ABSTRACT
In the past, voice-based applications have been accessed using unintelligent telephone devices through Voice Browsers that reside on the server. The proliferation of pervasive devices and the increase in their processing capabilities, client-side speech processing has been emerging as a viable alternative. In SiMPE 2008, the third in the series, we will continue to explore the various possibilities and issues that arise while enabling speech processing on resource-constrained, possibly mobile devices.

In SiMPE 2007 [2], the focus was on developing regions. Given the importance of speech in developing regions, SiMPE 2008 will include “SiMPE for developing regions” as a topic of interest. As a result of discussions in SiMPE 2007, we plan to invite and encourage Speech UI designers to participate in SiMPE 2008. We will also review the progress made over the last two years, in the areas and key problems identified in SiMPE 2006 [3].

Categories and Subject Descriptors
H.5.2 [User Interfaces]: Voice I/O

General Terms
Design, Human Factors, Experimentation, Performance

Keywords
Speech Processing, Mobile Computing, Pervasive Computing

1. BACKGROUND AND MOTIVATION
Over the years, mobile devices have been becoming more and more pervasive, owing partly to their continuous reduction in size, and perhaps largely to 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.

The proliferation of mobile pervasive 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.

Questions: How does one adapt speech applications 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.

Questions: 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.

Questions: What kind of interfaces offer a seamless experience to the user? Which modes or combination of modes are suitable for input vs. output?

Given the penetration of mobile phones in emerging economies (such as Africa, China and India), and that Voice-driven interfaces to applications have been found to have immense appeal for semi-literate and illiterate users [7], the need for designing flexible, adaptive and robust voice UIs is imminent. Since the design point, the end user, is completely different from the ones we have ever considered, several interesting questions arise.

Questions: How can SiMPE be made more intelligent and usable for such targets? Are multiple modalities confusing? People in developing regions seldom own a mobile on an individual basis. Most families now own a cell phone that is shared among family members. How do we develop speech applications that can provide a customized profile for every user? Since the purchase power in developing countries is not high, expensive and sophisticated interfaces may not be the right choice. How do we build speech systems on mobile...
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 multi-disciplinary 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.

1.1 Goals of SiMPE
SiMPE has only two ambitious goals:

- To provide a platform that brings together researchers from speech processing, algorithm design, application development and UI design to fuel faster growth of this multi-disciplinary area.
- To pose interesting problems to this community that will foster cross-pollination of ideas and hopefully define the course that SiMPE research should take over the coming years.

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:

- Speech interfaces/applications for Developing Regions
- Multilingual Speech Recognition
- 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
- Speech User Interface Design

2.1 Seed Questions
1. How to do speech recognition in noisy environments ?
2. What are the usability issues with speech on mobile devices ?
3. How to make voice UIs flexible and adaptive ?
4. Are there any novel and easier ways to handle multiple languages and dialects ?
5. How do we construct speech systems with small footprints of memory and power consumption ?
6. 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 ?
7. How do we make such devices adapt automatically to the user, task and environment ?
8. What kind of components and frameworks should be built to enable rapid application creation ?
9. How can we leverage context (such as location) to make more intelligent UIs that reduce the ‘cognitive burden’ of semi-literate/illiterate users ?

3. INTENDED AUDIENCE
This multi-disciplinary burgeoning area invites researchers interested in any aspect of the intersection of Speech processing and Mobile computing — speech recognition, speech synthesis, multimodal interfaces, mobile HCI, distributed speech processing, mobile applications, voice user interface design, memory/energy efficient algorithms, UI design — to meet and pave the way forward. We anticipate a good mix of international industrial and academic participation which should lead to lively discussions.

4. 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. air@cs.cmu.edu
- Markku Turunen, University of Tampere, Finland. mturunen@cs.uta.fi

5. PROGRAMME COMMITTEE

- Lou Boves, University of Nijmegen, The Netherlands
- Matt Jones, Swansea University, UK
- Yoon Kim, Novauris Technologies, USA
- Lars Bo Larsen, Aalborg University, Denmark
- Gary Marsden, University of Cape Town, South Africa
- Michael McTear, University of Ulster, Ireland
- Shrikanth S. Narayanan, University of Southern California, USA
- Sharon Oviatt, Inca Designs, USA
- Tim Paek, Microsoft, USA
- David Pearce, Motorola, UK
- Antonio M. Peinado, Universidad de Granada, Spain
- Mike Phillips, Vlingo, USA
- Yaxin Zhang, Motorola, China

6. REFERENCES