EU-BRIDGE
Bridges Across the Language Divide

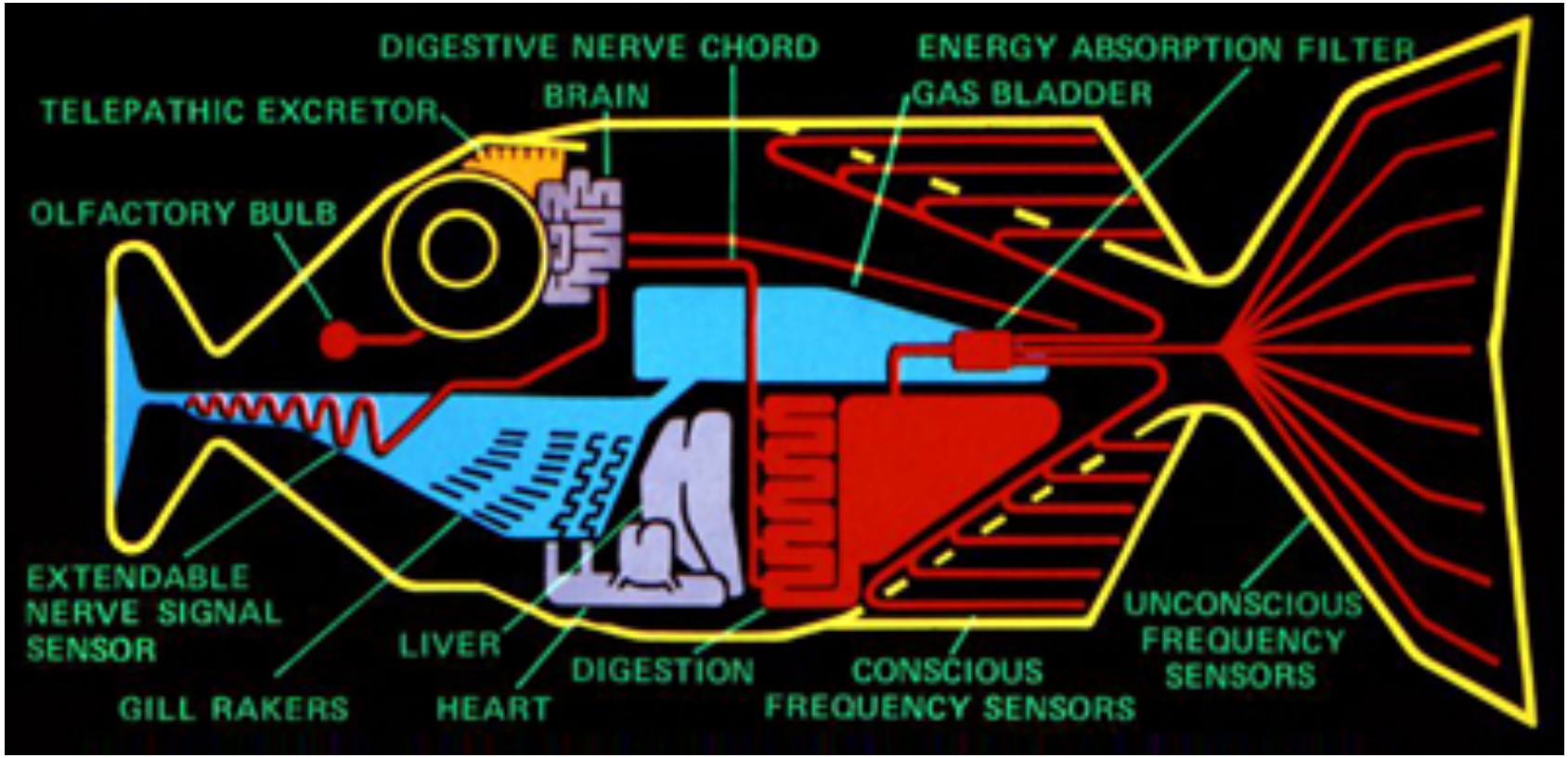
Sebastian Stüker, KIT

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


Languages in Europe
Icelandic
Faroese
Scottish Gaelic
Irish
English
Welsh
English
Britton
French
Basque
Spanish
Portuguese
Galician
Catalan
Dutch
Frisian
German
Luxembourgish
Danish
Swedish
Norwegian
Sami
Sami
Karelian
Finnish
Estonian
Latvian
Lithuanian
Russian
Russian
Belarusian
Polish
Ukrainian
Kazakh
Czech
Slovak
Hungarian
Romansh
Slovenian
Croatian
Bosnian
Serbian
Montenegrin
Albanian
Macedonian
Italian
Romanian
Gagauz
Bulgarian
Greek
Turkish
Turkish
Caucasian and Altaic languages
Ossetic
Georgian
Azeri
Armenian

Not this project, + different architecture

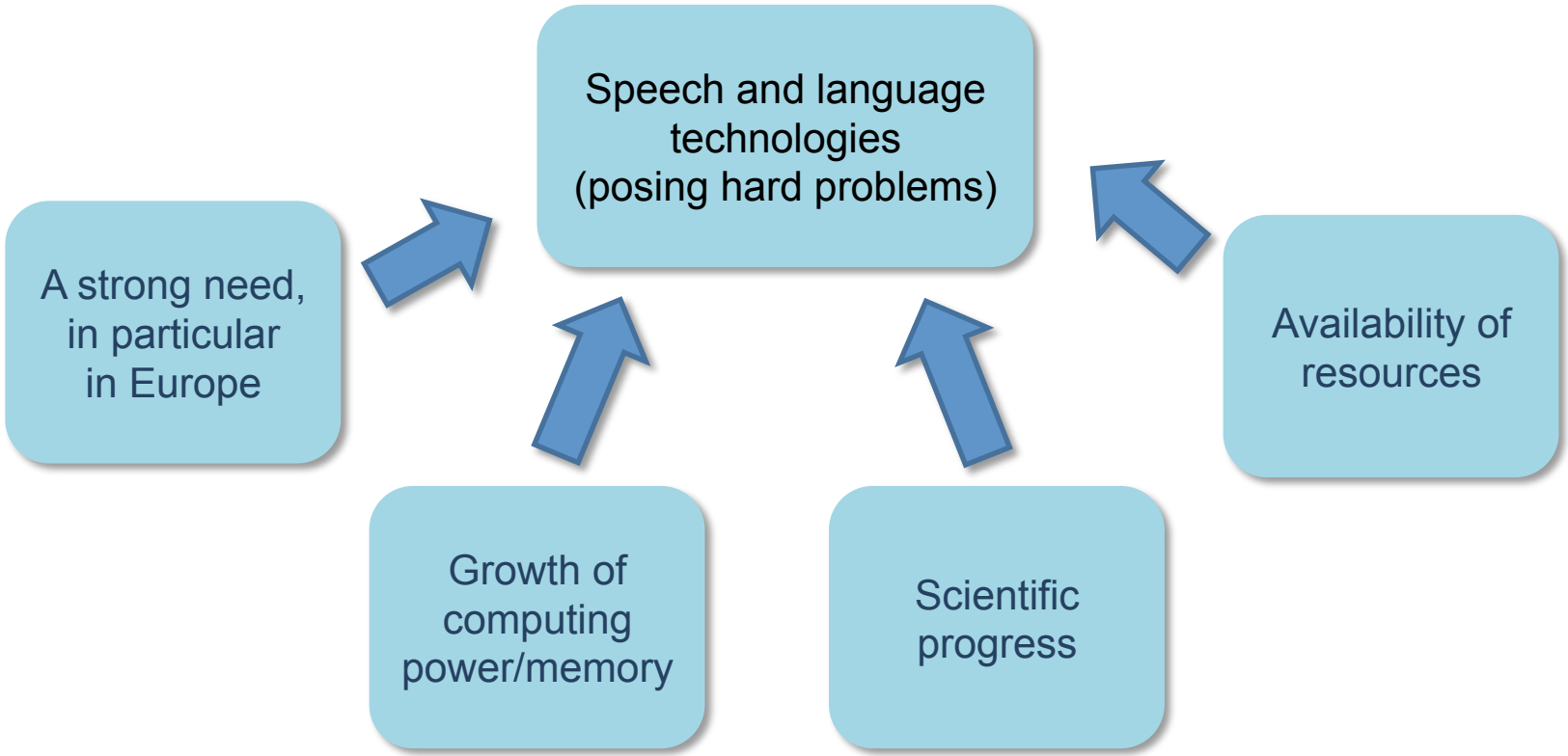
not EU★BRIDGE



BABEL FISH

from "The hitch-hiker's Guide to the Galaxy" (TV series)
1981 © BBC
orig. animation artwork by Rod Lord

It‘s Happening Right Now



The Vision

- Bridges Between Languages in Europe:
  - Technology to bridge the European language
  - Building/maintaining a sustainable European language infrastructure
  - Language divide is expressed in language and speech
  - Easily accessible services
- Bridges Between Science and Society
  - Language technology requires continued scientific attention
  - Exploitation and insertion requires suitable adaptation
  - Metrics:
    - WER, BLEU, TER, F-Scores
    - User Friendliness, Productivity, Sales, Distribution Channels, Customer Support
  - Identify use cases and applications
  - Effective transition and insertion

EU-BRIDGE in a Nutshell

- Here: Not science fiction, but results coming soon (2014/2015)
- FP7 IP EU-BRIDGE: Bridges Across the Language Divide
- Development of a speech translation service infrastructure
- Targeted to make progress in particular regarding market insertion
- Project footprint:
  - Feb 2012 - Jan 2015
  - Budget € 10.5m, EC funding € 7.8m
  - 10 partners
  - 1 service infrastructure, 4 use cases

Goals and Project Plan

Four use cases

1. Captioning and translation of subtitles for TV programs
2. Simultaneous translation of academic lectures
3. Speech translation services for the European Parliament
4. Translation of webinars

Four major objectives

1. Performance
2. Language portability
3. Reduction of dependency on data
4. Rapid technology transition and market insertion

EU-BRIDGE Partners

EU-BRIDGE Partners

EU-BRIDGE Partners

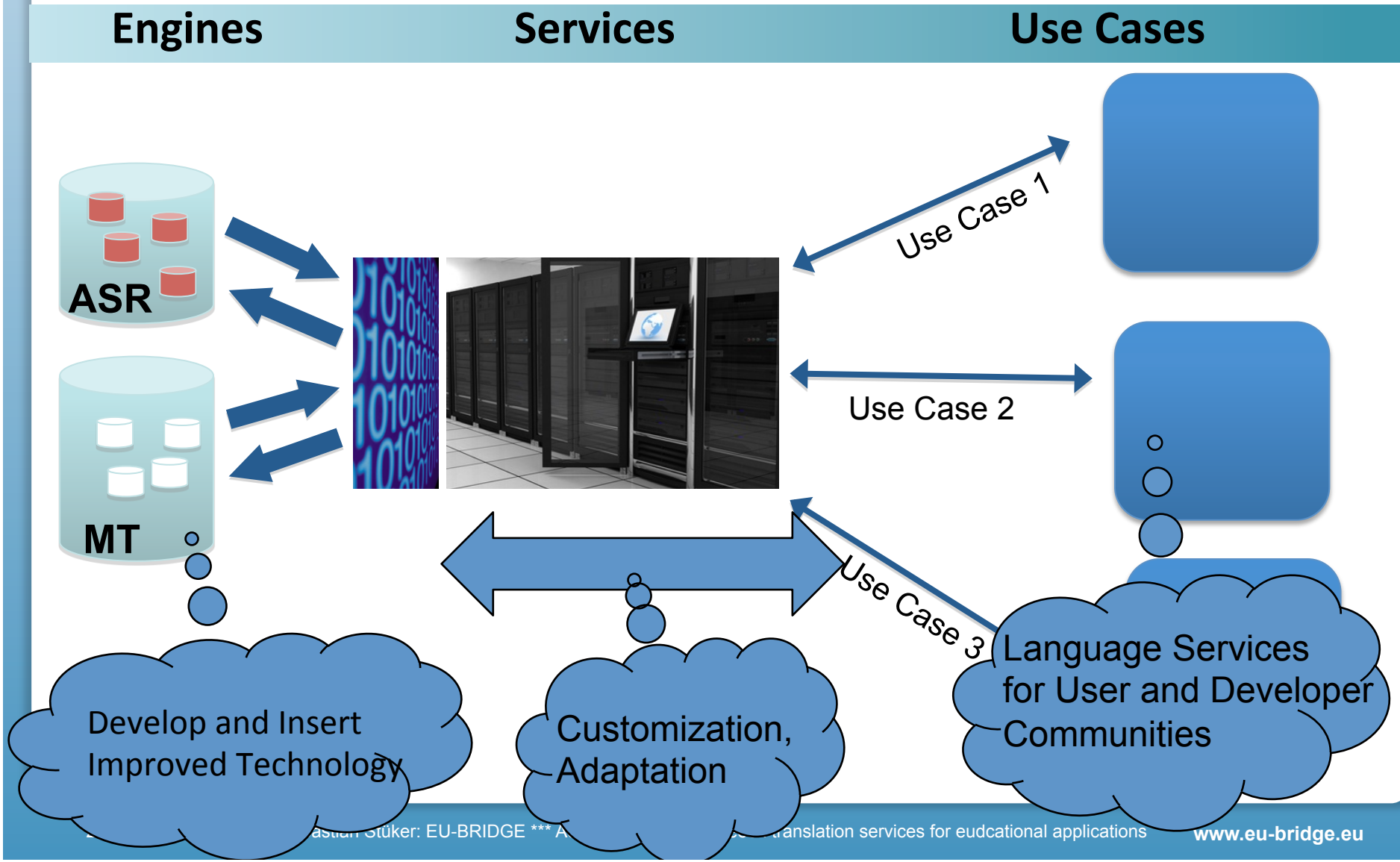
EU★BRIDGE

EU-BRIDGE Partners

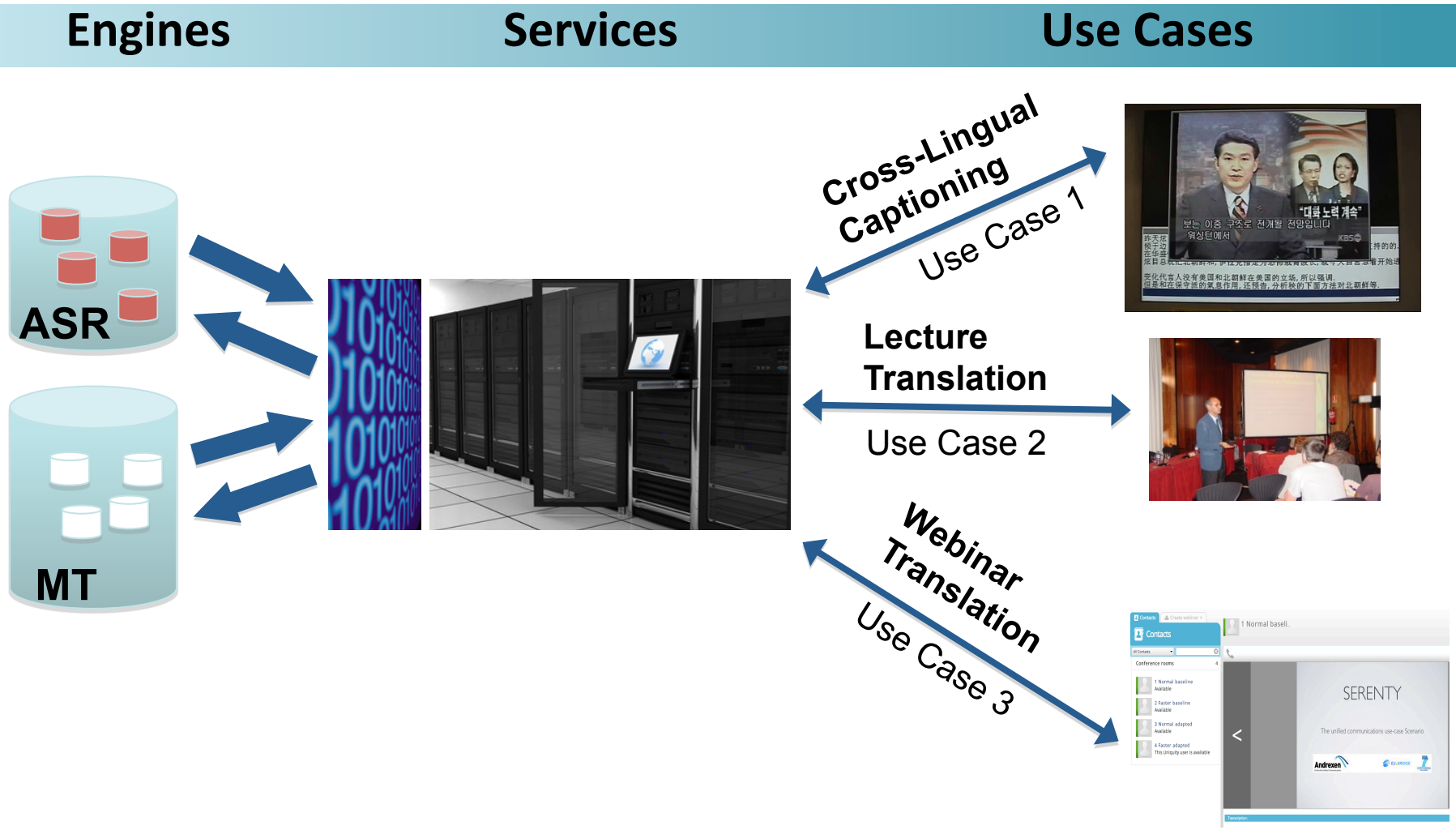
EU-BRIDGE Partners

ACCIPIO PROJECTS

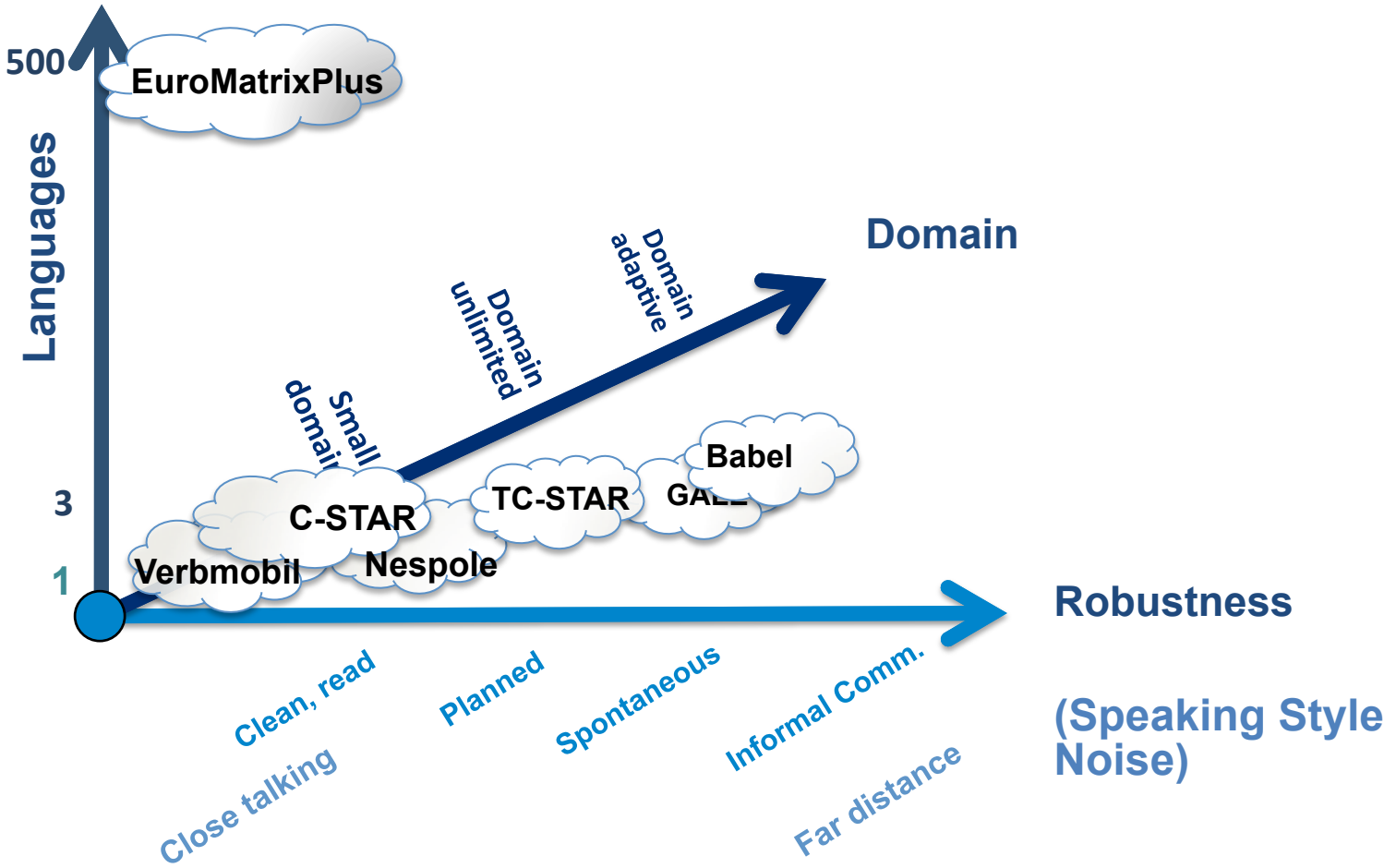
Language Service



Language Service



Technical Challenges



Project Organization

Laboratory

Research	**WP 1**	**WP2**	**Standarized, inhouse tests:** WER, BLEU
build & improve	Transcription Translation	Data Dependency Data Costs	
Engines	**WP 3**		**External campaigns**
run and schedule	Deployable Engines		
Services	**WP 4**		
utilize	Transcription and Translation Services		
Products	**WP 5**		**Field testing:** dev costs cost reduction, usability
	Caption Transl. RBM	Lecture Transl. EUP/RBM/KIT	Webinar ADX

WP6: Evaluation

Market — Market Insertion

WP 3
Engines

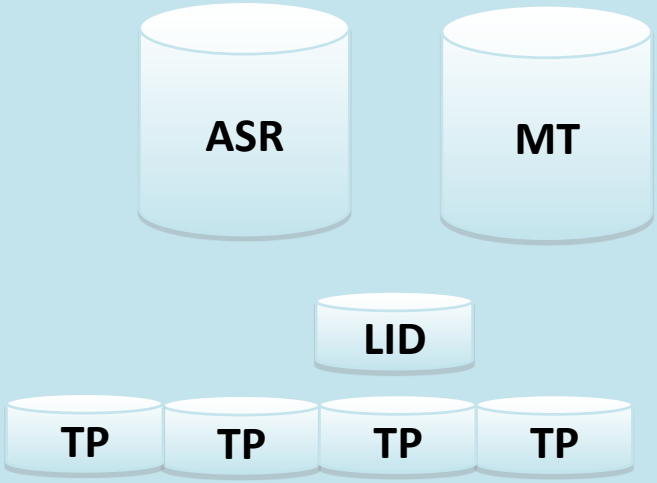
Task 3.1 ASR and MT

Task 3.2 Multithreaded+Stream Decoding

Task 3.3 Language Dependent Linguistic Processing

Task 3.4 LID

Task 3.5 Text Processing

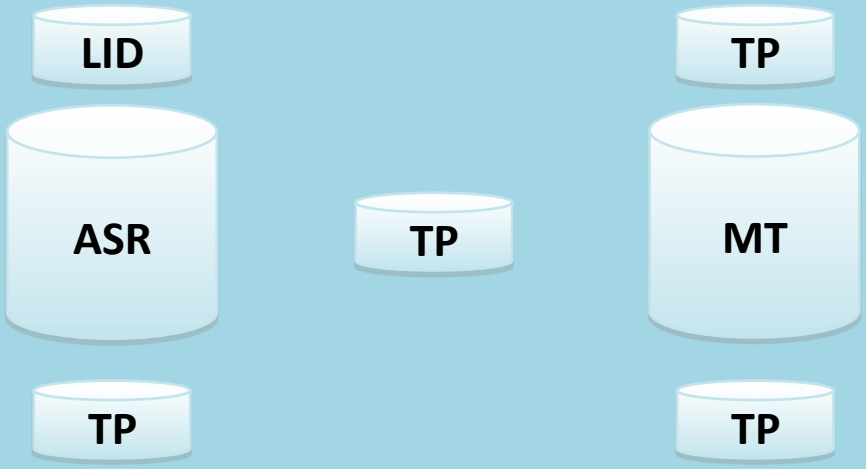


WP 4
Services

Task 4.1 Core Services: ASR, MT, Text Processing
Design and implementation of core service APIs and integration of core engines

WP 4
Services

Task 4.1 Core Services: ASR, MT, Text Processing
Design and implementation of core service APIs and integration of core engines



Task 4.2 Running Services in the cloud
Design and implementation of client-server architecture + APIs

WP 5
Products

Task 5.1
Caption Translation

Task 5.2
Lecture Translation

Task 5.3
Multilingual Tele- and Video Conferencing

Evaluations

“You Improve what you Evaluate”

- Evaluations organized around Use Cases
- Align Metrics with Product/Service Goals

Co-opetition

- Partners engage in friendly competition
- Goal is progress/complementarity, not site-to-site “horse-race”
- Partners participate with standard or optimized/tuned systems

Evaluations

- Standard open benchmarks
  - Calibrate technology internationally in open competition
  - IWSL, WMT: cover lecture task and multilingual MT
  - EU-Bridge partners are (co-) organizers of these evals and can thus influence the process to suit EU-BRIDGE's needs
  - Don't need to create/market a new campaign
- Internal EU-Bridge Tasks
  - Need to evaluate & optimize technology for EU-BRIDGE use cases
  - Different goals/sub-goals pursued (e.g. captioning, NE, shortening, ..)
  - Evals/Assessment need to change frequently during the project
  - Comparisons with outside teams not needed and not helpful
- Field Tests and Measures:
  - Evaluate on Living online 'Organism'
  - Optimize for extrinsic measures, not intrinsic ones

IWSLT 2014

- Evaluation of Speech Translation Technology on Talks and Lectures
- Working on TED data, because of efficiency of creating training and evaluation data

Use Cases

Captioning and Translation of Multimedia Content

- BBC Weather Data
- Euronews
- Skynews

Multilingual Lectures, Meetings

- TED Lectures
- University Lectures
- European Parliament Voting Sessions

European Parliament Interpreter Support

- Terminology Support: Tool Field Tests
- Named Entities: Eval, Tool Integration & Tests

Webinar Translation

- Integration into the Andrexen platform

Use Cases

Captioning and Translation of Multimedia Content

- BBC Weather Data
- Euronews
- Skynews

Multilingual Lectures, Meetings

- TED Lectures
- University Lectures
- European Parliament Voting Sessions

European Parliament Interpreter Support

- Terminology Support: Tool Field Tests
- Named Entities: Eval, Tool Integration & Tests

Webinar Translation

- Integration into the Andrexen platform

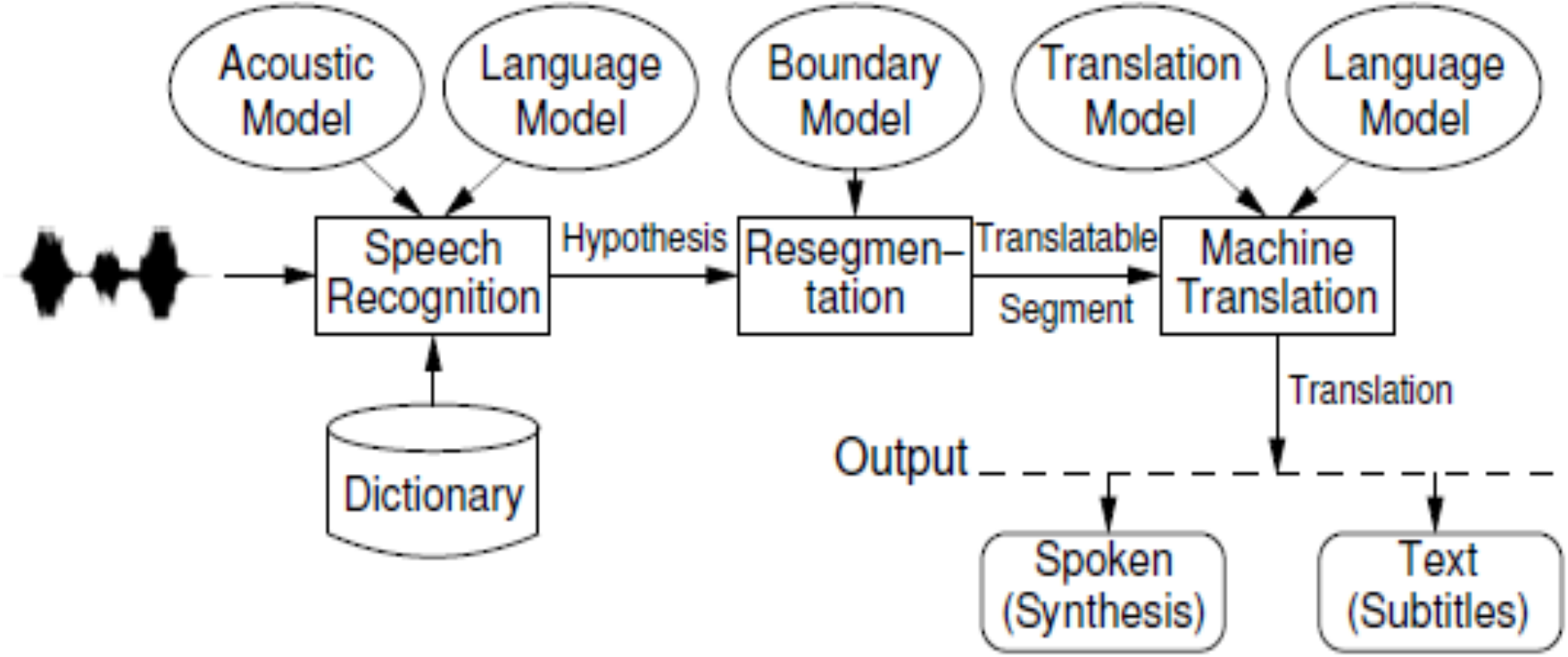
Lecture Translation

- Continuous Monologue
- Speaking-Style
  - Planned, not read, not memorized
  - Fast, spontaneous, fragmentary, and no punctuation!
  - Noises, coughing, non-verbal events (e.g., singing)
- Vocabulary
  - Very large
  - Specialized terms
  - Foreign Words
- Speed, Rea-ltime
- Service-Infrastructure
  - Many parallel lectures;
  - Automatic, robust assignment of compute power

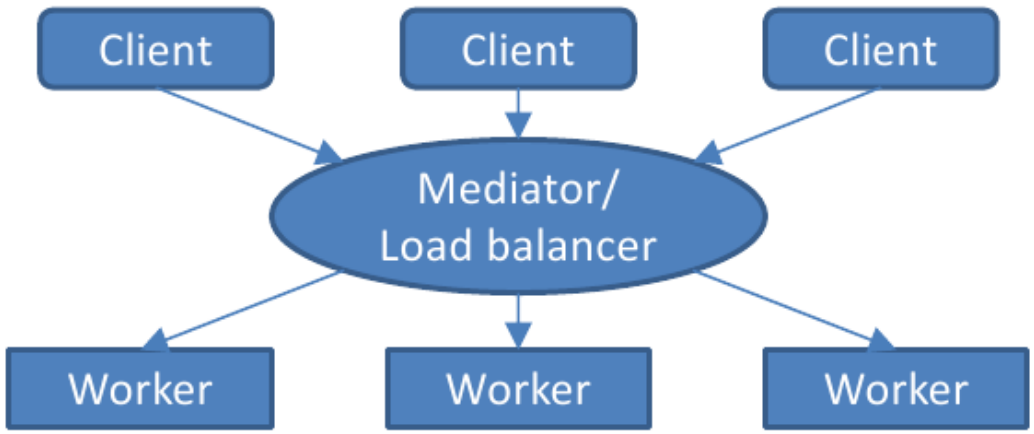
Lectures

Speech-Translation Systems

- Human Interpreters are too expensive
- Automatic Speech Translation an affordable solution:
  - Still lots of errors, room for improvement
  - But, better than nothing



Service Infrastructure



Service request

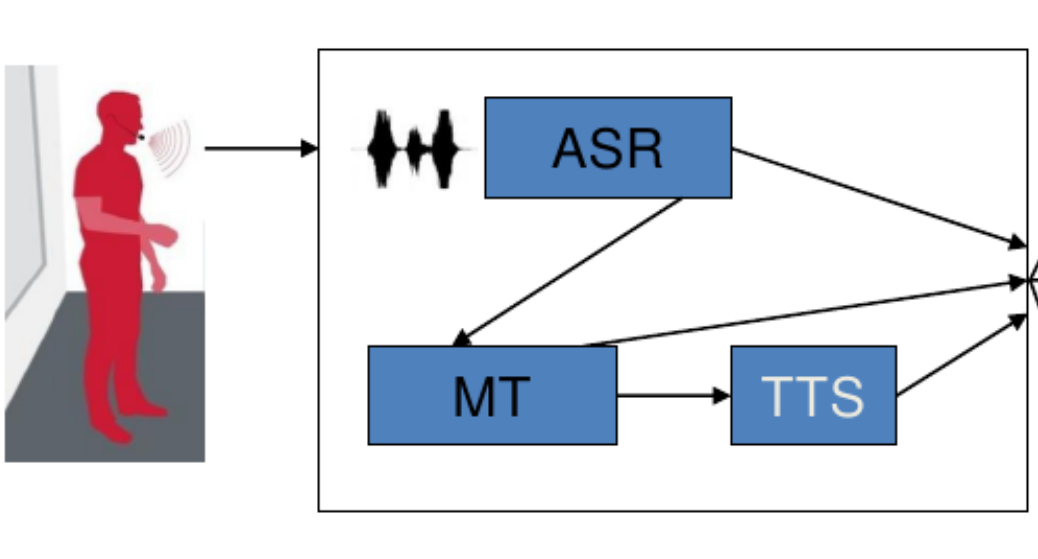
Distribution

Process request

Receive results

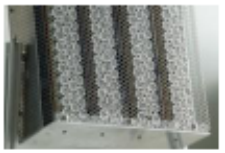
Combination/ Re-distribution

Return results



Mobile Devices

Web Browsers



Loudspeakers

Launch, June 11 2012

Mehr als nur Bahnhof verstehen - Weltweit erster Vorlesungsübersetzer

Karlsruhe. **In einer deutschen Vorlesung sitzen und wegen der Sprachbarriere nur Bal verstehen - sowas könnte für ausländische Studenten bald Vergangenheit sein. Der w erste automatische Vorlesungsübersetzer ermöglicht Studierenden künftig, dem Vortr**

von Dozenten auf Englisch zu folgen - schriftlich übersetzt in Echtzeit. Am Mont

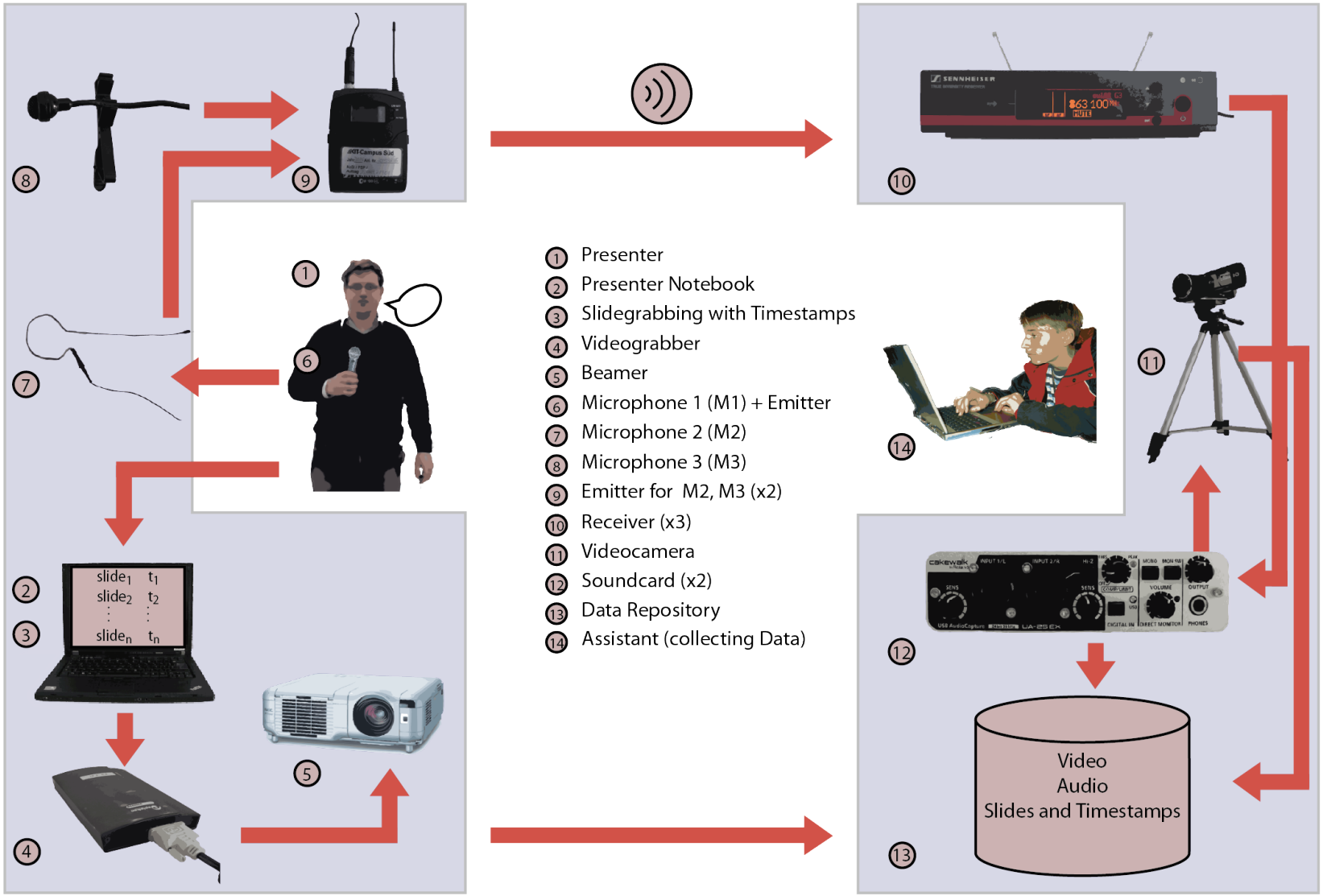
Deployment

Data Resources

KIT lecture corpus:

- Collected to fit the needs for training the systems:
  - ASR AM: Large amounts of in-domain audio data with careful transcriptions
  - MT Translation Model: Parallel sentences of in-domain data for the translation
  - ASR+MT LM: Large amounts of monolingual sentences in the required domain
  - Any kind of meta data that might help (e.g., lecturers' slides)
- Data collection took place at KIT's lecture halls:
  - Started with computer science lectures
  - Gradually spread to lectures from all departments at the university

Data Resources



1. Presenter
2. Presenter Notebook
3. Slidegrabbing with Timestamps
4. Videograbber
5. Beamer
6. Microphone 1 (M1) + Emitter
7. Microphone 2 (M2)
8. Microphone 3 (M3)
9. Emitter for M2, M3 (x2)
10. Receiver (x3)
11. Videocamera
12. Soundcard (x2)
13. Data Repository
14. Assistant (collecting Data)

Automatic Speech Recognition

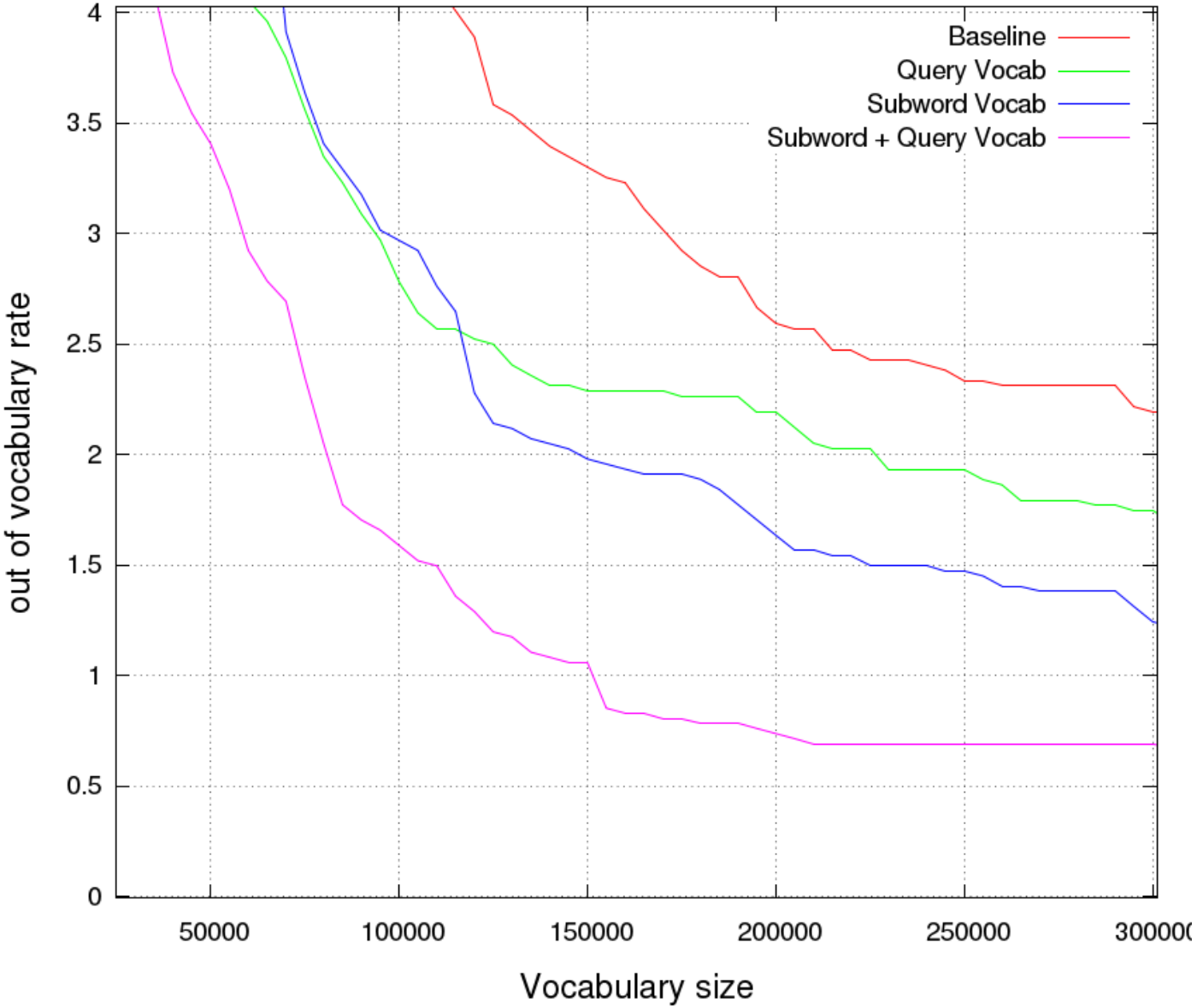
- Janus based ASR system: HMM/GMM based quinphone system with 4,000 models, MVDR front-end, 4gram LM
- Acoustic Model:
  - Trained on all lecture data in order to get a speaker independent model
  - Created 5 speaker adapted AMs: speaker independent model + Viterbi training and bMMIE training on the speaker dependent data
- Language Model:
  - Interpolation from 28 text corpora
  - Interpol. weights tuned on random selection from AM training data
  - 300k vocab selected by ML count estimation method
- German has a lot of compounds:
  - Sub-word vocabulary for compounds
- Vocabulary adaptation by deriving queries from slides (OOV 2.25%➔0.75% at 300k vocab size)

ASR Performance

- Tested the ASR system on the dev set of six lecturers

Lecturer	1	2	3	4	5	6
Speaker Independent AM	34.8%	21.1%	28.4%	22.9%	22.7%	19.0%
Speaker Dependent AM	–	18.9%	27.6%	22.6%	21.5%	17.8%
Adapted LM+Vocabulary	23.9%	17.3%	–	18.1%	–	15.4%

OOV Reduction



ASR Post-Processing

- Numbers are converted to digit sequences
- Common symbols are substituted, e.g., “Prozent”➔’%’
- Punctuation, i.e. ‘.’ and ‘,’ is inserted:
  - Prediction via a 4gram model
  - Pause information used to adjust the priors
- Simple, rule based conversion of equations:
  - “F of x”➔ ‘f(x)’
- Punctuation is used to structure text into sentences and chunks for translation

Machine Translation

Statistical Phrase Based MT System

- Trained on different out-of-domain data and the lecture corpus
- Applied the same compound word splitting as for the ASR system for consistency; names etc. were excluded from splitting by applying a named entity tagger
- Used a discriminatively trained word alignment approach
- Specific models for short and long range reorderings (also on the training data)
- Online system:
  - Phrase table filtered with ASR vocabulary
  - Simplified POS tagging

Machine Translation

Adaptation to the lecture domain

- Domain independent translation model trained on all data
- In-Domain TM only on the lecture data (re-use alignments)
- Combined via log-linear combination
- For LM log-linear combination of large LM, in-domain LM and TED LM
- Translations or special terms learned from Wikipedia and Wiktionary

MT Results

Lecturer	BLEU
Lecturer 1	13.80
Lecturer 2	22.58
Lecturer 3	14.24
Lecturer 4	20.83
Lecturer 5	24.50
Lecturer 6	24.13

Thoughts on the Output Modality

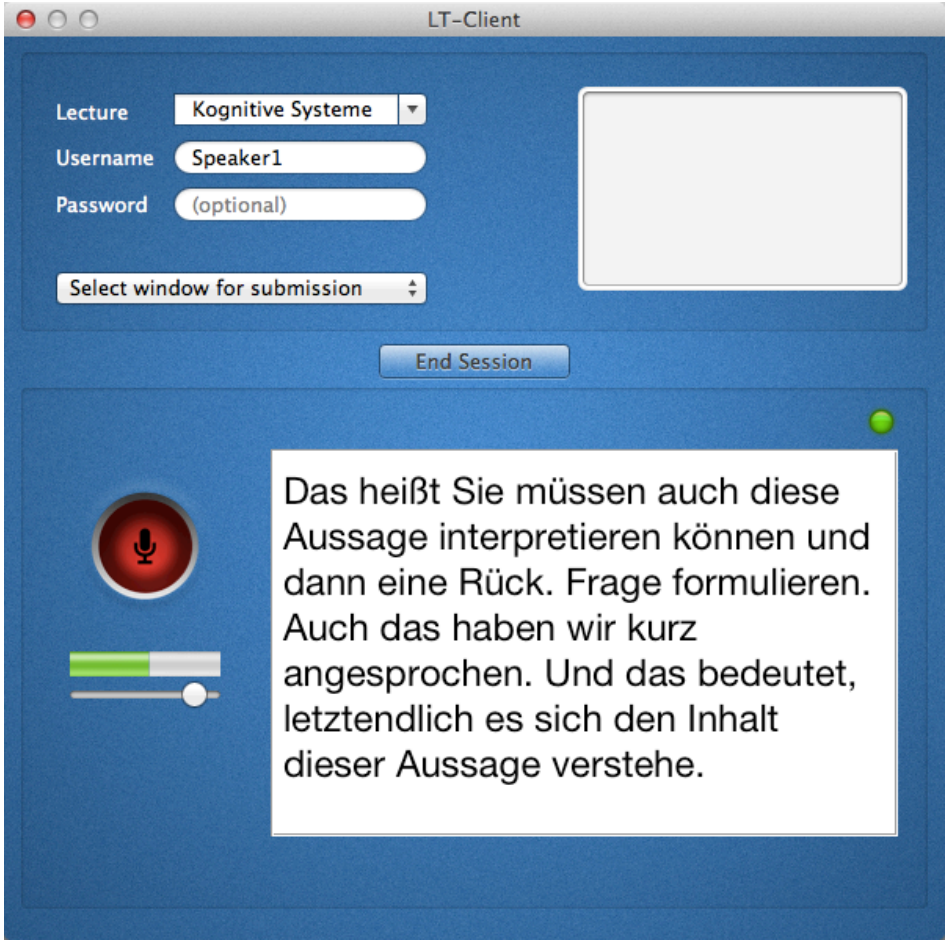
Text instead of synthesized speech

- Text can be easily distributed over the WWW
  - Laptops, smartphones, tablets
  - Nowadays ubiquitous
  - No proprietary software, just a browser
- Listening to synthesized speech can be very tiresome
  - Artificial voices not perfect
  - Original speech present in addition
- Translation system commits errors
  - Translated text contains errors; synthesis quality suffers from that
- Temporal Navigation
  - Once translation has been heard, it is gone
  - Text enables users to move back and forth in the translation; supports understanding the translation

Interface

Client for the lecturer:

- Must be as simple as possible!
- Client needs to know:
  - Who is speaking?
  - What lecture is it?
- Lecturer needs to know:
  - Is it turned on?
  - Is it doing something?



Interface

Displaying the results:

- Translation result + ASR result (?!)
- Scrolling back and forth in time
- Make text readable !!!

Where do we go?

- Interfaces need to get simpler!
- Systems need to become an omnivore:
  - Get all meta information as early as possible (before the lecture!)
  - Slides, papers, web sites, text books, etc.
  - Find a way for obtaining comparable corpora
- Make system self maintaining and autonomous
  - Automatic unsupervised acoustic model adaptation after every lecture
  - Automatically detect speaker, lecture and language: Access the university information system
- Offer additional services
  - Archive of the lectures for the students (for search)
  - Translation of the slides
  - Summary of the lecture
- Get the students in the loop
  - Automatic corrections by the students: during the lecture and afterwards
  - Make it a game

Andrexen - Webinar

- Role definition (Speaker / Listener)
  - Functionality was defined and implemented, and relevant MCloud services integrated
- Data flow (Architecture / Slideshow / Voice / Text)
  - Features implemented to support the parallel transmission of slides, speaker audio, translation text and intelligent user location to provide integrated UC experience

Andrexen - Webinar

Adapting to industry requirements:

- Real time streaming
- Support legacy systems:
  - Training narrow bandwidth (8KHz) acoustic model
- Specific vocabulary:
  - Vocabulary and language model adaptation

Andrexen - Webinar

- Translator integrated as virtual participant
- Is “chatting” the translation results