

The Virtual Human Journalist

ARIA VALUSPA eNTERFACE 2016 project



Abstract:

In this project you will have the chance to help construct a Virtual Human Journalist, using the latest Virtual Human technology developed in the ARIA-VALUSPA EU project. The ARIA technology is designed for building agents that have social, emotional, and linguistic skills, and to interface with a structured knowledge base to function as an information retrieval agent. In this workshop, we challenge the participants to turn this information retrieval functionality around and make an agent that can extract information from a human expert and store this in a structured knowledge base for future use. Thus the Virtual Human Journalist is born. We are looking for enthusiastic, motivated students who are keen to learn more about virtual human technology, and who thrive in working in teams. We are particularly interested in students who are able to take a fresh and unexpected look at what may at first sight seem straightforward problems.

PI: Michel Valstar (Nottingham)

CI: Elisabeth André (Augsburg), Matthew Aylett (Cereproc), Chloe Clavel (CNRS ParisTech), Laurent Durieu (Cantoche), Dirk Heylen (Twente), Mariet Theune (Twente) Catherine Pelachaud (CNRS ParisTech), Björn Schuller (Imperial College London)

Project Objectives

The main aim of the project is to create a toolkit for instantiating a Virtual Human Journalist (VJ) that can extract knowledge from experts through structured interviews. In more detail, the objectives of the VJ project is to:

1. Create a set of questions that the VJ should ask, possibly by building a dedicated tool
2. Build a tool to initialise the knowledge base of the VJ (domain knowledge)
3. Develop the experimental scenario in which a VJ will extract knowledge from an expert
4. Incept a new algorithm for generating follow-up questions or entirely new questions on-the-fly, i.e. during the conversation, possibly drawing on the existing knowledge of the VJ
5. Create algorithms for extracting knowledge from the human interlocutor's answers
6. Create a set of performance measures to evaluate the VJ
7. Evaluate the VJ's performance

Some stretch goals could be:

8. Create a semi-automatic version of the VJ with the ability of a wizard to control some aspects of the VJ's behaviour
9. Create a fully automatic version of the VJ

Background information

The ARIA-VALUSPA project (<http://aria-agent.eu>) is a 3-year, 3M€ EU project that started in January 2015. Its aim is to create a disruptive new framework that will allow easy creation of Affective Retrieval of Information Assistants (ARIA agents) that are capable of holding multi-modal social interactions in challenging and unexpected situations, with the goal to provide information to the user within a specific domain. The system can generate search queries and return the information requested by interacting with humans through virtual characters. These virtual humans will be able to sustain an interaction with a user for some time, and react appropriately to the user's verbal and non-verbal behaviour when presenting the requested information and refining search results. Using audio and video signals as input, both verbal and non-verbal components of human communication are captured. A sophisticated dialogue management system decides how to respond to a user's input, be it a spoken sentence, a head nod, or a smile. The ARIA uses special designed speech synthesisers to create emotionally coloured speech and a fully expressive 3D face to create the chosen response. Back-channelling, indicating that the ARIA understood what the user meant, or returning a smile are but a few of the many ways in which it can employ emotionally coloured social signals to improve communication.

The following partners make up the consortium:

Participant no.	Participant organisation	Investigators	Country
1 (Coordinator)	University of Nottingham	Michel Valstar	United Kingdom
2	Imperial College London	Björn Schuller	United Kingdom
3	Centre National de la Recherche Scientifique	Catherine Pélachaud, Chloe Clavel	France
4	Universität Augsburg	Elisabeth André	Germany
5	Universiteit Twente	Dirk Heylen	Netherlands
6	Cereproc	Matthew Aylett	United Kingdom
7	Cantoche	Laurent Durieu	France

Detailed Technical Description

Below we provide a detailed technical description of the project.

Technical Description

The technical description/functionality specification is derived from the project objectives. The VJ shall consist of a single situated PC, on which all software runs. The VJ shall obtain input from a video camera (webcam) and microphone. The VJ will express itself to the human interlocutors by means of a standard PC monitor and a set of speakers situated next to the monitor.

The user shall sit or stand in front of the VJ, displayed on the monitor, at a distance of between 0.5m and 2m. The user shall speak English, and be a known expert on a pre-defined domain. The user shall have given informed consent.

The audio-visual data shall be stored for analysis both directly and off-line. The VJ shall interview the user using a set of initial questions and by generating new questions as the interview progresses. The VJ shall adopt a social stance that encourages information transfer from the expert to the VJ. The VJ shall use social and emotional expressions to do this.

The VJ shall turn the expert's responses into a form of knowledge that can be stored and retrieved for later use and analysis.

The VJ shall graciously greet the expert, ask for their name and store it, use the name where appropriately in the interview, and thank the expert when the interview is over. The VJ will keep time and ensure all standard pre-defined questions are asked within the allotted time.

Required Resources

Software:

The VJ should build on the existing technology of ARIA-VALUSPA. Developer and demo machines should have:

- The latest stable version of ARIA-VALUSPA from GitHub
- Compilers for Visual Studio 2013 should be installed on developer machines
- JDK for Java 8, 64 bit
- Windows 8.1 or 10 should be installed on developer machines
- Developers should have GitHub accounts

Hardware:

Developer machines may not need all hardware at all times, but for the demo PC we have the following requirements:

- a standard research-grade demo PC equipped with an CUDA-enabled graphics card
- a high-grade wearable microphone
- a recent consumer-grade webcam
- a standard monitor
- speakers

Other:

- For the evaluation, a studio or otherwise reasonably quiet and empty room is required to conduct the human-VJ interviews.
- Sufficient local storage (app. 1 TB) is required to hold temporary data recordings before they are stored on a central ARIA-VALUSPA server.
- A few experts should be recruited to interview on a particular topic/domain.
- Ethics need to be cleared before the project commences

Project Management

The project will be managed by the PI and the team of CIs. Alex Ghitulescu will be present during the first and last week of eINTERFACE, and other members of the ARIA-VALUSPA developers' team will spend some time there as well to help the students on the project. It is expected that at most times there will be 1-2 members of the team there who will actively participate on the project with their expertise of the ARIA framework.

eINTERFACE participants who chose this project are responsible for their own progress, although the PI and/or Alex will always be available to give feedback where necessary. If the participants wish to adopt this, we would suggest they use an agile project management methodology, with daily morning scrums and weekly sprints, culminating in a demo of the latest version every Friday.

Workplan and implementation schedule

The timetable and workplan is to be created by the participants, ideally starting this before arrival in Twente and in collaboration with the PI and Alex Ghitulescu. This is so that everyone is somewhat prepared for the event and that the time together can be used for maximum efficiency. It would be particularly useful if participants create e.g. a slack forum to connect with each other before arrival.

In terms of planning, roughly, we would expect the following stages to be followed:

1. Familiarise the team with the brief and the existing ARIA-VALUSPA code
2. Create a new instance of an ARIA
3. Split team in groups to work on e.g. ARIA design, interview question input and generation, and knowledge extraction.
4. Design the interview experiment, and contact experts
5. Conduct the interviews
6. Evaluate the VJ
7. Write up a paper with the results

Impact

The outcome of the project is a new virtual human resource that can be used by any non-commercial entity to conduct knowledge acquisition interviews. We can see a large potential impact in such areas as conducting routine medical health questionnaires, new product and media evaluation, and any other domain where it is time-consuming to interview users. We expect the tool to be widely adopted, and as this is part of an EU project that's half-way through its life-cycle, we expect the VJ tool to continue to evolve in terms of functionality and ease of use. Hopefully what you'll make will have a lasting legacy!

We also expect the evaluation study to generate enough data of interest to allow for the writing of a conference paper of mid-level quality, e.g. IVA. The paper may not be finished by the end of the project, and as such it depends on the participants and the supervision team to ensure a suitable venue is identified, work is continued on this potentially after eNTERFACE itself ends, and the paper is ultimately submitted.

Participant Profile

This is an interdisciplinary project and we welcome participants with skills or interests in any of the following areas:

- JAVA/C++ programming
- Spoken Dialogue Systems
- Experimental Psychology
- Knowledge Representation and Database Systems
- Artificial Intelligence and Planning
- Dialogue Management Systems
- Automatic Speech Recognition or Synthesis
- Affective Computing and Social Signal Processing
- Information retrieval

While we will provide guidance on initial tasks, we strongly encourage participants to engage in the design process and implement their own ideas.

Team Profile

Principal and Co-Investigators:

Michel Valstar (Nottingham, PI)

Michel Valstar works in the field of machine learning and computer vision, where his expertise is in face and facial expression analysis. Before his position at Nottingham, he was a visiting researcher at the MIT Media Lab with Prof Rosalind Picard, and before that a postdoctoral researcher with Prof Maja Pantic at Imperial College London. He was the organiser of the first facial expression recognition challenge (FERA 2011) and its follow up FERA 2015 and the audio-visual emotion recognition challenge series (AVEC 2011-2015), as well as the first conference for researchers in affective computing and social signal processing in the UK (AC.UK 2012). To date he has published his work in more than 60 peer-reviewed articles, with a total citation count of over 3500 and an H-Index of 27. He has contributed significantly to the objective analysis of facial expressions by automating FACS, pioneered research on the analysis of facial dynamics, and recently contributed to the field of facial structure analysis, taking the state of the art in facial point localisation to a next level. He was a research fellow working on the video analysis components on the awarded SEMAINE project funded by the European Community's Seventh Framework Programme (FP7/2007-2013).

Elisabeth André (Augsburg)

Elisabeth André is a Full professor of Computer Science at Augsburg University, and Chair of the Research Unit Human-Centered Multimedia. She has a long track record in multimodal

interfaces, embodied conversational agents and social signal processing. She is on the editorial board of various renowned international journals, such as Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS), IEEE Transactions on Affective Computing (TAC), ACM Transactions on Intelligent Interactive Systems (TIIS), and AI Communications. In 2007 Elisabeth André was nominated Fellow of the Alcatel-Lucent Foundation for Communications Research. In 2010, she was elected a member of the prestigious German Academy of Sciences Leopoldina, the Academy of Europe and AcademiaNet. She is also an ECCAI Fellow (European Coordinating Committee for Artificial Intelligence).

Matthew Aylett (Cereproc)

Matthew Aylett has substantial experience in speech technology research and commercialisation, and project management of both research and commercial software development projects. He has worked as CereProc's Chief Technical Office since the company was founded, recently becoming Chief Scientific Officer to reflect his research remit. As well as developing his research career he has publishing widely at an international level. He currently holds a Royal Society Industry Fellowship which allows him to put more time and resources into innovation both within CereProc, but also to collaborate with other academic institutions. Dr Aylett holds a BA in Computing and Artificial Intelligence from the University of Sussex, and an MSc (Distinction) and PhD in speech and language technology from the University Of Edinburgh.

Chloe Clavel (CNRS ParisTech)

Chloe Clavel is Assistant Professor at Telecom Paristech. She owned a PhD on acoustic analysis of emotional speech. Before joining Telecom ParisTech she worked as a researcher at Thales Research and Technology where she focused on emotion analysis; then she became a researcher at EDF R&D working on sentiment analysis and opinion mining. She has participated in several collaborative projects and has coordinated one national project. Funding for her time is allocated to third-party Institut Mines – Télécom- ParisTech.

Laurent Durieu (Cantoche)

Laurent Durieu is the R&D Manager at Cantoche. After graduating as a software engineer from Louis de Broglie (Ecam Rennes) in 2008, Laurent worked during 2 years at Thomson and Collins Engineers, Inc. (US) before to join Cantoche in 2009.

Dirk Heylen (Twente)

Dirk Heylen is Professor Socially Intelligent Computing at the University of Twente, and co-leader of the Human Media Interaction group. He has a PhD in computational linguistics from Utrecht University He has been involved in several European projects such as HUMAINE, AMI and AMIDA, COST 2102, SEMAINE, SSPNET, and SERA. Recent Dutch national projects he was involved in include BrainGain on brain computing interfaces and GATE dealing with cognitive models of virtual agents in training applications. Currently he is the project leader of the COMMIT project "Interaction for Universal Access" exploring the use of virtual agents and

interactive storytelling for training. He is the current president-elect of the Association for the Advancement of Affective Computing (formerly known as the HUMAINE association).

Rieks op den Akker (Twente)

Rieks op den Akker is a lecturer in the Human Media Interaction group. His research interests include computational linguistics, artificial intelligence, conversational analysis, and multi-modal dialogue systems. In the Dutch national project IMIX (NWO) he worked on information retrieval dialogues. In AMI and AMIDA (EU FP6), he worked on machine classification of speech acts and on various dialogue aspects of mediated multi-party communication. He participated in the SSPNet (EU Network of Excellence) on the identification of patterns of conversational behaviour and their relation with interpersonal relations. In Smarcos (EC Artemis) he worked on responsive behaviour generation for virtual characters on mobile platforms. In SERA (EU FP7) he worked on the dialogue system for a robot interacting with elderly people. Currently, he is a work package leader in the COMMIT project "Interaction for Natural Access" in which he works on building police interview training support technology. In the "Let Him Speak" project he cooperates with the Police Academy on a review system for training witness interviewing.

Mariet Theune (Twente)

Mariet Theune is a lecturer in language technology at HMI. Her background is in computational linguistics; in particular, automatic language generation. At HMI she coordinates the research on interactive narrative, previously in the context of the GATE project and currently as part of the COMMIT project "Interaction for Universal Access". Besides interactive storytelling and natural language generation, her research interests include multimodal information presentation and conversational agents. She has over 90 peer reviewed publications in these fields and has (co)organized several conferences and workshops. These include the Computational Linguistics in Netherlands (CLIN) conference in 2001 and 2013, workshops on Multimodal Output Generation (MOG) in 2007, 2008 and 2010 and the European Workshop on Natural Language Generation (ENLG) in 2009.

Catherine Pelachaud (CNRS ParisTech)

Catherine Pelachaud is Director of Research at CNRS in the laboratory LTCI, TELECOM ParisTech. Her research interest includes modeling of nonverbal communication (face, gaze, and gesture), expressive behaviors and socio-emotional agents. She has been involved and is still involved in several European projects related to multimodal communication, to expressive agents, emotion and social behaviours. She was part of the Humaine Association committee. She is associate editors of several journals among which IEEE Transactions on Affective Computing, ACM Transactions on Interactive Intelligent Systems and Journal on Multimodal User Interfaces. She has co-edited several books on virtual agents and emotion-oriented systems.

Björn Schuller (Imperial College London)

Björn Schuller received his diploma in 1999 and his doctoral degree for his study on Automatic Speech and Emotion Recognition in 2006 and his habilitation in 2012, all in electrical

engineering and information technology from TUM. He is senior lecturer at the Imperial College London's Department of Computing in the UK and full professor at the University of Passau in Germany. Best known are his works advancing Machine Intelligence for Affective Computing. He is the president of the Association for the Advancement of Affective Computing (AAAC), editor in chief of the IEEE Transactions on Affective Computing, and elected member of the IEEE Speech and Language Processing Technical Committee (SLTC), IEEE, ACM, and ISCA and (co-)authored five books and more than 420 publications in the field leading to more than 10k citations – his current h-index equals 49.

Other Researchers:

Alexandru Ghitulescu graduated as a BSc in Computer Science at the University of Nottingham in 2014. He is now the lead developer in the ARIA-VALUSPA project.

Tobias Baur graduated as a Master of Science in Informatics and Multimedia from Augsburg University in 2012. Afterwards he joined the Research Unit Human Centered Multimedia. He is actively involved in the development of the SSI Framework, with a focus on social cue detection, and the lead developer of the NonVerbal Behavior Analyzer (NovA) tool, which aims to analyze signals received by the SSI Framework.

Johannes Wagner graduated as a Master of Science in Informatics and Multimedia from Augsburg University in 2007. Afterwards he joined the Research Unit Human Centered Multimedia. He brings a significant amount of experience in signal processing and is one of the lead developers of the Social Signal Interpretation framework (SSI).

Angelo Cafaro is the lead researcher and developer for ARIA-VALUSPA at CNRS ParisTech.

Blaise Potard is the lead researcher and developer for ARIA-VALUSPA at Cereproc.

Andry Chowanda is a PhD student at the University of Nottingham. His thesis is on the creation of emotionally and socially skilled characters for computer games. In the ARIA-VALUSPA project he works on the dialogue management since September 2015.

Eduardo Coutinho received his diploma in Electrical Engineering and Computer Sciences from the University of Porto (Portugal, 2003), and his doctoral degree in Affective and Computer Sciences from the University of Plymouth (UK, 2008). He was a postdoctoral researcher at the University of Sheffield (UK) and the coordinator of the Music and Emotion research Focus at the Swiss Center for Affective Sciences (CISA, Switzerland). Currently, Coutinho is a postdoctoral fellow in the group of Dr. Schuller, an Associate Researcher at CISA, and an Honorary Fellow at the Department of Music from the University of Liverpool. In his research, Coutinho focuses on the link between low-level acoustics and the expression, perception and induction of emotion in speech and music from a transdisciplinary perspective.

Jelte van Waterschoot started his PhD in dialogue management systems at the University of Twente in February 2016. As such, he is the newest member on our team!