

# MeetEval, Show Me the Errors!

## Interactive Visualization of Transcript Alignments for the Analysis of Conversational ASR

Thilo von Neumann<sup>1</sup>, Christoph Boeddeker<sup>1</sup>, Marc Delcroix<sup>2</sup>, Reinhold Haeb-Umbach<sup>1</sup>  
<sup>1</sup>Department of Communications Engineering, Paderborn University, Germany <sup>2</sup>NTT, Kyoto, Japan  
 {vonneumann, boeddeker, haeb}@nt.upb.de, marc.delcroix@ieee.org

### Reasons for a Visual ASR Analysis

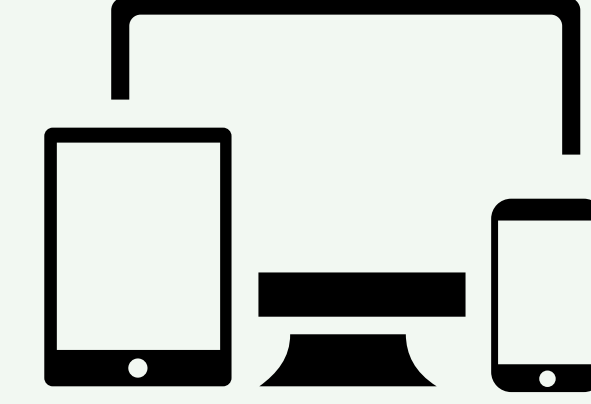
- WER doesn't tell you where the errors are
  - ▶ Burst errors?
  - ▶ Speaker confusions?
  - ▶ Duplicated transcripts?
- Easily spot errors in your ASR output with our visualization!
- Works for hour-long recordings

### Open Source



[github.com/fgnt/meeteval](https://github.com/fgnt/meeteval)

### Try it for Yourself!



[https://fgnt.github.io/meeteval\\_viz](https://fgnt.github.io/meeteval_viz)

### WER Breakdown

	Count	Relative	Counts by Speaker	
			P37	P38
Ref. Words	558	100.0%	259	299
Correct	243	43.5%	147	95
Substitution	205	36.7%	83	125
Insertion	187	33.5%	155	30
Deletion	110	19.7%	29	79

- Hover to see numbers of errors per speaker

### Alignment Variants

- cp: Speaker permutation
- tcp: cp + time constraint
- Word-level timings estimated from segments
- (tc)orc: Speaker-agnostic alignment

### Audio Playback

- Play recording file or separated audio
- Local files or file server required

### Navigation

#### Minimap(s)

- Top: Brief summary of errors
- Histogram of errors over time
- Bottom: Segment timeline
- Shaded area: Active selection
- Long sessions: multiple hierarchical minimaps

#### In Details View

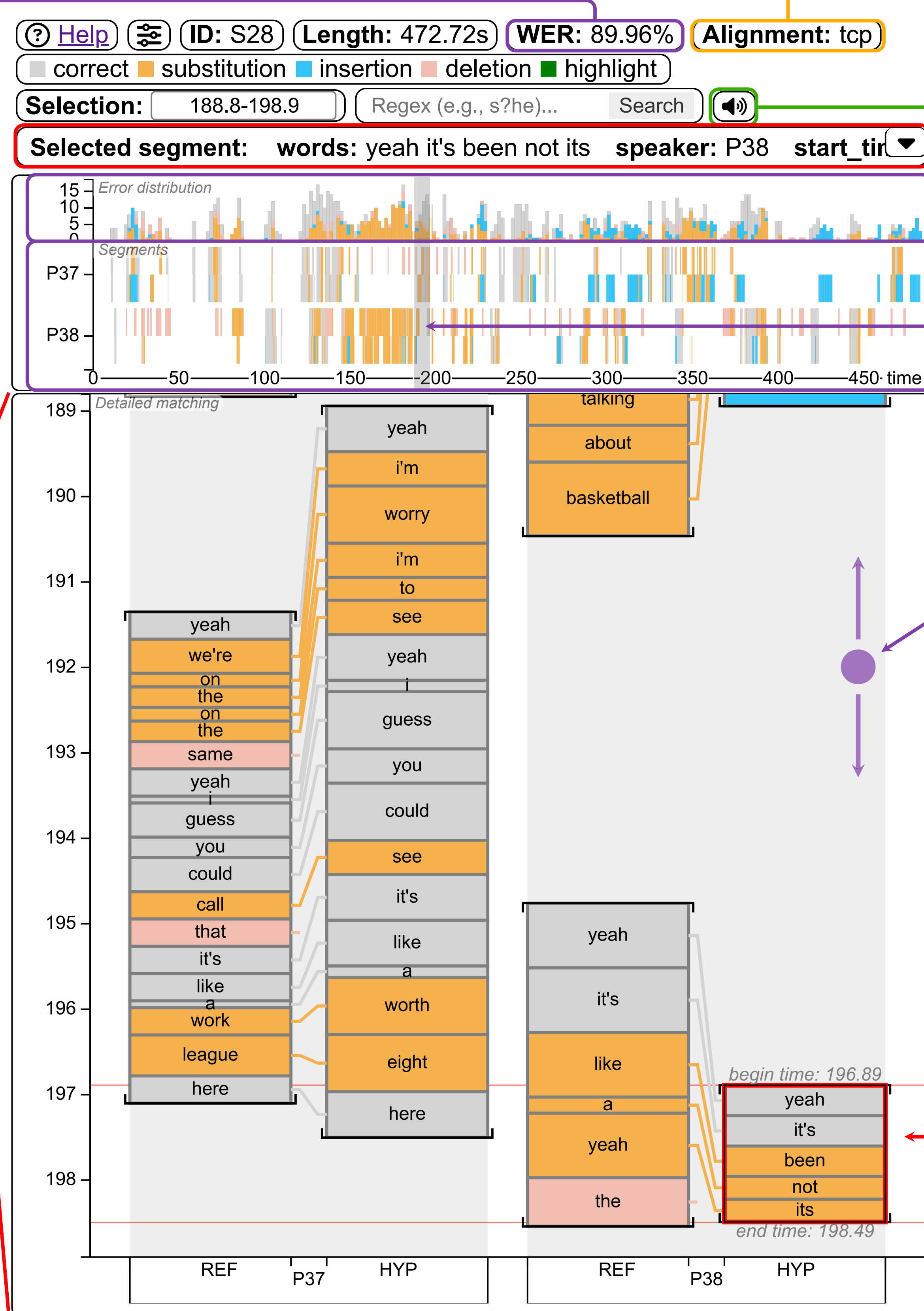
- Click, scroll, touch, drag
- On Tablet & Desktop!

### Detailed Matching View

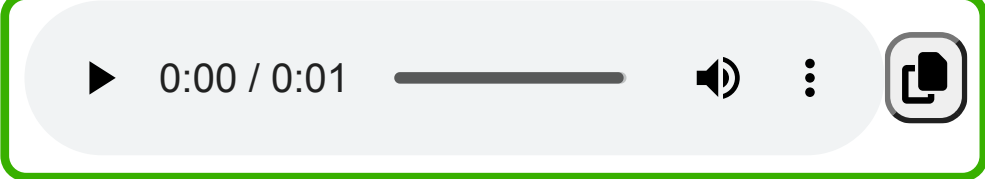
- Time flows from top to bottom
- One column for each speaker
- Left: Reference, Right: Hypothesis
- Matching words are connected by lines
- Colors represent errors
  - cor
  - sub
  - ins
  - del

### Other Features

- Regex search for words
- Synced side-by-side view for system comparison
- Offline, no server, sharable HTMLs
- Session Browser



### Selected Segment

words:	yeah it's been not its
speaker:	P38
start_time:	196.89
end_time:	198.49
audio_path:	
# words:	5

- Detailed information about selection
- Custom keys are displayed

### Input Format

SegLST (from CHiME challenges). List of segments with the following fields:

```
[
  {
    "words": "yeah it's been not its",
    "speaker": "1",
    "session_id": "S28",
    "start_time": 196.89,
    "end_time": 198.49,
    "audio_path": "audio/S28-1-019689_019849.flac"
    # Custom fields are shown in details view
  },
  ...
]
```

Supports common formats (STM, CTM, ...)

### Run it on Your Own System!

Install

```
> pip install meeteval[cli]
```

CLI # Creates a standalone HTML file

```
> python -m meeteval.viz html -r ref.json -h hyp.json
```

Python / Jupyter Notebook

```
import meeteval
ref = meeteval.io.load('ref.json').groupby('session_id')['S28']
hyp = meeteval.io.load('hyp.json').groupby('session_id')['S28']
meeteval.viz.AlignmentVisualization(ref, hyp).dump('viz.html')
```

JupyterLite

Runs in the browser, no installation necessary!

