

ALTCAI: Enabling the Use of Embodied Conversational Agents to Deliver Informal Health Advice during Wizard of Oz Studies

Maria J. Galvez Trigo
maria.galveztrigo@nottingham.ac.uk
University of Nottingham
United Kingdom

Martin Porcheron
m.a.w.porcheron@swansea.ac.uk
Swansea University
United Kingdom

Joy Egede
joy.egede@nottingham.ac.uk
University of Nottingham
United Kingdom

Joel E. Fischer
joel.fischer@nottingham.ac.uk
University of Nottingham
United Kingdom

Adrian Hazzard
adrian.hazzard@nottingham.ac.uk
University of Nottingham
United Kingdom

Chris Greenhalgh
chris.greenhalgh@nottingham.ac.uk
University of Nottingham
United Kingdom

Edgar Bodiaj
edgar.bodiaj@nottingham.ac.uk
University of Nottingham
United Kingdom

Michel Valstar
michel.valstar@nottingham.ac.uk
University of Nottingham
United Kingdom

ABSTRACT

We present ALTCAI, a Wizard of Oz Embodied Conversational Agent that has been developed to explore the use of interactive agents as an effective and engaging tool for delivering health and well-being advice to expectant and nursing mothers in Nigeria. This paper briefly describes the motivation and context for its creation, ALTCAI's various components, and presents a discussion on its adaptability and potential uses in other contexts, as well as on potential future work on extending its functionality.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**;
Natural language interfaces.

KEYWORDS

embodied conversational agents, virtual human, wizard of oz

ACM Reference Format:

Maria J. Galvez Trigo, Martin Porcheron, Joy Egede, Joel E. Fischer, Adrian Hazzard, Chris Greenhalgh, Edgar Bodiaj, and Michel Valstar. 2021. ALTCAI: Enabling the Use of Embodied Conversational Agents to Deliver Informal Health Advice during Wizard of Oz Studies. In *3rd Conference on Conversational User Interfaces (CUI '21)*, July 27–29, 2021, Bilbao (online), Spain. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3469595.3469621>

1 MOTIVATION AND CONTEXT

ALTCAI was developed to study the use of Embodied Conversational Agents (ECAs) to deliver health and well-being advice to pregnant and nursing mothers in Nigeria. Nigeria is considered

Africa's most populous country. It's estimated that maternal mortality is more than 60 times that of the 47 most developed countries and the risk of a woman in Nigeria dying of pregnancy-related causes is over 200 times more than the risk in those most developed countries [7]. The World Health Organisation stresses that this reflects the inequities in access to health services, and highlights how important antenatal care such as receiving support and information on healthy lifestyles, preventing diseases, and family planning is for pregnant women in this context [6, 7]. The use of Conversational Agents is gaining in popularity in many areas (e.g., [4]), including in the delivery of health and well-being advice with positive results among users who have low literacy [2, 5]. However, some of these studies also highlighted the limitations of voice-only interfaces or chatbot-based systems in this area [2, 4]. Medhi et al. found that, in their interaction with healthcare and banking systems, people with low-literacy levels using graphical interfaces and speech-based interfaces performed significantly better than those using text-only interfaces [4], which inspired our adoption of an Embodied Conversational Agent (ECA, i.e., "virtual human") for ALTCAI. Fadhil explored the use of chatbot-based conversational agents in the healthcare sector, more specifically with elderly users, and they found that common chatbot interfaces lack the emotional support that a human presence can provide, and that this is something considered essential in healthcare [2]. Furthermore, a study by de Visser et al. has shown that designing autonomous systems to mimic basic human characteristics, such as those that can be found in avatars or virtual humans, has an effect on the level of trust driven by neurological processes typically observed in interactions between humans, suggesting that the right level of anthropomorphism can help in achieving the desired level of trust in the system [1], something that is important for systems delivering health advice that their users should trust.

ECAs can provide a conversational agent with a human-like presence along with graphical means for users to progress through the interaction, something that with the implementation of emotive responses and an adequate level of anthropomorphism could help

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CUI '21, July 27–29, 2021, Bilbao (online), Spain
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8998-3/21/07.
<https://doi.org/10.1145/3469595.3469621>

in tackling the limitations of other types of conversational agents. ALTCAl was developed as a tool that would allow us to explore the use of ECAs for the delivery of health and well-being advice to expecting and nursing mothers in Nigeria, whilst giving us the flexibility and control during the interaction that Wizard of Oz approaches offer. Presently, there is nascent support for local Nigerian dialects or languages with Natural Language Processing technologies. Designing this software as a Wizard of Oz solution allows to collect training data needed to build future interactive AI technologies to support emerging users in the developing world, while also providing researchers and designers with valuable insight into how emerging users approach and interact with such systems, allowing us to iterate and refine our design and user experience.

2 ALTCAl

ALTCAl is a multi-platform desktop system built in Unity 3D that can be used to run Wizard of Oz studies with participants. A participant ostensibly interact with the system through spoken language, with the system represented as an ECA on screen. The ECA is imbued with features such as intent detection, affect recognition and progression through the interaction controlled by a human operator without the knowledge of the participant.

Interaction with the ALTCAl system is centred around a script—necessitated by the need to ensure only correct health and well-being advice is given to participants. Content is prepared as a number of pre-scripted content packages relating to specific topics. Although the system is presented to users as an interactive AI, a human operator retains the control to manage progression through the script. The human operator also has the option to make the virtual human perform predefined sentences and gestures, as well as to type text to be spoken by it. This is to give the participants the impression that the ECA is interacting with them itself, rather than it being controlled by a human operator.

As the COVID-19 pandemic spread across the world, we specifically designed the system to be used during both in-person and remote online studies that use video conferencing software and screen-sharing.

ALTCAl stores timestamped logs with the actions triggered by both the wizard and the participant whilst using the relevant interface. However, the collection of data to train future interactive AI systems should be set-up separately, with the participant's video and audio, and the screen for ALTCAl's interface being recorded via third-party software (e.g., Microsoft Teams, QuickTime). This, especially in the case of remote studies, gives the research team more control over what data is being collected and how, as to be compliant with the relevant ethical clearance and informed consent, as well as access to that data without the need for the participant's intervention post-study.

2.1 Components

ALTCAl, as a desktop application, is composed of multiple application windows representing: 1) the participant's interface, which is the window that the participant sees and interacts with, 2) the wizard's interface, which is the window that the human operator uses to control the progression through the interaction and the behaviour of the participant's interface.

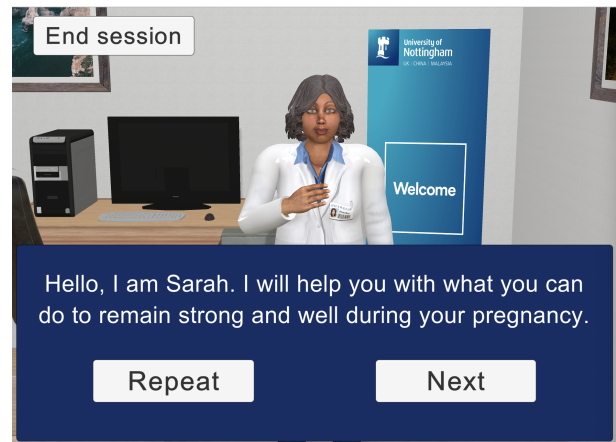


Figure 1: ALTCAl's participant's interface in virtual-human based modality.

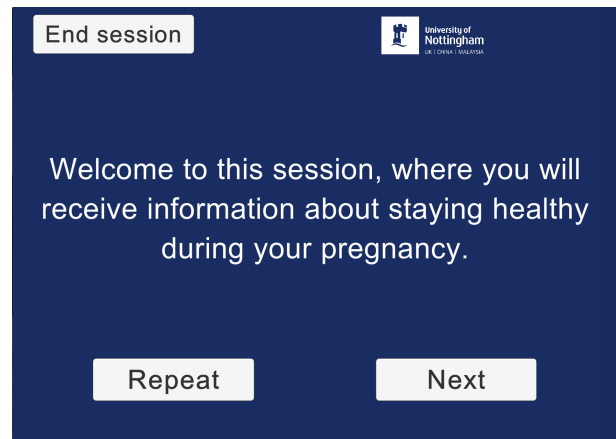


Figure 2: ALTCAl's participant's interface in text based modality.

2.1.1 The participant's interface. The participant's interface has two modalities: 1) Virtual-human based, 2) Text based. Both modalities use a Text-To-Speech (TTS) engine to produce synthesised speech, and offer the participants the opportunity to interact with it through speech too. The controller—i.e. the wizard—can choose which interface is presented to the participant.

In the *Virtual-human based* modality (Figure 1), the interface shows a virtual human in a given environment, with a banner underneath showing subtitles for the virtual human's words, as well as three optional buttons that allow the participant to progress through the interaction. The first button forces the virtual human to repeat the last sentence spoken, the second button makes the virtual human move onto the next sentence, and the third button can be used to finish the interaction and quit the application.

In the *Text based modality* (Figure 2), the whole interface is covered by a blue coloured banner with only the text spoken shown on top of it. The same optional buttons as in the Virtual-human based modality can be enabled.

2.1.2 The wizard's interface. The wizard's interface has two screens (Figure 3). The first screen is a configuration screen where the human operator can select the modality for the participant's interface (virtual-human based or text only), the country in which it is being used, the script or topic the content will cover, and the number of buttons to show on the participant's interface.

The second screen is the main screen, where the interaction will be controlled. This screen offers the means for the human operator to see the text corresponding to the piece of information that is being delivered in the participant's interface, as well as a series of buttons to control the progression through the script, to make the virtual human say a sentence typed by the operator, and to trigger the delivery of a series of predefined responses relevant to the context (in our case the delivery of health and well-being advice to expecting and nursing women).

2.2 Underlying technologies

ALTCAl has been developed in Unity 3D¹. It is based on the Greta ECA², created within the MagiCster project, which allows for an easy integration and coordination of a range of facial expressions and gestures with speech on the virtual human [3]. For the generation of speech, the system uses the TTS engine CereVoice³, and for the communication between the participant's and the wizard's interface, ALTCAl uses queues of messages delivered through an ActiveMQ⁴ channel.

3 INTERACTION

To interact with ALTCAl, participants can be in the same room as the human operator controlling the wizard's interface or in a different location, interacting with the technology remotely during a video conference call with the human operator. In either mode of interaction, the human operator should be able to see and hear the participant to be able to progress through the interaction accordingly (e.g., using a webcam and a microphone).

If interacting in person, the participant would use the participant's interface on a computer, with the human operator using the wizard's interface on another computer with network access to the one used by the participant.

If interacting remotely, the human operator would open both interfaces and share with the participant the window that shows the participant's interface. In this mode of interaction, the participant can be using any device that can join the video conference call, as long as the device has a front-facing camera and a microphone and speakers. The human operator can resize the window with the participant's interface accordingly to offer them the best experience possible.

In both modes, participants can interact with the virtual human using speech, asking it to move onto the next sentence, to repeat, or simply asking questions about the content being delivered. Those interacting with ALTCAl in person can use in addition the buttons shown on screen.

During the interaction, the human operator will observe the participant and make the virtual human progress through the script according to what the participant says or does. For instance, if the participant says 'Next', the human operator will make the virtual human deliver the next piece of content, and if the participant shows signs of being distracted, such as looking away, the human operator can make the virtual human intervene and ask if it shall repeat.

The interaction will continue until the end of the content package is reached or the participant requests to end the session.

4 DEMOING ALTCAl

ALTCAl has been designed to support both in-person and remote online studies. In fact, due to the restrictions imposed globally during the COVID-19 pandemic, it has been used within our project to support a remote online study with participants from the UK and Nigeria joining a session with the research team via Microsoft Teams or Zoom.

In our online demo, we will present ALTCAl in a way that mirrors what was done during these sessions, with the interaction happening as described in the Interaction section of this paper. However, we will not only share with attendees the window containing the participant's interface, but we will also share the window with the wizard's interface and navigate them through the initial steps that during our study were invisible to the participants.

After this initial set-up, we will offer attendees the opportunity to volunteer for a short real-life interaction with the system. During this interaction they will be able to experience the different ways in which the virtual human reacts depending on the actions triggered by a member of our team acting as the human operator.

This will provide a way for attendees to the demo session to experience what the use of our system is actually like.

5 ADAPTABILITY

Although ALTCAl has been developed within a specific context: the delivery of health and well-being advice to expecting and nursing women in Nigeria, the system can be easily adapted for its use in other contexts.

The use of Unity 3D as the development platform provides a way, without needing to code, for customising the physical appearance of the virtual human or ECA, as well as the background for the interfaces and the layout of other elements in it.

The scripts can be modified according to the topic in hand, as well as the facial expressions and the gestures made by the virtual human.

As part of our project, the system will be made available in the summer of 2021 as an open source tool, along with a manual for its use and documentation on how to adapt its components for other contexts. This will provide the CUI community with a tool that can be used in studies in a range of areas where a Wizard of Oz controlled ECA is chosen for the delivery of a scripted content package.

¹<https://unity.com/>

²<https://github.com/isir/greta>

³<https://www.cereproc.com/en/products/sdk>

⁴<https://activemq.apache.org/components/classic/>

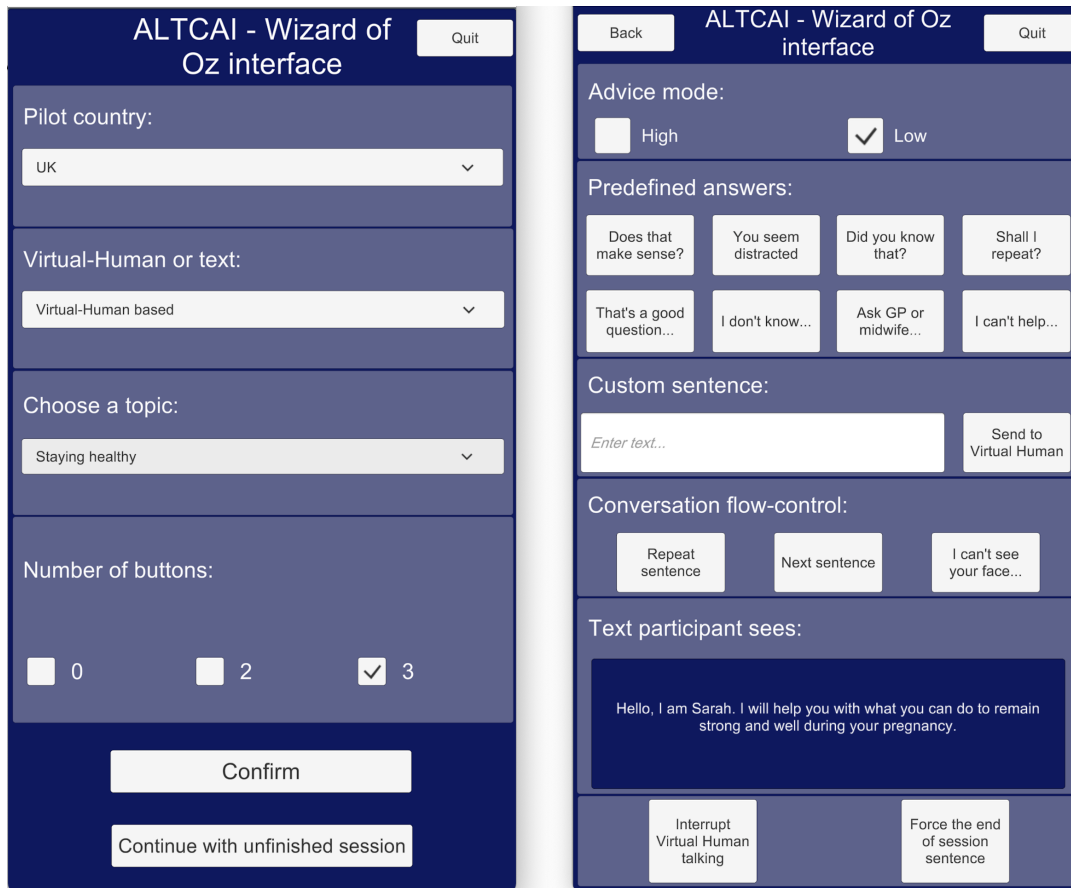


Figure 3: ALTCAI's wizard interface. To the left, the first screen; to the right, the second and main screen.

6 CONCLUSION AND FUTURE WORK

In this paper we have presented ALTCAI, a tool to support Wizard of Oz studies using ECAs for the delivery of pre-scripted health and well-being advice.

We are currently analysing the data obtained after using it within our project during a remote study with expecting and nursing women in the UK and Nigeria. A next step for us is to use the data and feedback collected during the study and to reflect on our experiences as human operators or wizards to refine its design. However, although outside our current scope, there is room for further future work outside its Wizard of Oz uses.

ALTCAI can be used as a base for creating an Artificial Intelligence (AI) led system that would allow the completion of studies exploring the use of ECAs with participants interacting with them throughout a period of time, in real-life situations, instead of them being restricted to a session connected with a researcher or human operator. This would enable the collection of different types of data and insights that could inform the design of future systems using ECAs available to the general population.

ACKNOWLEDGMENTS

This work was supported by the UK Department for International Development.

REFERENCES

- [1] Ewart J. de Visser, Samuel S. Monfort, Kimberly Goodyear, Li Lu, Martin O'Hara, Mary R. Lee, Raja Parasuraman, and Frank Krueger. 2017. A Little Anthropomorphism Goes a Long Way: Effects of Oxytocin on Trust, Compliance, and Team Performance With Automated Agents. *Human Factors* 59, 1 (2017), 116–133. <https://doi.org/10.1177/0018720816687205> arXiv:<https://doi.org/10.1177/0018720816687205> PMID: 28146673.
- [2] Ahmed Fadhil. 2018. Beyond Patient Monitoring: Conversational Agents Role in Telemedicine & Healthcare Support For Home-Living Elderly Individuals. *CoRR* abs/1803.06000 (2018), 7 pages. arXiv:1803.06000 <http://arxiv.org/abs/1803.06000>
- [3] Mary Ellen Foster. 2007. Enhancing Human-Computer Interaction with Embodied Conversational Agents. In *Universal Access in Human-Computer Interaction. Ambient Interaction*, Constantine Stephanidis (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 828–837.
- [4] Indrani Medhi, Somani Patnaik, Emma Brunskill, S.N. Nagasena Gautama, William Thies, and Kentaro Toyama. 2011. Designing Mobile Interfaces for Novice and Low-Literacy Users. *ACM Trans. Comput.-Hum. Interact.* 18, 1, Article 2 (May 2011), 28 pages. <https://doi.org/10.1145/1959022.1959024>
- [5] Jahanzeb Sherwani, Nosheen Ali, Sarwat Mirza, Anjum Fatma, Yousuf Memon, Mehtab Karim, Rahul Tongia, and Roni Rosenfeld. 2007. HealthLine: Speech-based access to health information by low-literate users. In *Proceedings of the 2007 International Conference on Information and Communication Technologies and Development (Bangalore, India) (ICTD '07)*. IEEE, New York, NY, USA, 1–9. <https://doi.org/10.1109/ICTD.2007.4937399>

[6] The World Health Organisation. 2017. 10 Ways to improve the quality of care in health facilities. Retrieved from <https://www.who.int/en/news-room/feature-stories/detail/10-ways-to-improve-the-quality-of-care-in-health-facilities>. (Accessed on 2021-04-13).

[7] The World Health Organisation. 2019. Maternal health in Nigeria: generating information for action. Retrieved from <https://www.who.int/reproductivehealth/maternal-health-nigeria/en/>. (Accessed on 2021-04-13).