Hype or Ready for Prime Time? Speech Recognition on Mobile Handheld Devices (MASR)

Dongsong Zhang, Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, USA

Hsien-Ming Chou, Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, USA

Lina Zhou, Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, USA

ABSTRACT

The pervasiveness of mobile handheld devices and advancement in real-time continuous speech recognition technology has opened up a wide range of research opportunities in human-computer interaction for those devices. On the one hand, there has been an increasing amount of research on developing user-friendly speech recognition solutions and applications for mobile handheld devices. On the other hand, there are many distinct challenges in mobile speech recognition. Aiming to gain a good understanding of this emerging yet challenging area and provide a research map, this paper presents a state-of-the-art overview of this field. We will discuss three main architectures of mobile speech recognition systems, analyze their strengths and weaknesses, introduce some major research issues in the field, and highlight a number of major applications of speech recognition on handheld devices. The authors will also shed some light into important future research issues as a road map for researchers and practitioners.

Keywords: Acoustic Model, Mobile Handheld Devices, Multimodal Interface, Speech Recognition, Voice Input

INTRODUCTION

The proliferation of mobile handheld devices (e.g., cell phones and PDAs) and significant advancement of wireless technologies and infrastructures have become a strong driving force of many mobile applications, such as ubiquitous information access and mobile healthcare services. According to Morgan Stanley Research, the growth of Web access through mobile handheld devices has outpaced the growth of desktop Internet access. Today, mobile handheld devices are no longer just tools for communication and personal information management, but also for social activities, entertainment, health monitoring (Marti & Delgado, 2004), location-based services, etc. They have been permeating into daily routines of many people.

Despite the tremendous affordability, flexibility, accessibility, and portability of mobile
handheld devices, interacting with them has not always been effective and pleasant, which often suffers from a variety of significant usability problems attributable to their inherent physical constraints, such as small screen size, restricted interaction mechanisms, and low memory space. In particular, the physical or soft keypads of those devices are considerably small, clumsy to interact with, and error prone. Even with the latest touch-sensitive screens of cell phones such as Apple’s iPhones and HTC’s Android phones, the interaction through a soft keyboard remains a challenge. Users have to frequently zoom in or out in order to resize the content and to switch from one application to another, which increases users’ cognitive load, especially when users are on the move. In addition, users’ ability to interact with devices by hand could be severely hampered in situations when users have one or both hands occupied by something else, or are physically impaired with hand or vision problems. As a result, developing alternative user-friendly interaction mechanisms for handheld devices is essential.

One natural way to solve or alleviate the problem of interaction with a handheld device is to enable speech input. Empowered by automatic speech recognition (ASR) technology, speech input can be much faster and easier than manual input/interaction through buttons, stylus, and keyboards. Speech is a natural skill of human beings. Speech input does not require much attention or direct observation to use a device; it can reduce or even completely eliminate the need for hand use during interaction, making mobile devices much more accessible. In particular, speech recognition technology for mobile handheld devices (MASR) could fundamentally help users who are physically challenged or visually impaired interact with those devices.

In the meantime, this research field is still in its infancy. The growth of research interest is also accompanied by the increasing awareness of the actual intricacies of MASR, which often involves large vocabularies and complex processing. MASR not only inherits the challenges in recognizing natural speech from ASR systems on desktop computers, such as co-occurring ambient noise, continuous utterance, diversity in individual speakers’ pronunciation, and out-of-vocabulary words (Zhou et al., 2006), but also faces significant resource barriers caused by the unique constraints of handheld devices and wireless networks. In addition, the often noisy environment could also pose challenges to the performance of speech recognition. This raises a key question about whether or not the ultimate goals of achieving full-fledged automatic speech recognition on mobile handheld devices can be truly achieved.

This research attempts to answer the following research questions through an overview of the state-of-the-art MASR research, technology, and major applications: What are the unique requirements for MASR technologies to achieve their goals? What are alternative architectural designs and technical developments to better meet those requirements? What are major challenges and issues that need to be addressed in future research in order to maximize the potential of MASR? We searched for papers on MASR technology and applications published in IEEE Xplore and ACM digital libraries, as well as a couple of other related journals such as Personal and Ubiquitous Computing and Pervasive and Mobile Computing. We also used Google Scholar as a supplementary source for finding relevant papers. The keywords used in search included mobile computing, mobile devices, speech recognition, voice/speech input, and multimodal interface. Most papers we reviewed were published within the past six years, aiming to reflect the most recent research.

The rest of the paper is organized as follows. First, we will introduce the concept of MASR and its unique challenges and describe three major types of MASR system architecture and compares their strengths and weaknesses,
Related Content

Trends in the United States Mobile Wireless Industry and the Impact of a Merger of AT&T and T-Mobile on the Trends and Overall Competitiveness of the Wireless Industry
[www.irma-international.org/chapter/trends-united-states-mobile-wireless/68650/](www.irma-international.org/chapter/trends-united-states-mobile-wireless/68650/)

Ubiquitous and Pervasive Application Design
[www.irma-international.org/chapter/ubiquitous-pervasive-application-design/17201/](www.irma-international.org/chapter/ubiquitous-pervasive-application-design/17201/)

Reducing Network Overhead with Common Junction Methodology
[www.irma-international.org/article/reducing-network-overhead-common-junction/55867/](www.irma-international.org/article/reducing-network-overhead-common-junction/55867/)

A Novel Prediction-Based Location Management Technique for Mobile Networks
[www.irma-international.org/article/a-novel-prediction-based-location-management-technique-for-mobile-networks/103967/](www.irma-international.org/article/a-novel-prediction-based-location-management-technique-for-mobile-networks/103967/)

Smart Prosthetic Hand with Object Slippage Detection, Measurement, and Control