

FlaReNet Launch 12.-13. Feb 2009 Wien

Session 3

Title: Towards more effective LR validation

Author: Florian Schiel, Bavarian Archive for Speech Signals (BAS), University of Munich, Schellingstr. 3, 80799 München, Germany, schiel@bas.uni-muenchen.de, www.bas.uni-muenchen.de

The BAS is performing LR evaluations internally and for other institutions on a regular basis. Therefore, the following regards only to LR validation and evaluation.

1. 'Automatic' vs. 'manual' validation

The distinction 'formal' vs. 'content' is impractical; we prefer for easier budgeting to distinguish between established methods that can be carried out automatically and methods that require human intervention. Furthermore we deem the validation of the documentation and meta data a separate task. A typical LR validation therefore consists of the following points:

- identification of reference from
 - specification
 - documentation
 - established good practise (e.g. SPEX or BAS guidelines)
- validation of documentation and meta data
- automatic validation
- manual validation
- quality assessment

What is to be done here?

- *development of more automatic tools for validation; examples are: automatic fault detection (clipping, noise,...), parsers for established annotation formats, detection of segmentation errors (using statistical methods).*
- *establish 'good practise' recommendations for manual validation*
- *minimize effort for manual validation*

2. Quality of an LR

A validation can result in an overall quality assessment, but these are not comparable across heterogeneous LRs. In practise a validation report contains a weighted summary of found problems (priorized list), which allows the producer/maintainer of the LR to take action or the prospective user of an LR to decide whether it is suitable for his/her purposes.

Keep in mind that a validation is in reference to a specification: if a specification allows wide margins for a certain error, the validation does not evaluate this specification.

What is to be done here?

- *establish 'good practise' recommendations about priorities*
- *establish 'good practise' recommendations for specifications (e.g. BAS guidelines)*

3. Evaluation of an LR

Evaluation of an LR makes only sense in reference to a specific objective.

This can be done, is difficult and expensive, but does not give any overall quality assessment for the LR, since the LR might be perfectly suited for another purpose than the original objective, which cannot be foreseen by the evaluator. In practise evaluations of a LR are hardly ever done because

- it is hard to agree on evaluation criteria
- a meaningful evaluation requires that the evaluator acts as a real 'user' of the LR

What is to be done here?

- *nothing*

For detailed guidelines regarding validation of speech corpora please refer to:
www.bas.uni-muenchen.de/forschung/BITS/TP2/Cookbook

References:

Schiel, F. (2003). *The Validation of Speech Corpora*. Bavarian Archive for Speech Signals, (ISBN: 3-8330-0700-1).

Schiel, F., Draxler, Chr. (2003). *The Production of Speech Corpora*. Bavarian Archive for Speech Signals, (ISBN: 3-8330-0700-1).

Henk van den Heuvel, L. Boves, E. Sanders: Validation of Content and Quality of Existing SLR: Overview and Methodology. ELRA/9901/VAL-1 Deliverable 1.1, Jan 2000.