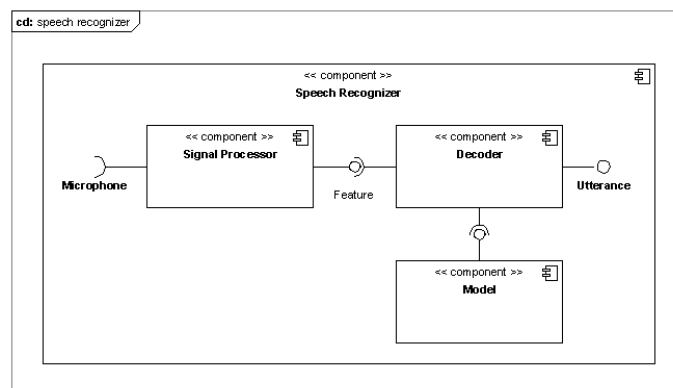
Figure 1. Speech recognition

Figure 2. General architecture of a speech recognizer

Related Content

Wireless Technologies for Mobile Computing and Commerce
David Wright (2009). *Mobile Computing: Concepts, Methodologies, Tools, and Applications (pp. 1175-1182).*
www.irma-international.org/chapter/wireless-technologies-mobile-computing-commerce/26579

Ambient Learning Conceptual Framework for Bridging Digital Divide in Higher Education
Simon Nyaga Mwendia, Peter Waiganjo Wagachaand Robert Oboko (2015). *Promoting Active Learning through the Integration of Mobile and Ubiquitous Technologies (pp. 235-264).*
www.irma-international.org/chapter/ambient-learning-conceptual-framework-for-bridging-digital-divide-in-higher-education/115479

A Local Statistical Information Active Contour Model for Image Segmentation
Shigang Liu, Yali Peng, Guoyong Qiuand Xuanwen Hao (2014). *International Journal of Mobile Computing and Multimedia Communications (pp. 33-49).*
www.irma-international.org/article/a-local-statistical-information-active-contour-model-for-image-segmentation/128999

LTE-A Implementation Scenarios: RF Planning Comparison
Mohammed Jalounand Zouhair Guennoun (2012). *International Journal of Mobile Computing and Multimedia Communications (pp. 31-42).*
www.irma-international.org/article/lte-implementation-scenarios/63049

FCVW: Experiments in Groupware
Ivan Tomekand Elhadi Shakshuki (2009). *International Journal of Mobile Computing and Multimedia Communications (pp. 29-52).*
www.irma-international.org/article/fcvw-experiments-groupware/4062