

SiMPE : 2nd Workshop on Speech in Mobile and Pervasive Environments

A. A. Nanavati, N. Rajput
IBM India Research Lab,
4, Block C, Vasant Kunj
New Delhi - 110070, India.
namit,rnitendra@in.ibm.com

A. I. Rudnicky
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA.
air@cs.cmu.edu

M. Turunen
Dept. of Computer Science
University of Tampere
Tampere, 33014, Finland.
mturunen@cs.uta.fi

ABSTRACT

Traditionally, voice-based applications have been accessed using unintelligent telephone devices through Voice Browsers that reside on the server. With proliferation of pervasive devices and the increase in their processing capabilities, client-side speech processing is emerging as a viable alternative. As in SiMPE 2006 [2], we will further explore the various possibilities and issues that arise while enabling speech processing on resource-constrained, possibly mobile devices.

In particular, this year's theme will be SiMPE for developing regions. The workshop will highlight the many open areas that require research attention, identify key problems that need to be addressed, and also discuss a few approaches for solving some of them — not only to build the next generation of conversational systems, but also help create the next generation of IT users, thus extending the benefits of technology to a much wider populace.

Categories and Subject Descriptors

I.2.7 [Artificial Intelligence]: Natural Language Processing—*Speech recognition and synthesis*; C.3 [Computer Systems Organisation]: Special-purpose and Application-based Systems—*Real-time and embedded systems*

General Terms

Algorithms, Performance, Design, Reliability, Human Factors, Standardization, Languages, Theory

1. BACKGROUND AND MOTIVATION

The growth of mobile devices has exceeded the Internet penetration by a significant margin. This difference is even larger for developing regions where people can not afford to own a PC and/or are not literate to work on computers. At the same time, mobile devices have been becoming more pervasive owing to their continuous reduction in size along with the monotonic increase in the features they offer. However, with reduced form factor of the device, the available

input mechanisms of the device have been extremely limited. For such devices, speech provides a natural and ideal input mechanism without the requirement of any additional increase in the device size. Moreover, such devices are often used in settings where hands/eyes may be occupied in other activities. Thus, speech provides an easier means of rendering information to the user, without requiring attention of the other human senses. In developing regions, speech provides a much more user friendly interface for the illiterate masses.

The proliferation of mobile devices has stimulated the development of applications that support ubiquitous access via multiple modalities. Since the processing capabilities of pervasive devices differ vastly, device-specific application adaptation becomes essential. How does one do speech application adaptation for pervasive devices with different resource (memory, power) constraints? How does one devise efficient algorithms for speech recognition and synthesis in resource-constrained devices operating in noisy environments? To provide high quality of speech recognition on a hand-held device, Distributed Speech Recognition is used as an alternative. In this setting, the initial speech processing of the user utterance is performed on the client device and the processed signals are then passed to the Voice Server. This approach is quite restrictive; it does not adapt to a particular client's capabilities. Device adaptation of speech applications seems to be a viable approach – how does one do flexible and efficient speech application adaptation? What efficient architectures, protocols and standards should be developed to support application flexibility and the variation in client capabilities? A mobile user accesses a pervasive device in various environments, requiring her to use multiple modalities. What kind of interfaces offer a seamless experience to the user?

The questions above give only a hint of the various issues that arise. Enabling conversational systems on pervasive devices will require new models, algorithms, systems that are robust across a variety of mobile and ubiquitous devices and dynamic and noisy environments. This multidisciplinary problem invites the attention of software architects, algorithm designers, speech recognition and synthesis experts, interface designers and modellers. Designing evaluation measures, benchmarks and performance modelling of mobile speech systems will be important for supporting the advancements in the above technologies. This workshop aims to provide answers to some of these questions by invit-

ing papers, talks and panel discussions. In particular, this year's theme will be SiMPE for developing regions. There are three compelling reasons for this:

1. The penetration of mobile phone in emerging economies,
2. The importance of speech for semi-literate and illiterate users, and,
3. The completely novel HCI issues that arise when the target population is not tech savvy.

2. TOPICS OF INTEREST

All areas that enable, optimise or enhance Speech in mobile and pervasive environments and devices. Possible areas include, but are not restricted to:

- Robust Speech Recognition in Noisy and Resource-constrained Environments
- Memory/Energy Efficient Algorithms
- Multimodal User Interfaces for Mobile Devices
- Protocols and Standards for Speech Applications
- Distributed Speech Processing
- Mobile Application Adaptation and Learning
- Prototypical System Architectures
- User Modelling
- Applications for Developing Regions

3. SEED QUESTIONS

- How to do speech recognition in noisy environments ?
- How to make voice UIs flexible and adaptive ?
- Are there any novel and easier ways to handle multiple languages and dialects ?
- How do we construct speech systems with small footprints of memory and power consumption ?
- How can we distribute processing more efficiently given the increased available computing power on handhelds ? How do we trade this off with a remote server to conserve energy ?
- How do we make such devices adapt automatically to the user, task and environment ?
- What novel applications and services can be deployed on such devices ?
- What are the right learning models for such settings ?
- What does the architecture of future systems looks like ?
- How can we design scalable speech applications ?
- How can we leverage context (such as location) to make more intelligent UIs that reduce the cognitive burden of semi-literate/illiterate users ?

- What are the acoustic implications ? The major challenge in deploying ubiquitous ASR is that the operating environments may change rapidly leaving the ASR system very vulnerable. To increase ASR system robustness in varying environmental noise context, we discuss schemes that are applicable to a specific environment context.

4. INTENDED AUDIENCE

This cross-disciplinary burgeoning area requires that the people from various disciplines – speech, mobile applications, user interface design, solutions for emerging economies – meet and discuss the way forward. It would be particularly relevant and interesting to have the perspective of practitioners (application builders) and the theorists to discuss the right models and principles upon which such applications should be based and evaluated.

SiMPE 2006 [2] was a common meeting ground to bring together small isolated groups of people working in this area. To continue exchanges further beyond SiMPE 2006, we created the SiMPE wiki [1], which currently has 27 participants. A follow-up in SiMPE 2007 will further strengthen this community and pave the way for more fruitful exchanges and collaborations. The focus on developing regions and its relevance to SiMPE is timely and compelling.

5. ORGANISING COMMITTEE

- Amit Anil Nanavati, *IBM India Research Laboratory*.
namit@in.ibm.com
- Nitendra Rajput, *IBM India Research Laboratory*.
rnitendra@in.ibm.com
- Alexander I. Rudnicky, *Carnegie Mellon University*.
arl@cs.cmu.edu
- Markku Turunen, *University of Tampere*.
mturunen@cs.uta.fi

6. PROGRAMME COMMITTEE

- Abeer Alwan, UCLA, USA.
- Péter Boda, Nokia Research Center, Finland.
- Nick Campbell, ATR, Japan
- Nobuo Hataoka, Tohoku Inst. of Tech., Japan
- Matt Jones, Swansea University, UK
- Gary Marsden, Univ. of Cape Town, South Africa.
- Shrikanth S. Narayanan, USC, USA.
- David Pearce, Motorola, UK.
- Yaxin Zhang, Motorola, China

7. REFERENCES

- [1] Wiki: SiMPE: Speech in Mobile and Pervasive Environments. <http://simpe.wikispaces.com/>.
- [2] A. A. Nanavati, N. Rajput, A. I. Rudnicky, and R. Sicconi. Workshop: SiMPE: Speech in Mobile and Pervasive Environments. *MobileHCI 2006*, Espoo, Finland, September 2006.