

## [Interspeech 2010 tutorial]

### **Multilingual Speech Processing - Rapid Language Adaptation Tools and Technologies**

**Tanja Schultz (Karlsruhe Institute of Technology, Carnegie Mellon University)**

**Alan W Black (Carnegie Mellon University)**

#### **Summary:**

The performance of speech and language processing technologies has improved dramatically over the past decade, with an increasing number of systems being deployed in a large variety of applications, such as spoken dialog systems, speech summarization and information retrieval systems, and speech translation systems.

Most efforts to date were focused on a very small number of languages spoken by a large number of speakers in countries of great economic potential, and a population with immediate information technology needs. However, speech technology has a lot to contribute to those languages that do not fall into this category. Firstly, languages with a small number of speakers and few linguistic resources may suddenly become of interest for humanitarian, economical or military reasons. Secondly, a large number of languages are in danger of becoming extinct, and ongoing projects for preserving them could benefit from speech technology.

With more than 6900 languages in the world and the need to support multiple input and output languages, the most important challenge today is to port or adapt speech processing systems to new languages rapidly and at reasonable costs. Major bottlenecks are the sparseness of speech and text data, the lack of language conventions, and the gap between technology and language expertise. Data sparseness results from the fact that today's speech technologies heavily rely on statistically based modeling schemes, such as Hidden Markov Models and n-gram language modeling. Although statistical modeling algorithms are mostly language independent and proved to work well for a variety of languages, the parameter estimation requires vast amounts of training data. Large-scale data resources are currently available for less than 80 languages and the costs for these collections are prohibitive to all but the most widely spoken and economically viable languages.

The lack of language conventions concerns a surprisingly large number of languages or dialects. The lack of a standardized writing system for example hinders web harvesting of large text corpora or the construction of dictionaries and lexicons. Last but not least, despite the well-defined process of system building it is very cost- and time consuming to handle language-specific peculiarities, and it requires substantial language expertise. Unfortunately, it is extremely difficult to find system developers who simultaneously have the necessary technical background and significant insight into the language in question. Consequently, one of the central issues in developing speech processing systems in many languages is the challenge of bridging the gap between language and technology expertise.

In this tutorial on "Multilingual Speech Processing - Rapid Language Adaptation Tools and Technologies" we will introduce state-of-the-art techniques for rapid language adaptation and will present existing solutions to overcome the ever-existing problem of data sparseness and the gap between language and technology expertise. We will describe in detail the building process for speech recognition and speech synthesis components for new unsupported languages and introduce tools to do this rapidly and at low costs. The tutorial will consist of several sections covering information ranging from database collection, to model building and system evaluation. Furthermore, the tutorial will include explicit instructions on the following issues:

- Designing databases for new languages

- Collecting text and speech databases at low costs

- Selecting appropriate phoneme sets for new languages efficiently

- Generating pronunciation lexicons for new languages rapidly

- Developing acoustic and language models for speech recognition for new languages

Developing models for text-to-speech for new language  
Integrating the built components into an application  
Evaluating and tuning the created components for this application