Karlsruher Institut für Technologie

Lecture Translator

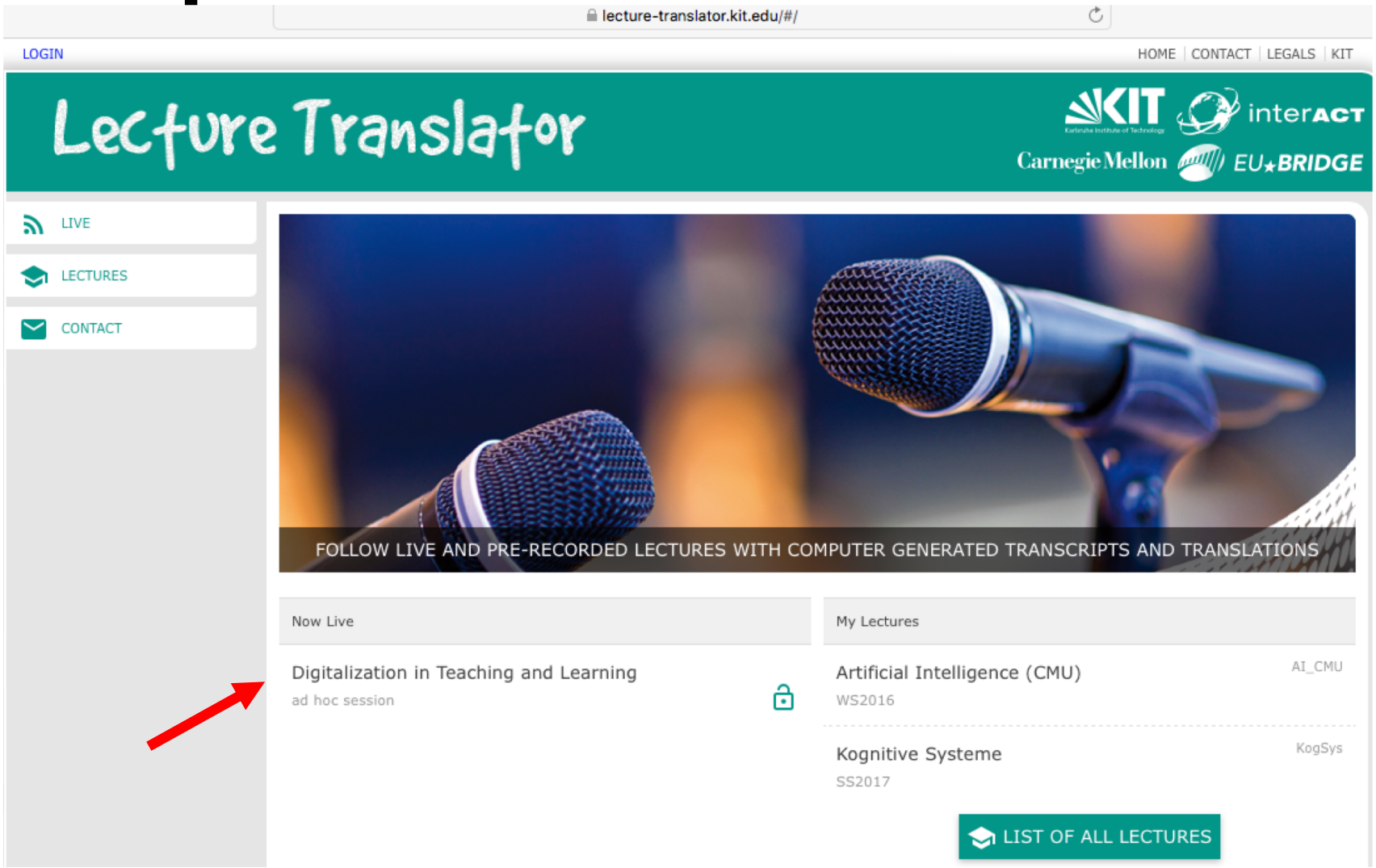
A System for the Automatic Simultaneous Translation of Lectures

Sebastian Stüker

Institut für Anthropomatik

Using the System

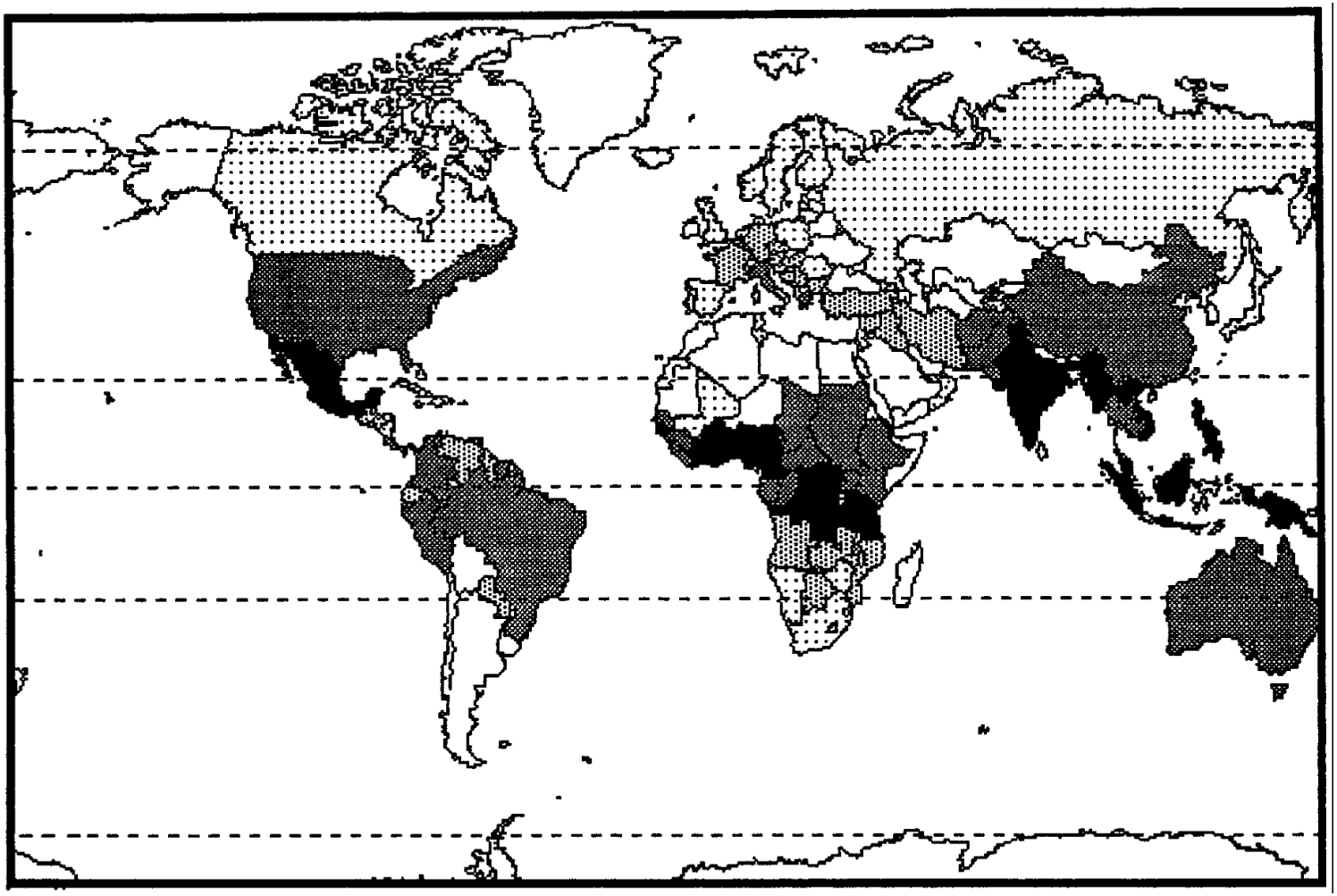
https://lecture-translator.kit.edu



Language Diversity

- 5.000 to 7.000 languages exist in the world today
- Languages are frequently dying
- ASR+MT Systems exist only for a fraction of the languages in the world
- Languages Selected: Rich, populous, politically relevant
- Training of ASR+MT systems requires large amounts of transcribed audio data
- ASR+MT now perform well enough for use in real-life applications
- Danger of creating a digital divide between languages with ASR systems and those without one

Language Diversity



[courtesy of Academic Press, Daniel Nettle. Explaining global patterns of language diversity 1998]

Importance of Language Diversity

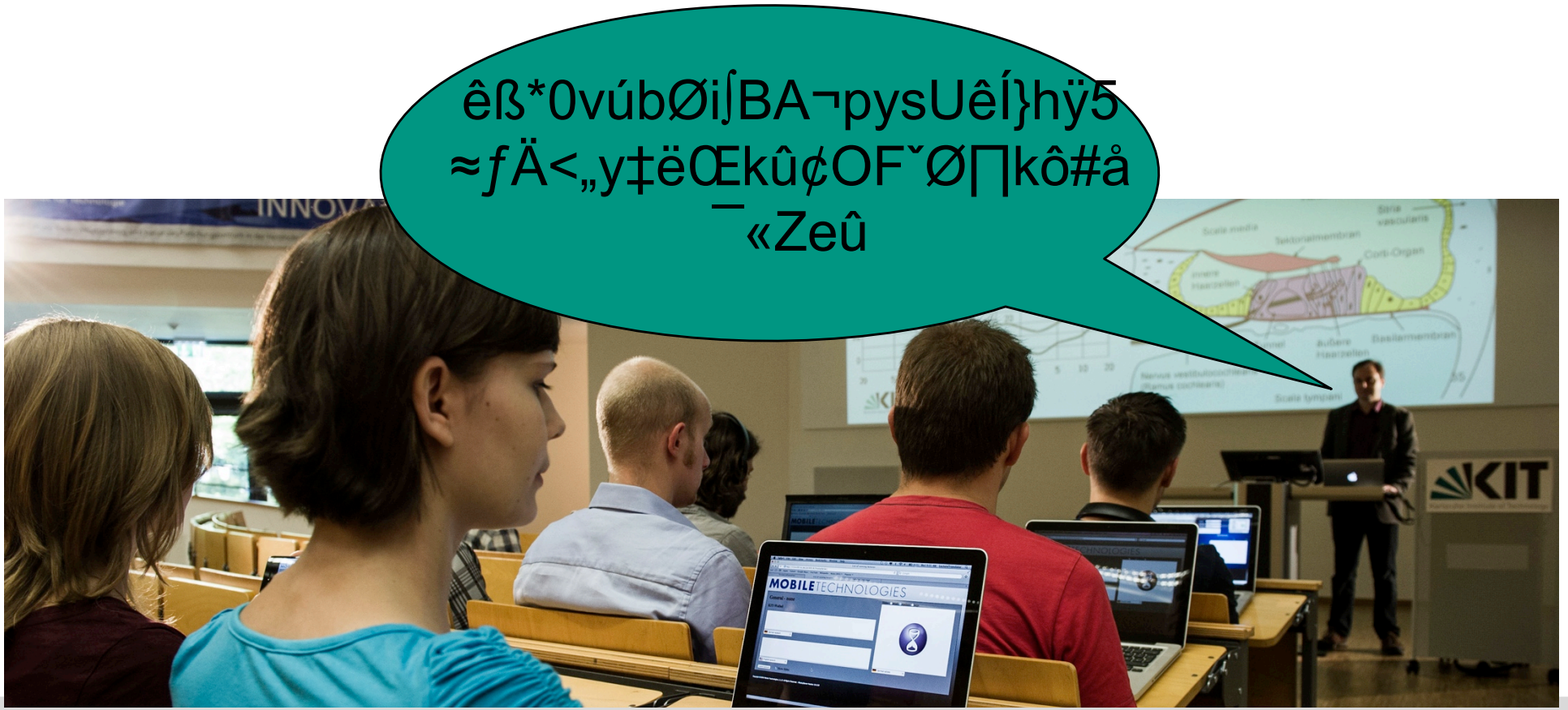
- “Language is the pinnacle achievement of man-kind” (Vivianne Redding, European Commission)
- Language is inherently linked with culture
- The way we speak / the way our language is structured influences the way we think (e.g., see research on cognation by Lera Boroditsky at UCLA)
- Language Diversity as important for human well-being and prosperity, as biodiversity in nature
- Languages are dying at an alarming rate (up to 90% of all languages extinct within a few generations?)
- Technology can make a difference

Lectures and the Language Divide

- Academic lectures and talks often do not reach full potential audience due to the language barrier
- Speech translation technology can provide a solution
- Academic lectures happen daily at our universities
  - Acoustic etc. conditions are good enough so that current technology is in reach of addressing them
  - Speech-to-Speech translation systems need to autonomously adapt themselves
  - Speech-to-Speech Translation Technology in its current form needs to be more than the bare sequence of words in the target language

Foreign Students at KIT

Karlsruher Institut für Technologie



Foreign Students at KIT

- Foreign Students
  - International competition for talented researchers
  - 15,7% at KIT in 2011 (45% from Europe, 35% from Asia)
  - 3% masters students
  - 49% at CMU
- How shall a German university react?
  - Teach in English?
  - Students need to learn German?

Everybody learns German?

- KIT has large efforts in teaching German
  - Excellent courses in languages, especially for foreign students
- But problems persist
  - Enrollment only with initial knowledge of German
  - 1-2 years of further language studies
  - In computer science English is of high importance as well.


Studentin aus Tunesien

Simultaneous Translation by Machines: a Solution

- Human interpreters are very expensive
  - European Parliament: 300 million € per annum
  - EU: 1.3 billion € per annum
  - Too expensive for universities

- Machine translation can fill a gap:
  - Important material will always be translated by humans
  - Where humans would never be employed, the machine can take over the task
    - The machine makes significantly more errors than humans
    - But alternative would be to have no translation at all

Machine Translation

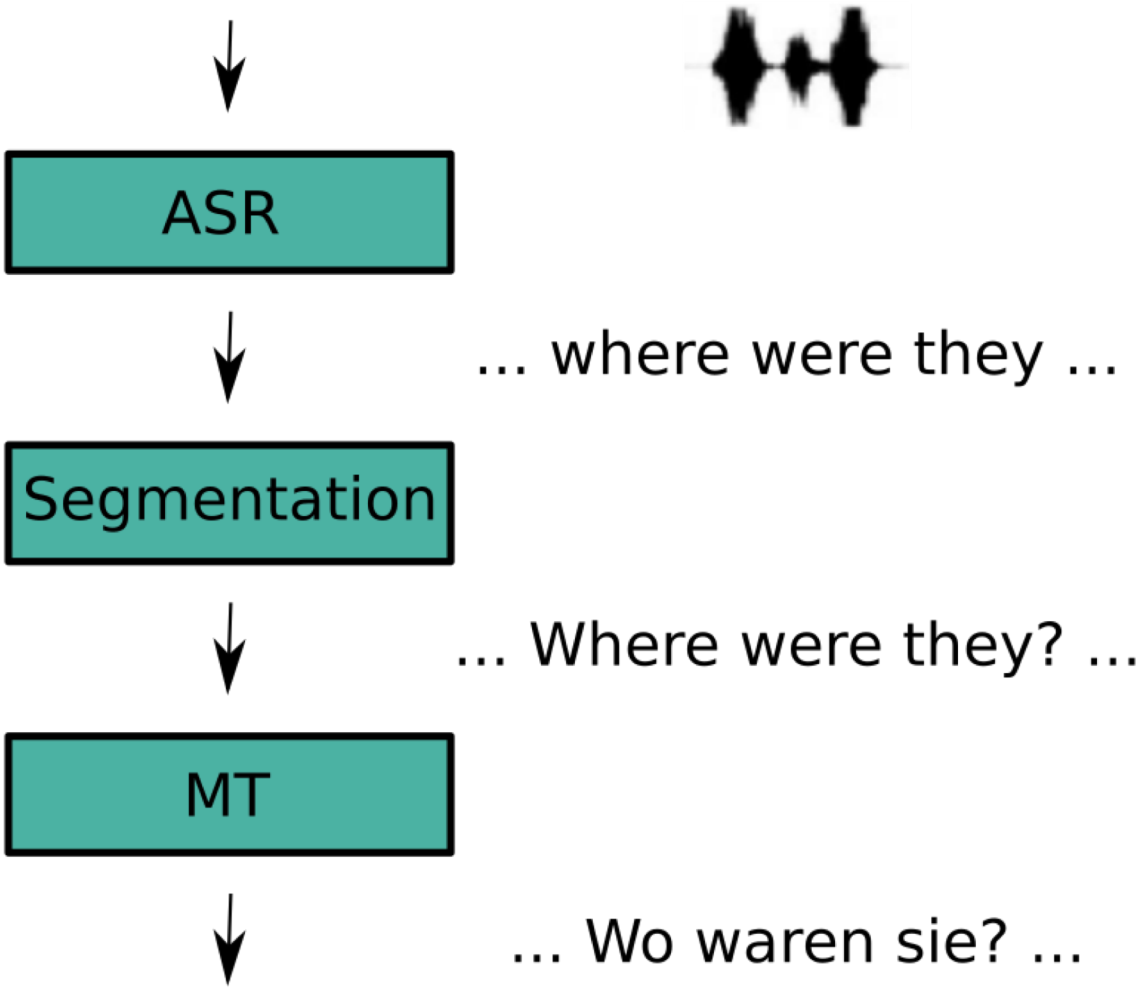
Speech-to-Speech Translation

- Combines automatic speech recognition, machine translation and speech synthesis
- Application areas: lectures, talks, videos, meetings, conversations.
- Simultaneous, turn-based, offline
- Needs to present translation result in an appropriate manner

Sample Applications:

- Simultaneous lecture translation
- Offline translations of speeches in the European Parliament
- Portable translation devices, for travellers, doctors etc.

Components


ASR
... where were they ...
Segmentation
... Where were they? ...
MT
... Wo waren sie? ...


Let's eat grandma!
Let's eat, grandma!
PUNCTUATION
SAVES LIVES!

Simultaneous Lecture Translation at KIT

History

2005: First Prototype

- English->Spanish talks/lectures
- Limited domain (technical talks in the area of natural language processing)
- Speaker dependent
- Prototype integration on one to two laptops (portable for demos)
- Experiments on the output modality
  - Targeted audio
  - Heads-up displays in glasses (Google Glasses is not as novel as advertised)
  - Special applications for displaying subtitles

Simultaneous Lecture Translation at KIT

2005-2007: Continuous Improvement of Base-Technologies:

- Automatic speech recognition
- Machine Translation
- Project TC-STAR: Speeches in the European Parliament

2007-2010: Preparation for Extension of Domains and Languages:

- German->English, French
- Initial data collection at the KIT computer science department
- Improvements in MT for German phenomena (word order)
- Improvements in ASR for German phenomena (Compositа, English/German mixed special terms)

Simultaneous Lecture Translation at KIT

Karlsruher Institut für Technologie

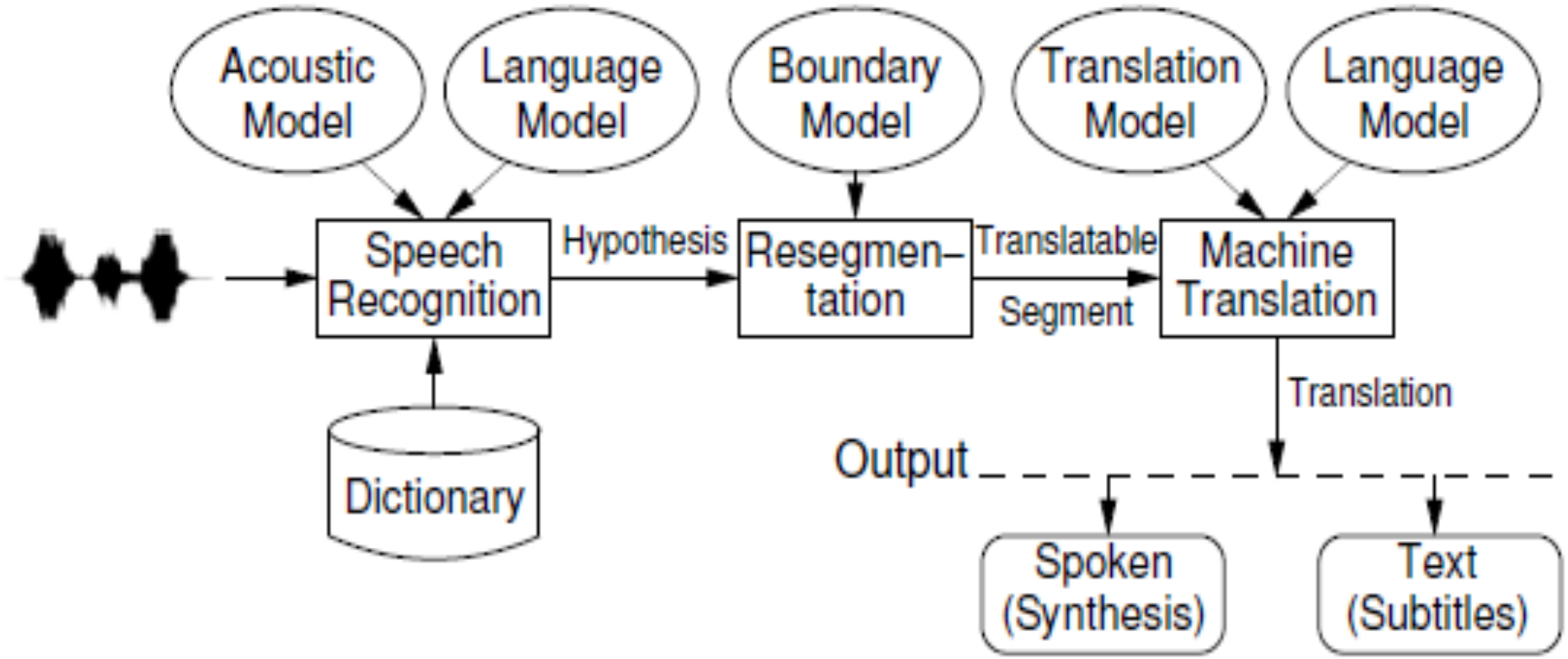
Since 2010: Funding through instruments from the Excellence Iniative: RG 3-01: Multilingual Speech Recognition

- Research in the area of automatic speech recognition for lecture translation
- Extended data collection efforts
- Research and experiments for unsupervised training and adaptation
- Research for integration of a multitude of information sources
- Established an international evaluation campaign for lecture translation (IWSLT)

Since 2012: Integrated in the FP7 Project EU-BRIDGE:

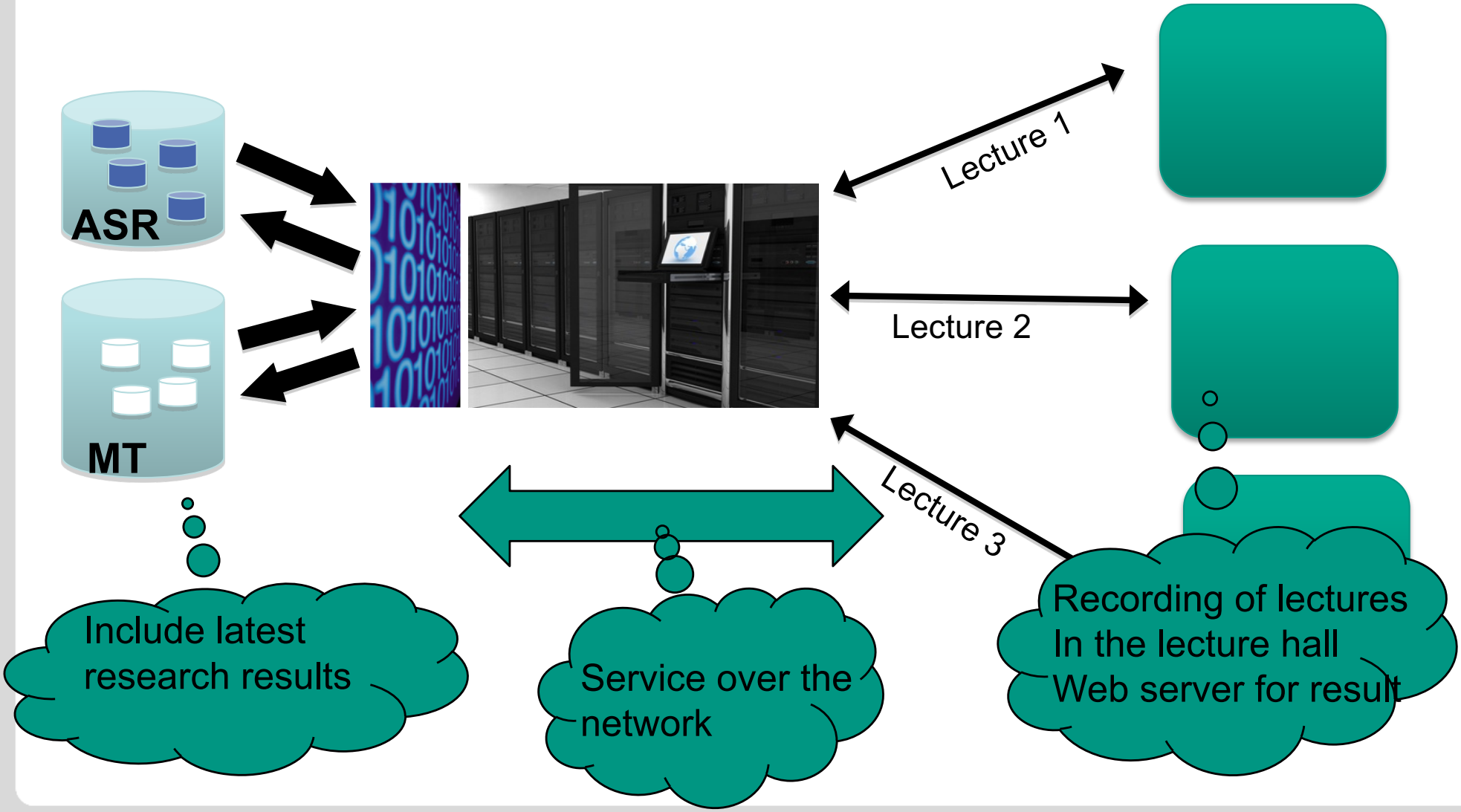
- Use of an advanced system architecture
- Additional funding in the area of unsupervised training and adaptation
- Transfer to additional use cases (TV broadcasts, webinars, etc.)

System Overview



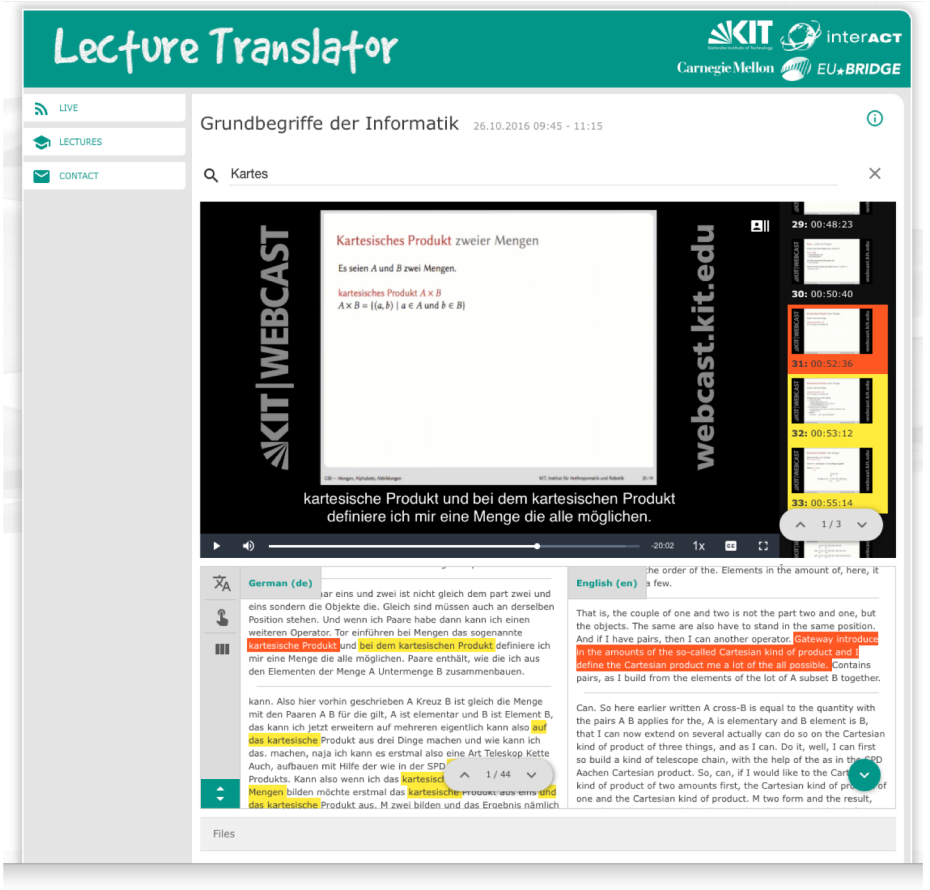
System Overview

Components Services Lecture



Archive

- Lecture streamed from the Digital Video Audio Archive (DIVA) of KIT
- Shows transcriptions and translations
- Searchable
  - Slides (OCR)
  - Transcriptions
  - Translations



Thoughts on the Output Modality

Text instead of Speech Synthesis

- Text can be easily distributed through the WWW
  - Laptops, smart phones, tablet PCs
  - Nowadays ubiquitous
  - No specialized software necessary, web browser is enough
- Following synthesized speech is tiresome
  - Artificial voices are imperfect
  - Original speech is still audible
- Systems commits errors
  - Translation contains errors, this leads to quality degradation fr speech synthesis
- Temporal Navigations
  - Once spoken, the translation is gone
  - Text makes skipping back and forth in time possible, supports understanding the lecture

Speech Translation is Difficult

- Speech is ambiguous and variable
- Homophones in speech recognition:
  - How much wood would a woodchuck chuck if a woodchuck would chuck wood?
  - The plane was plain white.
- Homonyms in translation:
  - The dog‘s bark gave the bark of the tree goose bumps.
  - Time flies like an arrow.
  - The down from the jacket fell down to the ground.
  - The bench at the river bank was donated by the local bank.

Speech Translation is Difficult

- Signal variations:
  - Speaker specifics: Accent, dialect, voice
  - Signal variations: Noise, microphone, room
  - Inherent variations
- Language Peculiarities
  - Word order:
    - Ich möchte mich zu der Konferenz über Maschinelle Übersetzung anmelden
      I want to register for the conference on Machine Translation
  - Compositа:
    - Worterkennungsfehlerrate
      Word Recognition Error Rate
  - Inflections and Congruence:
    - Zu **der** nächst**en** wichtig**en** interessant**en** Vorlesung

Specialized Vocabulary

- Technical terms normally not in the standard vocabulary
  - Mel frequency scaled cepstral coefficients
  - Roller bearings
  - Sub-space
- Technical Terms with Special Meanings
  - Exam: Prüfung (not “Untersuchung”)
  - Sign: Vorzeichen (not “Schild”)
- Equations:
  - Eff of lx: f(x)

Specialized Vocabulary

- Foreign words in German speech
  - Computer science lectures, e.g., contain many English terms
  - Official speeches often contain Latin phrases
- Pronunciation of foreign words in German difficult
  - “Cloud”, “iPhone”, “iPad”, “Laser”
- Inflections of these words (declination, conjugation)
  - Web-ge-casted, down-ge-loaded
- Composita:
  - Cloudbasierter Webcastzugriff

How are Speech Translation Systems Evaluated?

- Automatic vs. human metrics
- ASR: Word Error Rate
  - Single, unique result
  - Intuitive, objective and easy to calculate
- Machine Translation
  - No unique reference result
  - No simple comparison between reference and system output possible
- Automatic Metrics for MT
  - Use multiple references
  - Calculate number of occurrences of matching word sequences
  - Several problems (meaning of absolute value, comparison of different approaches)
- Human Evaluations
  - Bilingual humans judge quality of translations
  - Correctness of content, grammar, fluent and natural, etc.

Research Topics for the Lecture Translator

- Domain in theory limited: lectures
- De facto unlimited: philosophy, physics, chemistry, computer science, history, law, biology etc.
- Impossible to manually adapt the systems to all topics:
  - System has to work autonomously!
- Many lectures need to be processed in parallel
- System needs to be easy to use
- System must be easy to maintain
- Needs to work in many, heterogeneous lecture halls

Sources of Information for the System

- Titel of the Lecture
  - Reveals important key words
  - Can act as seed for searching further
- Lecture Slides:
  - Often electronically available beforehand
  - Can be used to adapt the vocabulary and language model
  - Can be used to tune the system on a fine time scale
- Speaker Identity:
  - Can be used to apply speaker specific models
  - Can be used to search for further information

Conclusion

- Automatic system for simultaneous translation of lectures
- Implemented as a service that scales to processing many lectures
- Automatic Translation Systems
  - Are they perfect? Match human performance? No.
  - But offers a service, where human interpreters would never be used
- Accesible University
  - Aid to foreign students taking classes at KIT
  - Bridge between cultures and languages

People


AlexWaibel
YuqiZhang
MarkusMüller
ChrisitanSaam
ChristianMohr
TeresaHerrmann
MohammedMediani
MatthiasSperber
ThiloKöhler
Thai-SonNguyen
KayRottmann
ChristianFügen
SebastianStüker
MichaelHeck
KevinKilgour
Thanh-LeHa
JanNiehues
EunahCho