Priority Theme 3:
Socially Aware Interactive Assistant

Joseph Mariani

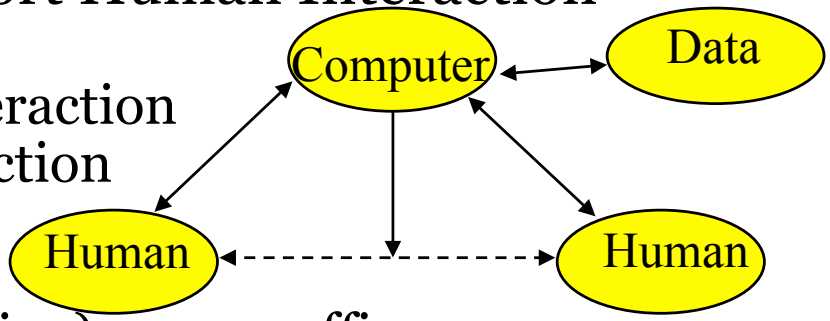
CNRS-LIMSI & IMMI, France
Contributors: M. Pantic, H. Bourlard, A. Waibel,
K. Jokinen, G. Riccardi, S. Renals, J. Mariani
META-NET Interactive Systems Vision Group, META-Council

META FORUM 2012 **A Strategy for Multilingual Europe**
Brussels, Belgium, June 20/21

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

Socially Aware Interactive Assistants

- Multilingual Assistants which support Human Interaction
  - Human-Computer Interaction,
  - Human-Artificial Agent (robot) interaction
  - Computer-mediated Human Interaction



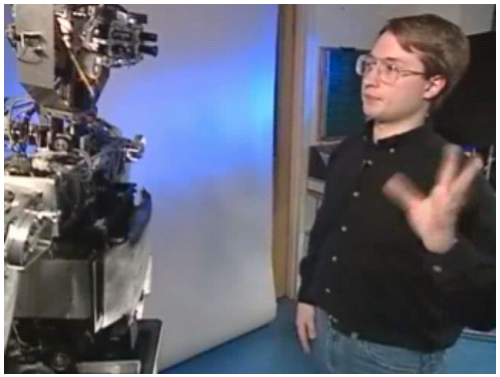
- Acting in various environments
  - Instrumented environments ((meeting) rooms, offices, apartments)
  - Instrumented open environments (streets, cities, transportation, roads)
  - Web, Virtual / Augmented worlds (incl. (serious) games)
- Personalized to user's needs and environment
- Learning incrementally and individually from all sources and interactions

Vision

- The Socially Aware Interactive Assistants can:
  - Interact naturally with you, wherever you are, in any environment
  - Interact naturally with your relatives, wherever they are
  - Interact in any language and in any communication modality
  - Adapt and personalize to individual communication abilities (handicap)
  - Transcribe into text any fluent speech, pronounce fluently any text
  - Self-Assess its performances and recover from errors
  - Learn, personalize & forget through natural interaction
  - Act on objects in instrumented spaces (rooms, apartments, streets)
  - Assist in language training and education in general
  - Provide a synthetic multimedia information analysis (*knowledge*)
  - Recognize people's identity, & their gender, accent, language, style, age
  - Move, manipulate objects, touch people (Robots)

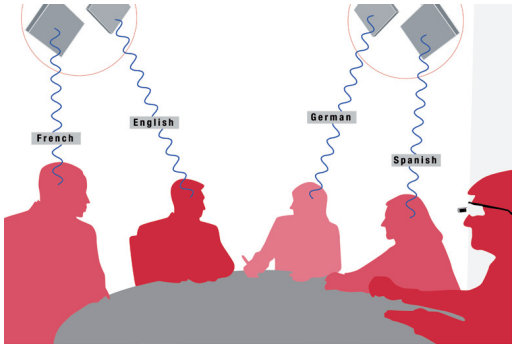
Global Abilities

- Interact naturally with Agents (terminals, ECA, robots, chatbots, humans, things)
  - In games, entertainment, education, communication, instrumented spaces, Call Centers,etc.
- Communicate everywhere
  - Mobile applications, Augmented Reality
- In society
  - Social networks and forums, Multiparty communication including several humans, several artificial agents/robots
- In a personalized way
  - Person, Gender, Style, Age
  - Accent, Language



Domain-specific Abilities

- Exhibit language proficiency
  - Speech-to-Speech Translation, Interpretation in meetings or in videoconferences,
  - Cross-lingual information access
- Overcome handicap obstacles
  - Crossmodal/crossmedia, Assistive applications, Sign Language
  - Adapted communication (cars, meetings)
- Refer to written support
  - Speech transcription, Subtitling
  - Reading machine
- Provide personalized training
  - Computer Aided Language Learning
  - Incl. dialects
  - Education, (self-)assessment



Tentative roadmap

- Where are we?
- Try to define time to deliver
  - Should reach “good enough” quality
- Socially Aware Interactive Assistants
  - Global Abilities
  - Domain-specific Abilities
    - Aid to Multilingualism
    - Aid to the “authors” (speech-text)
    - Aid to the handicapped
    - Aid to education and training
  - Resources and Evaluation

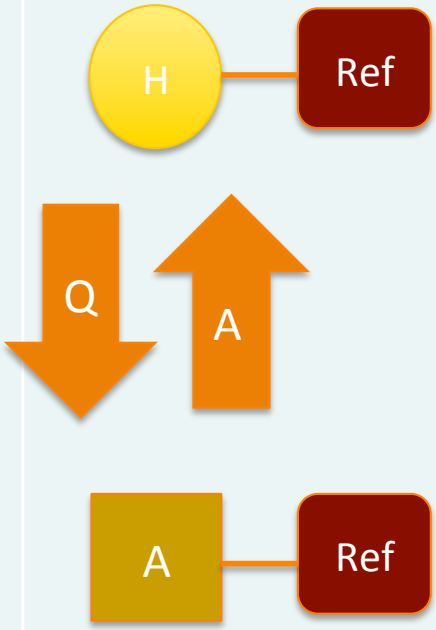
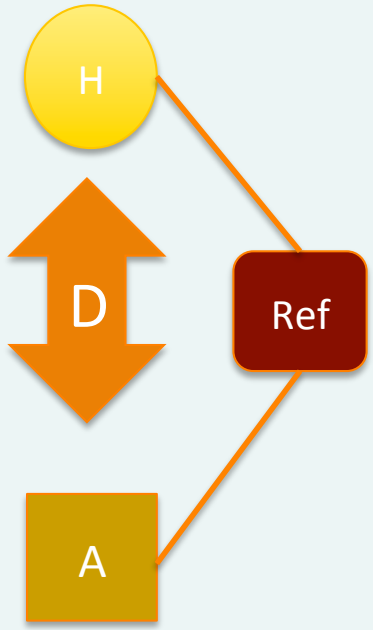
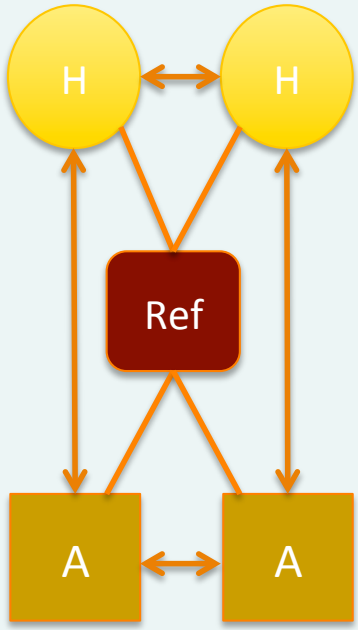
Where are we?

- What are the performances offered by the technologies?
- What are the performances needed by the application?
- Automatic Speech Recognition
  - NIST ASR evaluation
  - Voice Command and Voice Dictation achieve performances close to humans (2-4% WER)
  - Radio/TV broadcast transcription (10% WER) doesn't achieve human performances but sufficient for automatic indexing
  - Conversational speech and Meeting transcription performances are not sufficient (50% WER), especially for languages other than American English
- Machine Translation
  - Euromatrix+
  - MT achieve performances getting close to human translators for few EU language pairs (Maltese-to-English) (Machine: 70% – Human:80% BLEU)
  - Far behind for most EU language pairs
  - For Text translation: Speech translation ?

Where are we?

- New US Babel program (2012-2016)
  - Funded by US IARPA
  - ASR for other languages than American English
  - 22 languages in many different language families
    - Afro-Asiatic, Niger-Congo, Sino-Tibetan, Austronesian, Dravidian, Altaic,…
  - 4 “surprise languages”: the goal is to be able at final (2016) to develop an ASR system for a new language within a week
- Apple SIRI
  - Intelligent personal assistant and knowledge navigator
  - Initially (Fall 2011): 4 languages / 6 varieties
  - Next (iOS 6): 9 languages / 20 varieties
  - Consider countries not languages: also includes cultural dimension

Socially Aware Interactive Assistants

Research Priorities	Phase 1: 2013-2014	Phase 2: 2015-2017	Phase 3: 2018-2020
Interacting naturally with agents (terminals, ECA, robots, things) (in games, entertainment, education, communication, instrumented spaces, Call Centers, etc.)			

Roadmap (excerpt)

Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Interacting naturally with agents (terminals, ECA, robots, things)	**Provide usable human interface,** Reliable speech recognition, natural and intelligible speech synthesis, limited understanding and dialog capabilities	**Provide usable dialog interface,** Context and dialog aware speech recognition and synthesis. Recognize and produce emotions, understanding capabilities, context aware dialog, using other sensors (GPS, RFID, cameras, etc.)	**Provide multiparty (Human-Agents) interface**, multiple voices, mimicking, advanced understanding and advanced personalized dialog (indirect speech acts, incl. prosodics (lies, humor))

Roadmap (excerpt)

Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Using language but also other modalities, in parallel or together	**Multimodal interaction** (speech, facial expression, gesture, body postures)	**Multimodal dialog,** fusion and fission.	**Fleximodal dialog**, identification of best suited modalities
Conscious of its performing capacities	Confidence in hearing/ understanding, interactively recovering from mistakes	Ability to learn continuously and incrementally from mistakes by interaction	Unsupervised learning/ forgetting.

Roadmap (excerpt)

Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Exhibiting multilingual proficiency (speech-to-speech translation, interpretation in meetings and videoconferences, crosslingual information access) *(Theme 1)*	Ensure availability or portability to major EU languages (30). Recognize which language is spoken. Multilingual access to multilingual information	More languages (migrants, foreign languages), accents and dialects. Recognize dialects, accents. Exploit limited resources. Crosslingual access to information.	Speech translation in human-human interactions (multiple speakers speaking multiple languages). Cross-cultural support. Learn new language with small effort.

Roadmap (excerpt)

Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Resources *(Theme 4)*	**Install infrastructure** Collection of multi-task benchmark data. Collaborative production of semantically annotated data (multimodal). Incremental production of dialog data. In all EU languages	**Use infrastructure** More data. More languages	**Use infrastructure** More data More languages

Roadmap (excerpt)

Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Evaluation *(Theme 4)*	Multi-task benchmark evaluation. Measures and protocols for automated speech synthesis, dialog systems and speech translation evaluation. For all EU languages.	Measure of progress / Phase 1 More languages	Measure of progress / Phase 2 More languages.

Q/A

Thank you.

office@meta-net.eu

http://www.meta-net.eu
http://www.facebook.com/META.Alliance