

# Anvil

## A Generic Annotation Tool for Multimodal Dialogue

Michael Kipp

University of the Saarland & DFKI

Germany

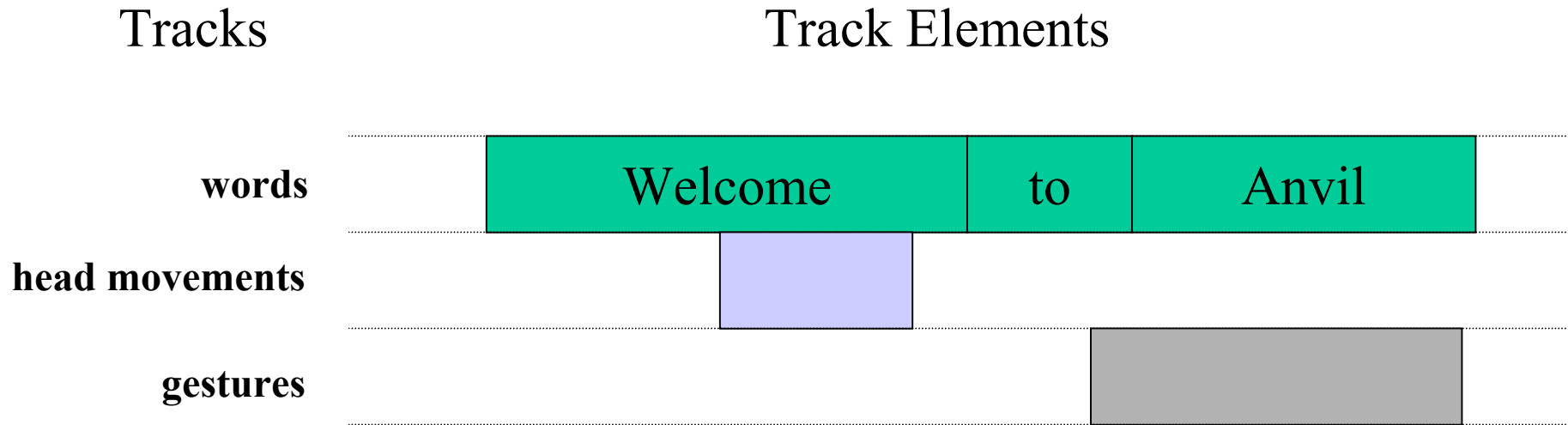
[kipp@dfki.de](mailto:kipp@dfki.de)

# **A***nnotation of* **V***ideo* *and Spoken* **L***anguage*

# Motivation

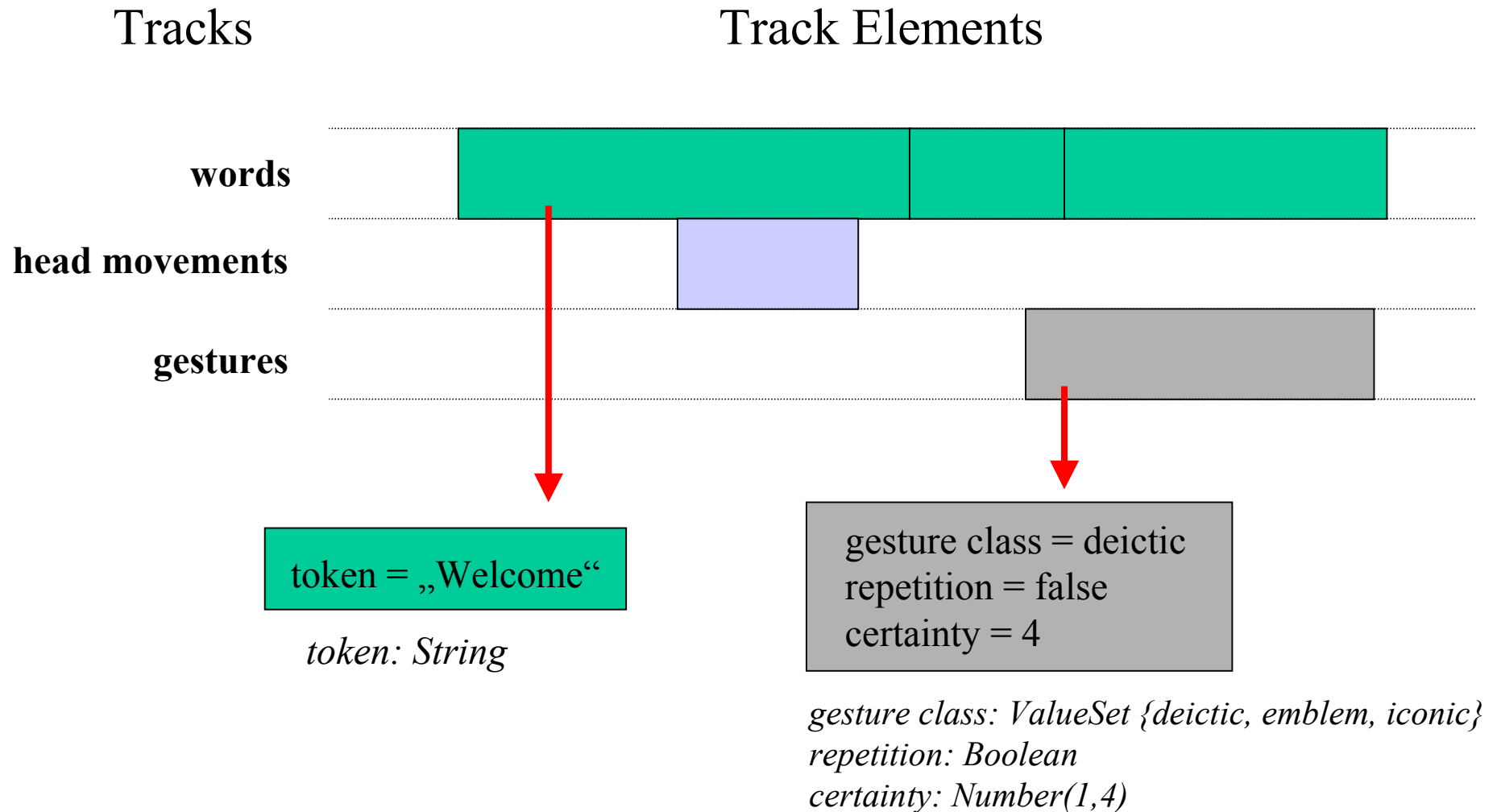
- Find new **gestures** and their semantic and temporal relation to **speech**
- Analysis of empirical data on video files
- **Notepad** for parallel events (video synchronized):
  - speech
  - head movements
  - gestures
- Parallel information layers = *Tracks*
- Notes for one video = *Annotation*

# Annotation



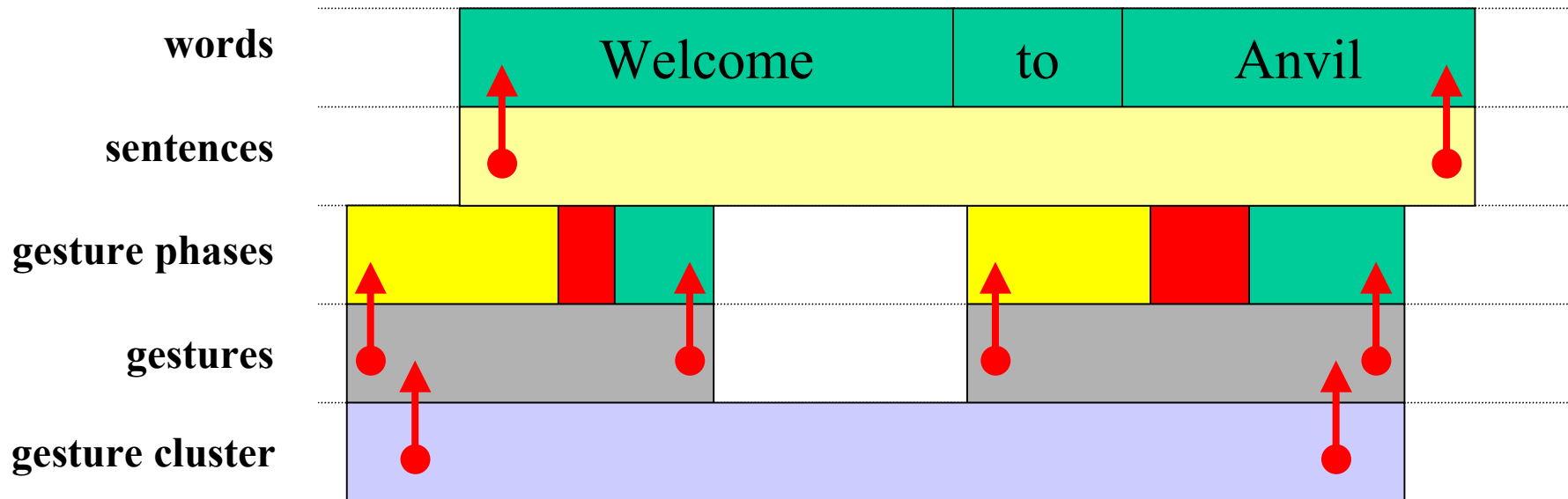
Elements within one track *must not* overlap!

# Annotation



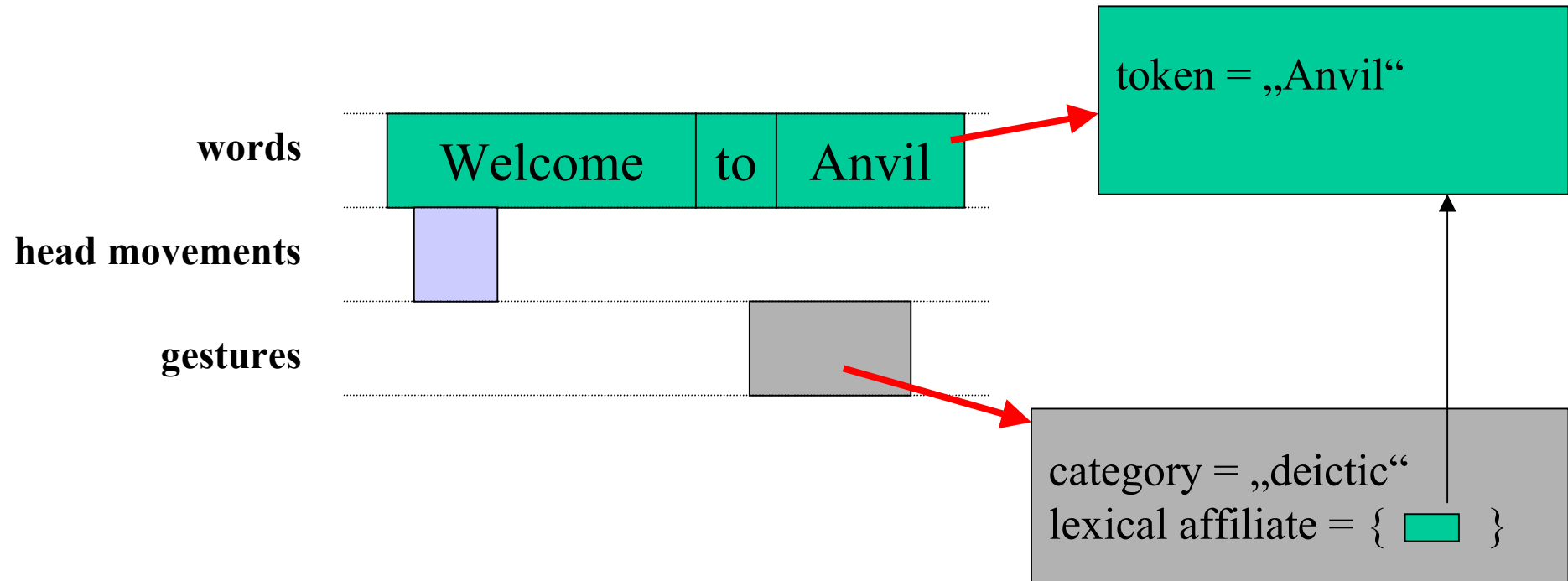
# Track Hierarchies

*primary and secondary tracks*

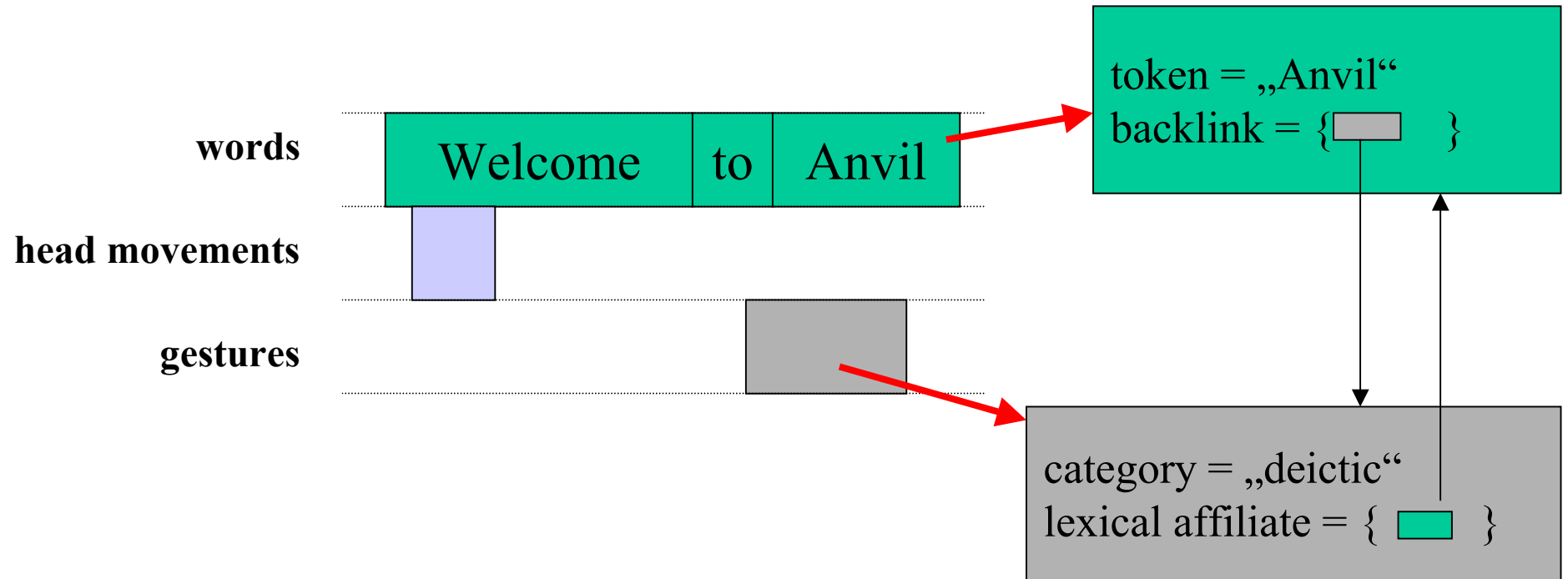


- Track hierarchies express a very specific kind of **relation**
- How to code **general interdependencies** between elements in the same or different tracks?
- Put logical **links** into a new type of attribute

# Links



# Reciprocal Links



*lexical affiliate: ReciprocalLink(backlink)*

# Specification

- Specify
  - Tracks *name + type + reference track*
  - For each track: Attributes *name + type*
    - String, Boolean, Number(x, y)
    - ValueSet {set of user-defined labels}
    - MultiLink, ReciprocalLink(A)
- This is the *annotation scheme*  
= specification file (XML)

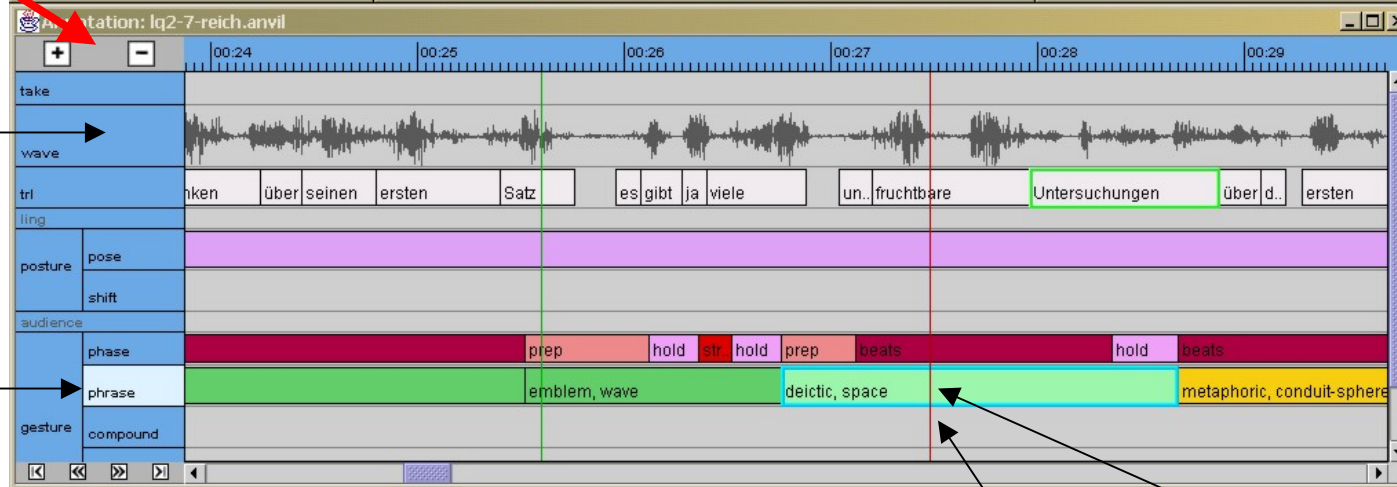
main window

video window  
(frame-accurate control)

track window



annotation board



waveform

active track

tracks

track elements

playback line

selected element

# Project Tool

- Anvil only good for **one annotation** at a time
- A *project* is a **set of annotations** with equal specification (same scheme)
- What can you do with projects?
  - **overview** (coder, date, progress)
  - **search & find** across annotations
  - **export** found units for analysis

# Project Tool



search engine

annotation files

annotation info

- coder
- last modification time
- summary/notes
- each track's size

Anvil 3.1

File Edit View Tools Bookmarks ?

Audio format: LINEAR, 22050.0 Hz, 8-bit, Mono  
 Computed frame rate = 25 frames per sec.  
 Duration: 01:10:48 (1761 frames)  
 Open first player  
 Closed (last) video lq1-2-reich.avi  
 Loading video: IV50, 384x288, FrameRate=  
 Audio format: LINEAR, 22050.0 Hz, 8-bit, Mono  
 Computed frame rate = 25 frames per sec.  
 Duration: 03:15:92 (4897 frames)  
 Open first player

Current specification:  
 D:\Promotion\spec\litqua2.xml

Video: lq1-4-reich.avi

speed 100 90 80 70 60 50 40 30 20 10 0

Track: gesture.phrase  
 Referenced track: gesture.phase  
 Start: 00:47:40  
 End: 00:48:56 (29 frames)

Attributes  
 emblem type: attention  
 copy gesture: true  
 timing: span  
 attitude: importance  
 certainty: 0

Comment

start edit end  
 cut extend del

Frame: 1200  
 Time: 00:48:00

Annotation: lq1-4-reich.anvil

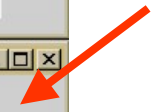
Search Hits (Project)

Searched Track: gesture.phrase  
 260 elements found

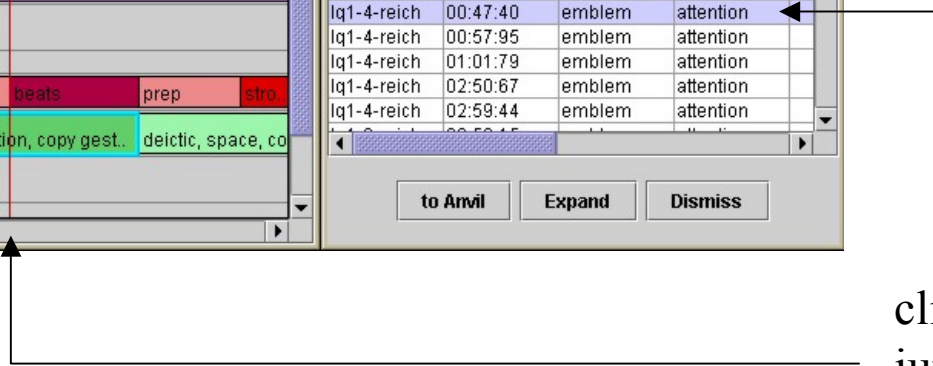
file name	start time	category	emblem type	me
lq2-1-reich	02:24:63	emblem	anticipation	
lq2-1-reich	02:59:39	emblem	anticipation	
lq3-2-reich	01:01:24	emblem	anticipation	
lq1-2-reich	00:43:59	emblem	attention	
lq1-4-reich	00:41:27	emblem	attention	
lq1-4-reich	00:43:72	emblem	attention	
lq1-4-reich	00:47:40	emblem	attention	
lq1-4-reich	00:57:95	emblem	attention	
lq1-4-reich	01:01:79	emblem	attention	
lq1-4-reich	02:50:67	emblem	attention	
lq1-4-reich	02:59:44	emblem	attention	

to Anvil Expand Dismiss

list of hits



click to jump/load



# Import/Export

- Anvil is not suited for speech transcription or statistical analysis
- **Import** PRAAT tier to Anvil track
- **Import** XWAVES speech transcription
- **Export** single track of an annotation as ASCII text
- **Export** single track of a project to text tables
- **Import/Export** by working with Anvil XML annotation files (e.g. using XSL)

# Extending Anvil using Plug-ins

- File reader/writers/converters  
(other formats!)
- Different views on the annotation:  
tables, matrices, trees, 3D graphics, etc.
- Analysis modules

# Gesture Analysis Plug-in

*single gesture*

adaptor  
beat  
deictic.addressee  
deictic.audience  
deictic.person  
deictic.self  
deictic.space  
emblem.anticipation  
**emblem.attention**  
emblem.block  
emblem.chide  
emblem.concession  
emblem.despair-threat  
emblem.dismiss  
emblem.doubt-shrug  
emblem.finger-ring  
emblem.hand-clap  
emblem.hand-rub  
emblem.hands-up  
emblem.indignation  
emblem.invitation  
emblem.knee-slap  
emblem.more-or-less  
emblem.number  
emblem.one-hand-other-hand  
emblem.pray  
emblem.purse  
emblem.refusal  
emblem.so-what  
emblem.surprise-hands  
emblem.two-sides  
emblem.wave  
emblem.wipe  
iconic.away-motion  
iconic.circle

**emblem.attention (34)**  
**av. num of phases: 2.7058823**  
num of single strokes: 10  
num of mult. strokes: 20  
num of indep. holds: 1

RH = 11  
LH = 18  
2H = 2

*handedness*

average init/span [sSpeech-sGest] timing (13): -0,31 sec +/- 0,32 sec  
average direct/indir [sSpeech-sGest] timing (13): 0,05 sec +/- 0,26 sec

*relative timing*

Lexical affiliates:  
aufmerksam  
wichtig  
einen einzigen  
eins ist sicher  
und  
warum  
nie  
gezeigt  
interessant  
halt  
gute  
nur  
ganz besonders  
aber noch etwas  
Höhepunkten

*list of lexical affiliates*

Distribution of connected argument elements:  
toulmin (19)  
claim: 9 (47,4 %)  
claim-preparation: 4 (21,1 %)  
data: 3 (15,8 %)  
qualifier-certain: 1 (5,3 %)

Dismiss

# Technical Data

- Developed on PentiumIII PC *500MHz 256MB*
- Platform-independent (Java + JMF)
- XML-based
- Supports QuickTime and AVI
- Small (1.2MB) and easy to install
- Documented *43 page manual + papers + demo files*

**FREEWARE**

# Circulation of Anvil

- Papers + presentations at
  - Agents 2001 (Workshop)
  - Oralité et Gestualité (Orage) 2001
  - EuroSpeech 2001
- Part of ISLE (formerly EAGLES) survey
- Anvil mailing list: 80 entries  
*(HCI, Ling., Phonetics, CA, Psycholing., Ethology, Animation, Tools/Corpora - Research & Teaching)*
- Model function for NITE, LDC, Observer and TalkBank project

# Summary

- Anvil is a free video annotation tool for
  - coding single annotations
  - browsing multiple annotations
- Clean separation of tool and coding scheme
- Leaves special tasks to other programs  
*speech transcription, statistical analysis*
- Extensible via plug-in architecture

**[www.dfki.de/~kipp/anvil](http://www.dfki.de/~kipp/anvil)**