Implementing an Accessible Conversational User Interface

Applying feedback from University students and disability support advisors

Francisco Iniesto†
Institute of Educational Technology
The Open University
Milton Keynes, MK7 6AA
francisco.iniesto@open.ac.uk

Tim Coughlan
Institute of Educational Technology
The Open University
Milton Keynes, MK7 6AA
tim.coughlan@open.ac.uk

Kate Lister
Faculty of Wellbeing, Education & Language Studies
The Open University
Milton Keynes, MK7 6AA
kate.lister@open.ac.uk

ABSTRACT

In the ADMINS (Assistants for the Disclosure and Management of Information about Needs and Support) project, we have implemented a virtual assistant which is designed to enable students to disclose disabilities and to provide guidance and suggestions about appropriate accessible support. ADMINS explores the potential of conversational user interfaces (CUIs) to reduce administrative burden and improve outcomes, by replacing static forms with written and spoken dialogue. For the beta version of the assistant, we have carried out a trial to evaluate its accessibility and user experience (UX). Following the project's participatory-design approach, the trial sample included university students with accessibility needs and disability support advisors for its evaluation. The results included both qualitative and quantitative feedback from the participants (students and advisors) which helped to identify accessibility and UX barriers for improving the assistant's design in the next stages of the project.

CCS CONCEPTS

• Computing methodologies~Intelligent agents • Information systems~Expert systems • Social and professional topics~People with disabilities

KEYWORDS

Conversational user interfaces, chatbots, virtual assistants, artificial intelligence, accessibility, trial

1 Introduction

Research suggests that conversational user interfaces (CUIs) present opportunities for users with accessibility needs [1-3]. CUIs can perform tasks for users based on commands through online chat or interpreting human speech and responding via synthesised voices; allowing flexibility, personalisation of the experience and alternative modes of communication. CUIs could enable more efficient and effective access to support for people with accessibility needs. However, there is little research to date that explores how to design CUIs to be accessible [4-6], or how best to use them to support people with accessibility needs [7].

The Assistants for the Disclosure and Management of Information about Needs and Support (ADMINS) project is developing a CUI assistant in support of students enrolling in studies at the Open University (OU), UK. The OU is distance learning university which currently supports more than 20,000 students who have declared disabilities. Generally, the OU uses a combination of online forms and conversations with advisors to assess the needs of thousands of students who declare a disability, a process that is challenging and time-consuming both for the students and the institution. The form is designed to support students to report their needs and any existing strategies and technologies they use, but students have reported that they find it challenging [8] and form responses often lack detail.

ADMINS follows a participatory-design approach [9]. This involves working with students to understand their accessibility needs and preferences and working with disability support advisors as expert stakeholders. Student consultants have been employed to provide substantial guidance to the project and take part in ad-hoc testing, and this is supplemented with workshops and formal trials with students who have diverse accessibility needs [10]. Disability support advisors have also been involved as expert stakeholders throughout the project, thus ensuring that the assistant is designed to support the disability advisor role and not replace it.

The assistant design provides multiple communication modalities and alternative media options. Offering that personalisation allows students to adapt their experience when interacting with the assistant and therefore supports accessibility. The assistant can, for example, provide multiple means of communicating the same information, including explanatory videos, text, or spoken words [11]. It can also offer extra information through links to university or government resources, where students can expand their understanding and find additional support outside the assistant's body of knowledge. The assistant enables the users to provide information and access support via spoken or written dialogue to build the profile providing support suggestions.

In this paper, we detail the initial project trial. This took place in summer 2020 as a forerunner to the main project trial and aimed to introduce improvements in the assistant design using stakeholders' feedback (students and university advisors). This initial trial helped to disclose key aspects to consider when

evaluating and improving accessibility and UX in CUIs design and implementation processes.

2 Background

There is little research to date that explores how to design CUIs to be accessible and usable [12]. When reviewing similar evaluation processes in the literature, several authors have explored the potential of UX in CUIs in a range of domains such as industry [13], home devices [14], providing IT support [15] or educational experiences [16-18]. Other researchers have explored the comparison between CUIs and proposed frameworks [19-21]. These studies have commonly used a combination of methods and captured qualitative and quantitative data. Table 1 shows that many of the studies used a combination of questionnaires with task-driven or free iteration with the CUI, direct observation, or included interviews in their methodologies. Broadly speaking, these studies tend to omit specific consideration of accessibility aspects to be evaluated and included in the design [7].

Table 1. Methods used in CUI evaluation

	Method	Reference
1	Questionnaire	[13] [14] [15] [16] [18] [19] [21]
2	Task driven interaction	[13] [14] [15] [16] [17] [18] [21]
3	Free interaction	[16] [18] [19] [20]
4	Direct observation	[14] [16] [18]
6	Interview	[16] [20]

Questionnaires are the most common method, Kocaballi et al. [22] in their literature review identified several standardised questionnaires to evaluate the UX in CUIs such as AttrakDiff, Subjective Assessment of Speech System Interfaces (SASSI), the Speech User Interface Service Quality (SUISQ), the Mean Opinion Scale (MOS) and The System Usability Scale (SUS). Kocaballi et al. [22] suggest the combination of more than a single CUI questionnaire in evaluation processes; that is because they are usually designed to identify strengths and weaknesses, therefore the combined use of several can cover broader aspects which are relevant when evaluating the design.

3 Methodology

This trial took place in summer 2020, within the context of the COVID-19 pandemic. The impacts of the pandemic on the trial design are discussed in the subsequent section.

3.1 Trial methods

For the testing process in ADMINS a mixed methods methodology included free and direct observations of users working with the assistant; pre-and-post activity questionnaires; and semi-structured interviews. Methods included the combination of two quantitative questionnaires: SUS and Speech User Interface Service Quality Reduced (SUISQ-R), being this last one a simplified version of SUISQ. SUS has been widely used in usability evaluation at design stages [23] while SUISQ-R has been highlighted as a reliable questionnaire to evaluate CUIs UX [24]. Including two questionnaires follows Kocaballi et al.'s approach of drawing on different questionnaires' strengths [22]. For the observations, questionnaires and interviews, accessibility-focused open questions to flag accessibility barriers were designed. Interviews were employed to seek in-depth information from the participants, reflecting on the experience of interacting with the assistant and improvements.

With the current pandemic context and to understand the individual situations students have when interacting with the assistant and their varied accessibility needs, a Person-Centred Planning (PCP) approach has been used. PCP is a combination of approaches designed to empower people with accessibility needs to make their own choices and decisions [25]. PCP was used to allow students to choose their preferred way to interact with ADMINS assistant.

A four-step protocol was used:

1. **Step 1**. Online pre-questionnaire
    a. Project sheet and consent form
    b. Demographic information
    c. Previous experience using virtual assistants
    d. Access preferences when using virtual assistants
    e. Preference for the trial (PCP)
2. **Step 2.** Online trial – Interaction with ADMINS
    a. Free interaction (Option A)
    b. Direct observation with the team (Option B)
3. **Step 3**. Online follow-up questionnaire
    a. SUS
    b. SUISQ-R
    c. Open questions about the experience (only for option B)
4. **Step 4**. Online interview (Optional for selected students). Open questions to cover the experience and improvements.

Ethical approval for the research was granted by the OU's Human Research Ethics Committee (HREC). The version used for the trial was accessibility assessed by expert testers from beyond the project team.

3.2 Sample

A sample of students in the four British nations, who have declared a disability was facilitated by the OU's Student Research Project Panel (SRPP). This included a split across the disability categories and ensured students had agreed to be contacted for research purposes. Students were invited in batches in order to balance the sample for diversity, including students declaring different disabilities and using several modalities (text\speech) and channels (device\operating system\ web browser).

In total, 550 students received an email advertising the trial, 40 (7.2%) filled in the prequestionnaire while a total of 22 (4%) finished the trial. Of those, 13 (60%) were female. The disabilities more frequently declared were long-term medical conditions (12), mental health (9), fatigue or pain conditions (8) and specific learning difficulties such as dyslexia (7). Other needs included restricted mobility (3), autistic spectrum conditions (3), restricted manual skills (1), impaired speech (1) and visual

impairment (1). Therefore, the sample was successful in covering diverse needs.

The participants' ages were concentrated between 26 and 55 (82%). Most participants had prior experience of interacting with CUIs, using text or combination of speech and text (Table 2).

Table 2. Previous experience

Previous experience using CUIs		Interaction with a CUI	
Never	9% (2)	Only speech	4.5% (1)
1-2 times	18.5% (4)	Only text	49.5% (11)
3-5 times	31.5% (7)	Speech and text	45% (10)
More than 5	40% (9)		

The devices mostly utilised by students in their interactions included laptops (18), mobile phones (14), tablet (10) and desktop computer (6) with a predominance of Microsoft Windows (20) and equal use of Android and Apple's iOS (10). Google Chrome was the most extended browser (15) followed by Internet Explorer and Edge (9). Assistive technologies used by the students included screen readers, reading rulers and colour overlays, as well as spellcheckers or speech recognition software. Other configuration options comprised the use of transcripts, subtitles and audio description for videos and to change the size and colour of the text.

The sample was completed with 3 advisors from the OU's disability support team to ensure a range of stakeholder views were present in the ADMINS assistant trial. In step 2, from a total of 25 participants, 18 (72%) selected option A, free interaction with the assistant rather than direct observation. From the sample of students who completed the trial, 5 took part in an optional post-trial in-depth interview (7 were invited).

4 Results and discussion

Results of the trial included quantitative data from the questionnaires and qualitative data from the observations, questionnaires and interviews. Despite the small sample size, results were rich in feedback, perhaps due to the design of the trial and the variety of accessibility needs represented. From the trial, we created a total of 163 log instances for implementation (111 from students, 22 from advisors and 30 from the interviews). Those have been agreed between the research and development team, considering the agile software development perspective of this project and to fulfil a participatory approach [9].

The results of the SUS questionnaire (Table 3) indicate a score of 72.3, which is classified as good (B) usability and over the average score of 68.

SUISQ-R questionnaire is designed to offer an overall score (4.67) indicating the assistant got a fairly good evaluation (Table 4). The questionnaire offers a psychometric evaluation of four variables:

- **User goal orientation** (UGO) and **Customer service behaviour** (CSB) were 4.93 and 5.64 respectively, therefore the assistant was correctly identified to support disability disclosure and offered the expected service for the OU brand. Participants indicated agreement the assistant used everyday words, was polite, courteous and friendly, and the conversation was organised and logical, (questions 5, 6 and 8).
- **Speech characteristics** (SC) and **Verbosity** (V) had lower scores of 4.23 and 3.89. Messages were therefore repetitive and too talkative, the assistant was providing more details than needed (questions 9, 13, and 14).

Table 3. SUS scores

	Question	SUS	SD
1	I think that I would like to use ADMINS frequently.	64	0.96
2	I found ADMINS unnecessarily complex.	75	0.82
3	I thought ADMINS was easy to use.	78	0.88
4	I think that I would need the support of a technical person to be able to use ADMINS.	78	1.24
5	I found the various functions in ADMINS were well integrated.	67	0.85
6	I thought there was too much inconsistency in ADMINS.	58	1.49
7	I would imagine that most people would learn to use ADMINS very quickly.	82	0.89
8	I found ADMINS very cumbersome to use.	63	1.29
9	I felt very confident using ADMINS.	77	0.91
10	I needed to learn a lot of things before I could get going with ADMINS.	81	1.05
	SUS Score	**72.3**	

Table 4. SUISQ-R scores

	Question	M	SD
	User goal orientation (UGO)	**4.93**	
1	I would be likely to use the assistant again	5.28	1.40
2	I felt confident using the assistant	5.20	1.41
3	I could find what I needed without any difficulty	4.68	1.91
4	The assistant made me feel I was in control	4.56	1.78
	Customer service behaviour (CSB)	**5.64**	
5	The assistant used everyday words	5.80	1.71
6	The assistant seemed polite	6.04	1.02
7	The assistant seemed professional in its speaking style	5.04	1.06
8	The assistant seemed friendly	5.68	1.22
	Speech characteristics (SC)	**4.23**	
9	The assistant voice sounded like a regular person	4.12	1.62
10	The assistant voice sounded natural	4.24	1.16
11	The assistant's voice sounded enthusiastic or full energy	4.32	1.52
	Verbosity (V)	**3.89**	
12	I felt like I had to wait too long for the system to stop talking so I could respond	4.48	1.69
13	The messages were repetitive	4.00	1.35
14	The assistant was too talkative	3.84	1.40
	Overall	**4.67**	

Some questions reported positively in both questionnaires indicating that the assistant was easy to use and intuitive (7 and 10 SUS, 6 SUISQ-R). Meanwhile, the low score in other questions indicate inconsistencies in the assistant (7 SUS, 14 SUISQ-R) aspect that is aligned with the open questions, observations and interviews. In these, several barriers reported had to do with the speed and time it took the assistant to load at the

beginning. Sometimes the assistant got into repeating the same question, or some links or videos were not working.

> On the iPad, the textbox wasn't always possible to type into, without clicking in and out of it. It seemed to have an issue where it lost focus while the cursor was still in the box. *A010030A*

Another aspect was the space included to provide the answers, the way questions were presented and the interaction speed.

> The reading time of the text was not proportionate to the length of each text sentence or section which caused me to have to re-read some sentences. The movement of the text rolling down the screen caused me problems to focus on the text I needed to read. *A010023A*

Regarding the conversation flow and the information provided by the assistant and requested to the participants, reported incidences were mostly about the differences in functionality between the text and speech versions, and complexities to work with the second one.

> I was unable to expand on things and could get into a loop. It also kept picking up its own voice. I think there are some issues around clarity. It is not clear how to go onto voice-activation mode instructions for that would be useful. *A010024A*

The conversation should allow students to expand the topics, improving the clarity and quality of the information required and provided.

> This was yes/no question with no point to elaborate or another question to ask if I had proof/paperwork. Other questions would benefit with a yes/no/maybe or even just a chance to elaborate. *A010026A*

It was appreciated it was possible to scroll all the conversation and the quality of the information provided in the summary at the end but there was disagreement on how the information was displayed during the conversation.

> The rapid buffering dots and the length of writing that appears and moves the screen down meant I had to wait for it all to calm down before I could scroll back up to the top of the question to be read. *A010004A*

The use of the language was reported as needing improvement since the assistant did not always understand the participants and could return confusing sentences or complex definitions.

> The language used doesn't help for dyslexics, is too technical, for example, assistive technology, if I ask the definition you get very long definition and uses assistive continuously, we need shorter definitions. *S010003S*

Even though the functionality of the assistant was clear, we noticed the assistant needs to improve the user expectations of the actions that are possible to achieve through its use [20]. Improvements agreed included providing more space in the text boxes for users to provide longer answers, and fixing different behaviours between browsers, as well as the consistent use of colours and fonts sizes. Since it was noticed participants required extra time to answer questions, it was decided to add a new feature to enable speed to be adjusted by the user, providing more time to answer.

The importance of user engagement while designing the conversation in a CUI is well documented [13]. Suggestions considered from the trial included improving the flow of the conversation since the role of the assistant is a mediator in the broader process of creating a profile about the student and providing them with appropriate support. For that purpose, reflections were made in how the assistant's interpretations need to be presented back to the users at appropriate points. We noticed the conversation in some cases was triggering anxiety in participants (mostly with the use of the speech version). Improving the flow can help mitigate the risk that unexpected outcomes could emerge from the system misinterpreting user statements. In that sense improvements in the detected deficiencies were considered since they can potentially lead to conversation breakdowns.

Some improvements consider emphasizing empathy with the student who is filling in the information and facing communication difficulties. As well as recognising the use of different communication channels which need to be equally usable (text and speech) since not everyone has the same experience or needs while interacting with assistants. In general terms, the voice and accent were well appreciated but it is important to consider the gender neutrality and its pace.

Forms are often designed to help staff to gain the information they need, (i.e. the language and order in which the questions are framed) while the assistant has been designed to be student oriented. To be engaging and acceptable, the assistant needs to present a friendly, empathetic and calm virtual 'personality' [21]. At the same time, it represents the values of the institution, it needs to accommodate the accessibility needs in the audience, avoid bias, serve all its users, and use language with which the user is comfortable. Improvements agreed therefore cover the improvement of the use of language (i.e. wording in questions), including incidental pleasantries, enrichment of the QnA maker for better and more definitions and keeping a gender-neutral personality.

5 Conclusions and future work

The trial provided rich feedback, catalysing reflection within the team in three key design aspects to move the project forward and that generally apply to implement accessible CUIs:

- **Conceptual design.** Manage user expectations and clear limitations of what the CUI can achieve, as well as allowing personalisation.
- **Conversational design.** Keep the user engaged and avoid unnecessary conversation breakdowns.
- **Personality design.** Ensure empathy with the use of the language and the way the information is presented.

At this time, we have run the project's main trial with more than 130 newly registered students. This used a new version of the assistant that responded to feedback provided from this beta trial. For the protocol, students interact with both ADMINS assistant and

the disability support form. We have included the Technology Acceptance Model (TAM), SUISQ-R and a new ad-hoc accessibility questionnaire; open questions and interviews to compare the experiences of using the assistant and the form, and to understand the accessibility and UX of the assistant towards scaling this solution up.

From the start of the project, we have engaged in explorations with other institutions about their disability support processes, to understand how ADMINS could be adapted and utilised in other institutions. CUIs like ADMINS are not a replacement for existing disability advisors but an additional tool to enable them to have a more beneficial role in facilitating learners with accessibility needs participating in the university experience more fully. Therefore, the approach taken in ADMINS could have much wider applicability filling the gap to make CUIs more accessible and adapted to diverse user needs and preferences in administrative processes.

ACKNOWLEDGEMENTS

The ADMINS project has been funded by Microsoft through an AI for Accessibility grant

REFERENCES

[1] Armando Barreto and Scott Hollier. 2019. Visual Disabilities. In Yesilada and Harper (eds) Web Accessibility; A Foundation for Research. Washington, USA: Springer.

[2] Alisha Pradhan, Kanika Mehta and Leah Findlater. 2018. "Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

[3] Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. 2014. Hands free-care free: elderly people taking advantage of speech-only interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (pp. 203-206).

[4] Stacy M.Branham and Antony Rishin Mukkath Roy. 2019. Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 446-458).

[5] Pradhan, Alisha, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

[6] Brewer, Robin N., Leah Findlater, Joseph'Jofish Kaye, Walter Lasecki, Cosmin Munteanu, and Astrid Weber. 2018. Accessible Voice Interfaces. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 441-446).

[7] Kate Lister, Tim Coughlan, Francisco Iniesto, Nick Freear and Peter Devine. 2020. Accessible conversational user interfaces: considerations for design. In *Proceedings of the 17th International Web for All Conference* (pp. 1-11).

[8] Tim Coughlan and Kate Lister. 2018. The accessibility of administrative processes: Assessing the impacts on students in higher education. In *Proceedings of the Internet of Accessible Things* (pp. 1-10).

[9] Andrea Cornwall and Rachel Jewkes. 1995. What is participatory research? *Social science & medicine*, *41* (12), pp.1667-1676.

[10] Francisco Iniesto, Tim Coughlan, Kate Lister and Wayne Holmes. 2020. Designing an Assistant for the Disclosure and Management of Information about Needs and Support: the ADMINS project. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1-4).

[11] Anne Meyer, David Howard Rose and David T. Gordon. 2014. *Universal design for learning: Theory and practice*. CAST Professional Publishing.

[12] Clauirton Siebra, Walter Correia, Marcelo Penha, Jefte Macedo, Jonysberg Quintino, Marcelo Anjos, Fabiana Florentin, Fabio QB da Silva, and Andre LM Santos. 2018. Virtual assistants for mobile interaction: a review from the accessibility perspective. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (pp. 568-571).

[13] Julia Saenz, Walker Burgess, Elizabeth Gustitis, Andres Mena, and Farzan Sasangohar. 2017. The usability analysis of chatbot technologies for internal personnel communications. In *IIE Annual Conference. Proceedings* (pp. 1357-1362). Institute of Industrial and Systems Engineers (IISE).

[14] Jingya Guo, Da Tao and Chen Yang. 2019. The Effects of Continuous Conversation and Task Complexity on Usability of an AI-Based Conversational Agent in Smart Home Environments. In International Conference on Man-Machine-Environment System Engineering (pp. 695-703). Springer, Singapore.

[15] Dario Fiore, Matthias Baldauf and Christian Thiel. 2019. "Forgot your password again?" acceptance and user experience of a chatbot for in-company IT support. In Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (pp. 1-11).

[16] Ranci Ren. 2019. Experimental study of socio chatbot usability (Master's thesis).

[17] Rongjia Liu and Zhanxun Dong. 2019. A study of user experience in knowledge-based QA chatbot design. In International Conference on Intelligent Human Systems Integration (pp. 589-593). Springer, Cham.

[18] Jhon Rodrigo Carrizales Mamani, Yelithza Janerth Ramirez Álamo, Jimmy Alexander Armas Aguirre and Elizabeth Eliana Grandón Toledo. 2019. Cognitive services to improve user experience in searching for academic information based on chatbot. In 2019 IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing (INTERCON) (pp. 1-4). IEEE.

[19] Carla, Jessica Tin Gordon, Jeremy Brown, Elisabeth Fritzsch and Shirley Gabber. 2016. Wochat chatbot user experience summary. Proc. WOCHAT.

[20] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5286-5297).

[21] Tuva Lunde Smestad. 2018. Personality Matters! Improving the User Experience of Chatbot Interfaces-Personality provides a stable pattern to guide the design and behaviour of conversational agents (Master's thesis, NTNU).

[22] Ahmet Baki Kocaballi, Liliana Laranjo and Enrico Coiera. 2019. Understanding and measuring user experience in conversational interfaces. Interacting with Computers, 31(2), pp.192-207.

[23] James R Lewis. 2018. The system usability scale: past, present, and future. International Journal of Human–Computer Interaction, 34(7), pp.577-590.

[24] James R Lewis. 2016. Standardized questionnaires for voice interaction design. *Voice Interaction Design*, *1* (1).

[25] Jaimee Lewis and Helen Sanderson. 2011. A practical guide to delivering personalisation: person-centred practice in health and social care. Jessica Kingsley Publishers.