UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Ask Alice: an artificial retrieval of information agent

Michel Valstar, Soumia Dermouche, Catherine Pelachaud, Eduardo Coutinho, Björn Schuller, Yue Zhang, Dirk Heylen, Mariët Theune, Jelte van Waterschoot, Tobias Baur, Angelo Cafaro, Alexandru Ghitulescu, Blaise Potard, Johannes Wagner, Elisabeth André, Laurent Durieu, Matthew Aylett

Angaben zur Veröffentlichung / Publication details:

Valstar, Michel, Soumia Dermouche, Catherine Pelachaud, Eduardo Coutinho, Björn Schuller, Yue Zhang, Dirk Heylen, et al. 2016. "Ask Alice: an artificial retrieval of information agent." In *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016, Tokyo, Japan, November 12-16, 2016*, edited by Yukiko I. Nakano, Elisabeth André, Toyoaki Nishida, Louis-Philippe Morency, Carlos Busso, and Catherine Pelachaud, 419–20. New York, NY: ACM Press. https://doi.org/10.1145/2993148.2998535.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

Ask Alice: An Artificial Retrieval of Information Agent

Michel Valstar
University of Nottingham, UK

Tobias Baur
University of Augsburg, DE

Angelo Cafaro
CNRS, FR

Alexandru Ghitulescu
University of Nottingham, UK

Blaise Potard
Cereproc, UK

Johannes Wagner
University of Augsburg, DE

ABSTRACT

We present a demonstration of the ARIA framework, a modular approach for rapid development of virtual humans for information retrieval that have linguistic, emotional, and social skills and a strong personality. We demonstrate the framework's capabilities in a scenario where 'Alice in Wonderland', a popular English literature book, is embodied by a virtual human representing Alice. The user can engage in an information exchange dialogue, where Alice acts as the expert on the book, and the user as an interested novice. Besides speech recognition, sophisticated audio-visual behaviour analysis is used to inform the core agent dialogue module about the user's state and intentions, so that it can go beyond simple chat-bot dialogue. The behaviour generation module features a unique new capability of being able to deal gracefully with interruptions of the agent.

CCS Concepts

•**Human-centered computing** → **Interaction devices;** •**Computing methodologies** → **Machine learning;** *Activity recognition and understanding;*

Keywords

Virtual Humans, Technology Demonstrator, Affective Computing, Social Signal Processing

1. INTRODUCTION

Task-specific AI is attaining super-human performance in an increasing number of domains. In the near future, virtual humans (VHs) will be the human-like interface for increasingly capable AI systems, in particular information retrieval systems. However, there remains a large gap in the smoothness of interaction between either a current VH or another human being. In the Horizon 2020 project ARIA-VALUSPA we aim to drastically reduce this gap.

This means first and foremost that interacting with the ARIA-agents should be engaging and entertaining. They should display interactive, believable behaviour. They should be adaptive to the user at various levels, from adapting to a user's appearance, age, gender, and voice, to sudden changes in the dialogue initiated by the user. As part of ARIA-VALUSPA, we have developed an interactive virtual human that can hold a prolonged dialogue about the book 'Alice in Wonderland', by Lewis Caroll.

Some particular challenges that we have addressed in the project and that we wish to demonstrate here are to deliver a reusable framework that can be used to create VHs with different personalities, behaviours, and underpinning knowledge bases. The framework is in principal independent of the language spoken by the user. Another important challenge that we set ourselves and will demonstrate here is to be able to deal with unexpected situations, in particular interruptions initiated by the user. This is a hard problem that has not been addressed previously.

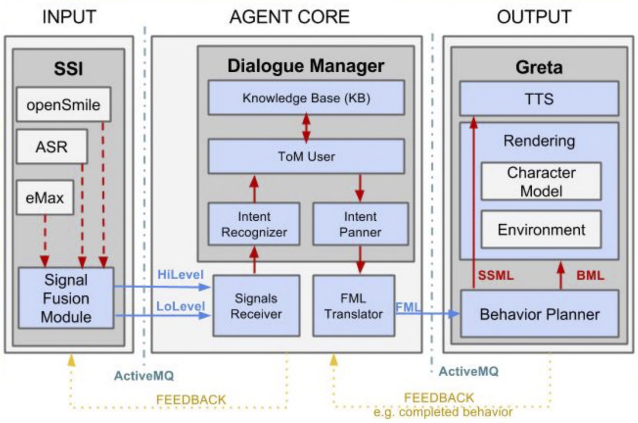

Figure 1: ARIA framework architecture, composed of modular Input, Agent Core, and Output blocks.

2. ARIA FRAMEWORK

The 'Ask Alice' demonstration is built on top of the ARIA Framework. This is a modular architecture with three major blocks running as independent binaries whilst communicating using ActiveMQ (see Fig. 1). Each block is in turn modular at a source-code level. The Input block processes audio and video integrated in SSI [1, 6] to analyse the user's expressive and interactive behaviour and does speech recognition. The Core Agent block maintains the agent's information state, including its goals and world representation.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution.

Copyright is held by the owner/author(s).

ICMI'16, November 12–16, 2016, Tokyo, Japan
ACM. 978-1-4503-4556-9/16/11...$15.00
http://dx.doi.org/10.1145/2993148.2998535

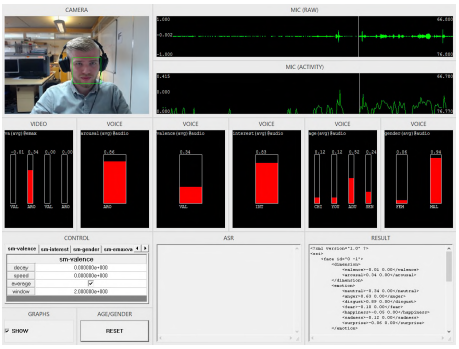

Figure 2: Input Block Visualisation

It is responsible for making queries to its domain-knowledge database to answer questions. Once all goals and states are taken into account it decides on which intents the agent should express and which information it will give, generating an FML message. The Output block generates the agent behaviour, that is, it synthesises speech and visual appearance of the virtual human. The ARIA Framework makes use of communication and representation standards wherever possible. For example, by adhering to FML and BML we are able to plug in two different visual behaviour generators, Greta [4] or Cantoche's Living Actor technology.

The Input block includes state of the art behaviour sensing, many components of which have been specially developed as part of the project. From Audio, we can recognise gender, age, emotion, speech activity and turn taking [7], and a separate module provides speech recognition [3]. Speech recognition is available for the three languages targeted by the project. From Video, we have implemented face recognition, emotion recognition [2], detailed face and facial point localisation [5], and head pose estimation. Fig. 2 shows a visualisation of the behaviour analysis.

3. DEMONSTRATION

In the demonstration, a single user will be invited to face Alice, who is displayed on a large screen. The invitation will either be done by a researcher or by Alice herself, if it detects the presence of a new face. Once Alice has detected that the user is engaging with her, she will initiate a greeting process, and then introduce the topic of the book, Alice in Wonderland. Alice will first try to establish if the user has any domain knowledge, for example by determining whether they've read the book, seen the original animation film or the later hollywood film of the book. Depending on this domain knowledge, she will either elaborate on some more background information about the book or will dive straight into offering her views on the book, and allowing the user to ask questions and provide their own opinion. This interaction lasts until the user is satisfied with the interaction, or until Alice gets bored with the user. Fig. 3 shows the Living Actor output of the demo, i.e. Alice, and Fig. 2 shows the user interacting with Alice.

Because of the framework's ability to adapt to people who interact with it, Alice will be able to recognise a user and pick up a conversation with someone who previously visited the demo. She will also be able to deal with common issues related to interruptions that occur during a typical conference demo setting, i.e. multiple people interacting with it, or a user suddenly addressing one of their colleagues instead of Alice, or a user simply walking away from the interaction mid-interaction.

Figure 3: Samples of Alice expressive posing.

4. ACKNOWLEDGMENTS

Funded by European Union Horizon 2020 research and innovation programme, grant agreement No 645378.

5. ADDITIONAL AUTHORS

Elisabeth André (University of Augsburg, DE), Laurent Durieu (Cantoche, FR) Matthew Aylett (Cereproc, UK), Soumia Dermouche, Catherine Pelachaud (CNRS, FR) Eduardo Coutinho, Björn Schuller, Yue Zhang (Imperial College London, UK), Dirk Heylen, Mariët Theune, Jelte van Waterschoot (University of Twente, NL).

6. REFERENCES

[1] S. Flutura, J. Wagner, F. Lingenfelser, and E. André. Mobilessi - asynchronous fusion for social signal interpretation in the wild. In *Proc. of ICMI*, 2016.

[2] S. Jaiswal and M. Valstar. Deep learning the dynamic appearance and shape of facial action units. In *IEEE Winter Conf. Applications of Computer Vision (WACV)*, pages 1–8, March 2016.

[3] A. E.-D. Mousa and B. Schuller. Deep bidirectional long short-term memory recurrent neural networks for grapheme-to-phoneme conversion utilizing complex many-to many alignments. In *Proc. Interspeech*, 2016.

[4] F. Pecune, A. Cafaro, M. Chollet, P. Philippe, and C. Pelachaud. Suggestions for extending saiba with the vib platform. In *W'shop Architectures and Standards for IVAs, Int'l Conf. Intelligent Virtual Agents*, pages 16–20. Bielefeld eCollections, 2014.

[5] E. Sánchez-Lozano, B. Martinez, and M. Valstar. Cascaded regression with sparsified feature covariance matrix for facial landmark detection. *Pattern Recognition Letters*, 73, 2016.

[6] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André. The social signal interpretation (ssi) framework: multimodal signal processing and recognition in real-time. In *Proc. ACM Int'l Conf Multimedia*, MM '13, pages 831–834, New York, NY, USA, 2013. ACM.

[7] Z. Zhang, F. Ringeval, B. Dong, E. Coutinho, E. Marchi, et al. Enhanced semi-supervised learning for multimodal emotion recognition. In *IEEE ICASSP*, pages 5185–5189, 2016.