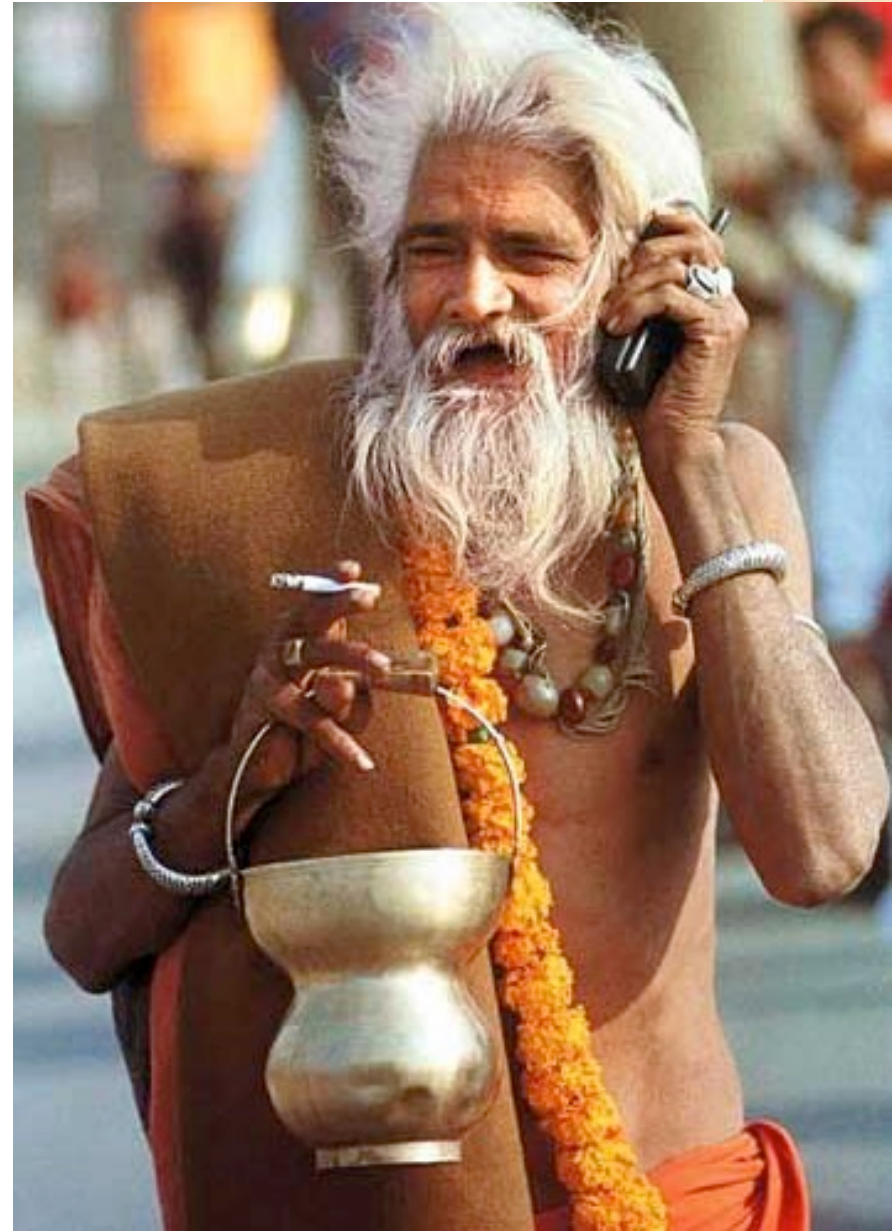


Technologies for Multilingual Human Communication

Alex Waibel and the InterACT Team
Carnegie Mellon University
Karlsruhe Institute of Technology

alex@waibel.com
waibel@cs.cmu.edu
waibel@kit.edu

- Living in the Global Village
 - **6 Billion Phones** (1 Billn PC's)
 - Unprecedented Exchange and Communication
 - Global/European Integration
- Cultural Diversity:
 - **6000 Languages**
 - Identity, Language, Culture,...
 - Pride and Individualism
- Challenge:
 - Access and Integration
 - $\leftrightarrow$ Diversity/Identity



1.) Communication is Multimodal

Communication is not only by Text, but:

- Speech, Gesture, Gaze..

- Images



- Ill-formed Text

“lol-jah I want hr to be like dat...”, Hppyyyyy BD, CU, LMK



2.) Communication is Multilingual



3.) Language is Ambiguous

„Give me a new display“

„Time flies like an arrow“

„The Spirit is willing but flesh is weak“

„Zentraleuropa“

„If the baby does not like the milk, boil it“

3.) Language is Ambiguous

„Give me a new display“ → „Give me a nudist play“

„Time flies like an arrow“ → 6 Parses

„The Spirit is willing but flesh is weak“ →

„The vodka is good but the meat is rotten“

„Zentraleuropa“ → „Headquarter Great Grandpa“

„If the baby does not like the milk, boil it“
→ the milk? the baby?

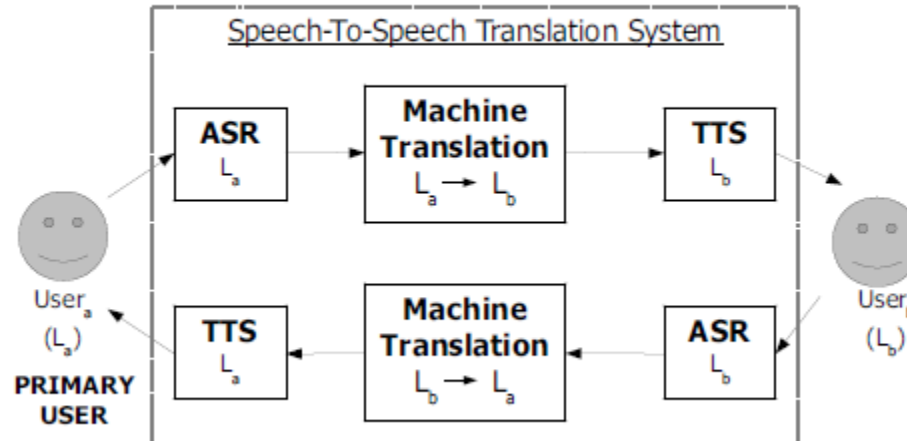


3.) Language is Ambiguous

Language Problems can only be Conquered,
if Machines Embrace, Represent, Process:

- **Ambiguity:**
Scores, Statistics, Neural Activations, ..
- **Machine Learning:**
Build Models, Extract Knowledge from
Human Data & Interaction, Automatically

→ Performance Depends on Data & Computing

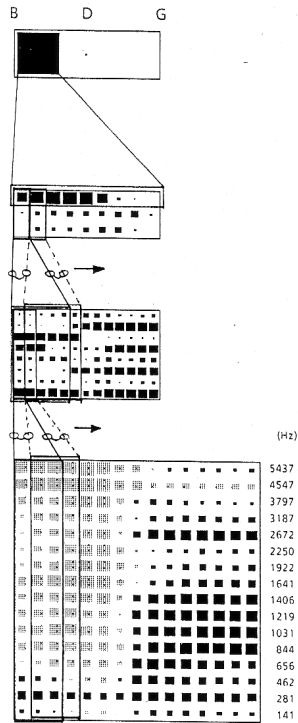


To Build a Language Communicator

- 6 Component-Engines: Automatic Speech Recognition, Machine Translation, and Text-to-Speech Synthesis
- Each is in Principle Language Independent, but Requires Language Dependent Models
- Models are Automatically Trained but Require Large Corpora
- Certain Language Dependent Challenges still Persist

Neural Nets 3.0: Bigger, Deeper, Faster

(1987)



TDNN: Shift-Invariance, Waibel '87

(1989)

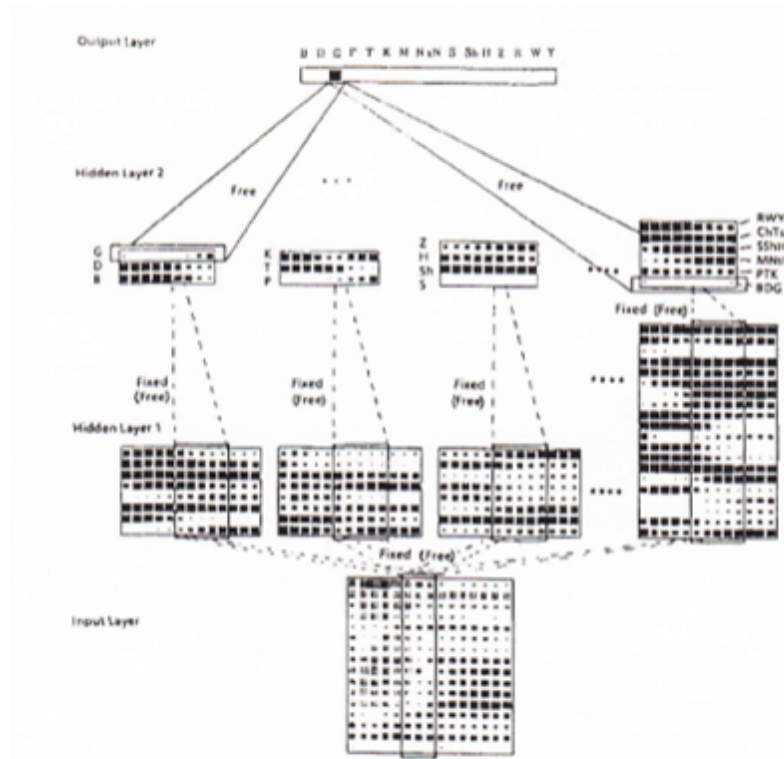
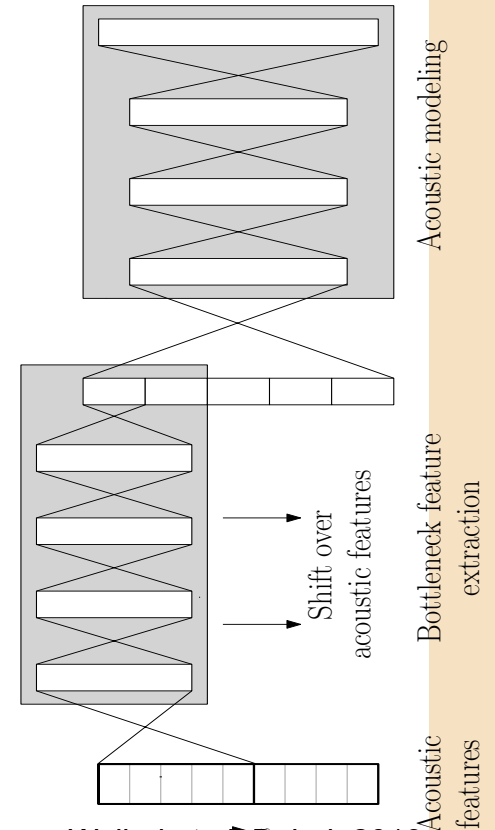


Fig. 7. Modular construction of an all consonant network.

Modular (deep) TDNN: Waibel '87

(2013)



Waibel et al. Babel, 2013

Weights: ~6,000

TrnData[hrs]: ~0.1

Time[weeks] ~1

~40,0000

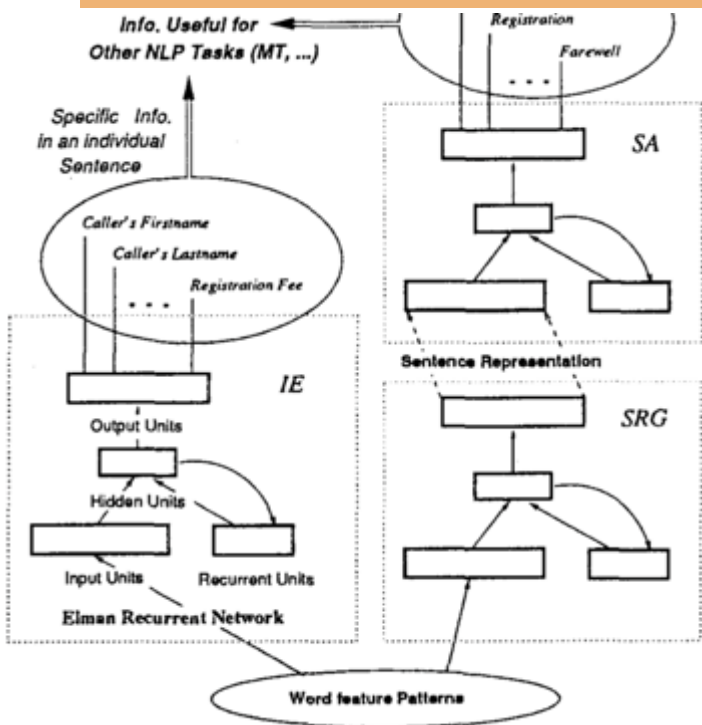
~1

~1

~33,000,000

~1,000

~1



- Incremental Parsing by Modular Recurrent Nets (Jain, Waibel '89-92)
- NN for Dialog Processing (Wang, Waibel '89-'91)
- Connectionist Transfer in MT (Wang, Waibel '95)

Translation Model:
Phrase Table e -> f

Translation Model:
Phrase Table f -> e

Translation Model:
Lexicon e -> f

Translation Model:
Lexicon f -> e

Language Model

POS LM

Distortion Model

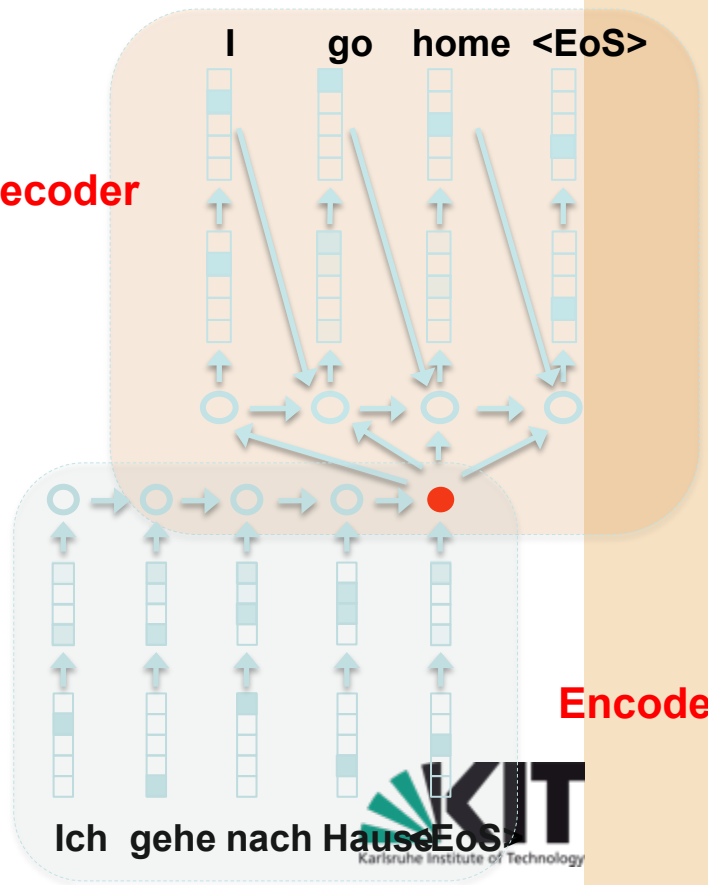
Word Count

Phrase Count

Cohesion Constraint

Global Search

Decoder



Encode



interACT

First Speech Translation VideoCall '91-92

- 1992 – C-STAR Consortium for Speech Translation Advanced Research
- 1993 – Public C-STAR Demo, ATR-CMU-UKA-Siemens



2009



JIBBIGO



Hello. Nice to meet you!
(Hello, nice to meet you.)



Hola, encantado de conocerle.



*Pocket Translators for Travelers
offline embedded on iPhone - Jibbigo*





Free App

⚙️ This app is designed for both iPhone and iPad

Category: Travel

Updated: Oct 12, 2012

Version: 2.1

Size: 15.0 MB

Languages: English, Arabic, Chinese, French, German, Italian, Japanese, Korean, Spanish, Thai

Seller: Jibbigo LLC

© 2012

Rated 4+

Requirements: Compatible with iPhone, iPod touch, and iPad. Requires iOS 5.0 or later. This app is optimized for iPhone 5.

Top In-App Purchases

- | | |
|-------------------------|--------|
| 1 Spanish-Englis... | \$4.99 |
| 2 European Bund... | \$9.99 |
| 3 Chinese-Englis... | \$4.99 |
| 4 French-English... | \$4.99 |
| 5 Asian Bundle fo... | \$9.99 |
| 6 Japanese-Engli... | \$4.99 |
| 7 Italian-English (...) | \$2.99 |
| 8 German-Englis... | \$4.99 |
| 9 Thai-English (I... | \$4.99 |
| 10 Filipino (Tagalo... | \$4.99 |

Jibbigo Translator

Description

★★★★★ Finally! A FREE online voice translator. Unlimited Online Use ★ Download Offline Translations

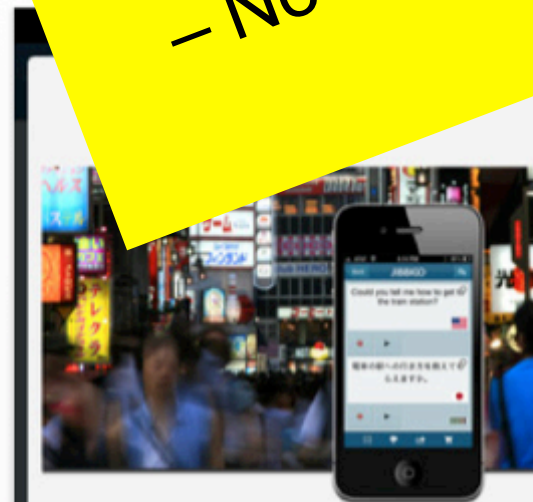
Jibbigo LLC Web Site > Jibbigo Translator Support

What's New in Version 2.1

- * Support for iPhone 5 Resolution
- * Better Download UI
- * Bugfixes / Improved Error Handling...

Screenshots

iPhone iPad



Jibbigo brings state-of-the-art speech-to-speech translation technology right to the palm of your hand!

Jibbigo:

- Real-Time Translation
- 15++ languages
- 40,000 Words
- All on the Phone
- No Server Necessary!

Jibbigo on Apple Commercials



The Economist Jan. 2013. “...The most sophisticated technology currently belongs to Jibbigo.....”



Claudio Munoz



Humanitarian Relief



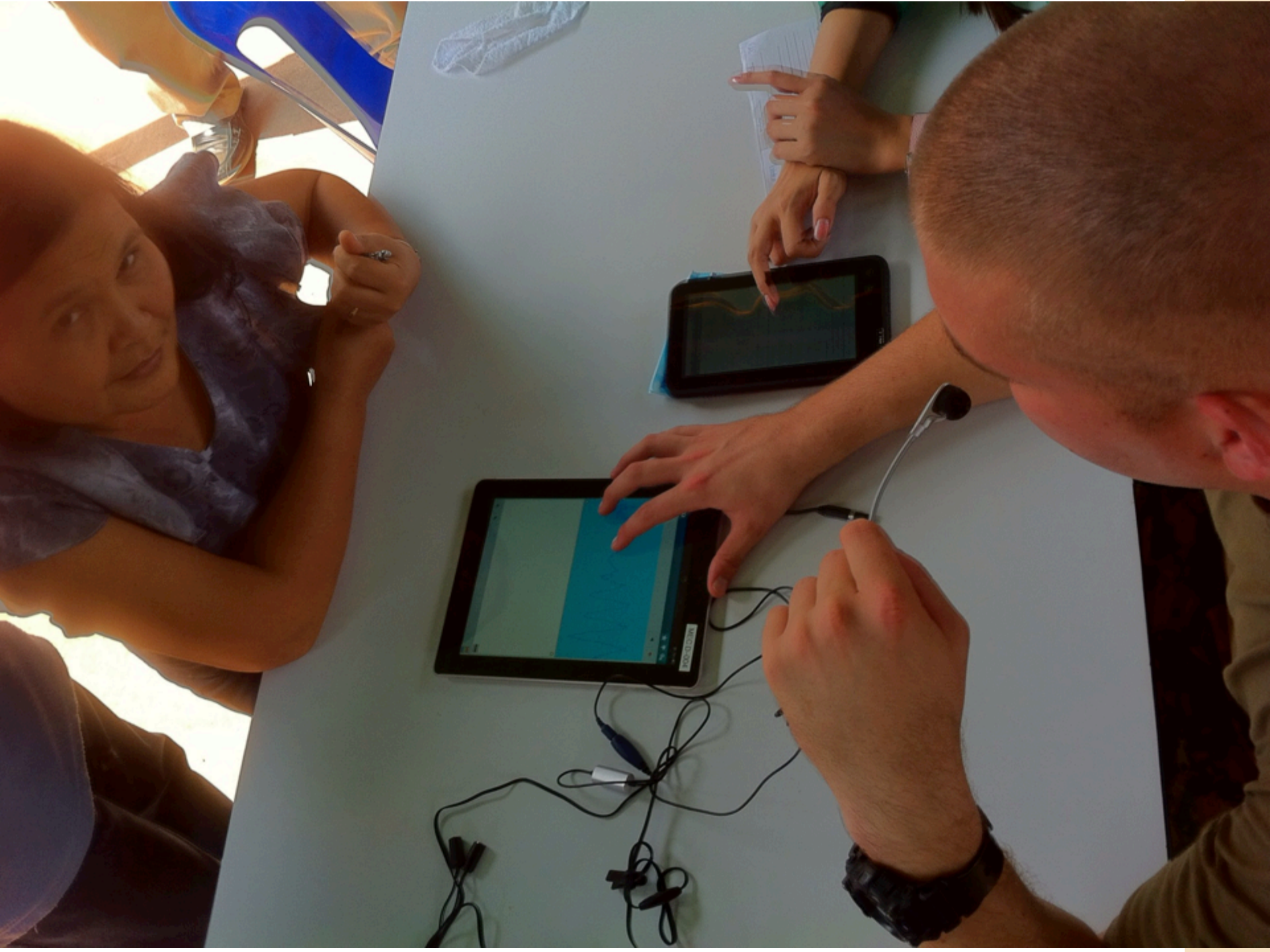
interACT

Humanitarian Deployment ('07-'13)



Thailand









Cambodia



San Jose , Honduras





interACT

Flüchtlingskrise 2015



Unlimited Domain Simultaneous Speech Translation Technologies



interACT

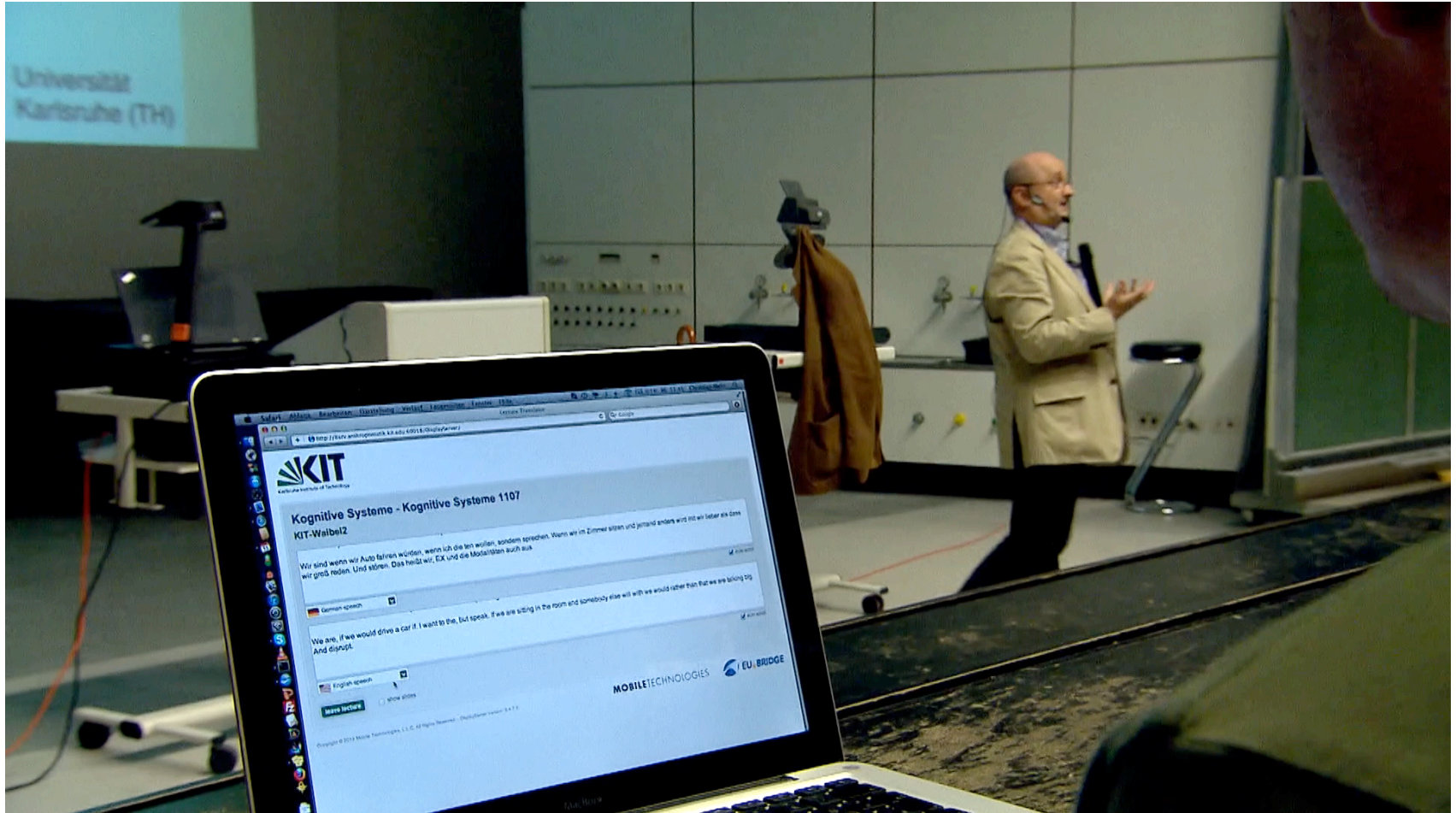
Lecture Transcription/Translation at KIT



- Speech more Spontaneous than TED
- Real-Time Requirement
- Specialist Vocabularies

Studentin aus Tunesien

die Vorlesungen waren unverständlich



- Students bring their own Devices
- Transcription/Translation Output is Delivered via Web Page
- Interpretation Done on Server
- User Can Select Languages
- Launched 2012 as KIT Student Service
- Data Collected in Use and Evaluate

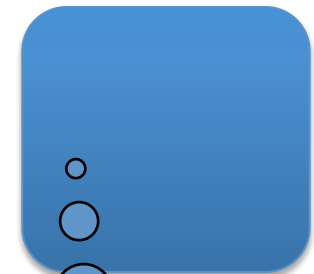
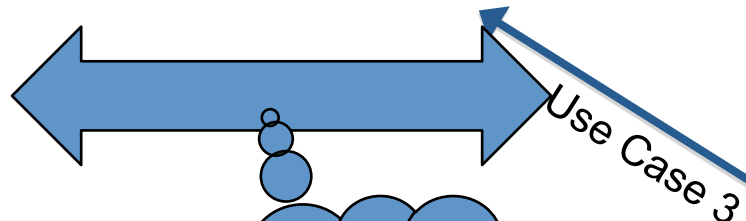
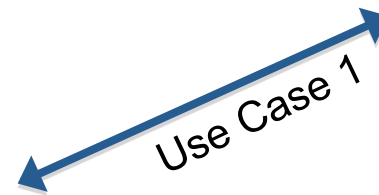
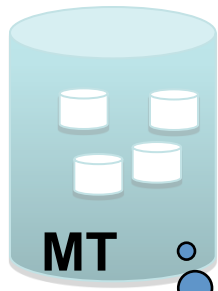


Language Services

Engines

Services

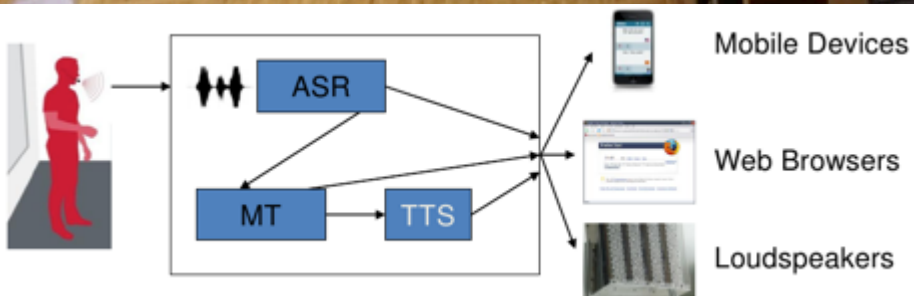
Use Cases



Develop and Insert
Improved Technology

Customization,
Adaptation

Language Services
for User and Developer
Communities



English (en)

French (fr)



Three Use Cases:

- Terminology Support
- Named Entity Support
- Interpreter's 'Cruise Control'

22:41 - 22:43
three by roll call the vote is op

22:48 - 22:49
vote is closed

22:53 - 22:54
and that you rejected

22:55 - 22:59
I put the joint motion for resolu

23:00 - 23:00
those in favour

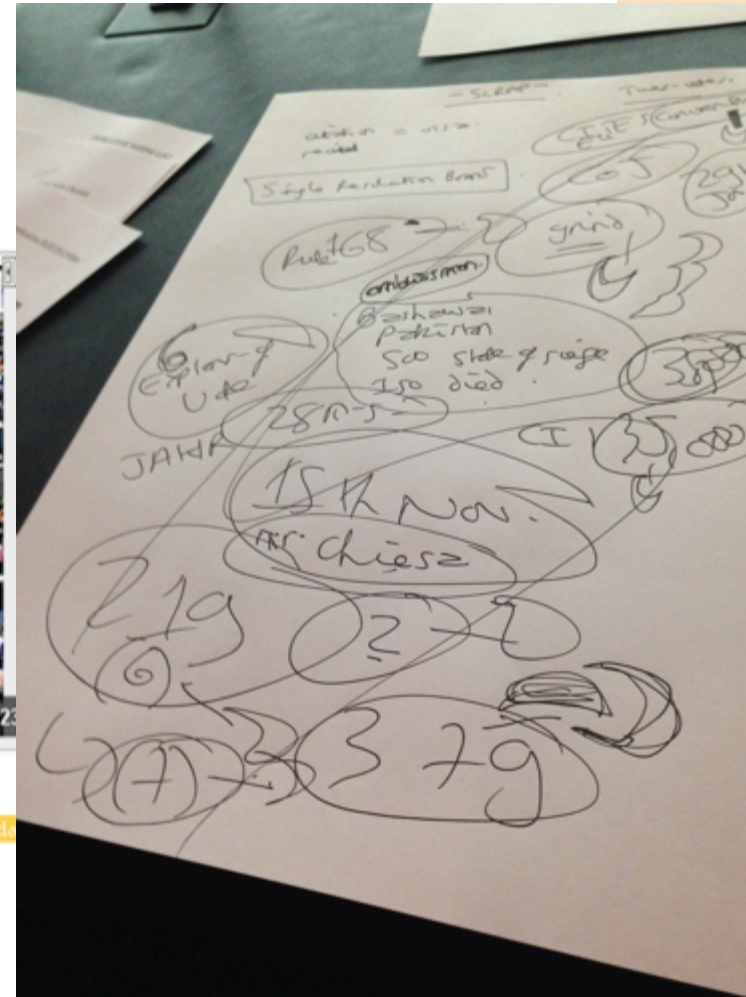
23:02 - 23:06
those against abstentions that is

23:07 - 23:08
and that concludes that

23:11 - 23:18
commission work programme but we now move to the vote on the report by Missus yannakouda
tissues and cells

23:20 - 23:22
K paragraph three

23:24 - 23:25
by split vote in two parts



Observations:

Interpreting Voting Sessions is...

- Boring and Repetitive
- Still Stressful, and Demanding
- Many Numbers
and Named Entities



Continuing Research Issues

- Neural MT
- All the usual issues:
 - Morphology, Pronoun Resolution, Word Order
- Learning in Practice
 - Incremental Learning
 - Crowd Source
- Improving Quality of Speech Translation
 - Low Latency
 - Dealing with Errors and Disfluencies
 - Put the Drama back
 - Suitable Intonation or Textual Realization
 - Emotion and Emphasis
 - Contextually Aware MT
 - Formality/Informality in Translation
 - Male Female
 - Plausibility in MT

Die Fehlerstromschutzschalterprüfung
Die Wirtschaftsdelegationsmitglieder
Die Bankwirtschaftsfreigabeerklärung
Die Lehrverpflichtungserklärungen
Die Schiffskommunalschuldverschreibungen
Die Vorkaufsrechtverzichtserklärung
Das Mehrzweckkirschentkerngerät
Die Gemeindegrundsteuerveranlagung
Die Nummernschildbedruckungsmaschine
Der Mehrkornroggenvollkornbrotmehlzulieferer
Die Verkehrsinfrastrukturfinanzierungsgesellschaft
Die Feuerwehrrettungshubschraubernotlandeplatzaufseherin
Das Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz

Zentraleuropa:

Zentral-Europa → Central Europe

Zentrale-Ur-Opa → Headquarter-Great-Grandpa

Dramatisch:

drama-t-isch → dramatic

drama-tisch → drama table

Asiatisch:

asia-t-isch → asian

asia-tich → asia table



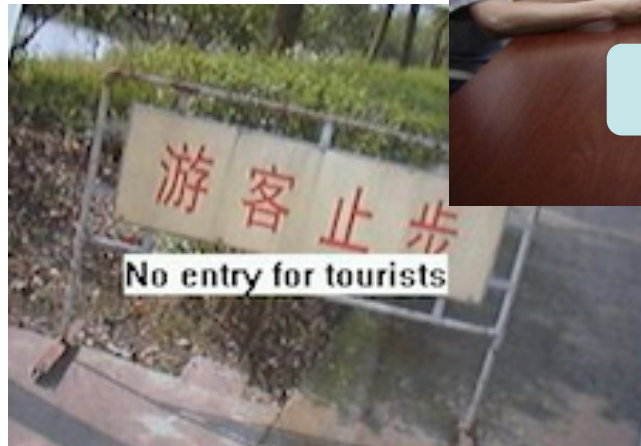
- Foreign Words in German Language
 - Computer Science, English Expressions
 - Political Speeches, Latin Proverbs
- Accent
 - “Würfelkalkül” (Asfour)
- Foreign Words in German Language
 - “Cloud”, “iPhone”, “iPad”, “Laser”
- Inflections & Declinations of these Words
 - Web-ge-casted, down-ge-loaded
- Formation of Compounds:
 - Cloudbasierter Webcastzugriff

1.) Communication is Multimodal

Communication is not only by Text, but:

- Speech, Gesture, Gaze..

- Images

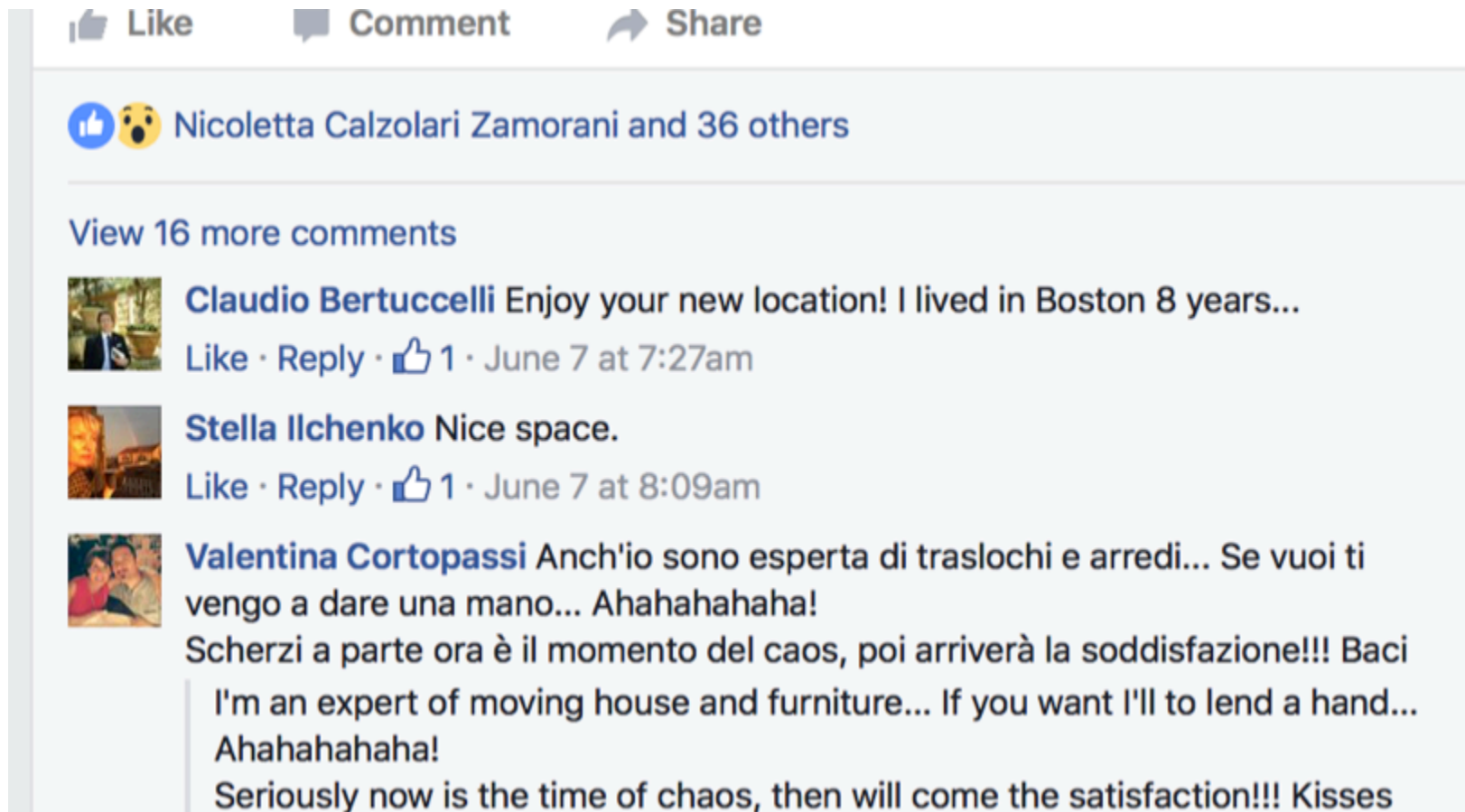


- Ill-formed Text

“lol-jah I want hr to be like dat...”, Hppyyyyy BD, CU, LMK



- Robust Text Translation for Posts, Texting at Facebook:
Use posts data to expand “vocabulary” and language use

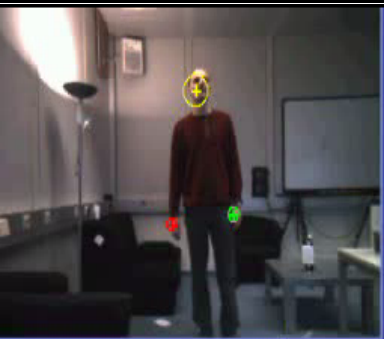


Road Sign Translator (2001)



A PDA-Based Sign Translator, ICMI 2002, Zhang, Yang, Chen, Waibel

Localization



Identification



Tracking & Gesture



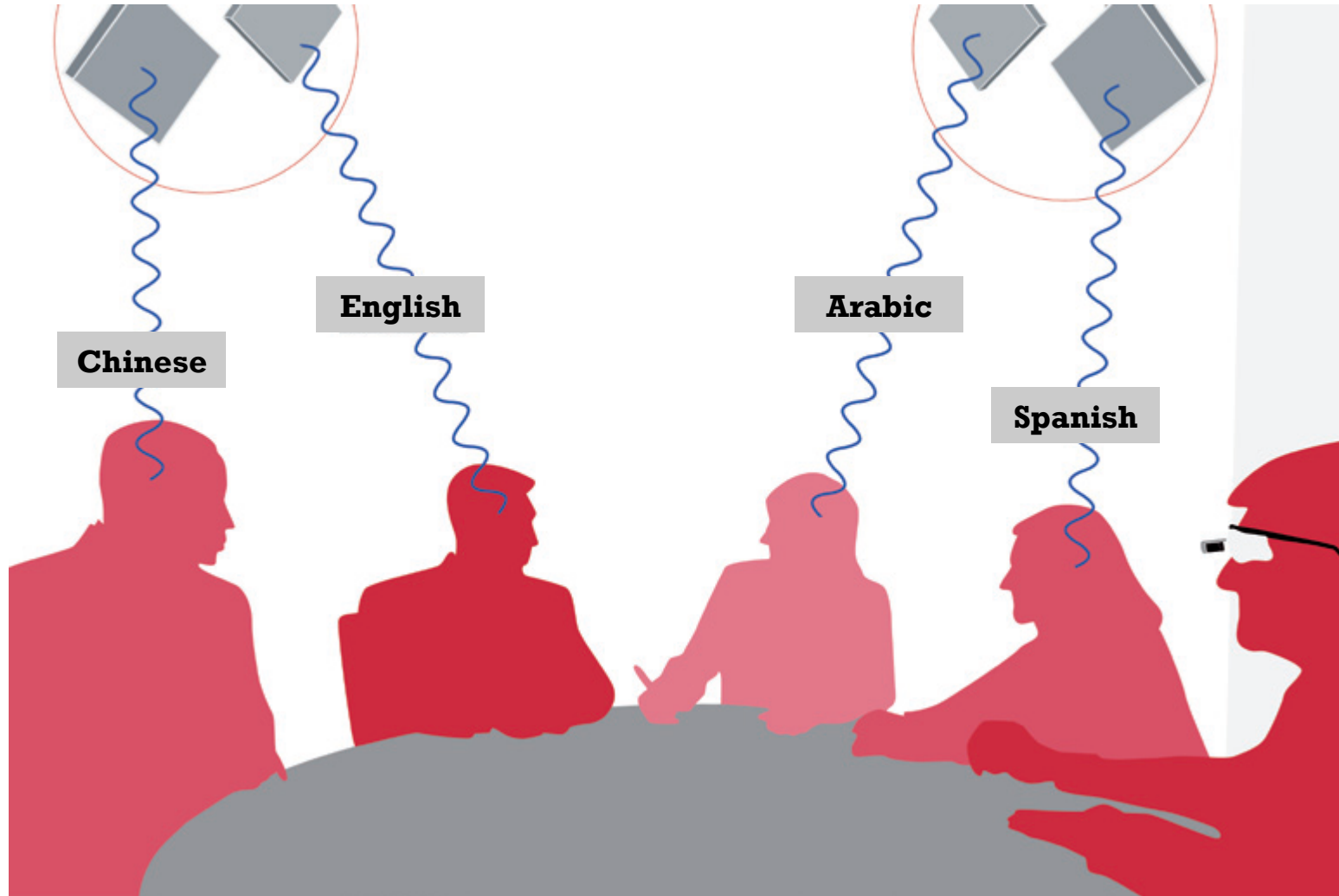
Focus of Attention





interACT

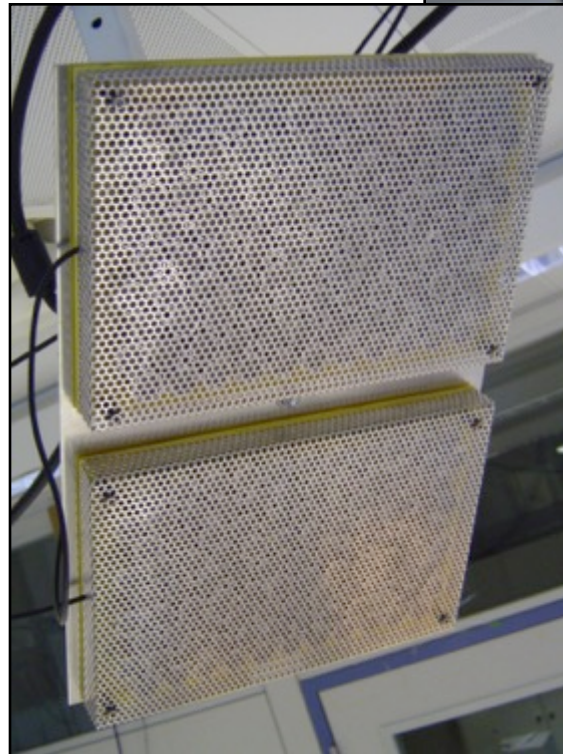
Language Transparent Meetings





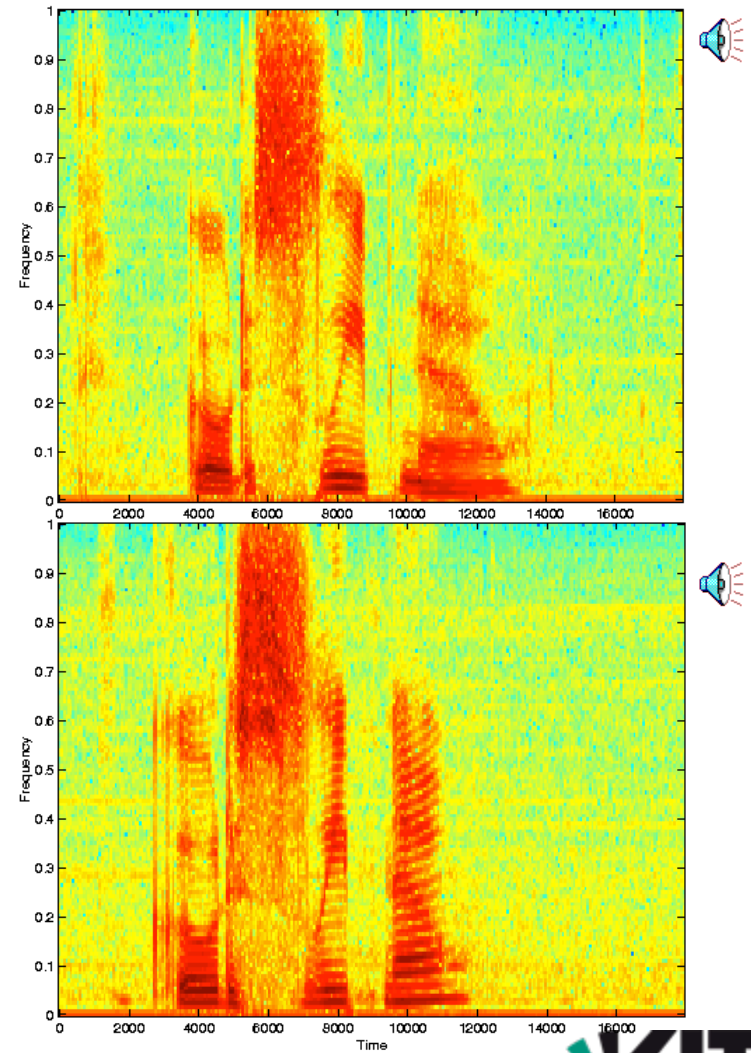
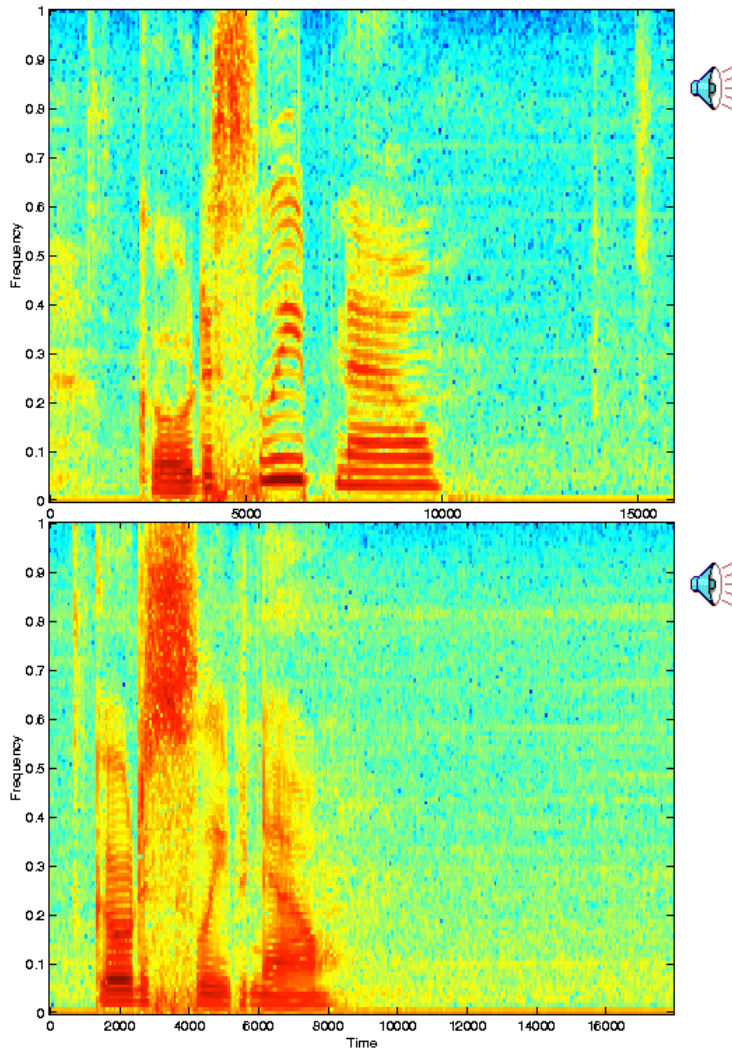
interACT

Seeing/Hearing Personal Translation





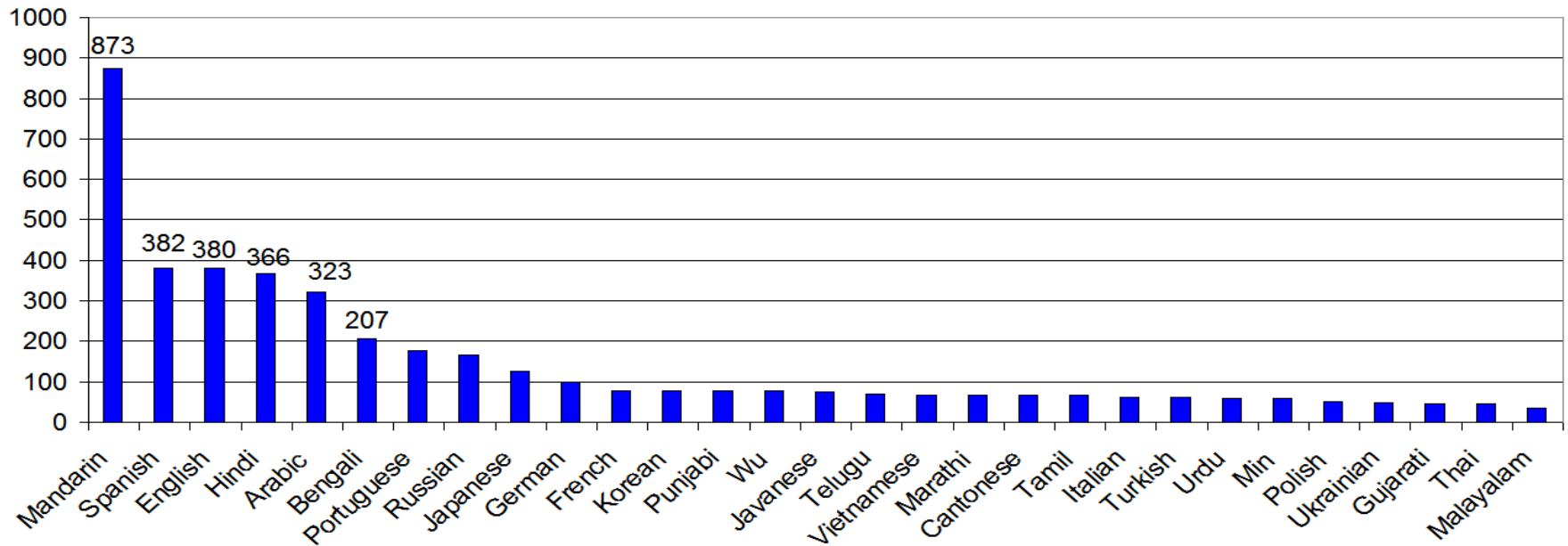
Emotional Speech (Dellaert, Polzin, Waibel '96-'99)



The Long Tail of Language

- Languages:
 - Only a Few Languages are Currently Addressed (~20)
 - Development of Technology Takes Long & Is Expensive
 - Cost ~.5M \$ and DevTime more than 1 Year per Language
 - 6, 000 languages, 36 M potential language pairs, Plus Dialects
 - Technology is Always a Step (or Two!) Behind Deployment

Languages by Million Native Speakers



From Babylonian
Confusion...





interACT

... to World-Understanding.. In our Lifetime!



