

# AVATech audio detectors

Fraunhofer IAIS

AVATech Workshop 21.04.2010 Nijmegen, Netherlands  
Daniel Schneider, Sebastian Tschöpel

© Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS



# Contents

- Introduction
- AVATech Corpus
- Annotation scenarios
- Semi-automatic workflows for the annotation scenarios
- AVATech audio detectors: State of the art
- Outlook
- Demonstration

© Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS



# Fraunhofer IAIS Speech Group

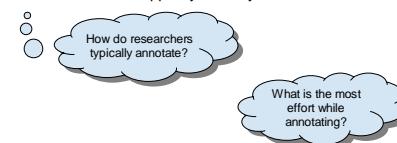
- Working on Spoken Document Retrieval since 2001
  - ASR, speech search, structural audio analysis
- Involved in public research projects and industry cooperations
- So far: mainly work on Broadcast data
  - Focus on language-dependent solutions for German
- But also specialized work
  - ASR on motorcycles, Animal sound discovery

© Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS



# Introduction

- What has been done?
  - Reviewed AVATech corpora provided by MPI Nijmegen
  - Derived examples for *annotation scenarios*
  - Improved analysis *algorithms on difficult AVATech data*
  - Developed concepts for detectors that *exploit user-feedback*
- Open problem: Definition of more annotation scenarios
  - how can we support your daily work?



© Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS



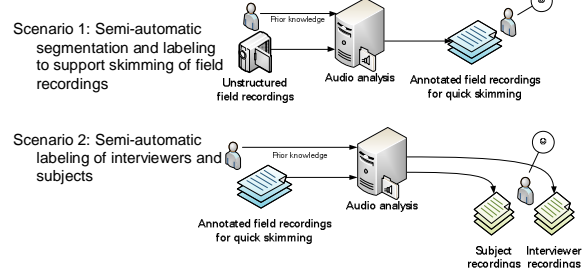
# AVATech Corpus

- Material in various MPI corpora is
  - varying in audio quality (office experiments vs. field recordings) disqualifies fixed analysis models
  - varying in language disqualifies language-dependent approaches
  - varying in genre (interviews, monologues, ...) disqualifies specialized solutions
  - not necessarily carrying information of interest in audio
- Flexible solutions needed that are able to cope with a large variety of annotation problems
- Initially we focus on two general annotation scenarios

© Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS

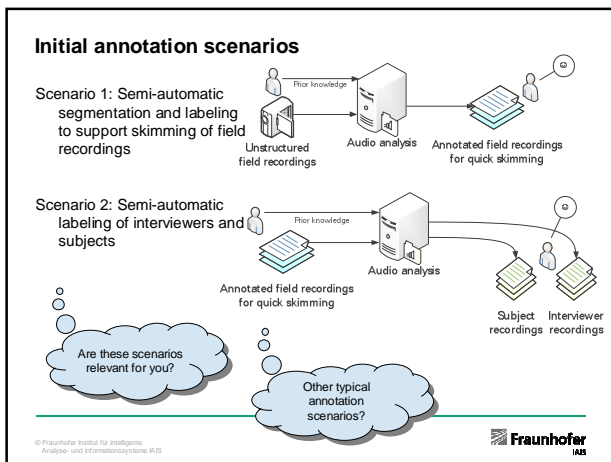
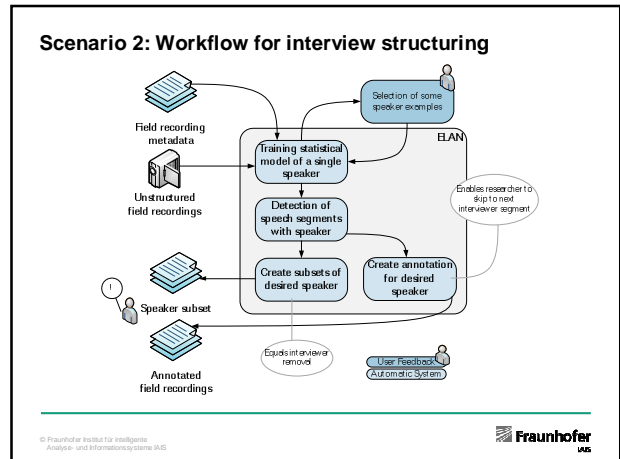
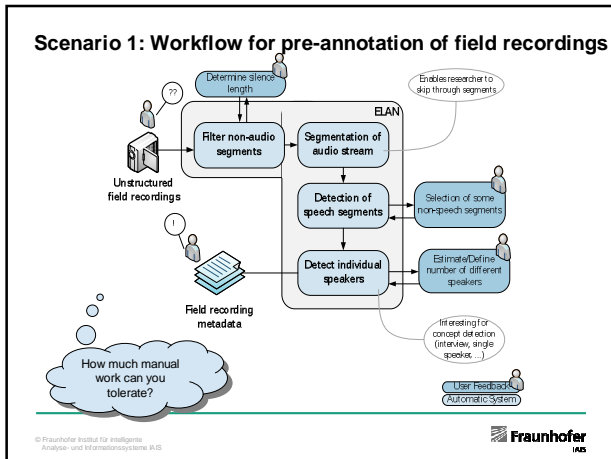


# Initial annotation scenarios



© Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS



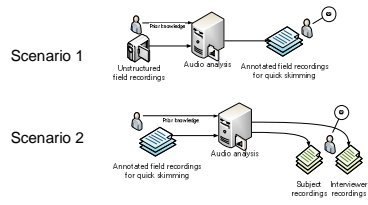


- ### AVATeCh audio detectors: State of the art
- Audio segmentation
    - *Autonomously splits audio stream into homogeneous segments*
    - Using Dynamic Programming / Bayesian Information Criterion (BIC)
    - Baseline with MFCC features
    - We investigate noise-robust features using spectral auto-correlation (SAC)
    - Essential pre-processing, works well on non-noisy data
  - Speech/Non-speech detection
    - *Detects whether a segment contains speech or not*
    - Based on GMMs with MFCCs/SAC
    - Works well with in-domain training data
    - Integrate user-driven feedback mechanism for adaptation
    - Similar: Gender Detection
- © Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAS

- ### AVATeCh audio detectors: State of the art
- Speaker clustering
    - *Joins and labels segments with the same speaker*
    - Based on Bayesian Information Criterion
    - Works well on Broadcast data, e.g. for detection of anchor person
    - Poor results on most AVATeCh corpora, robustification needed
    - How to integrate user-feedback?
    - High time complexity of clustering – what about large collections?
  - Speaker Identification
    - *Identifies segments with known speakers in a given corpus*
    - Using spectral and pronunciation features
    - Plan to integrate user-driven mechanisms to automatically train new speaker models
- © Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAS

- ### Outlook
- Language Independent Alignment
    - Approach: Top-Down method (from paragraph to word level) using different language-independent features
      - Histogram-like matching of repetitive patterns in text and audio
      - Optional anchor points available through user-feedback
    - Core difficulties: Lack of language model & noisy data
  - Acoustic Query-By-Example
    - Find repeated similar audio events by marking one example
    - Approach: Fast matching in pre-computed feature index
      - Detection and discrimination of linear and noise-like spectral features
      - Sparse point of interest encoding
      - Idea from animal sound discovery
- © Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAS

## Demonstration



Thank you for your attention!