Chapter LI
Speech Synthesis and Recognition Based on Mobile Computing Application

Diego Moreira Alves
Luteran University of Brasil, Brasil

Adriano Petry
Luteran University of Brasil, Brasil

Gilberto Fernandes Marchioro
Luteran University of Brasil, Brasil

ABSTRACT

The main objective of this chapter is to present a solution to the development of speech synthesis and speech recognition systems through mobile computing application. A system model that captures the audio data from a mobile device for further recognition is presented. This way, the mobile device is capable of receiving a word in plain text format and synthesizing it. These two processes are executed over a voice recognition and synthesis server, such as a typical client-server model. A simple mobile application example for learning and practicing the English language is used as a test environment.

INTRODUCTION

The evolution of mobile computing is related to the growth and maturity of the new (technological) equipment provided by manufacturers of mobile devices. As the use of mobile computing gets more intense every year, manufacturers are very concerned with constantly improving their technologies and innovating their devices with more features and functionalities. At the same

time, software providers aim to develop products (to do researches) in order to satisfy and support the demands of the market.

At the end of 2007, there were more than 110 million cellular phones in Brazil, representing a financial movement of BRL 80 million/US $40 million in the Brazilian market of mobile software applications (Morais, 2005). Furthermore, by the end of 2010, 10% of all bank transactions will be realized using mobile software applications (Junior, 2007). These numbers show that the demand for new solutions based on mobile computing will get higher and higher.

A great challenge for the new solutions based on mobile computing is to simplify the user interaction (eliminating the difficulties existing today). The mobile software applications' ease of access is one of the most important aspects when developing services based on mobile technology (Talukder & Yavagal, 2006). The concern about mobile devices and their ease of use is recurrent in researches such as in Mobile Computing Principles (B´far, 2005), where the author presents different types of mobile applications that all aim to offer a high level of user-friendliness. In order to increase the accessibility in mobile software applications, several means can be used, such as resizing the displays and use of extern keyboards. However, the use of speech interfaces is still very rare in these cases.

Due to the high innovation level, synthesis and speech recognition processes have been subjects in research during the last years. Speech synthesis is the computer's ability to produce sounds that resemble human speech, whereas speech recognition refers to computer systems that can recognize spoken words. These processes are very precise and give the reader an idea of the complexity of this technology that uses techniques based on artificial intelligence (Jurafsky & Martin, 2000; Rabiner & Juang, 1993).

In this context, this work presents an application model that by means of voice interface aims to simplify and facilitate user interaction with

mobile devices. Furthermore, this chapter has the purpose of integrating technologies using speech synthesis and recognition with the development based on mobile computing.

The reader can find a list of works related to this topic in section 2 and a complete description of the developed system in section 3. An example of the application is shown in section 4, and the conclusion of the work can be found in section 5.

Related Works: A Background

The use of voice interfaces in mobile applications is a big advance when it comes to minimizing the difficulties existing today with, for instance, the use of reduced-size keyboards or reduced-size displays. However, the author points out that the capacity of the mobile devices' processing is too poor in order to realize speech synthesis and recognition with an accepted quality (B´far, 2005).

The growth of applications is based on a client-server model in mobile network communication services, where a wide range of services available on the corporate market is presented (Fiorese, 2005).

A model for voice services in virtual environments is based on Java Speech API (JSAPI) technology. The model is described very technically with codes and a lot of practical examples. The research presents the use of engine synthesis and recognition developed by IBM, called IBM ViaVoice (Damasceno, Pereira & Brega, 2005).

Another study focuses on examples of the use of Java Speech Markup Language (JSML); in other words, the language for definition of labels markings in the development of voice applications used by JSAPI. Using the IBM ViaVoice as the engine, the author shows, as a result, an application capable of reproducing music from voice commands. Furthermore, the writer describes the grammatical rules and dictations used by JSAPI (Rodrigues & Moreira, 2001).

Related Content

Tissue Image Classification Using Multi-Fractal Spectra
Ramakrishnan Mukundanand Anna Hemsley (2010). *International Journal of Multimedia Data Engineering and Management (pp. 62-75).*
www.irma-international.org/article/tissue-image-classification-using-multi/43748

Content-Based Multimedia Retrieval Using Feature Correlation Clustering and Fusion
Hsin-Yu Ha, Fausto C. Fleitesand Shu-Ching Chen (2013). *International Journal of Multimedia Data Engineering and Management (pp. 46-64).*
www.irma-international.org/article/content-based-multimedia-retrieval-using-feature-correlation-clustering-and-fusion/84024

The Axis of Good and Evil
Jonathan Melenson (2011). *Designing Games for Ethics: Models, Techniques and Frameworks (pp. 57-71).*
www.irma-international.org/chapter/axis-good-evil/50731

Information Hiding, Digital Watermarking and Steganography
Kuanchin Chen (2005). *Encyclopedia of Multimedia Technology and Networking (pp. 382-389).*
www.irma-international.org/chapter/information-hiding-digital-watermarking-steganography/17273

Container Orchestration With Cost-Efficient Autoscaling in Cloud Computing Environments
Maria Rodriguezand Rajkumar Buyya (2020). *Handbook of Research on Multimedia Cyber Security (pp. 190-213).*
www.irma-international.org/chapter/container-orchestration-with-cost-efficient-autoscaling-in-cloud-computing-environments/253033