

META=NET

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

January 25, 2013 – Berlin, Germany



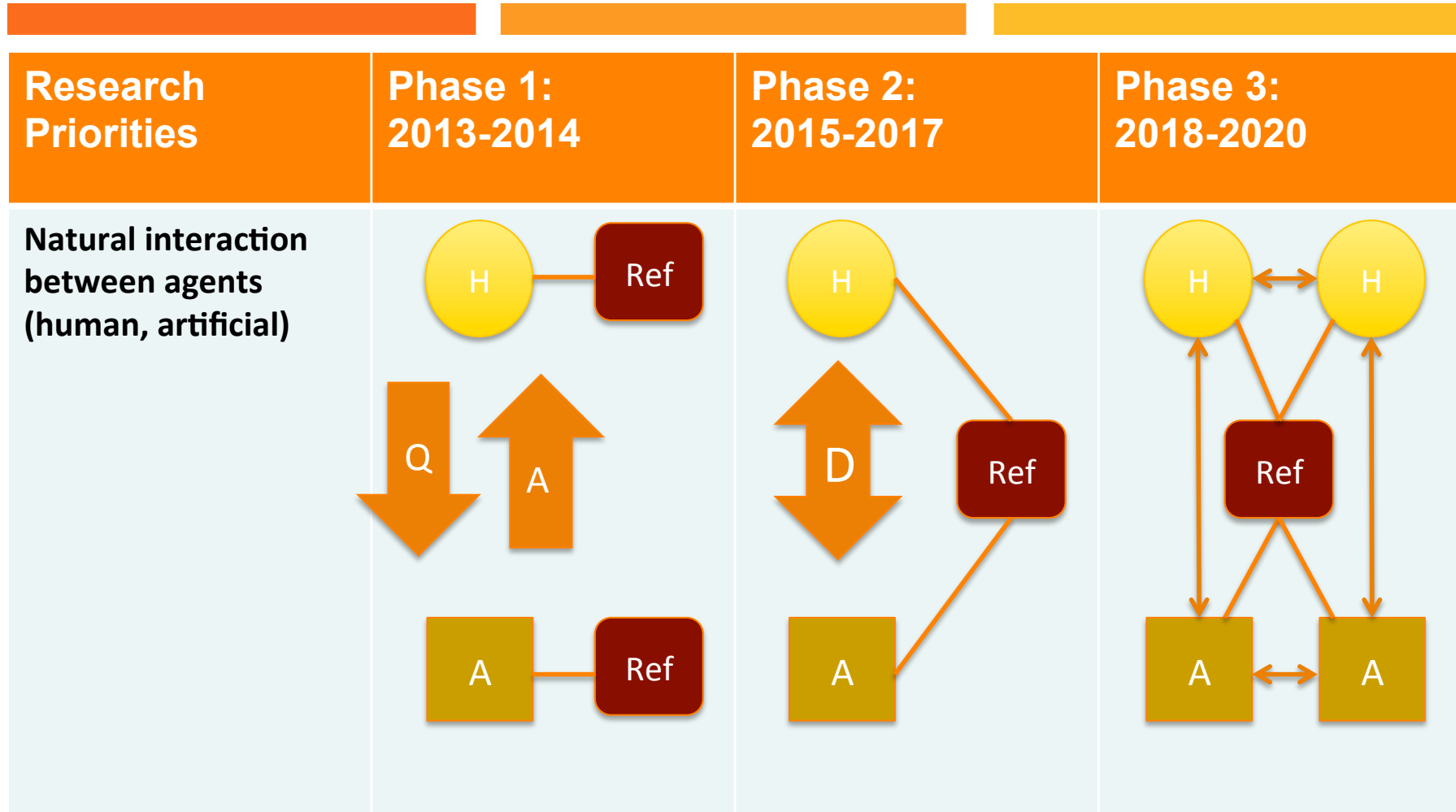
Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

Vision: Socially Aware Interactive Assistants

- 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 (incl. 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 (apartments, streets, transportations)
 - Act on behalf of humans over the Internet (social networks) and in Virtual/ Augmented Worlds
 - 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)

- ❑ Where are we?
 - State-of-the-Art ASR, SS, SLT, SDS,...
- ❑ Try to define time to deliver
 - Should reach “good enough” quality (Evaluation)
 - Core Abilities
 - Interacting naturally with agents, with and in groups
 - In a multimodal and multilingual way
 - Conscious of its performing capacities
 - Exhibiting robust performances everywhere
 - Domain-specific Abilities
 - Referring to written support (speech-text)
 - Overcoming handicap obstacles (crossmedia)
 - Providing personalized training
 - Resources and Evaluation

Roadmap



Roadmap (Core abilities)

Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Interacting naturally with agents (humans, smartphones, Embodied Conversational Agents (ECA), robots, things) In (serious) games, entertainment, education, communication, instrumented spaces, Call Centers, etc.	Provide usable human interface, Reliable speech recognition. Natural and intelligible speech synthesis. Limited understanding and dialog capabilities, from domain-specific to open domain. Person identification (gender, age, style, social, language, accent, identity...)	Provide usable dialog interface, Context & dialog aware speech recognition and synthesis. Recognize and produce non verbal signals (cough, laugh) and emotions. Extended understanding capabilities. Context aware, proactive dialog, also using other sensors (GPS, RFID, cameras)	Provide multiparty (Humans-Agents) interface, multiple voices for multiple agents, mimicking someone's voice. Advanced understanding and personalized dialog handling (indirect speech acts, anaphors, metaphors, paraphrases, incl. prosodics (lies, humor))

Roadmap (Core abilities)



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. Crossmodal dialog , ability to carry same information through different modalities	Fleximodal dialog , identification of best suited modalities. Context awareness
Conscious of its performing capacities	Confidence in hearing/ understanding , Interactively recovering from mistakes	Ability to learn continuously and incrementally from mistakes through interaction	Self assessment and self adaptation. Unsupervised learning/ forgetting.

Roadmap (Core abilities)



Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Exhibiting robust performances everywhere (indoor and outdoor environments, mobile applications, virtual/augmented reality)	Develop distant communication (remote microphone, room acoustics, directional loudspeakers)	Robust to noise and crosstalks. Detection that a human voice emission is in machine intention. Adapted communication (cars, meetings)	“Transparent” systems, with open microphone, dealing with multiparty conversations (humans ECAs, robots), Cocktail Party effect. Interpret a silence.
Interacting naturally with and in groups (in social networks and forums, with several humans/agents)			Socially aware dialogs Understanding of socio-emotional functions of communicative behavior.

Roadmap (Core abilities)



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) <i>(Theme 1)</i>	Ensure availability or portability to major EU languages (30). Recognize which language is spoken. Multilingual access to information	More languages (regional, 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 (Domain-specific)



Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Referring to written support (Voice dictation, subtitling of audiovisuals, reading machines, multimedia books,...)	Better transcription of dictated speech or audiovisual material. Out-of-Vocabulary words, New Named Entities, Any speaker. Better correction interfaces. Confidence measures. Better quality speech synthesis for reading machines. Improved naturalness. Expression beyond the sentence.	Better transcription of dictated speech or audiovisual material. Increased accuracy, larger use of syntax. Better quality speech synthesis for reading machines. Larger use of syntax. Multiple voices.	Better transcription of dictated speech or audiovisual material. Increased accuracy, larger use of semantics and pragmatics (text & context understanding) Better quality speech synthesis for reading machines. Larger use of semantics and pragmatics (text & context understanding)Multiple voices, adapted to characters.

Roadmap (Domain-specific)



Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
<p>Overcoming handicap obstacles By means of suitable technologies (sign language generation/understanding, assistive applications, adapted communication to suit cognitively impaired, etc.)</p>	<p>Provide crossmodal and crossmedia LT for the handicapped Usable ASR and intelligible speech synthesis for the blind and the deaf (HCI & Information access). Usable Sign Language generation for the deaf.</p>	<p>Usable Sign Language recognition and understanding for the deaf</p> <p>Speech to Sign Language Translation</p> <p>Oral description of images, scenes, videos for the blind.</p>	<p>Usable Sign Language to Speech Translation</p> <p>Translation across Sign Languages</p> <p>Multiparty (in social networks and forums, involving several humans with various handicaps)</p>

Roadmap (Domain-specific)



Research Priorities	Targeted Breakthroughs in Technology Development		
	2013-2014	2015-2017	2018-2020
Providing personalized training and Knowledge Access (CALL, CAE, self assessment in general) <i>(Theme 2)</i>	Computer Aided Language Learning (CALL) , including reliable ASR and intelligible Speech synthesis for many languages. On-line education , translation of lectures, incl. slides.	Q&A in on-line education. Automated students proficiency assessment (such as PISA)	International remote education , including collaborative activities of students and teachers, eventually through agents. Personalized training. Students Self Assessment.

Roadmap (Resources / Evaluation)



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

Roadmap (Resources / Evaluation)



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

- ❑ Are there any major missing issues?
- ❑ Do you consider the plan feasible and credible?
- ❑ Are there any obstacles to its implementation (e.g. some necessary technologies to implement the Roadmap are still too immature)?
- ❑ Are there any other prerequisites (technical, infrastructural, social, legal, psychological, ethical factors) that could impede the process?
- ❑ Is the breakdown into the three phases reasonable?
- ❑ Should the 4 Roadmaps be linked?
- ❑ What is the relationship with FP7 Call 10 Roadmaps?

office@meta-net.eu

<http://www.meta-net.eu>

<http://www.facebook.com/META.Alliance>

25.01.2013

META-NET Network Meeting - Berlin

14