ROCKIT: Roadmap for Conversational Interaction Technologies

Citation for published version:
Renals, S, Carletta, J, Edwards, K, Bourlard, H, Garner, PN, Popescu-Belis, A, Klakow, D, Girenko, A, Petukhova, V, Wacker, P, Joscelyne, A, Kompis, C, Aliwell, S, Stevens, W & Sabbah, Y 2014, ROCKIT: Roadmap for Conversational Interaction Technologies. in *RFMIR '14 Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges.* pp. 39-42. https://doi.org/10.1145/2666253.2666262

Digital Object Identifier (DOI):
10.1145/2666253.2666262

Published In:
RFMIR '14 Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges

ROCKIT

Roadmap for Conversational Interaction Technologies

S Renals, J Carletta, K Edwards
University of Edinburgh
Edinburgh, UK
s.renals@ed.ac.uk

H Bourlard, P Garner
A Popescu-Belis
Idiap Research Institute
Martigny, Switzerland

D Klakow, A Girenko
V Petukhova
Saarland University
Saarbrücken, Germany

P Wacker, A Joscelyne
LT-Innovate
Brussels, Belgium

C Kompis, S Aliwell
Vodera
London, UK

W Stevens, Y Sabbah
Europe Unlimited
Brussels, Belgium

ABSTRACT

ROCKIT is a strategic roadmapping action in the area of multimodal conversational interaction technologies funded as a support action by the EU during 2014 and 2015. We envisage a future in which human-human, human-machine, and human-environment communication are not hampered by differences in language capability, accessibility, or knowledge of the technology, and where security and privacy are built in. These future conversational interaction technologies will enable interaction, collaboration, creativity, and information access within a vast, dynamic, heterogeneous, and partly ephemeral information space. ROCKIT is developing a roadmap to achieve this vision, linking research and innovation activities, and connecting a broad range of stakeholders. In this paper we present the ROCKIT roadmapping process, together with five target scenarios, which we believe can form a basis for discussion and engagement at the ICMI workshop which can further progress the community roadmap.

Categories and Subject Descriptors

H.5.0 [**Information Interfaces and Presentation**]: General—*roadmap*

Keywords

Multimodal interaction, Conversational technologies

1. INTRODUCTION

Over the past few years significant progress has been made in fundamental aspects of multimodal, speech, and language processing. This has resulted in the deployment of significant conversational interactive systems, often with close ties to physical devices, such as Apple's Siri and Microsoft's Cortana. These systems form a significant landmark since it is the first time that conversational systems have had widespread consumer acceptance. However, despite their undoubted success, these systems have many limitations. They typically operate in a limited domain, the interaction is usually far from natural, and for broader domain systems — such as voice search — the interaction is very shallow. These systems are also highly limited in the way they are able to use richer sources of knowledge.

There are now significant research and innovation opportunities in the area of conversational interaction technologies – Figure 1 presents the strengths, weaknesses, opportunities, and threats related to this field. Allied to this, there are a number of technology trends such as the rise of cloud computing, the prevalence of social media, and the move towards open linked data, which may be enabled by the availability of open platforms and architectures of increasing scale and ubiquity.

Building on the META-Net Strategic Research Agenda [3], we have identified a number of societal drivers which offer significant challenges for conversational, interactive technologies:

- **Demographic change**: an ageing population requires natural, interactive technologies (such as smart homes) that provide support in a friendly, personalised way.
- **Personal information services**: as Siri and Google Now indicate, smart personal assistants have the potential to change the way people interact with personal devices, with the network, and with information.
- **Global computing and human communication**: smart video conferencing (eg with speech-to-speech translation) has the potential to reduce carbon footprints while enabling improved multilingual and multicultural communication.
- **Learning technologies**: interactive, conversational technologies, with a rich knowledge base, can provide personalised learning experiences.
- **Disabled population**: the development of assistive technologies including spoken and gestural interaction are key to allowing people disabilities to live independently and to have equal opportunities. 10% of European citizens have permanent disabilities.
- **Cultural heritage**: speech and multimodal technologies have the potential to make vast media archives accessible through search and translation.

RFMIR'14 November 16 2014, Istanbul, Turkey

ACM 78-1-4503-0615-7/14/11.
http://dx.doi.org/10.1145/2666253.2666262

STRENGTHS

- Conversational interaction systems have become mainstream (eg Siri, Google Now).
- Strong research base.
- Global reach of systems is a strong driver for multilingual conversational systems.
- Demand: No longer a need to create a market — voice search, Siri, etc have created a demand.
- There exists an enabling infrastructure: from networks and cloud computing, through to availability of open source software systems.
- Significant recent technology advances: reliable, multilingual ASR; intelligible multilingual speech synthesis; trainable spoken dialogue systems; initial systems for generation and recognition of affect and social signals; initial systems for multimodal processing.

WEAKNESSES

- Technologies are fragile and lack robustness to changing conditions.
- Complex software infrastructure required for conversational processing.
- The European innovation ecosystem is relatively small in this area.
- Limited data is available to develop systems in some areas, eg human-computer dialogue.
- Most systems currently have shallow interactions.
- Limitations of current technology: speech recognition makes errors in broad domains and realistic environments; unnatural-sounding speech synthesis, with very limited control of expression and emotion; dialogue systems operate in limited, simple domains; very few systems make use of multimodal, heterogeneous inputs.

OPPORTUNITIES

- New SMEs, fast tech transfer from research.
- Open data can start to enable conversational services.
- Social media is naturally conversational.
- Cloud computing makes it possible to provide sophisticated applications running on many scales of device.
- Open source systems make it possible to develop sophisticated systems quickly.
- Many drivers from societal challenges (eg demographics, global communication, security).
- Well-matched to computing trends (eg wearables).

THREATS

- Large corporations will dominate the field and make it difficult for SMEs to enter.
- Much data is still proprietary (eg recordings of human-computer dialogue system interactions), and serious privacy issues apply to data.
- Many conversational applications may need super-human performance to be acceptable to users, and may need to solve the complete AI problem.
- Privacy issues may lead to user distrust of conversational systems.

Figure 1: SWOT Analysis: conversational interaction technologies.

- **Security and law enforcement**: there is great potential for conversational technologies in areas as diverse as emergency response, homeland security, and capture and analysis of interviews and conversations.
- **E-participation**: conversational interaction technologies have the potential to analyse streams of multimodal, multilingual data in real-time.

The aim of the ROCKIT project is to develop a strategic roadmap for research and innovation in the area of natural conversational interaction. ROCKIT is funded as a Support Action by the European Union FP7 programme during 2014 and 2015, and is thus particularly concerned with action at the European level. Europe has a very strong research community in multimodal interaction, and speech and language processing, but the level of commercial and innovation activity over the past couple of decades has been small considering the quality and size of the research base. The research & innovation roadmap that is being developed by ROCKIT will integrate the vision and innovation agendas of a broad range of stakeholders in the field across Europe – including small and large companies, universities and research institutes, funding agencies, venture capitalists, and governmental organisations). One of the key goals of ROCKIT is to mobilise the innovation and research communities across Europe, bringing together universities and research institutes with commercial organisations at all scales. There is a specific focus on SMEs, which represent the majority of fragmented commercial activity and innovation in Europe.

Vision: ROCKIT is driven by a broad vision for conversational interaction technologies. We see a future, in 2024, where human-human, human-machine, and human-environment communication is not hampered by differences in language capability, accessibility, or knowledge of the technology, and where security and privacy are built in. These future conversational interaction technologies will enable interaction, collaboration, creativity, and information access within a vast, dynamic, heterogeneous, and partly ephemeral information space.

ROCKIT will develop a roadmap that will enable this 10-year vision to be realised. This will involve connecting the strong R&D base, commercial and industrial activity, and policy makers – at the EU and national levels. This requires:

1. An open roadmapping process, involving the full range of stakeholders;
2. The establishment of a broad-based community bridging the gap between research and innovation, with a very strong emphasis towards SMEs;
3. The development of a number of target scenarios as a focus for the vision;
4. Planning for an infrastructure to achieve the outcomes, including the development of an open platform for conversational interactive technologies.

In Section 2 we outline the process we are using to develop the ROCKIT roadmap, summarising the current state of the process. In Section 3 we discuss five target scenarios that have arisen from the roadmapping process. In Section 4 we set a possible framework to progress the discussion – and to improve the roadmap and scenarios – that we hope can take place at the ICMI workshop. Part of this will involved developing practical steps to build a research and innovation community in the area, and links to related activities.

2. ROADMAPPING

Strategic roadmapping is a process to communicate the development and integration of science and technology, and their incorporation into products or services to address business or societal opportunities. It can provide ways to identify, evaluate, and select alternatives that can be used to achieve a route to some vision of the future (derived from application or usage scenarios). Roadmapping highlights gaps between what is expected to happen anyway and what is needed thus helping plan resources and devise focused strategies to enable the vision to be realised. The value of roadmapping lies in the alignment of the different perspectives and the search for connections between them. Visually enhanced in a graphical format, a roadmap is an effective decision support tool.

In ROCKIT we are using a roadmapping methodology driven by a number of workshops that enable input across different types of stakeholder [2]. In contrast to many technology roadmapping efforts, ROCKIT will go beyond descriptive statements about the future. It will instead present a more structured and visual mapping of the various strands of activity that need to happen in order for a future vision to be realised. The structure emphasises distinct layers relating to markets and drivers, products and services, and enabling technologies that allow the reader to easily identify the aspects of most relevance to them. Essentially, it starts by summarising the key scenarios and future use cases, and the associated drivers and constraints. It then defines a vision of the future and assesses the position of the community and its ability to deliver that vision. The process then establishes the possible routes and required developments to fulfil that vision. With this mapping done it is then possible to highlight key enablers, technology gaps, risks, and resource gaps. In other words, the process is predominately about synthesising the knowledge in a community and bringing it together in an accessible format that allows easy identification of trends, gaps, opportunities, and resources.

The ROCKIT roadmap is structured in terms of three main layers: Vision (drivers and scenarios); Innovation (markets, products, and services); and Research (science and technology). A fourth layer concerns Resources, to which the other layers will contribute. Each layer is structured in terms of sub-layers, for example technology areas in the case of Research, and vertical market sectors in the case of Innovation.

The ROCKIT roadmapping process is discussed in greater detail in [1].

To enable inputs from different stakeholders, we organised a number of workshops for the different layers: Innovation and Vision workshops were held at the 2014 LT-Innovate Summit[1], and Research workshops took place at ICASSP–2014 and LREC–2014, as well as at the University of Edinburgh and Saarland University. In total, about 100 people participated in these workshops.

3. SCENARIOS

One of the main ways of developing the roadmap, and of making tangible links between research and innovation is through the development of a small number of *target scenarios*. Each scenario will include its societal and technological drivers, R&D aspects, market and business drivers, and potential testbeds. We identified a number of common themes coming out of the ROCKIT workshops, in particular accessibility, multilinguality, the importance of design, privacy by design, systems for all of human-human, human-machine, and human-environment interactions, robustness, security, potentially ephemeral interactions, and using the technology to enable fun.

Building on these themes, together with the different social, commercial, and technological drivers, we have identified five possible target application scenarios:

1. **Adaptable interfaces for all**
2. **Smart personal assistants**
3. **Active access to complex unstructured information**
4. **Communicative robots**
5. **Shared collaboration and creativity**

We envisage all these scenarios to be operating in multilingual and multisensor contexts.

3.1 Adaptable interfaces for all

Multimodal conversational interfaces have the potential to adapt automatically to the user, to the environment, and to the users state. For instance: interfaces adapted to older people will take account of cognitive, auditory, visual, and articulatory ageing; interfaces will adapt to what a user is doing (working in a noisy, hands-free environment, rushing for a train); interfaces will help to rehabilitate users with some kind of disability. By 2024 we envisage multimodal interfaces designed for *everyone*, and which can adapt to perform optimally in a given environment, and with respect to the available sensors.

Example use cases:

- Rehabilitation / assistive technology
- Access for older people
- Usage across different environments which may limit possible interface modalities
- Home and domestic applications
- Applications for travel and tourism
- Applications for children and young adults

3.2 Smart personal assistants

Smart personal assistants (such as Siri and Cortana) and applications such as Google Now have stimulated huge growth in the area of conversational interaction technologies – recent analyst reports from Gartner, Opus Research, and Grand View Research have predicted annual growth rates of over 30% for the market in speech/language driven smart personal assistants. However, current systems are still far from the capabilities that we envisage for 2024. Enriched per-

[1] `http://www.lt-innovate.eu/event/LTi-events/lt-innovate-summit-2014-brussels/presentations`

sonal assistants would be capable of deep(er) language understanding and generation, and able to make effective use of environmental, informational, and social context. In particular, this scenario focuses on systems which display social awareness, and are able to behave naturally in multiuser situations, and capable of proactive behaviour. They will also be able to integrate heterogeneous sources of knowledge.

Example use cases:

- Smart mobile agents which are capable of deeper natural language and multimodal interaction, possibly focused on specific domains, and capable of rich question answering
- Systems for devices where "traditional" interfaces (keyboard, mouse, trackpad, touchscreen, etc.) are not usable (eg small wearable devices) or not appropriate ("companion" systems)
- Interfaces to smart environments
- Many vertical market sectors, with domain-specific assistants: shopping, travel, social service planning, learning and tutoring

3.3 Active access to complex unstructured information

Linking knowledge to rich interaction will enable the development of agents which can search proactively and can make inferences from their (possibly limited) knowledge, to enable people to be notified of relevant things faster, and to help people reach understanding of complex situations involving many streams of information. By 2024, we envisage such systems which operate on huge, dynamic, heterogeneous data streams, and which also provide powerful approaches to navigation and visualisation. In many cases it will be important to consider issues such as data provenance, trust, privacy, security, and rights. A key issue for this scenario, in particular, relates to positive (democracy) and negative (surveillance) aspects of large scale multimodal knowledge integration and access.

Example use cases:

- Enterprise knowledge management
- Crisis management, with strong real-time and partial knowledge constraints
- Social computing - health, eco, energy

3.4 Communicative robots

Robotics is predicted to be a huge growth area, and many robots will operate in communicative settings. By 2024, we envisage robots that are capable of human-like multimodal embodied communication, with the ability to operate in social environments with personality and capable of generating and interpreting social signals, able to learn through communication.

Example use cases:

- Human like domestic assistants
- Robot-robot communication (eg in hostile environments)
- Therapy / education (eg autistic spectrum)
- Industrial robot management
- Surgical and medical robots

3.5 Shared collaboration and creativity

The previous scenarios have generally focused on different varieties of conversational human-computer interaction, although with the computing platform including "traditional" smart phones, small wearable devices, robots, and smart spaces. The final target scenario focuses on empowering and augmenting human-human communication. By 2024 we envisage applications which help people to be more creative more of the time (especially in group situations), new approaches to social sharing (across languages), design-enabling platforms which enable people to build their own tools, and systems to enable groups (at all scales) to collaborate with shared goals, facilitating problem solving, and providing powerful mechanisms for engagement.

Example use cases:

- Social platforms that incentivise people to share and collaborate;
- Enhanced meetings (remote and face-to-face) with speech translation, content linking, etc.;
- Applications for visual, multimodal and audio artistic creativity;
- Games (entertainment and serious games);
- Shared understanding and communication, for example in multi-discipline professional teams (eg medicine)
- Shared learning, MOOCs, peer-to-peer learning

4. DISCUSSION

The aims of this paper are to present the ROCKIT roadmapping process, and the target scenarios for multimodal conversational interaction, as well as to stimulate discussion about the future development of the roadmap. There are several questions which it would be good to discuss at the workshop:

- Are the scenarios well thought out?
- Do the scenarios match the key drivers?
- What are the emerging trends, markets, or technological developments that require particular attention?
- Have the key technological, societal and economical innovation drivers been identified?
- What could be the main R&D milestones?
- When should we expect major innovations?
- How should the scenarios rank in terms of readiness level, economic impact, and social impact?

Finally, an important goal of ROCKIT is to link the different stakeholders involved in this area. A key event will be the launch of the Conversational Interaction Technologies Innovation Alliance (CITIA, `http://citia.lt-innovate.eu`) in early 2015, in collaboration with existing organisations including META and LT-Innovate.

ACKNOWLEDGMENTS

This work is supported by the EU Seventh Framework Programme (FP7/2007-2013) under grant agreement No 611092 (ROCKIT Support Action).

5. REFERENCES

[1] C. Kompis and S. Aliwell. ROCKIT roadmap specifications. ROCKIT Deliverable D4.1, 2014. `http://groupspaces.com/ROCKIT/pages/project-results`.

[2] R. Phaal, C. Farrukh, and D. Probert. *Roadmapping for Strategy and Innovation: Aligning Technology and Markets in a Dynamic World*. University of Cambridge, Institute for Manufacturing, 2010.

[3] G. Rehm and H. Uszkoreit, editors. *Strategic Research Agenda for Multiligual Europe 2020*. Springer, 2012. `http://www.meta-net.eu/sra-en`.