

Interactive transcription and transliteration

A win-win opportunity for HSS and speech research

Jens Edlund
KTH Tal, Musik och Hörsel/CLARIN-SPEECH



This is a talk about tools and the people who use them

- Speech!
- ...and text, uncharacteristically
- Transcription of speech is tedious, error prone and *very* expensive
- Transliteration of text is almost as bad
- (Will likely use “transcription” for both...)
- Some examples...



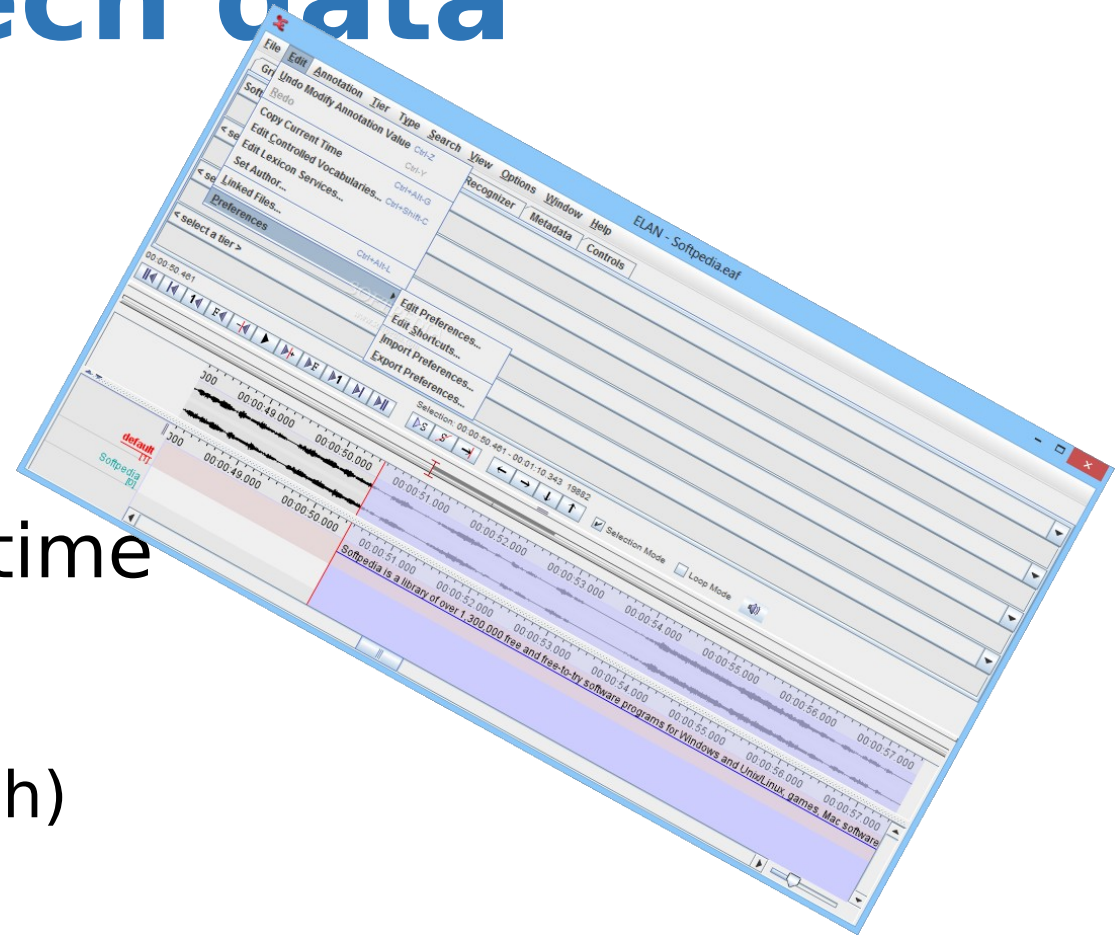
Example 1: subtitling

- ASR on the audio signal does not work
- Subtitles aren't transcriptions of what's said
 - they're summaries
- More to do than typing
 - Fit start time and end time
- Live subtitling very stressful
- Syllable keyboards take up to a year to master, and professionals
 - are worn out in a few years
- Legislation, accessibility requirements



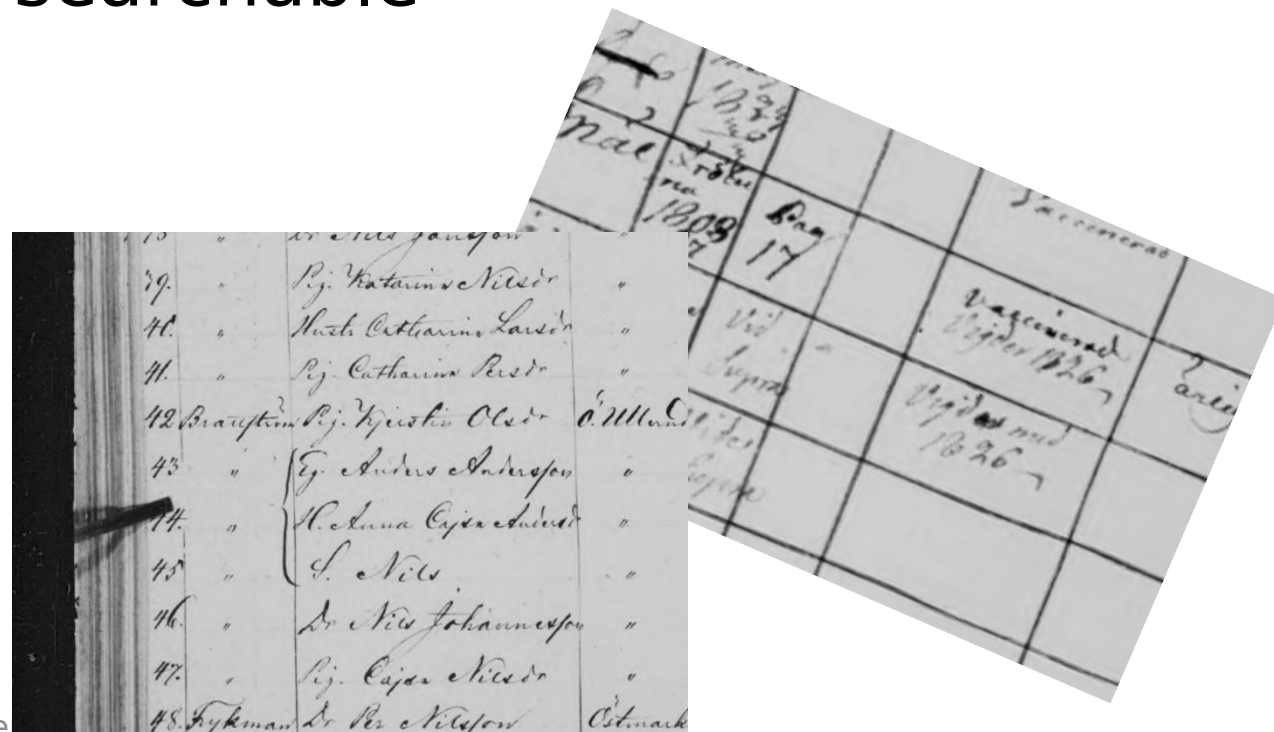
Example 2: transcription of speech data

- ASR will often not work
 - Audio quality
 - Multiple speakers
 - Noise
 - Language
 - At least 10 times real time
- (with good tools)
- Much needed
 - For research (of the speech)
 - For speech technology



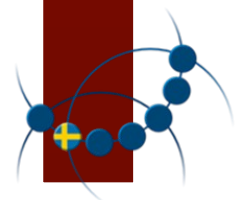
Example 3: translitteration

- Ergonomically bad
- Again, more to do than typing (fit to original text)
- Lots of materials aren't searchable
 - Hand written materials
 - Poor scans
 - Signs
 - etc.

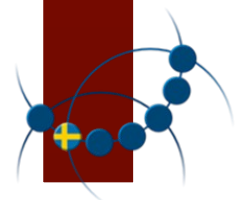
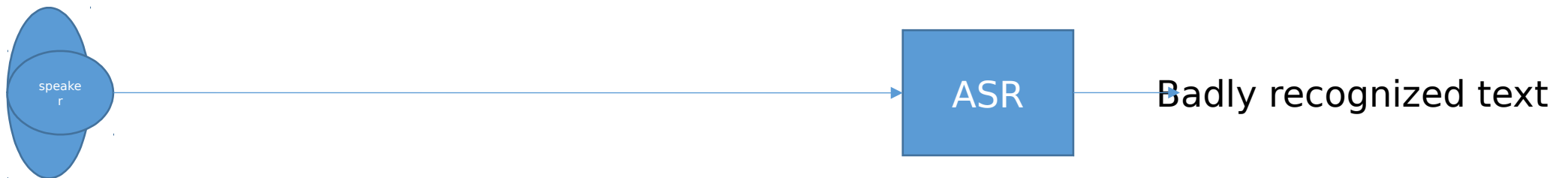


The solution

- Speech supported transcription/transliteration
 - Why is this better?
 - And how is this interaction design?

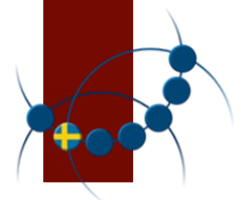
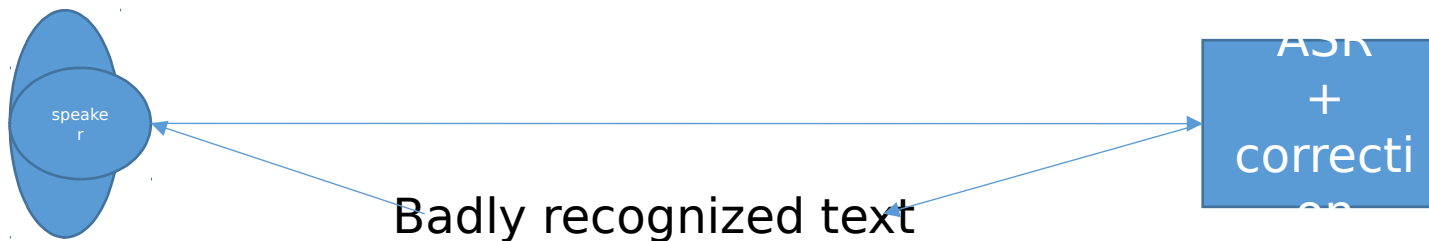


No interaction design



Interaction design!

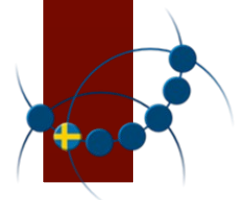
- This is where we're at now.
 - Learning iteratively
 - GUI and usage tests



Speech drawbacks

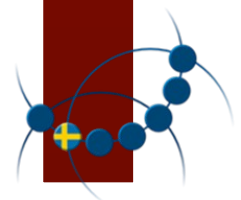
- OOV words
- Poor acoustic models
- Poor language models

- Solutions?



But win-win? Really?

- Really.
- Subtitles: Speed – Efficiency – Training data
- Transcription: Data all over
- Translitteration: Same, plus read texts (possible accessibility)



Thank you for your time!

Questions?



Jens Edlund, NCN Workshop on Interaction Design,
Göteborg

2017-05-03 - 04



What's this?

- We propose a speech technology based method to transcribe speech recordings and transliterate hand writing that brings significant and concrete benefit to two fields: HSS (including speech research) and speech technology. The method - interactive speech based transcription and transliteration - is intended as a collaborative effort between these areas, and a collaboration in which both parties win.
- In HSS research, a large amount of potential research materials consist of manuscripts that cannot easily be transliterated automatically with optical character recognition (OCR) for a host of reasons ranging from hand writing variability through poor print quality and

