Towards an Interaction Baselayer

LTI Summit, Brussels / 17 May 2016

Steve Renals / University of Edinburgh
Global Technology Trends
In Annual Letter, Sundar Pichai Says Computing, And Google, Will Be Driven By Artificial Intelligence

“We will move from a mobile first to an AI first world”
In Annual Letter, Sundar Pichai Says Computing, And Google, Will Be Driven By Artificial Intelligence

“We will move from a mobile first to an AI first world”

Mark Zuckerberg thinks AI will start outperforming humans in the next decade

Computer systems that can see, hear, and understand language better than their creators
Global Technology Trends

In Annual Letter, Sundar Pichai Says Computing, And Google, Will Be Driven By Artificial Intelligence

“We will move from a mobile first to an AI first world”

Mark Zuckerberg thinks AI will start outperforming humans in the next decade

Computer systems that can see, hear, and understand language better than their creators

Hollywood Goes Multilingual

“The market share of local-language films is expanding in territories around the world, currently topping 90% in India”
Conversational interaction technologies are at the heart of AI
Conversational interaction technologies are at the heart of AI

- **Technologies for**
  - Human-Human Communication
  - Human-Machine Communication
  - Human-Environment Communication
  - No language barriers
  - Transferable across domains
  - Accessible
  - Multimodal
  - Multidevice
- **Supporting** *natural communication*
• Connect the European research base with commercial & industrial activity – especially startups and SMEs
• Facilitating a research/technical infrastructure
• Multiple stakeholders
  • Industry – SMEs and large companies
  • Research – universities and research institutes
  • Investors and venture capital
  • Government, EC, and policy makers
• Build a community
• Provide an infrastructure for a multilingual world
CITIA Roadmap – SWOT

**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
- Available 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 / data
- Dominance by the big corps
- No coherent vision
- Privacy/trust/security concerns
CITIA Roadmap

- 12 workshops
- >200 experts consulted
- >1000 information items collected
CITIA Roadmap

- 12 workshops
- >200 experts consulted
- >1000 information items collected
CITIA Roadmap

- 12 workshops
- >200 experts consulted
- >1000 information items collected
12 workshops
>200 experts consulted
>1000 information items collected

http://citia.eu
• Rapid technological progress, including
  • Speech recognition
  • Spoken dialogue
  • Face analysis
• Improved robustness & adaptation to real-life contexts
• Focus on composed services, not single interfaces
• Interaction driven and informed by analytics
• Rich dialogues, social and cultural context
• Innovation across speech and language technologies requires more data
• Privacy concerns – what happens to my data?
• Data required to build multilingual systems
Roadmap – Target scenarios

Five Scenarios for Research & Innovation

Version 2.1 05-February-2015
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.
Adaptable Interfaces for All

Multimodal conversational interfaces have the potential to adapt automatically to the user, to the environment, and to the user’s state. Interfaces adapted to older people will take account of cognitive, auditory, visual, and articulatory ageing. Interfaces for mobile devices will adapt to what a user is doing – such as working in a noisy hands-free environment or rushing for a train. Interfaces for disabled people will help to rehabilitate them.

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.

Automatic interface subtitling
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.
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.

Smart assistants for kids
By 2024, multimodal interactive systems will operate on huge, dynamic, heterogeneous data streams, providing powerful approaches to navigation and visualisation.
Active Information Access

By 2024, multimodal interactive systems will operate on huge, dynamic, heterogeneous data streams, providing powerful approaches to navigation and visualisation.

Immersive Intelligence
Communicative Robots

By 2024, robots will be capable of human-like multimodal embodied communication, with the ability to operate in social environments.
Robotics is predicted to be a huge growth area, and many robots will operate in communicative settings. Such robots will display personality and will be able to generate and interpret social signals. By 2024, robots will be capable of human-like multimodal embodied communication, with the ability to operate in social environments.
Roadmap – Target scenarios

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.
Roadmap – Target scenarios

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.

Gaming Interfaces
Roadmap – SWOT

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
• Available 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 / data
• Dominance by the big corps
• No coherent vision
• Privacy/trust/security concerns
Roadmap – SWOT

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
- Available 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 / data
- Dominance by the big corps
- No coherent vision
- Privacy/trust/security concerns
Practical Steps

CITIA Multilingual Baselayer
“We need more data”

The multilingual digital single market needs to support

- 24 official languages
- 5 “semi-official” languages
- >100 further regional/minority languages
- Not to mention languages required for the global market!
- If we rank the 50 most used languages in Europe, then there are >50 million speakers of languages 26–50 (Finnish – Montenegrin)

Lack of multilingual speech resources is a major barrier to researchers, developers, and innovators
A Practical Vision

- Open multilingual infrastructure for speech technology
  Enable research, enable innovation, create the multilingual digital single market
- Develop a **core multilingual base layer**
  - **core open access data** across 40–50 languages
  - **baseline open systems** for speech recognition and speech synthesis across these languages
  - **practical recipes** for collecting & transcribing multilingual data for training & evaluating speech technology systems
  - **cloud-based reference services** to enable people to prototype and test new multilingual spoken applications
Achieving the Vision

- **Standard data collection processes**
  - data collection across a variety of genres
  - standard transcription procedures
  - ~100 hrs transcribed data for ASR
  - ~10 hrs to construct a male & a female synthetic voice

- **Open source baseline systems**
  - building on existing open source projects (e.g. Kaldi)

- **Cloud-based backend with open APIs**
  - enabling the rapid deployment of prototype multilingual systems that can be used to develop new applications

- **Open source** (e.g. BSD)
- **Open data** (e.g. CC-BY)
An Affordable Vision

• What’s required?
  • A core project team
    • Project management & support
    • Definition of data collection processes
    • Open source recipes (and extensions)
    • Building baseline systems
    • Deployment of cloud-based servers
  • Language teams
    • Data collection, transcription, quality control
    • 2–3 person-years / language
Multilingual Baseline Layer

- Open multilingual infrastructure across 40–50 languages
- Baseline open platform to incentivise people to share, collaborate, and innovate
- Practical step towards the multilingual digital single market
- Sustainability
  - open data and open source systems for 40-50 languages
  - virtuous cycle of apps-users-data
  - enough to seed a sustainable community
- Growth
  - more data, better software systems
  - widened coverage
Going forward

- Multilingual interaction technologies can enable and grow the digital single market – soon everything will be an interface
- Robust multilingual language & speech technologies / services will be at the heart of future interfaces
- How to move forward
  - Attack the **hard research problems** – best way to drive high-impact innovation
  - Create an open infrastructure – the **multilingual base layer**
  - Build a **community** around the infrastructure
• Multilingual interaction technologies can enable and grow the digital single market – soon everything will be an interface

• Robust multilingual language & speech technologies / services will be at the heart of future interfaces

• How to move forward
  • Attack the **hard research problems** – best way to drive high-impact innovation
  • Create an open infrastructure – the **multilingual base layer**
  • Build a **community** around the infrastructure

**Thanks!**