ELAN as Flexible Annotation Framework for Sound and Image Processing Detectors

E. Auer, A. Russel, H. Sloetjes, P. Wittenburg,
O. Schreer, S. Masneri, D. Schneider, S. Tschöpel

MPI for Psycholinguistics, FhG IAIS, FhG HHI

MPI for Psycholinguistics

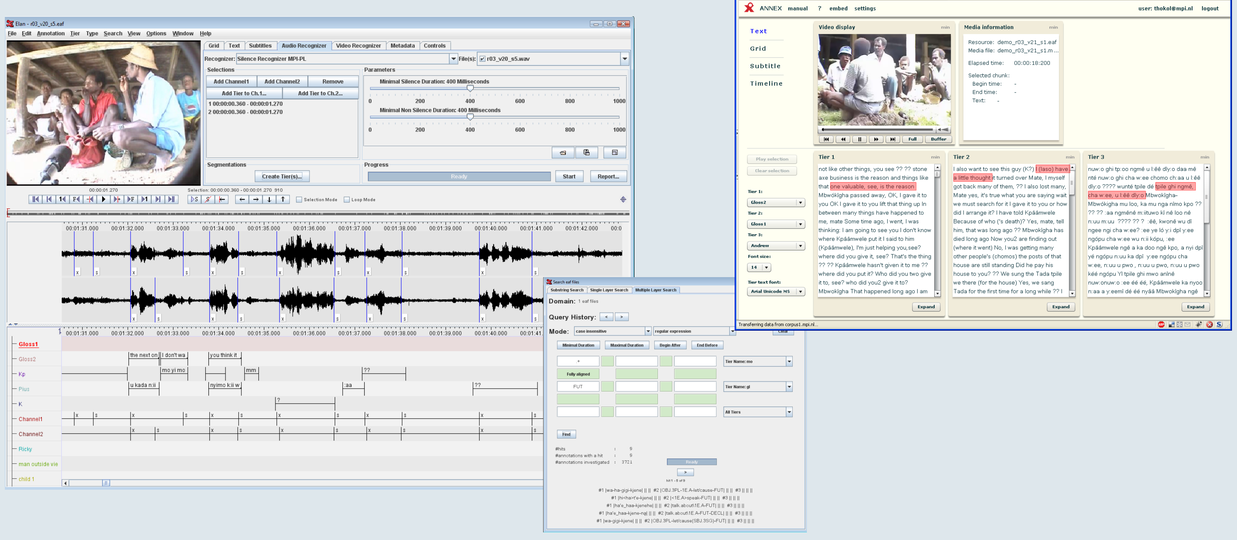
The Problem

- Large and growing archives in humanities research (at MPI > 50 Terabyte)
- Manual annotation can take from 35 up to 100 times real time (media duration)
- Many recordings are not annotated at all or very rudimentary
- Many recordings are therefore not analyzed and researchers forget about the content
- Huge variety in type and quality of recordings, many languages spoken
- Existing annotation tools like ELAN mainly support manual annotation
- Existing pattern recognition software is not capable of dealing with this level of variations

AVATecH Goals

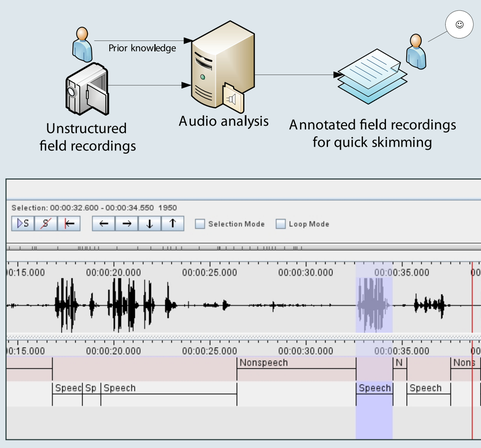
AVATecH (Advancing Video/Audio Technology in Humanities Research) is a project that aims to develop and implement audio and video technology for semi-automatic annotation of heterogeneous media collections as they occur in daily research practice.

- Develop pattern detection components that
  - can cope with the huge variety
  - allow user interaction on various levels
  - produce annotations and/or intermediate other results
- Start with relatively simple detectors based on existing technologies
- Develop search and filtering components for complex annotation structures created by the detectors
- Test set of 100 Gigabyte from the MPI for Psycholinguistics archive
- Define Interface Specification for recognizer components
- Component specifications in CMDI files (Clarin Metadata Infrastructure)
- Use ELAN as an interactive integration platform
- Develop ABAX for batch-wise, unattended processing



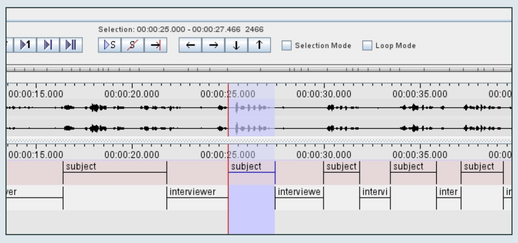
Low-Hanging Fruit Audio Detectors

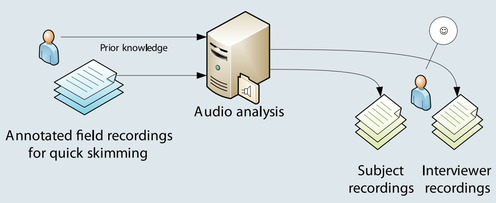
- Noise-robust segmentation of the audio stream into homogeneous segments
- Language-independent extraction of audio segments containing speech
- Language-independent intra-document speaker clustering, creates a speech tier for each speaker per document
- Pitch based vowel detector that produces annotations with pitch and intensity properties



Scenario 1: automatic speech/non-speech segmentation. The user can preselect non-speech segments to improve the algorithm.

Scenario 2: speaker identification in preprocessed recordings, detects all segments with a specific speaker.





Low-Hanging Fruit Video Detectors

- Shot and sub-shot boundary detection that identifies scene changes and creates a story board for quick inspection
- Motion detection that detects camera motions or, more importantly, motion of an object in the scene
- Skin color detection detecting head, arm and hand patterns
- Tracking of movements of head, arms and hands
- Face recognition detector, identifies the number of participants

A storyboard based on shot/sub-shot detection and a head/arm/hand tracking system based on skin color detection.

First Results and Conclusions

ELAN / ANNEX / TROVA

- Widely used in the field of linguistics for multimedia / multi-modal annotation
- Available for Windows, Mac OS X and Linux
- Support for up to 4 video files per transcription
- Support for unlimited number of hierarchically organized tiers
- Annotations stored in EAF format (XML)
- Already equipped with two audio detectors for silence and intonation pattern detection
- TROVA search engine to search for complex patterns
- ANNEX allows to visualize annotation and media via the web

- Manual annotation is utterly time consuming and does not scale with the amount of material
- First detectors for audio and video have been implemented building on existing technologies and algorithms
- First user feedback indicates that
  - ease of use of detectors is most important in humanities research
  - integration of detectors in tools like ELAN would be of great help
  - interaction of the user with the detectors is highly appreciated
- The next phase is the development of more complex detectors that operate cumulatively, i.e. using existing annotations
- Improve and streamline integration of the available detectors in interactive frameworks

LAT: http://www.mpi.nl/lat

AVATecH: http://www.mpi.nl/research/research-projects/language-archiving-technology/avatech/