Roadmap for Conversational Interaction Technologies (ROCKIT)

Support Action
FP7–611092
D6.4 – Final Public Report

Nature	P	Work Package	WP6
Due Date	30/11/2015	**Submission Date**	23/12/2015
Main authors	Steve Renals		
Co-authors			
Reviewers	all		
Keywords	Public report		
Version Control			
v0.1	**Status**	Draft	07/12/2015
v1.0	**Status**	Final Draft	23/12/2015
v1.1	**Status**	Submitted	09/01/2015

Roadmap for Conversational Interaction Technologies

Public Report

Imagine 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. Such a future could be enabled by **conversational interaction technologies** which will enable interaction, collaboration, creativity, and information access within a vast, dynamic, and heterogeneous information space.

This was the guiding vision of the ROCKIT Support Action, funded by the European Union FP7 programme from December 2013 for two years in order to address Objective ICT–2013.4.1: Content analytics and language technologies – Natural spoken and multimodal interaction. ROCKIT achieved its four main objectives:

1. The construction of a **strategic roadmap** for conversation interaction technologies which can form the basis of a research and innovation agenda in the area.
2. The formation of a **stakeholder community** including SMEs and corporations, universities and research institutes, European institutions, investors, and specialized media – CITIA, the Conversational Interaction Technologies Innovation Alliance, (http://citia.eu)
3. The development of a set of **target scenarios** as focus the strategic vision.
4. A plan for an **infrastructure** to facilitate these outcomes based on a shared layer of open access multilingual data, core open source software components, and a broader open software architecture.

During 2014 and 2015 we constructed a technology roadmap that to enable the ROCKIT vision to be realised. The roadmapping process was carried out at the European level, connecting the strong R&D base, with commercial and industrial activity, and policy makers at the EU and national levels. Over 200 experts were consulted, and the complete exercise is viewable at http://sharpcloud.com/ROCKIT.

As part of this roadmapping process, we constructed a SWOT analysis of conversational interaction technologies, based on the inputs we received from the experts. The SWOT Analysis (summarised on the next page) indicates that Europe has a strong research infrastructure, and there are many opportunities, but problems arise relating to the weak link between research and innovation, and owing to the increasing dominance of a small number of large corporates who operate in the area.

STRENGTHS
- European research community
- Strong research infrastructure
- Demand for CIT systems
- Cooperative SME culture
- 50 languages in use in EU

WEAKNESSES
- Weak innovation infrastructure
- Fragmented market
- Fragility of current systems
- Limited control of platforms
- Availability of multilingual data

OPPORTUNITIES
- Open data and infrastructure
- Digital single market
- Data science as a service
- Social drivers
- Many feasible R&I scenarios

THREATS
- "Solved problem"
- Proprietary platforms and data
- Google/Apple/etc dominance
- No coherent vision
- Privacy/trust/security concerns

Conversational Interaction Technologies in Europe:
SWOT Analysis

BRIDGING RESEARCH AND INNOVATION

CITIA was established in February 2015 at the launch of the ROCKIT roadmap. A key aim of CITIA is to connect the European research base with commercial & industrial activity – especially startups and SMEs, and those companies that are scaling-up. To do this CITIA is building a community that crosses research and innovation and includes multiple stakeholders: Industry – both SMEs and large companies; Research – universities and research institutes; Investors and venture capital; and Government, EC, and policy makers.

TRENDS

The roadmap identified a number of trends in conversational interaction technologies, including:
- Rapid **technological progress** across many key areas, including speech recognition, spoken dialogue systems, and face analysis. Many of these technological advances have been driven by recent steps forward in deep learning.
- Improved **robustness & adaptation** to real-life contexts, for example on smartphone apps
- Focus on **composed services, combining multiple component technologies and potentially multiple interface modalities** – this is particularly important for cloud-based services
- **Interaction** driven and informed by **analytics** – intelligent interaction relies on an analysis of the context of the interaction
- Rich dialogues taking account of **social and cultural context** – conversational technologies must taken into account the language of the users, and the place and situation in which they are being used
- **Privacy** concerns – people are becoming increasingly concerned about what happens to their data, something that is especially important given the personal nature of multimodal conversational data
- **Data** – innovation across speech and language technologies requires more data for the machine learning algorithms that define the state of the art. This is particularly true when multilingual systems are required; beyond a few major languages there is limited data to support the multilingual diversity of Europe.

SCENARIOS

The roadmap is organised around five target research and innovation scenarios.

1. **Adaptable Interfaces for All.**
   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 case: Automatic Interface Subtitling.
2. **Smart Personal Assistants.**
   By 2024 smart personal assistants will be capable of deep language understanding and generation, and able to make effective use of environmental, informational, and social context.
   ***Example use case*:** Smart Assistants for Kids.
3. **Active Information Access.**
   By 2024, multimodal interactive systems will operate on huge, dynamic, heterogeneous data streams, providing powerful approaches to navigation and visualisation.
   ***Example use case*:** Immersive Intelligence.
4. **Communicative Robots.**
   By 2024, robots will be capable of human-like multimodal embodied communication, with the ability to operate in social environments.
   ***Example use case*:** Industry 4.0
5. **Shared Collaboration and Creativity.**
   By 2024 systems for collaboration and creativity will help people to be more creative more of the time, especially in group situations.
   ***Example use case*:** Gaming Environments.

ACHIEVING THE VISION

During the roadmapping process, there was consensus across the experts that multilingual interaction technologies can enable and grow the digital single market – soon everything will be an interface, particularly when considering the internet of things. New interface technologies will be required, particularly for small devices not well served by current interfaces. Robust multilingual language & speech technologies and services will be at the heart of future interfaces.

These are ambitious goals: attacking the hard research problems in multimodal language understanding and generation is the best way to drive high-impact innovation in the area. However to create major impact, we need to further build the mutually reinforcing relationship between research & innovation, initially through building a strong research and innovation community in the area.

One of the main points, reiterated many times across the entire stakeholder community, was the need for multilingual data and baseline systems. Although there is now a strong software layer for building speech recognition, machine translation, and speech synthesis systems – arising from EU projects such as Simple4All and EU-Bridge – both commercial and research users have to put in significant effort, including basic data collection, when extending a system to new languages.

This is a serious issue if we are to build a multilingual digital single market in Europe. In order to crack the language barrier, European speech technology systems need to support the 24 official languages and the 5 “semi-official” languages, as well as the languages required for the global market. And this is not to mention the many regional and minority languages in use in Europe. If we take the 25 languages comprising the 26th to the 50th most used in Europe (Finnish – ¡ Montenegrin) we find there are over 50 million speakers of these languages residing in Europe.

THE CITIA BASELAYER

To address this, we propose the construction of an open multilingual infrastructure for speech technology. The aim of the **multilingual CITIA baselayer** is to enable and accelerate both research and innovation, and to provide one of the key blocks that will be essential to the multilingual digital single market. In a nutshell we propose:

1. The creation of **core, open access data**, initially across 40–50 languages, suitable for training broad-domain speech recognition and speech synthesis systems. The initial layer will require 100 hours per language of transcribed speech for constructing baseline ASR systems, plus about 10 hours per language to construct a male and a female synthetic voice per language.
2. The development of **baseline open source systems** for speech recognition and speech synthesis across these languages, building on existing open source projects (e.g. Kaldi)
3. The dissemination of **practical recipes** for collecting and transcribing multilingual data for training and evaluating speech technology systems.
4. The deployment of **cloud-based reference systems** to enable people to prototype and test new multilingual spoken applications.

This is an ambitious vision, but we believe it can be achieved if we coordinate at the European level, and take advantage of the strong and diverse community committed to multilingual speech technologies, **Achieving the vision** will require:

- **Standard recipes and processes for multilingual data collection and system development.** This needs to include protocols for data collection across a variety of genres (read speech, conversational speech, lectures, broadcast speech, …), combined with standard transcription procedures and recipes to build high quality open source baseline systems.
- **Cloud-based backend with standard APIs** to enable the rapid deployment of prototype multilingual systems that can be used to develop new multilingual applications.

The CITIA baselayer should be open source and open data, using permissive licenses such as CC-BY and the BSD-type open source license. Open source and open data is essential to stimulate and accelerate research in the field, and seed a sustainable research and innovation community.

Establishing the multilingual baselayer will require a **core project team** concerned with project management, definition of data collection process, open source recipes, building baseline systems, and the deployment of reference cloud-based systems. Each **individual language** will require data collection, transcription, and quality control, following the guidelines and processes defined by the core team, and will probably require 2-3 person years of effort per language.

We believe the project will rapidly become self-sustaining owing to the use of open data and open source software, which will seed a research and innovation community. The project will grow through a virtuous cycle of data-systems-apps-users-data-….

Join the community and help establish the CITIA Multilingual Baselayer!

ANNEX: THE ROCKIT CONSORTIUM

The ROCKIT consortium comprises six partners:

- University of Edinburgh, UK (coordinator); http://www.inf.ed.ac.uk/
  - contact: Steve Renals (s.renals@ed.ac.uk)
- Europe Unlimited, BE; http://www.e-unlimited.com
  - contact: Youssef Sabbah (youssef@e-unlimited.com)
- Idiap Research Institute, CH; http://www.idiap.ch
  - contact: Hervé Bourlard (bourlard@idiap.ch)
- LT-Innovate, UK; http://www.lt-innovate.eu
  - contact: Philippe Wacker (phw@lt-innovate.org)
- Saarland University, DE; http://www.uni-saarland.de
  - contact: Dietrich Klakow (dietrich.klakow@lsv.uni-saarland.de)
- Vodera, UK; http://www.vodera.com
  - contact: Costis Kompis (costis.kompis@vodera.com)

For more information about ROCKIT, please see http://rockit-project.eu

For more information about CITIA, please see http://citia.eu

ENDPAGE

ROCKIT

FP7-611092

D6.4 – Final Public Report