Real-Time Machine Translation for Software Development Teams

Rafael Prikladnicki
PUCRS

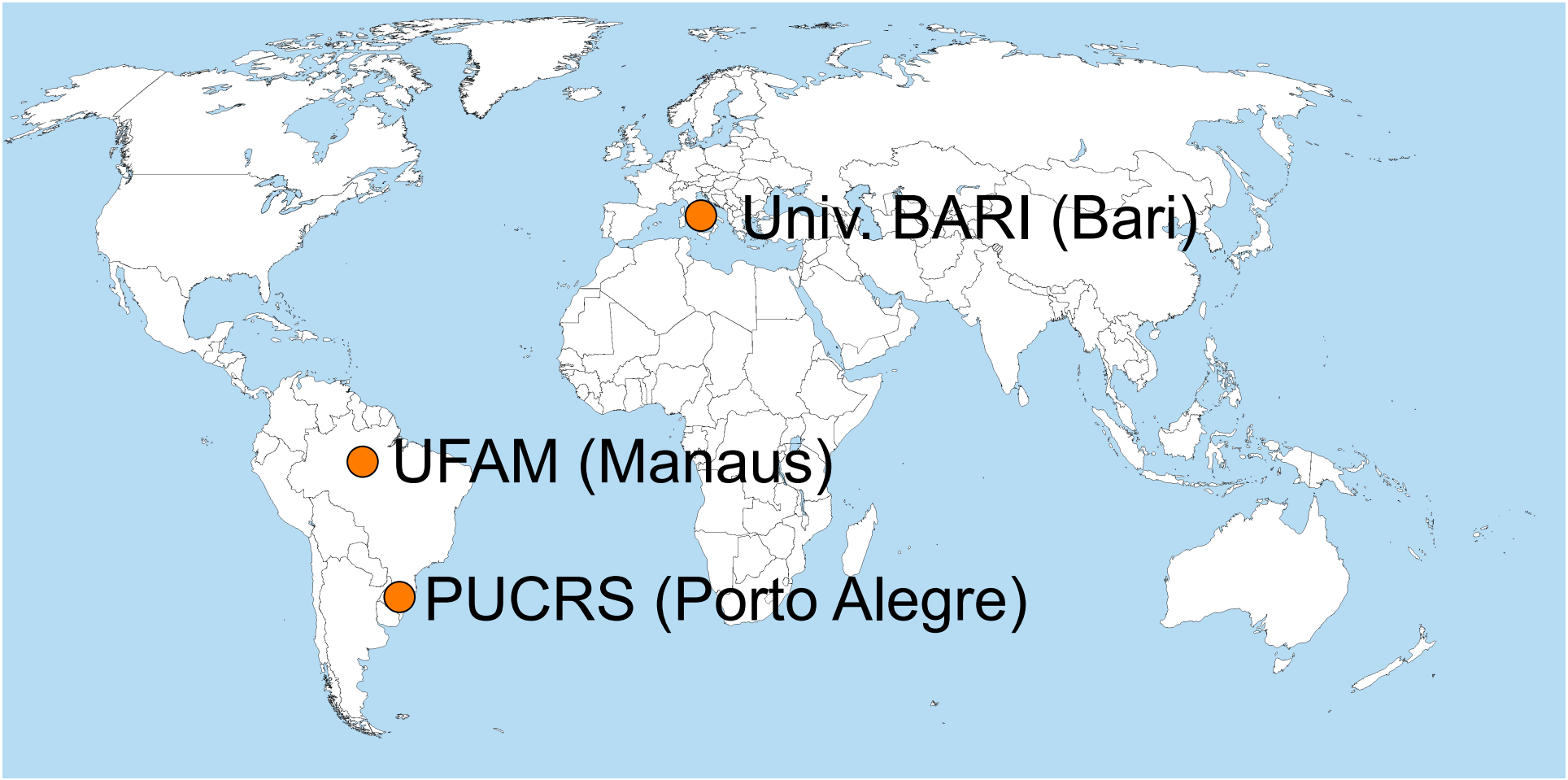
Tiago Duarte
PUCRS

Tayana Conte
UFAM

Fabio Calefato
U.BARI

Filippo Lanubile
U.BARI

Where are we located?



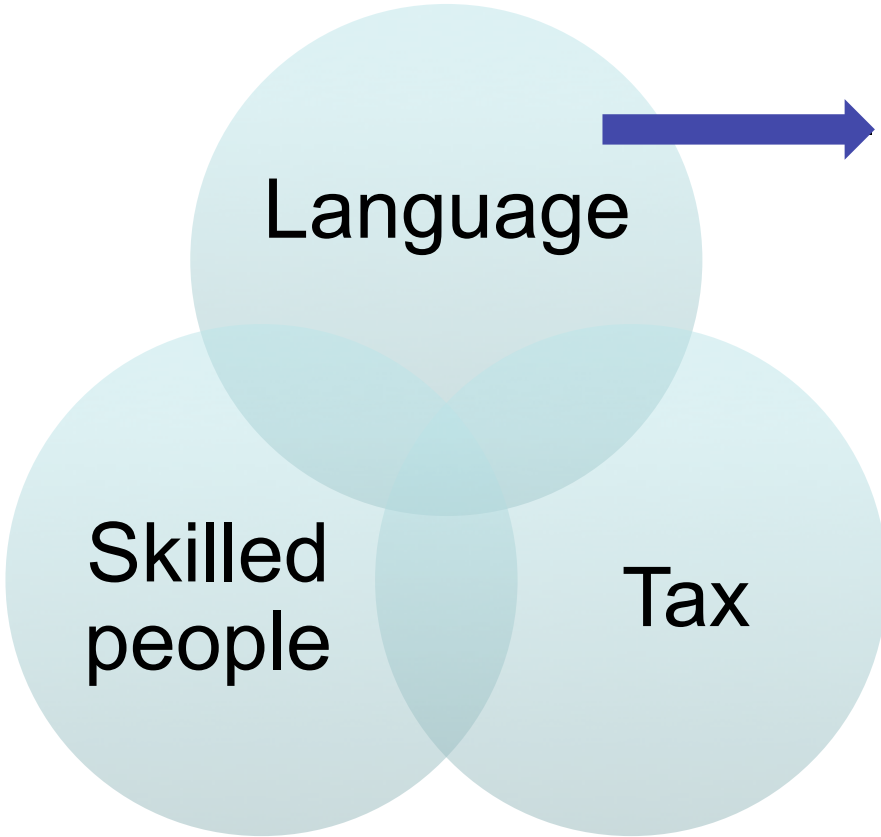
Outline

- Motivation
- Machine translation background
- Program of research

Motivation

- Global software projects suffer from **language distance**
  - **Shared understanding** challenged by language disparities
    - More severe for requirements engineering and activities intensive in communication
- Vision
  - Use **machine translation (MT)** technology for remote meetings in countries with
    - Opportunities for global software engineering (GSE) projects
    - **Lack of English speaking** professionals
    - **Text-based** and **voice-based** (automatic speech recognition) MT
- Goal
  - To investigate how **MT technology** could be used by **software development teams**

Brazil's challenges for global competitiveness



Limited number of English speakers

- Argentina: 9.8% (3M)
- Brazil: 5.4% (10M)
- Russia: 4.8% (7M)
- China: 0.8% (10M)

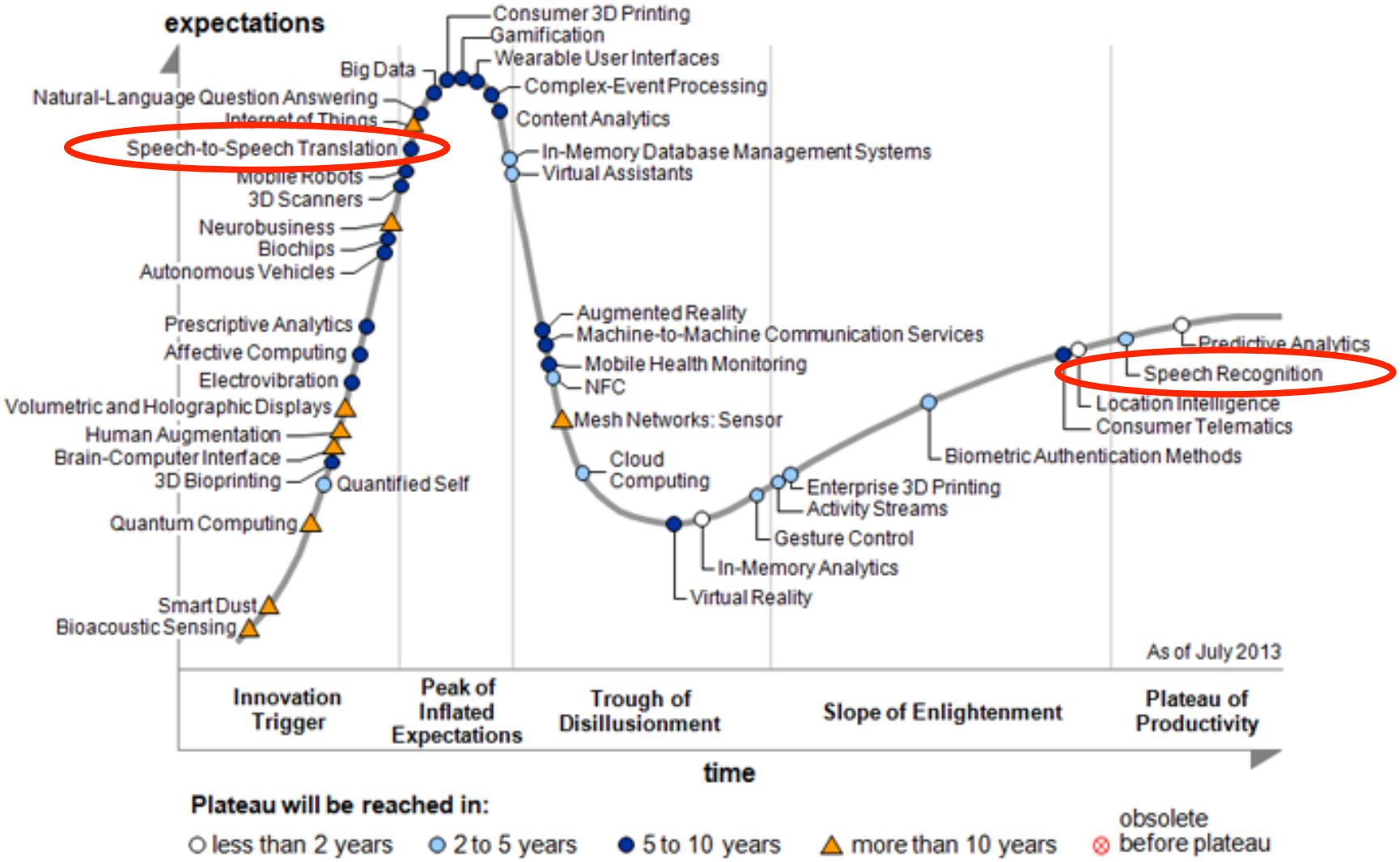
Source: Brasscom IT BPO Book, Technical Report

Machine translation background

- MT technology 50 years in the making
  - Goal: fully **automatic** translation of ordinary text from natural language A (source) into different natural language B (target)
  - Text-based or voice-based
- Ambitious goal, ambiguous task
  - Involves a **huge** amount of **human** knowledge to be coded into a **machine**-processable form
  - Still far from perfection
- Steadily growing in **interest** due to economic reason
  - EU currently spends over a billion euro per year to translate official docs
  - Speech-to-speech translation is included in the Gartner's 2013 hype cycle (http://www.gartner.com/newsroom/id/2575515)

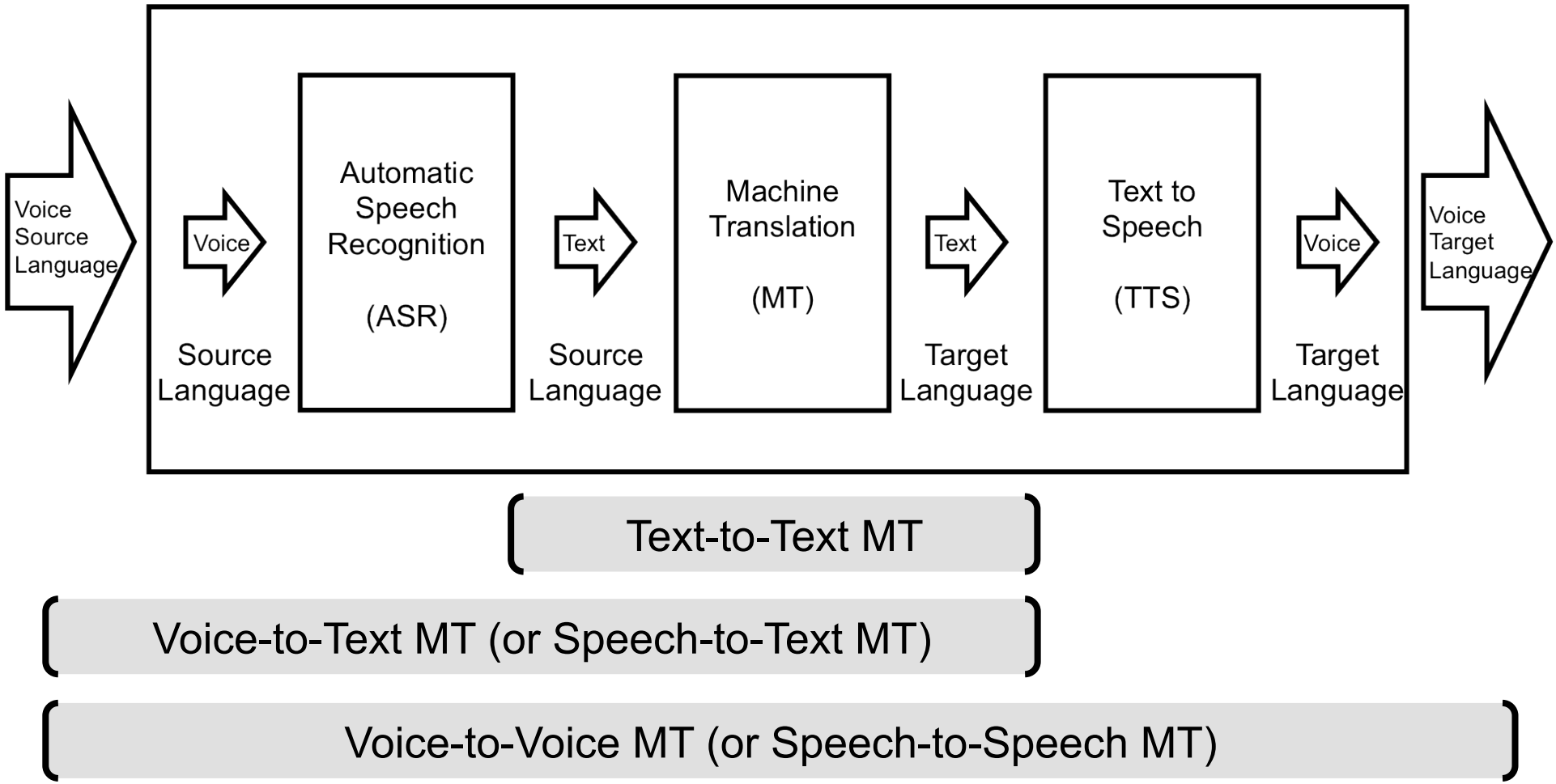
Machine translation technology

Gartner Hype Cycle 2013



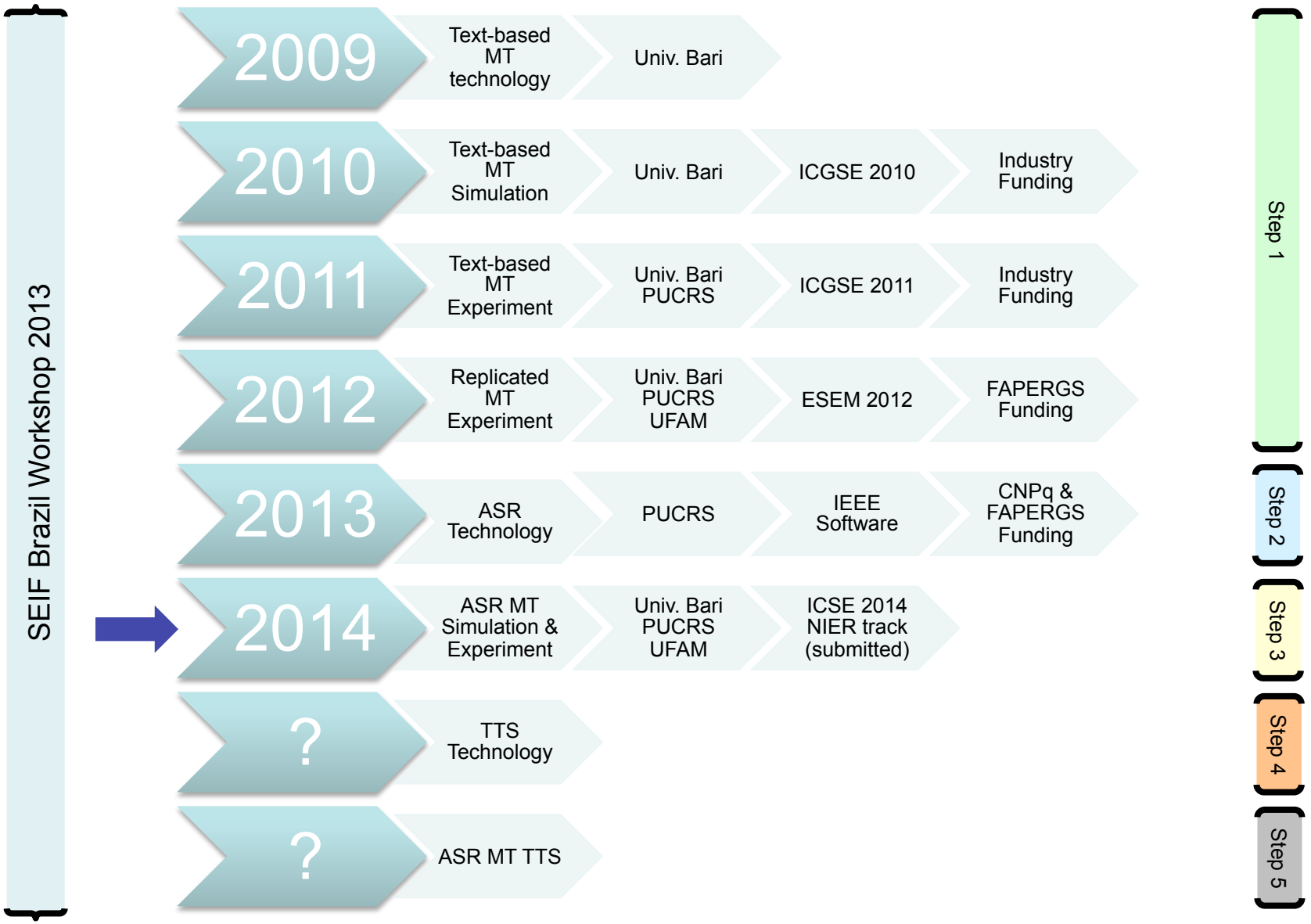
http://www.gartner.com/newsroom/id/2575515

Machine translation components

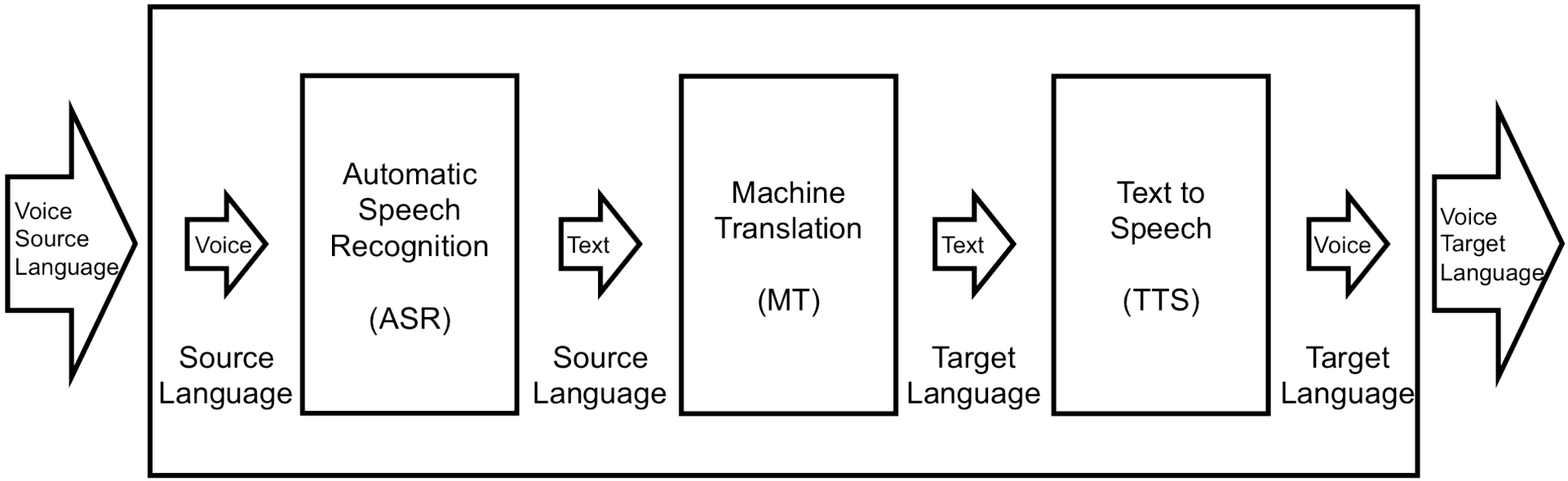


Source: Waibel, A.; Fugen, C. Spoken language translation. Signal Processing Magazine, 25(3): 70–79, May 2008.

Real-time MT Program of Research



Machine translation components



Step 1 - MT (2009-2012)

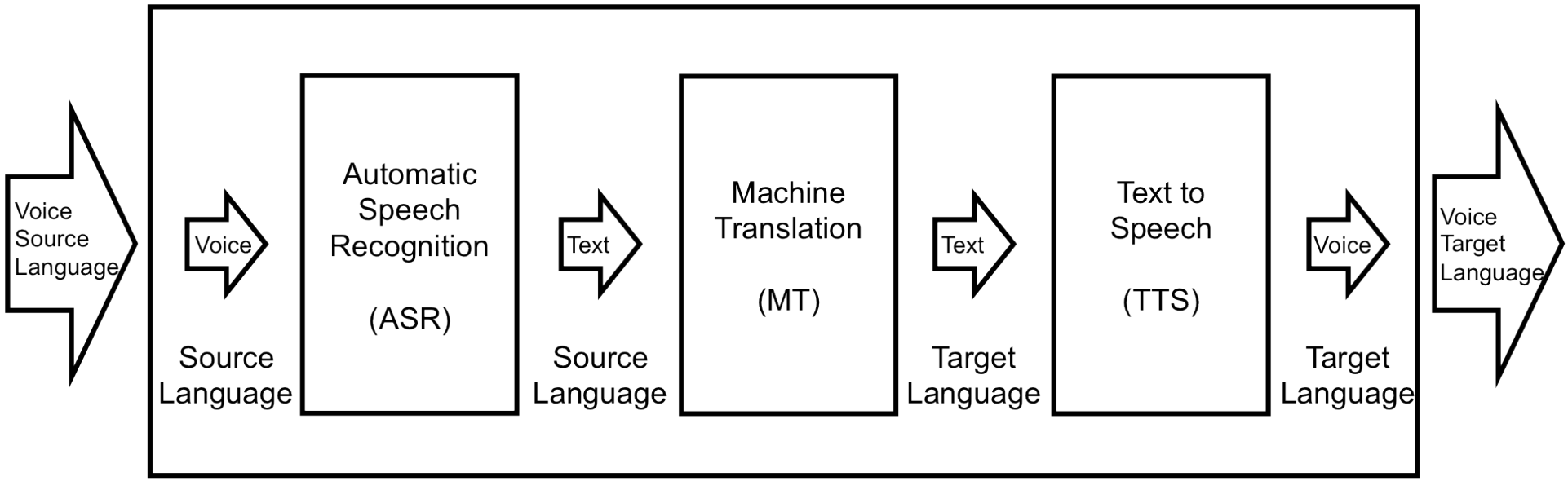
Step 2 - ASR (2013)

Step 3 – ASR / MT (2014)

Step 4 - TTS (?)

Step 5 – ASR / MT / TTS (?)

Text-based MT



Step 1 - MT (2009-2012)

Step 2 - ASR (2013)

Step 3 – ASR / MT (2014)

Step 4 - TTS (?)

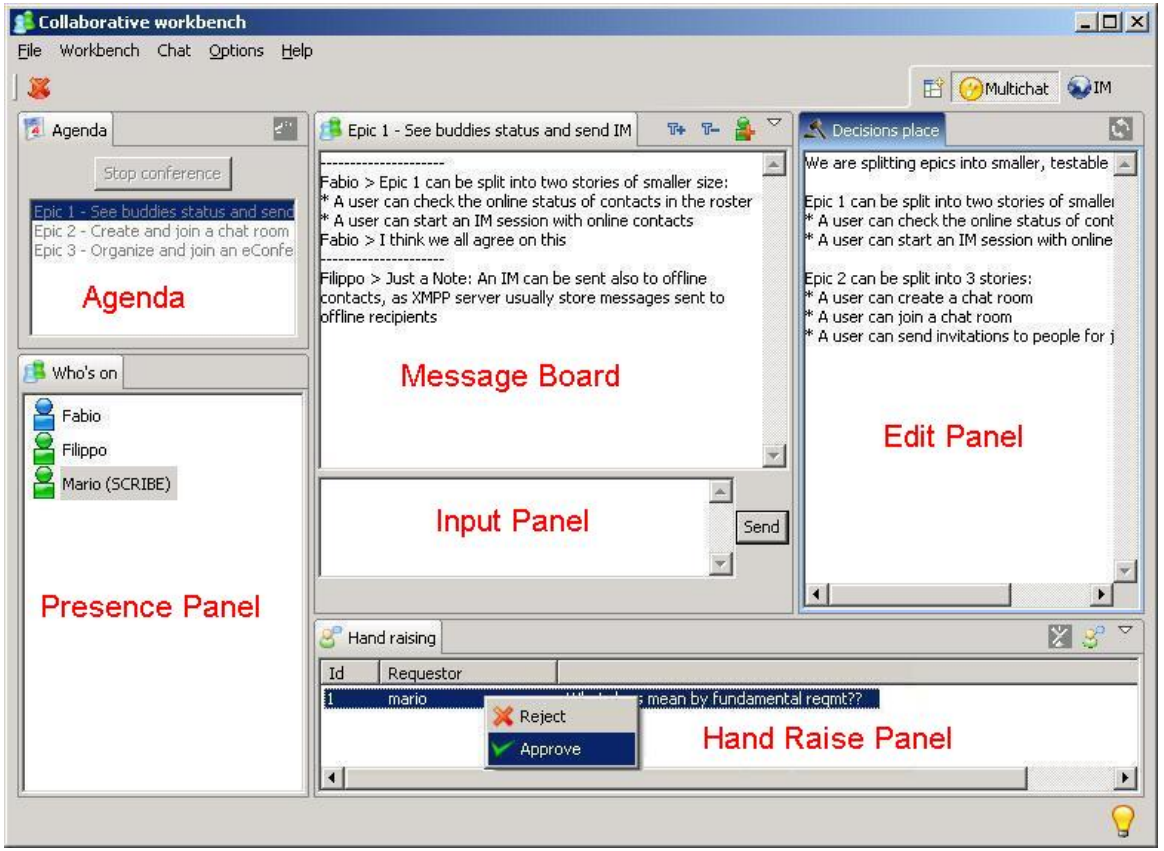
Step 5 – ASR / MT / TTS (?)

Text-based MT simulation

- MT Technology
  - Google translate
  - Apertium
- Text-based MT simulation
  - Simulating the adoption of a MT service in a cross-language, real time, text-based meetings
  - Assessment of translation quality and time performance of Google Translate and Apertium
- Test corpus
  - Chat logs (in English) collected from 5 requirements meetings during a RE course
  - 1h long meetings between clients and developers (5-8 participants)
  - 2000+ utterances exchanged overall

eConference MT plug in

- Extension of the eConference tool
- Conferencing tool built on Eclipse RCP platform
  - Textual communication based on XMPP (via GMail accounts)
  - Audio communication based on Skype



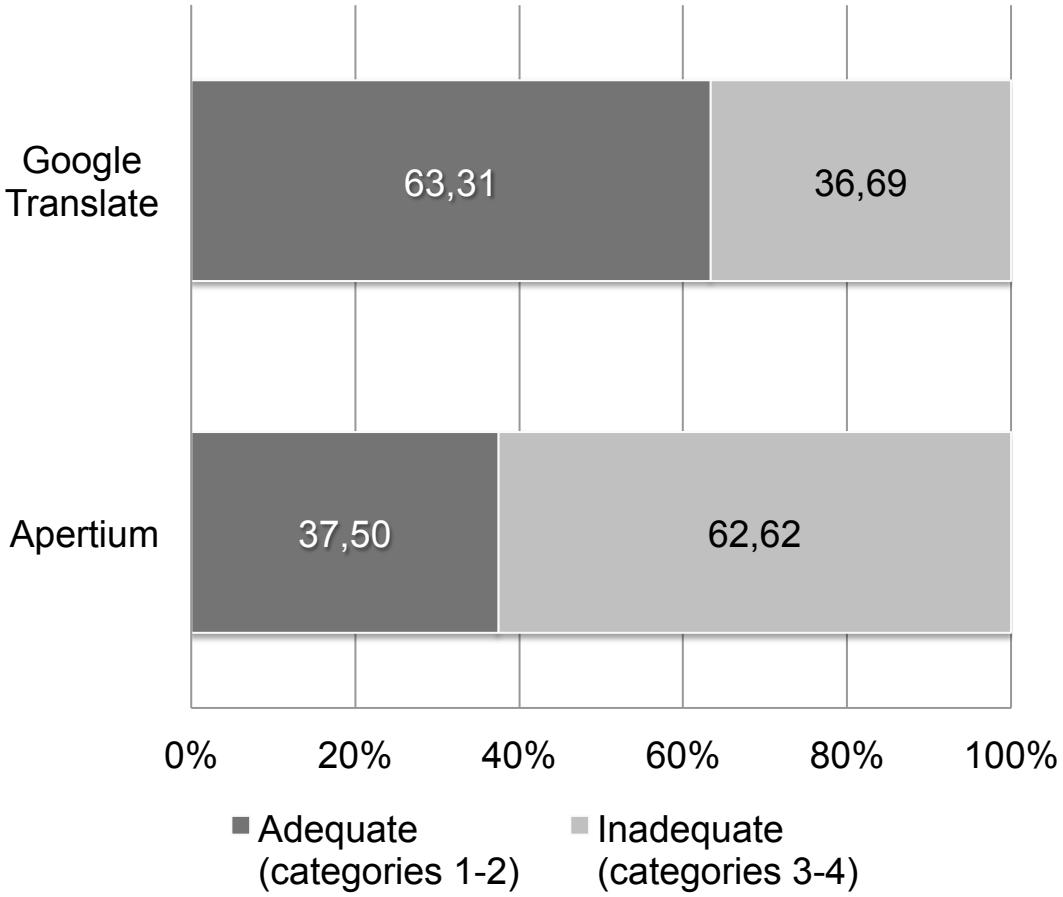
eConference : http://code.google.com/p/econference4/

MT plugin: http://code.google.com/p/econference-mt-plugin/

Results

- Google Translate produces more adequate translations than Apertium

- State-of-the-art MT services can be embedded into synchronous text-based chat without disrupting real-time interaction



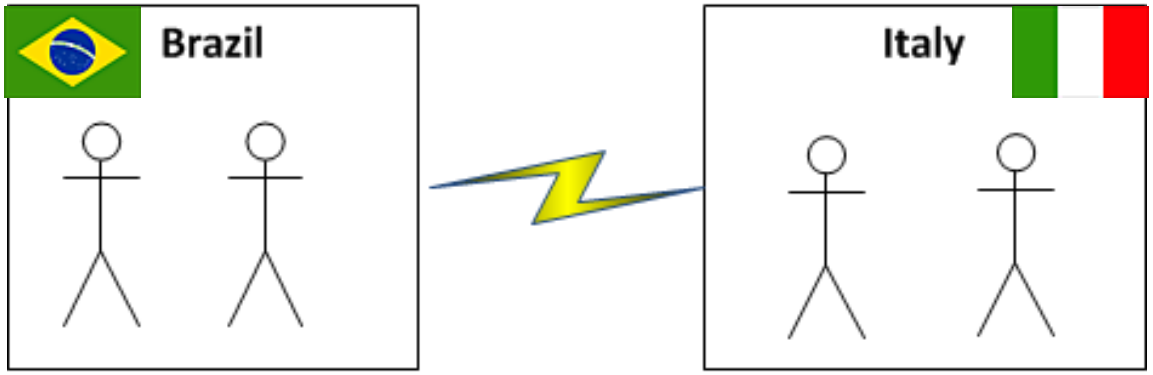
Text-based MT experiment

- RQ1: *Can machine translation services be used in distributed multilingual requirements meetings instead of English?*

- RQ2: *How does the adoption of machine translation affect group interaction in distributed multilingual requirements meetings, as compared to the use of English?*

Methodology

- Controlled experiment
- Participants: students from Brazil and Italy
- Multilingual groups involved in a Planning Game activity
- Analysis from questionnaires and chat logs



Experimental tasks

T1 – requirements prioritization (30 min.)

- Customer's perspective

1. Assign 16 mobile phone features to 3 piles: *very important, important, less important*
2. Rank the features within piles

T2 – release planning (60 min.)

- Developer's perspective

1. Distribute 1000 story points to each feature as an estimate of implementation costs
2. Plan 3 releases based on priorities (T1) and cost estimates

Experimental design

- 3 factors with 2 levels:
  - Communication mode: *MT, EN*
  - Task: *T1 prioritization, T2 planning*

- 8 distributed meetings executed
  - Gr1, Gr3: *MT – T1 / EN – T2*
  - Gr2, Gr4: *EN – T1 / MT – T2*
  - Only groups with high English proficiency (Cambridge questionnaire to assess English proficiency level)

Conclusions

RQ1: *Can machine translation services be used in distributed multilingual requirements meetings instead of English?*

- Yes, MT services can be used without disrupting the conversation flow
  - despite still far from 100% accuracy
- Generally accepted with favor

RQ2: *How does the adoption of machine translation affect group interaction in distributed multilingual requirements meetings, as compared to the use of English?*

- Not enough data to provide an answer
  - Just some clues: speed and participation
- Differences might be more evident with lower levels of English skills

Text-based MT replicated experiment

- RQ1: *Can machine translation services be used in distributed multilingual requirements meetings instead of English?*
- RQ2: *How does the adoption of machine translation affect group interaction in distributed multilingual requirements meetings, as compared to the use of English?*
- RQ3: *Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?*

Methodology

- Participants: 16 students from Univ. Bari (Italy) and Fed. Univ. of Amazonas (UFAM), Manaus (Brazil)

- Multilingual groups
  - Same tasks
  - Same instrumentation
  - **Lowly proficient in English**

Experimental design

	Original experiment (high proficiency)		Replicated experiment (low proficiency)	
	MT	EN	MT	EN
Run 1	Gr1, Gr3 execute T1	Gr2, Gr4 execute T1	Gr6, Gr8 execute T1	Gr5, Gr7 execute T1
Run 2	Gr2, Gr4 execute T2	Gr1, Gr3 execute T2	Gr5, Gr7 execute T2	Gr6, Gr8 execute T2

Data sources:

- post-task questionnaires
- meeting logs

Conclusions

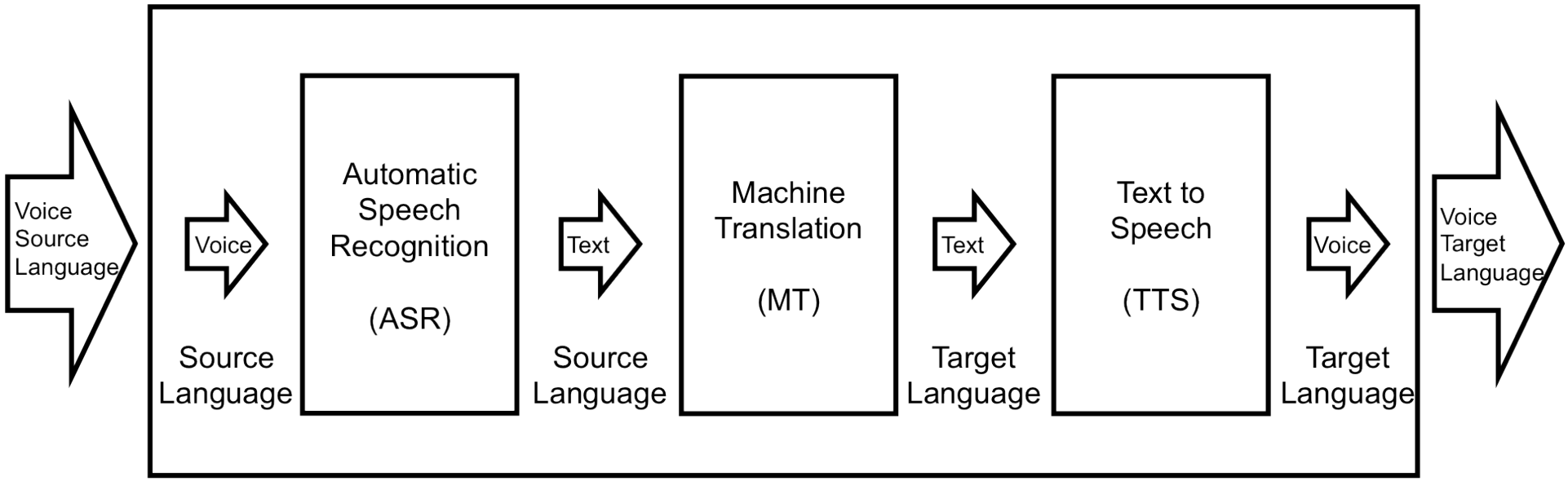
RQ3: Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?

so far, **NO**

however

- people with low English skills are more prone to use MT again
- messaging is easier than talking for a non-native English speaker

Technologies for Speech Recognition



Step 1 - MT (2009-2012)

Step 2 - ASR (2013)

Step 3 – ASR / MT (2014)

Step 4 - TTS (?)

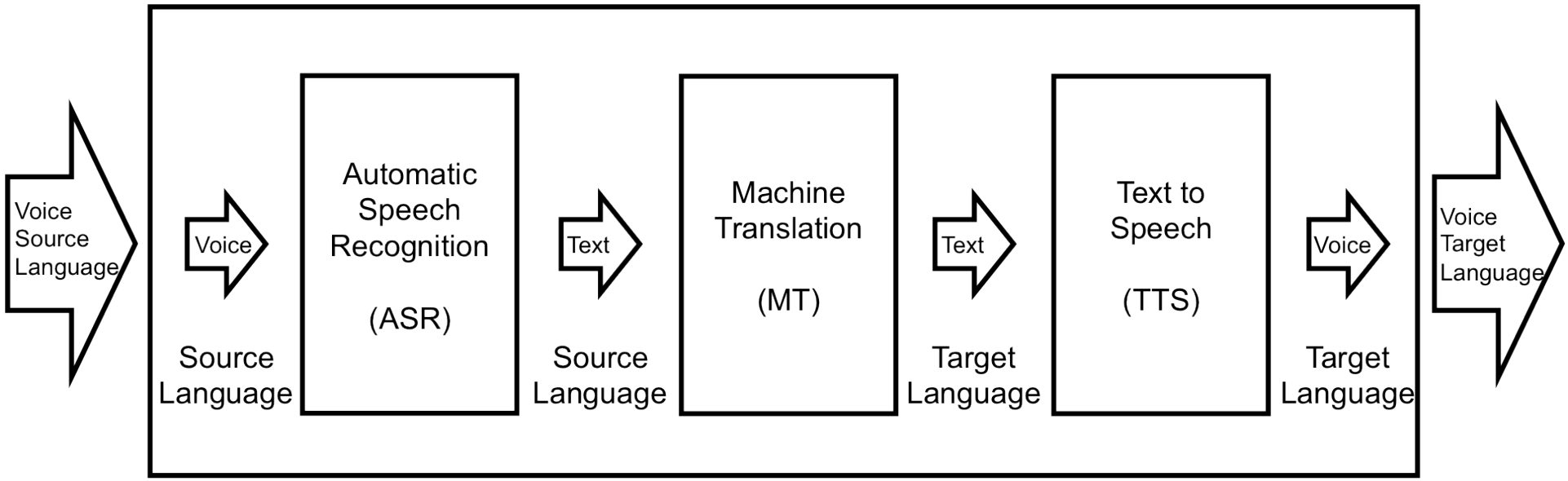
Step 5 – ASR / MT / TTS (?)

Technologies for Speech Recognition

- Systematic Literature Review (SLR)
  - Microsoft Speech API
  - Microsoft .NET System.Speech namespace
  - Microsoft Speech Platform
  - Microsoft Unified Communications API
  - CMU Sphinx
  - HTK
  - Julius
  - Java Speech API
  - Google Web Speech API
  - Dragon

Coming up in IEEE Software (Jan/Feb 2014)

Voice-based machine translation



Step 1 - MT (2009-2012)

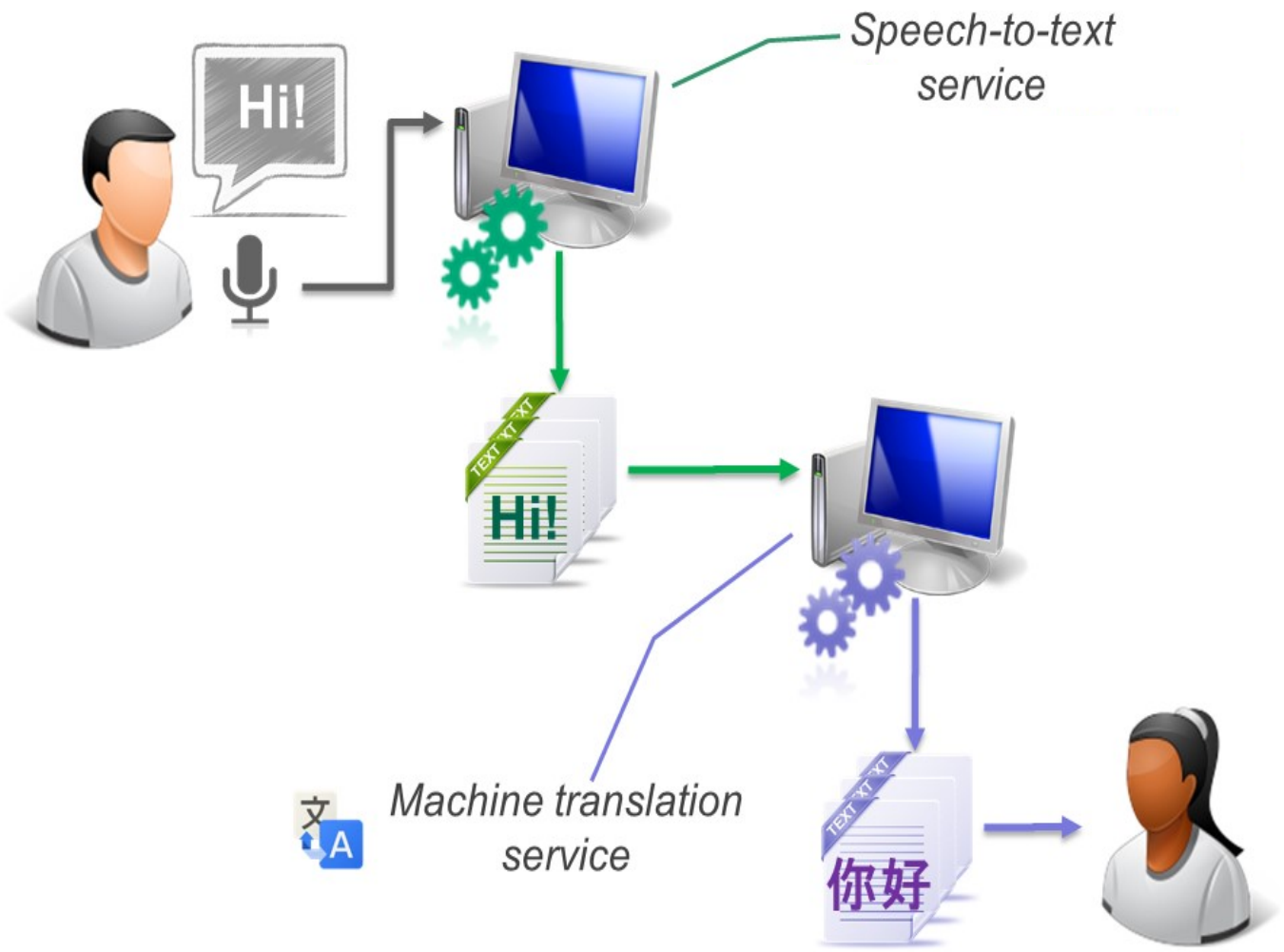
Step 2 - ASR (2013)

Step 3 – ASR / MT (2014)

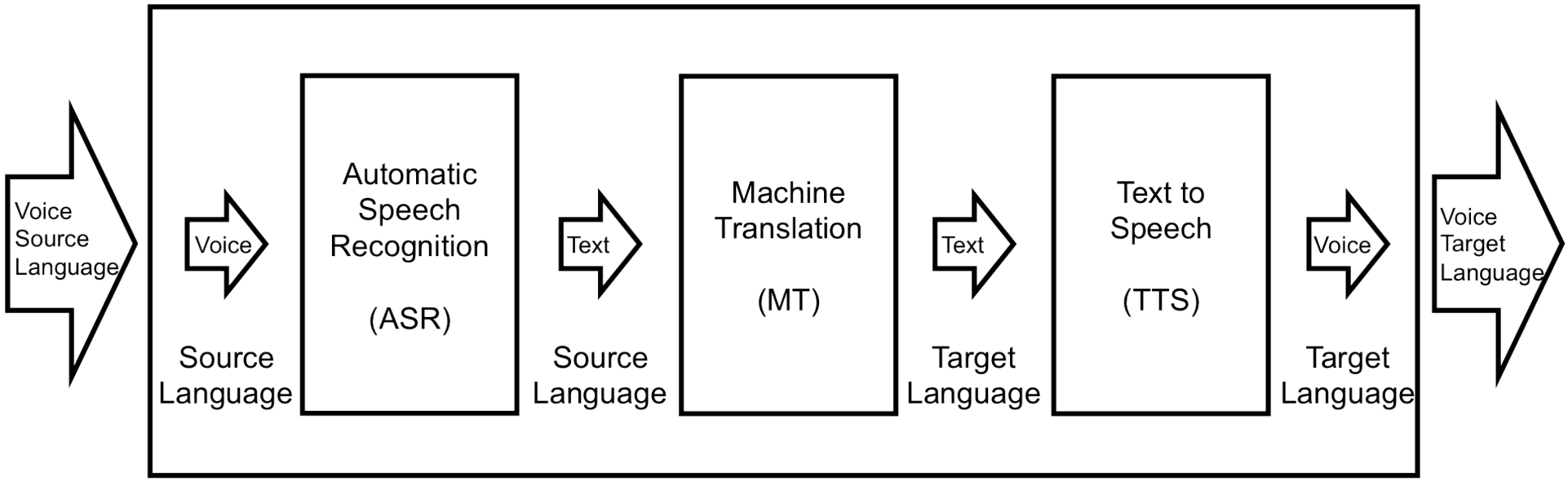
Step 4 - TTS (?)

Step 5 – ASR / MT / TTS (?)

Voice-based MT simulation



Future work



Step 1 - MT (2009-2012)

Step 2 - ASR (2013)

Step 3 – ASR / MT (2014)

Step 4 - TTS (?)

Step 5 – ASR / MT / TTS (?)

Conclusions

- The **advances** in the fields of speech recognition and machine translation have brought speech translation **close** to the **practical** level.
- Both research and development should be further **accelerated** for real-time speech translation to become a mainstream technology to be used by multilingual teams.
- Acknowledgments
  - All the participants in the studies (Brazilians and Italians)
  - Funding agencies and companies in Brazil and Italy

Further information

- F. Calefato, F. Lanubile, and P. Minervini, "**Can Real-Time Machine Translation Overcome Language Barriers in Distributed Requirements Engineering?**", *ICGSE'10.*

- F. Calefato, F. Lanubile, and R. Prikladnicki, "**A Controlled Experiment on the Effects of Machine Translation in Multilingual Requirements Meetings**", *ICGSE'11*.

- F. Calefato, F. Lanubile, T. Conte and R. Prikladnicki, "**Assessing the Impact of Real-Time Machine Translation on Requirements Meetings: A Replicated Experiment**", *ESEM'12.*

- R. Prikladnicki, T. Duarte, T. Conte, F. Calefato, F. Lanubile, **"Real-Time Machine Translation for Software Development Teams**", *Microsoft SEIF Brazil Workshop, 2013*.

- T. Duarte, R. Prikladnicki, F. Calefato, F. Lanubile, "**Speech Recognition for Voice-Based Machine Translation**", *Forthcoming in IEEE Software*, 2014.

- F. Calefato, F. Lanubile, R. Prikladnicki, T. Duarte, T. Conte, "**Real-Time Speech Translation for Software Development Teams**", *Submitted to ICSE'14 NIER track*.

Real-Time Machine Translation for Software Development Teams

Rafael Prikladnicki
PUCRS

Tiago Duarte
PUCRS

Tayana Conte
UFAM

Fabio Calefato
U.BARI

Filippo Lanubile
U.BARI

Methodology

Category	Description
1	**Completely adequate.** The translation clearly reflects the information contained in the original sentence. It is perfectly clear, intelligible, grammatically correct, and reads like ordinary text.
2	**Fairly adequate.** The translation generally reflects the information contained in the original sentence, despite some inaccuracies or infelicities of the translation. It is generally clear and intelligible and one can understand (almost) immediately what it means.
3	**Poorly adequate.** The translation poorly reflects the information contained in the original sentence. It contains grammatical errors and/or poor word choices. The general idea of the translation is intelligible only after considerable study.
4	**Completely inadequate.** The translation is unintelligible and it is not possible to obtain the information contained in the original sentence. Studying the meaning of the translation is hopeless and, even allowing for context, one feels that guessing would be too unreliable.

Adapted from: D. Arnold et al. "*Machine Translation: an Introductory Guide*" (1994)