Speech Database Processing Tools:
- the state of the art in automatic labelling of speech
Panel Session at LREC-2000

Nick Campbell

ATR SLT, Kyoto, 619-02, Japan
nick@itl.atr.co.jp

Abstract

The goal of this panel session is to determine the extent to which presently-existing speech database processing tools can be used to assist in the creation and annotation of large speech corpora. We hope that its outcome will be the creation of an open-source toolkit for the automatic segmentation and annotation of phonemic, prosodic, and voice-quality characteristics of large speech corpora. The post-panel report will be made available under the ISCA SynSIG web pages.

1 Objectives of the Panel

With the growth of speech technology applications, there is need for large and well annotated corpora of many kinds of speech that would be very difficult to collect and label manually. The panelists will suggest methods and tools for the automatization of this work.

2 Open-use software

Many efficient and well-designed speech processing software packages have recently been made freely available to the community. Examples of these can be found at the following internet sites: www.speech.kth.se/wavesurfer, www.etca.fr/CTA/gip/Projets/Transcriber, www.cstr.ed.ac.uk/projects/festival, http://tcts.fpms.ac.be/synthesis/mbrola.html, www.sil.org/computing/speechtools, cslu.cse.ogi.edu/get, and many more.

Each such group is developing tools for its own purposes and submitting them for public use, but (with some notable exceptions) there is little coordination of activities between the sites. There is need for an overview of the strengths and weaknesses of each piece of software and the creation of an index so that potential users can decide which best suits their requirements.

3 Common-use tools

The basic tools for the automatic labelling of speech may be the same as those for speech recognition, but extensions are required for the annotation of prosodic and speech-quality or speaking-style information. In order for common use to be made of such tools, there is need for some agreement on data formats and general-purpose data models, or standardisation of interfaces between which the data can be shared.

A secondary issue will be speech database management and tools for manual labeling or for visualisation and editing of automatic labeling. There is already much information about this at http://www.ldc.upenn.edu/annotation/.

In order to define some basic requirements and match them to abilities of the current technology, the panel will present concrete examples of data that has been annotated and will attempt to list up the full range of information that might eventually be required of an automatic system.

4 Future approaches

Although impressive progress has been made with automatic labeling and segmentation of speech, for high-end uses demanding very high temporal precision or accurate and elaborate prosodic labeling it may be necessary to go back to the fundamentals and design next-generation systems that will be based on new principles. This point-of-view will be presented in order to obtain a consensus view of the elementary design issues that might be required for future software development.

5 Participants

The names and affiliations of the participants are as follows:
- Alistair Conkie - AT&T
- Bruce Millar - Australian National University
- Dafydd Gibbon - University of Bielefeld
- Daniel Hirst - University of Aix-en-Provence
- Edouard Geoffrois - CTA/GIP
- Jan van Santen - CSLU, OGI
- Stephen Bird - LDC
- Vincent Pagel - MBROLA

6 Post-panel Discussion

Participants wishing to contribute further to this discussion or to suggest software tools and techniques are invited to contact the Panel Chair at the above address or submit their suggestions directly via SynSIG, the ESCA Special Interest Group for Speech Synthesis at synsig@isca.org.