

Large Vocabulary, Multilingual Speech Recognition: Session Overview

Lori LAMEL, Yoshinori SAGISAKA

Spoken Language Processing Group
LIMSI-CNRS, BP 133
91403 Orsay , FRANCE
lamel@limsi.fr

ATR Interpreting Telecommunications Res. Labs.
2-2 Hikaridai Seika-cho Soraku-gun
619-02 Kyoto Japan
sagisaka@itl.atr.co.jp

Over the last decade it has become increasingly apparent that the availability of large corpora, both spoken and written, and of lexicons, is critical for system development. Moreover, the advances made in technology developed for a given language provide crucial input to speech recognition technology world-wide. However, up until the present not too much effort in the research community has been devoted to directly associate knowledge on speaker-independent, large vocabulary, continuous speech recognition technology among different languages. For efficient research on recognition technologies and wider application, direct comparison of recognition methodologies and results across different languages and the clarification of language common/specific factors is quite important. In this session an overview of the state-of-the-art multilingual, large vocabulary speech recognition is provided as a bootstrap for a discussion on the use of common speech recognition technologies and multilingual speech recognition.

The first paper summarizes the major worldwide activities in corpora collection and distribution, and is the result of contributions from ELRA, LDC and COCOSDA. The next two papers focus on the current situation of large vocabulary, speaker-independent continuous speech recognition in European and Asian languages. These papers address the choice of recognition strategies, acoustic modeling, language modeling and lexical representation, as well as any language specific characteristics that must be taken into account in developing a recognition system for that language. The basic characteristics of European and Asian languages are summarized and compared, including, for example, the inventory of phonemes in the languages, the relevance of syllables, words, and tones for the languages, language specificities such as the occurrence of homophones and monophones, the use of compound words, lexical coverage of fixed-size vocabularies, etc.

The problems of localizing speech recognition technologies to a given language are addressed in the fourth paper. This paper briefly summarizes available multilingual products and discuss the requirements and issues involved in porting to a new language. The fifth paper summarizes research in language identification, a subject in which there recently has been revived interest, due in large part to the availability of multilingual corpora and to comparative evaluations. Language identification can be seen as a necessary first step for real-world use of multilingual speech recognition, where it may not always be possible to know in advance the language of the user.

The remainder of the session will be devoted to a discussion of outstanding research problems and to system evaluation, such as: is it reasonable (as is typically the case) to use the same signal processing for different languages?; what recognition units are most appropriate for a given language or class of languages?; what types of data and how much are needed to port from one language to another? how can the relative difficulties of speech recognition in different languages be assessed?; how can evaluations of different recognition technologies reported for different languages be compared?; a summary of (multi)lingual evaluation activities will be given, with reference to the multilingual evaluation carried out in the context of the LRE SQALE (Speech recognizer Quality Assessment for Linguistic Engineering) project, and possible extensions.