META-NET

Preliminary Findings of the Interactive Systems Vision Group

META-VISION

Joseph Mariani

LIMSI-CNRS & IMMI

META-COUNCIL meeting, Brussels

About the Speaker

- Joseph Mariani

- Senior Researcher at CNRS

- Director LIMSI-CNRS and Head Human-Machine Communication Dept (1989-2000)
- Director ICT Dept at the French Ministry for Research (2001-2006)
- Director IMMI (LIMSI, KIT, RWTH) (2007-)

- President ESCA (1988-1993) and ELRA (2002-2004)

- Founding Member of META-NET
- Convenor Interactive Systems Vision Group
- Member of META-COUNCIL

The Vision Group
Interactive Systems

META VISION

- **Chair**
  - Alex Waibel (KIT, CMU & Jibbigo, Germany/USA)
- **Rapporteur**
  - Volker Steinbiss (RWTH & Accipio, Germany)
- **Convenors**
  - Joseph Mariani (LIMSI-CNRS & IMMI, France)
  - Bernardo Magnini (FBK, Italy)
- **Meetings**
  1. Paris, September 10, 2010
  2. Prague, October 5, 2010

The Vision Group Interactive Systems

- **Fields:** Telephone and mobile communication, Call centers, Internet navigation, Social Networks, Videoconferencing, Interpretation and translation, E-commerce, Finance, Healthcare, (Autonomous) Robotics, Car navigation, Security, Entertainment (Games), Edutainment, CALL (Computer Aided Language Learning), etc.

- **Stakeholders:** Telecom and internet companies/operators, Network companies (videoconferencing), Software companies, Translation companies, E-commercial companies, Banks, Robotics companies, Automotive industry, Security companies, Edutainment and game companies, Audiovisual sector, Service providers, etc.

- **Technologies:** Speech recognition, synthesis, understanding, Spoken and Multimodal Dialog, Speaker and language recognition, Emotion analysis, Voice search, Information Retrieval (Question&Answer), Text analysis and synthesis, Topic identification, Speech Acts analysis, Summarization, Machine translation and speech translation, Sign Language Processing, Image and gesture analysis and synthesis, Computer graphics, Computer vision, Acoustics, etc

Situation Interactive Systems

- Very long deployment process (started in the 1950's)
- (Successful) applications now in many different areas:
  - ***SmartPhones:*** *Dialling, Control (Samsung,...), Voice search (Google, Nuance...), Speech translation (Jibbigo...), eMail answering, Service (SIRI), Voice Dictation (SMS) (Nuance)*
  - ***On line Information:*** *, Call Centers, Customer care and technical support, (public) Information access (such as train time table) and transactions, Museum guides and public information kiosks*
  - ***Car*** *interfaces (in particular navigation)*
  - *Spoken dialog in* ***Video games*** *(MS Kinect, MILO)*
  - ***Military*** *applications (translation and training)*
  - ***Aids to the handicapped*** *(Reading machines for the blind, Sign language in railway stations)*

Enabling and Prohibitive Factors

SOCIETY & ECONOMY

- + *Ageing*
- + *Globalization*
- + *Automatization of society and more efficiency*
- + *Reduced costs of hardware*
- + *Huge market*
- + *Online availability (App Store)*
- + *Green technologies (Videoconf.)*
- – *Cultural, political and economic*
- – *Psychological (Human Factors)*
- – *Privacy and Ethics*
- – *Price for personalized systems*
- – *Business Models*

TECHNOLOGY & SCIENCE

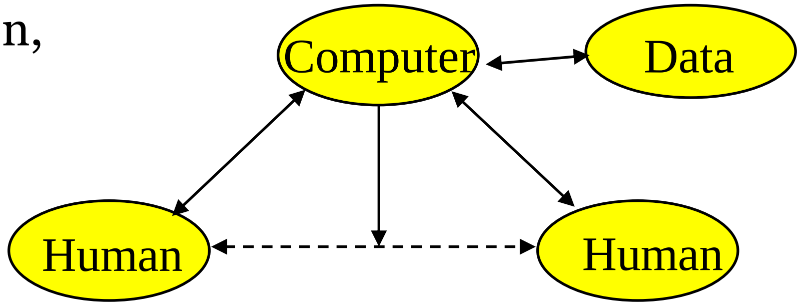
- + *Technology advances*
- + *Ubiquitous technology availability (at low cost)*
- + *Intelligent ambiance*
- + *User-centric, Crowd-sourcing*
- + *Low Barrier of Entry (Apps, Cloud)*
- + *LT Evaluation (TRL)*
- + *LR availability*
- – *Limited LT Evaluation*
- – *Limited LR availability*
- – *Limited knowledge*
- – *Technological complexity (//)*
- – *Server Cost*

Grand Visions 2020

The Multilingual Assistant

META-VISION

- Multilingual Assistants to Support Human Interaction
- Acting in various environments
  - Computer-Supported Human-Human Interaction, Human-Computer-Human Interaction, Human-Computer Interaction, Human-Artificial Agents (robots)



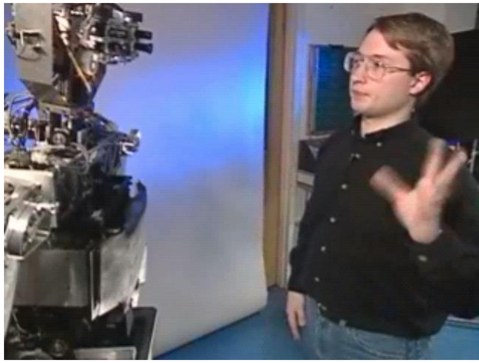
  - Personalized to user’s needs and environment
  - Learns incrementally and individually from all sources and interactions
  - Instrumented environments ((meeting) rooms, offices, apartments)
  - Instrumented open environments (streets, cities, transportation, roads)
  - World Wide Web, Virtual worlds (incl. (serious) games)

The Multilingual Assistant

- The Multilingual Assistant 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 all fluently speech, pronounce fluently written 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
  - Recognize people's identity, and their gender, accent, language, style
  - Move, manipulate objects, touch people (Robot)

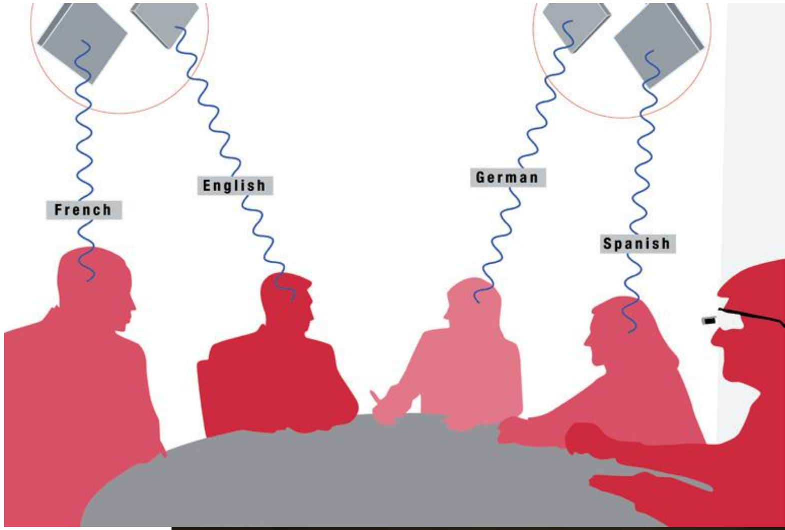
Domain specific visions

- Vision #1. Interacting naturally with Agents and Robots
  - Interaction with Conversational Agents (in games, entertainment, education, communication, etc), Interaction with robots, Spoken dialog, also in instrumented spaces
- Vision #2. Communicating everywhere
  - Mobile applications, Augmented Reality
- Vision #3. Technologies which help limitations
  - Crossmedia, Assistive applications, Sign Language
  - Adapted communication (cars, meetings)
- Vision #4. Community Building
  - Social networks and forums, Multiparty communication including several humans, several artificial agents/robots



Domain specific visions

- Vision #5. I speak your language!
  - Speech-to-Speech Translation, Interpretation in meetings / Videoconferencing, Cross-lingual information access
- Vision #6. Gutenberg still alive
  - Speech transcription, Close-captioning
  - Reading machine, Multimedia book
- Vision #7. My private teacher
  - Computer Aided Language Learning, Education
- Vision #8. I know who you are
  - Person, Biometrics
  - Gender, Style
  - Accent, Language



"대화 노력 계속"
보는 이중 구조로 전개될 전망입니다
워싱턴에서
KBS
变化代言人没有美国和北朝鲜在美国的立场，所以强调.
但是和在保守派的氣息作用，还预告，分析秧的下面方法对北朝鲜等.

Research/Technology Needs

Research/Technology Needs

- **Need #1. Better core Speech & Language Technologies**
  - More basic research (incl. physiological, perception and cognitive processes)
  - Better Speech Recognition
    - Lower the Word Error Rate, Accommodate noisy environment / far-field microphone, Open vocabulary, any speaker
    - Robustness: Noise, Cross-Talk, Distant Microphone
    - Lower Maintenance: Self-Assessment, Self-Adapting, Personalization, Error Recovery, Learning *and Forgetting* of New/Old
  - Better Speech Synthesis
    - Control parameters for linguistic/paralinguistic meaning, speaking style, voice conversion and emotion
  - Better Sign Language analysis / generation

Research/Technology Needs

META VISION

- **Need #2. From Recognition to Understanding**
  - Speech is Communication, not only STT / TTS
  - Communication should be Multimodal (text, speech, gestual, visual), Crossmodal and *Flexi*modal. Accept pragmatically best suited Modalities.
  - Semantic and pragmatic models of Speech and Language
    - Contextual Awareness: Model rapidly linguistic expression and domain
    - Self-Assessment: What is plausible?
  - Detect and recover interactively from mistakes
    - Learn continuously and incrementally from mistakes
    - Unsupervised or by interaction
  - Include paralinguistics (prosody analysis, visual cues): emotion, laughs
  - Necessitates cooperation with psychologists and communication experts
  - Production of adequate Language Resources, Annotation: Huge effort
    - Methods to better use massive amounts of poorly annotated data

Research/Technology Needs

- **Need #3. Going to Natural Dialog**
  - Spoken / Multimodal dialog
  - “Transparent” systems
    - Multiple microphones in (non-stationary) noise, Open microphone, Multiparty conversations (humans, artificial agents, robots), cocktail party effect, bi-modal communication (lip reading)
    - Use of other sensor-devices: RFID, motion capture, GPS, etc
  - Dialog models
    - Faster Dialog Models
    - Pro-active (not only reactive)
    - Detect that a voice emission is in machine intention, Interpret a silence
    - Process direct/indirect Speech Acts, including lies, humor…
  - Study of Human factors, and usability
  - Define dialog systems evaluation metrics / protocols
  - Produce LR (acquisition /annotation) from Real World
    - Incremental system design
    - Use of data available on internet (conversation, talks shows)

Research/Technology Needs

- **Need #4. Handling Multilingualism**
  - Interactive systems should cover, or be easily portable to all EU languages
    - 23 official languages + regional languages (catalan, basque, etc)
  - General Language Portability: From few to *Many* Languages
    - Language Support for European to/from non-European languages
  - Speech Translation in Human-Human interaction (e.g. meetings)
    - Speech translation among multiple human users, speaking different languages
  - Deal with Languages, Accents and Dialects effectively
    - Should recognize language, gender and accents
    - Cross-Cultural Support
  - Provide cross-lingual access to information and knowledge
  - Availability of Multilingual Resources (data, tools)
    - Taggers, Morph Decomposition, Lexica, etc.
  - Availability of Language Resources and Evaluation in all languages, or adaptability within a language family

Summing up: Topics with Strong Visionary Potential

META-VISION

- Domain-specific
  - The Multilingual Assistant
  - Provide interaction between humans, agents and “intelligent” spaces
  - Able to transfer information across medias and across languages
  - Demonstrate many functionalities
  - Which correspond to many application areas
- Domain-independent
  - Single European Information Space based on Multilingualism
  - As a guiding principle, all EU languages should benefit from LT