



## EU-BRIDGE Partners

Karlsruhe Institute of Technology  
(Germany), PEV

## Contacts

Karlsruhe Institute of Technology  
Adenauerring 2  
D-76131 Karlsruhe  
<http://isl.anthropomatik.kit.edu>  
Prof. Alex Waibel  
[Alexander.waibel@kit.edu](mailto:Alexander.waibel@kit.edu)

## Punctuation

### Description and Exploitable Knowledge

We add punctuation to text. This can be done either in an online mode which only uses a 4 word window and a simple language model to decide which if any punctuation mark to add to a stream of text or in an offline mode that uses 20 word context window and multiple models.

The online system is trained on text which has been double checked for correct punctuation. For the offline system, a phrase table trained on a large amount of text data as well as both maximum entropy and n-gram based language models are combined log linearly.

### Infrastructure

- Server – mediator – client setup
- The mediator receives transcription requests and audio from the client and forwards the audio to the corresponding ASR worker which then returns the transcription to the mediator.
- ASR workers run constantly on a server waiting for audio. They use up no CPU time when not receiving audio or when the audio only contains silence.
- As soon as a worker is selected by the mediator it starts to receive packets of audio data which it then decodes and returns text fragments to the mediator.
- These text fragments can then be combined into sentences by a separate segmentation/punctuation prediction component and then (if required) passed onto an MT worker.

### Application Sectors

anywhere we have text without punctuation, like:

- News (Sky and Euro)
- Webinars
- Lectures
- EPPS stuff
- Weather reports
- subtitles
- 2nd step in speech translation
- many many more

### Project Coordinator

Karlsruhe Institute of Technology  
Adenauerring 2  
D-76131 Karlsruhe  
<http://isl.anthropomatik.kit.edu>  
Prof. Alex Waibel  
[alexander.waibel@kit.edu](mailto:alexander.waibel@kit.edu)



The work leading to these results has received funding from the European Union under grant agreement n°287658



[www.eu-bridge.eu](http://www.eu-bridge.eu)

### EU-BRIDGE - the Project

EU-BRIDGE is a European Integrated Project that aims at developing automatic transcription and translation technology that will permit the development of innovative multimedia captioning and translation services of audio-visual documents between European and non-European languages.

### Technical Requirements

- Server with 8GB of RAM and 8 cores, 8 active systems can run in parallel on a modern server with 4 AMD Opterons (16 cores per CPU) as well as >>100 non-active but ready workers
- OS: Linux Ubuntu LTS 12.04 (Precise Pangolin) or similar
- Mediator connection: The workers can be accessed through the EU-BRIDGE Mediator service infrastructure

### Terms of Availability

Can be inquired at the Karlsruhe Institute of Technology (Prof. Alex Waibel)

### IPR Protection

Karlsruhe Institute of Technology (Prof. Alex Waibel)