Towards Automatic Transcription of ILSE – an Interdisciplinary Longitudinal Study of Adult Development and Aging

Jochen Weiner, Claudia Frankenberg, Dominic Telaar, Britta Wendelstein, Johannes Schröder, Tanja Schultz

jochen.weiner@uni-bremen.de

Motivation

- The population in Germany is ageing rapidly
- The most populous age group in

1950	10-year-olds
2000	40-year-olds
2050	60-year-olds

→ Causes changes to society and bears challenges

- The challenges are addressed in interdisciplinary projects including physicians, psychologists, gerontologists, sociologists, linguists, engineers, and computer scientists
- ILSE:
  Large-scale corpus with the goal to assess healthy and satisfying aging in middle adulthood and later life

ILSE

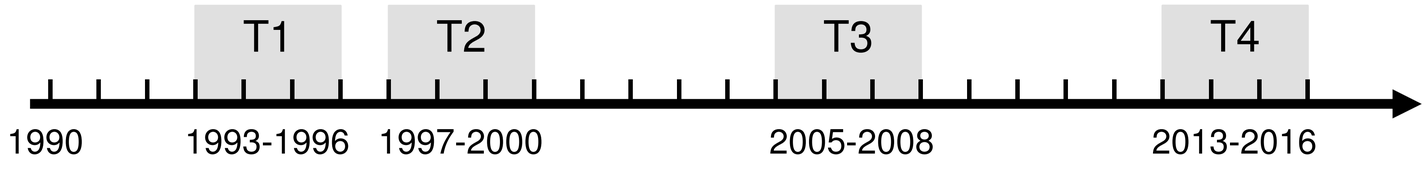
- Extensive medical examinations
- Biographic interviews
  - Participant was interviewed by one of 53 interviewers
  - Semi-standardized interview procedure

→ **over 8,000 hours of recordings**

- 1000 participants
- Two birth cohorts:

C30	1930 - 32
C50	1950 - 52

- Four measurements:

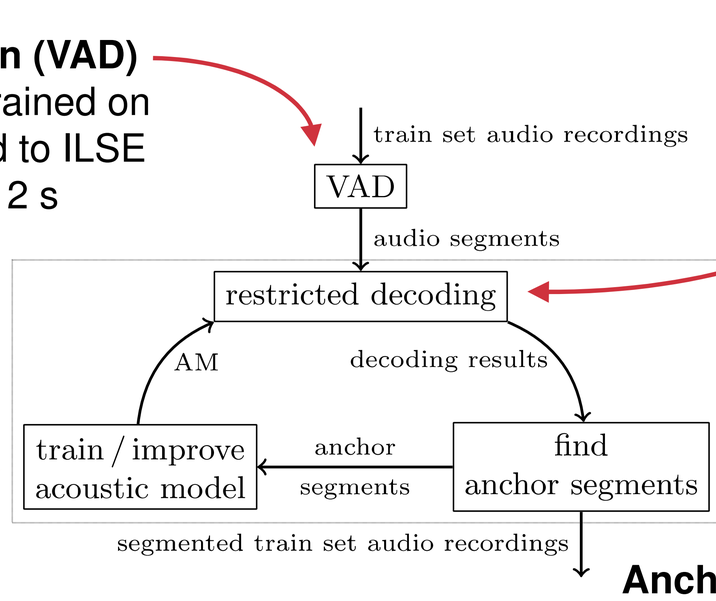


Corpus of Transcribed Interviews

Transcribed interviews: < 5%
(384 h / 74 participants)

Text processing
- Structure, Normalization
- 50k word types / 2,800 tokens
- Half of the words occur just once

Speech Data Processing
- Flat start or forced alignment not possible
- Iterative long audio alignment
  → 96 hours of data usable (three iterations)



Voice Activity Detection (VAD)
- HMM-GMM system trained on GlobalPhone adapted to ILSE
- Segment at silence > 2 s

Restricted Decoding
- Only use words of the transcription

Training
- Use anchor segments to train a new DNN
- Speaker independent

Anchor Segments
- Levensthein Alignment: Transcription ↔ Hypothesis
- > 2 correct words: transcription must be correct

ILSE Interviews

- Interview duration per participant (in hours):

T1	T2	T3	T4
6.0 ± 2.6	2.5 ± 0.7	1.8 ± 0.8	1.3 ± 0.3

- Recording: stereo voice recorder on the table

T1	T2	T3	T4
tape	tape	MP3	PCM

Challenges for Automatic Speech Processing

- Transcription quality
- Anonymized Transcriptions
- Recording Quality
- Speaking Style and Crosstalk
- Emotional Speech
- Dialectal Speech

Universität Bremen