

# Towards more effective LR Validation

Florian Schiel

Bavarian Archive for Speech Signals  
Institute of Phonetics and Speech Processing  
Ludwig-Maximilians-Universität München, Germany

The Bavarian Archive for Speech Signals (BAS) performs validations of LRs within the archive as well as for external partners on a regular basis.

Our guidelines are publicly available in the monograph *F. Schiel: 'The Validation of Speech Corpora'. 2004.*

**In this contribution we would like to make 3 points:**

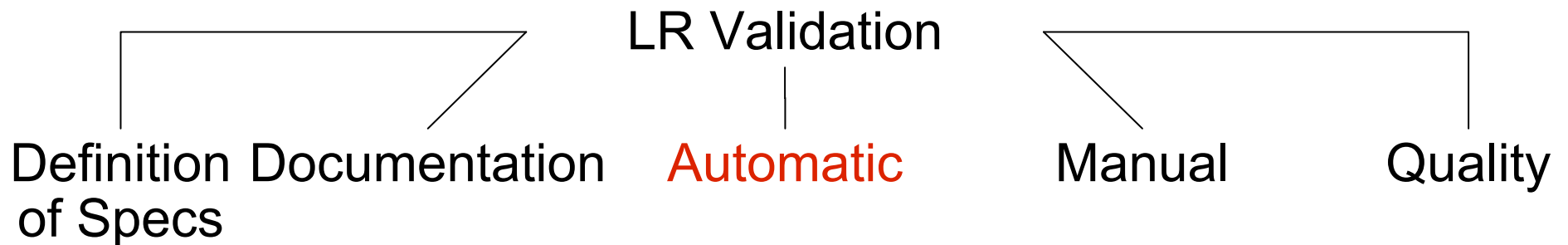
- LR validation structure
- LR quality assessment
- LR evaluation

# 1. Validation Structure

Most commonly used:



Bavarian Archive for Speech Signals:

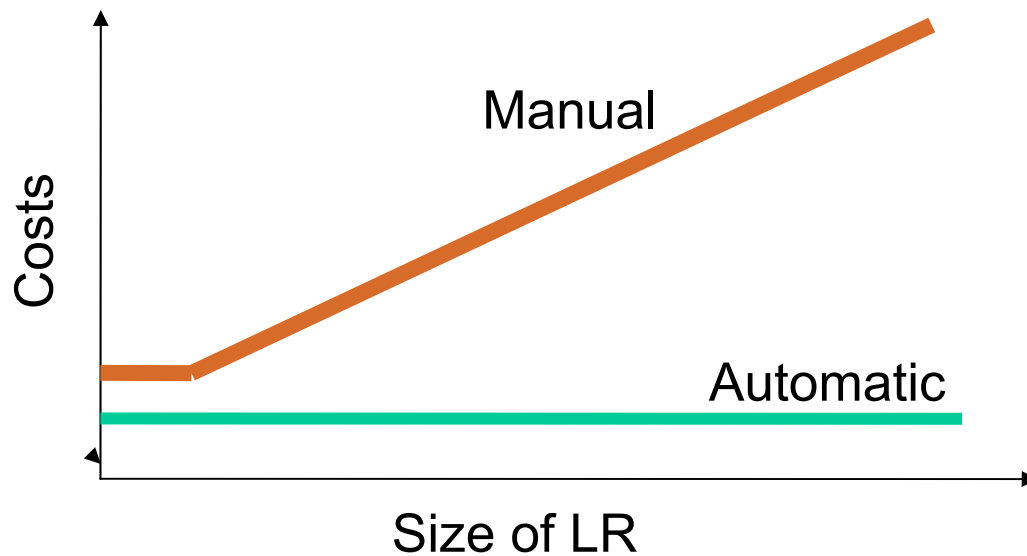


# Why?

Why structure into automatic and manual instead of formal and content?

Because...

- it's simpler (often distinction between formal and content not clear)
- it makes cost estimates easier



# 1. Validation Structure

*What is to be done here?*

- *development of more automatic tools for validation e.g. 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 Assessment

quality measure = (weighted) average over specific sub-assessments

but ...

- not comparable across heterogeneous LRs
- depends on the quality of the specification
- might mislead producers of LRs to achieve artificially high values

probably better:

prioritized fault list = top 10 most several faults found by validation (with remedy)

## 2. Quality Assessment

*What is to be done here?*

- *establish good practice rules for specifications*
- *establish good practice rules for priorities (dependent on type of LR)*

### 3. Evaluation

Evaluation of an LR is answering the following question:

*Does the LR fulfil its specified purpose and if so, to what extent? (= measure of usefulness)*

**The main characteristic of evaluation is that we never do it!**

The reason is probably that many LRs are useful with respect to many different aims.

(Good examples are the *TIMIT* and *Verbmobil* corpora.)

Therefore we think that a general requirement for evaluation of LRs is obsolete. It might be useful as a pre-study of certain projects where the aim is clearly defined and a decision of usage of an existing LR has to be taken.

## 3. Evaluation

*What is to be done here?*

- *probably nothing*