Exploring challenges for *Conversational Web Browsing* with Blind and Visually Impaired Users

Marcos Baez
mbaezpy@gmail.com
Université Claude Bernard Lyon 1
Lyon, France
Politecnico di Milano
Milano, Italy

Claudia Maria Cutrupi
Maristella Matera
Isabella Possaghi
Emanuele Pucci
Gianluca Spadone
Cinzia Cappiello
Antonella Pasquale
{name.surname}@polimi.it
Politecnico di Milano
Milano, Italy

ABSTRACT

Conversational AI is changing the way we interact with digital services. However, there is still a lack of conversational paradigms facilitating access to the Web. This paper discusses a new approach for *Conversational Web Browsing*, and introduces a design space identified through a user-centered process that involved 26 blind and visually impaired users. The paper also illustrates the conceptual architecture of a software framework that can automatically generate conversational agents for the Conversational Web.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **Web-based interaction**; **Accessibility technologies**; *Accessibility systems and tools.*

KEYWORDS

Conversational UIs, Conversational Web Browsing, Conversational Patterns

ACM Reference Format:
Marcos Baez, Claudia Maria Cutrupi, Maristella Matera, Isabella Possaghi, Emanuele Pucci, Gianluca Spadone, Cinzia Cappiello, and Antonella Pasquale. 2022. Exploring challenges for *Conversational Web Browsing* with Blind and Visually Impaired Users. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29–May 05, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3491101.3519832

1 INTRODUCTION

Blind and visually impaired (BVI) users have been long resigned to an interaction paradigm for accessing the Web based on the adoption of assistive voice technologies, e.g., screen readers. However, Web browsers and the information on the Web are optimised to make full use of user's visual perceptive capabilities. Additionally, the support offered by screen reading technology mainly consists in the sequential reading of the visually displayed content, and sometimes is ineffective due to the scant adoption of accessibility guidelines in the design of Websites. These and other limitations often result into user-experience problems that are well documented in the literature [18, 19].

Conversational AI is emerging as a promising technology for offering a more natural and accessible interaction with information and services on the Web [5]. Conversational Agents (CAs), such as Amazon Alexa, Siri and Google Assistant, offer benefits not only to BVI users, but also to other populations that in different usage situations may take advantage of voice-based interaction for accomplishing their tasks [8, 15]. Recent work is capitalising on the promise of this technology, for example to design voice-based CAs for searching the Web [7], to automatically generate CAs out of website content [16], or to enable end users to customize their CAs for the Web [11]. These contributions indicate the potential and the feasibility of adopting conversational AI for making the Web truly for everyone. Yet, there is still a lack of concrete guidance to inform designers and developers about the meaningful design choices to deliver usable and accessible conversational experiences [12]. Our work aims to fill this gap by understanding how the conversational paradigm can be adopted to enable Web access, and how the users would like the conversation to be designed to support their Web browsing experience. This paper in particular illustrates the results

CHI '22 Extended Abstracts, April 29–May 05, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…$15.00
https://doi.org/10.1145/3491101.3519832

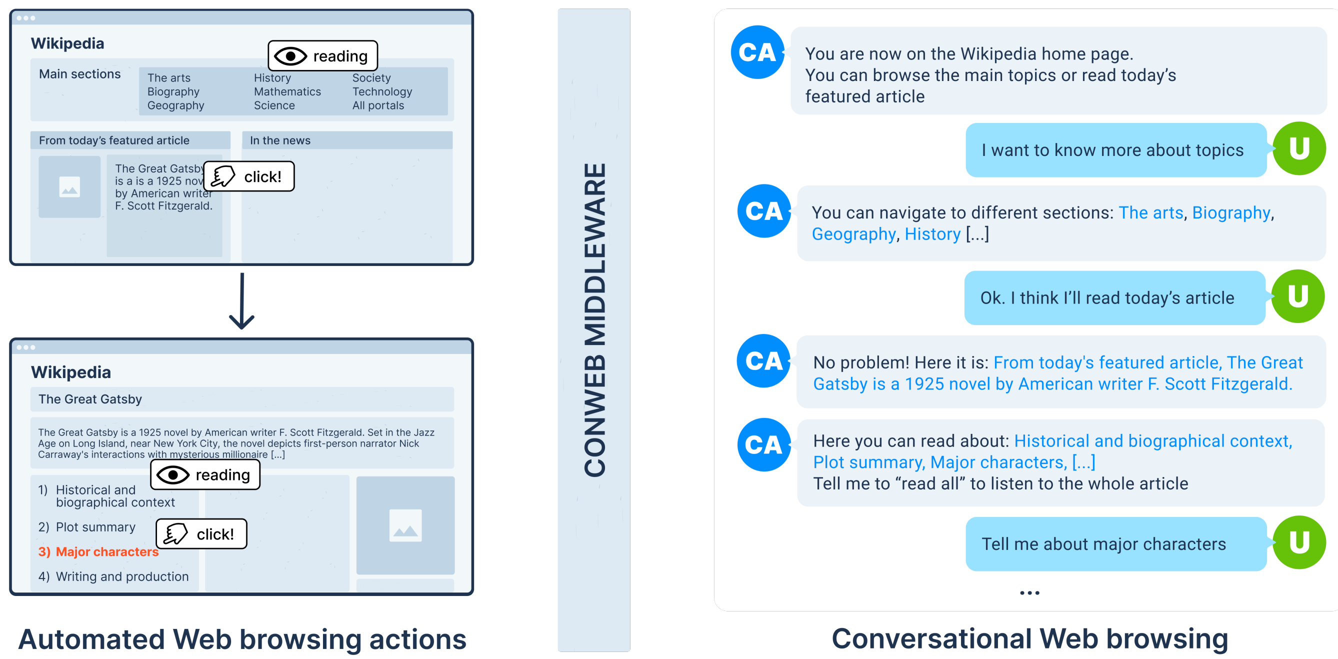

Figure 1: Illustrative example of a Conversational Web Browsing session on Wikipedia. On the right: A session of conversational Web browsing in which a conversational agent (CA) assists the user (U) accessing two Wikipedia pages. On the left: visual organization of the two Wikipedia pages and automated browsing actions activated by the Conversational Web (ConWeb) middleware.

of a user-centered process that we adopted to identify, with the help of a group of 26 BVI users, the *design space* characterizing the notion of Conversational Web Browsing. It also shows how related design patterns, which can respond to the needs identified with the users, can be integrated into a software framework that automatically generates CAs for Web browsing.

The paper is structured as follows: Section 2 introduces the notion of Conversational Web Browsing and discusses some notable related works. Section 3 illustrates the user-based activities that we conducted with a group of BVI users to analyze the main needs for accessing the Web through conversation. Section 4 outlines the resulting challenges that suggest interesting design dimensions. Section 5 describes the conceptual architecture of our *ConWeb* software framework supporting conversational Web browsing. Finally, Section 6 discusses some implications of the proposed approach and outlines our future work.

2 RATIONALE AND BACKGROUND

The idea behind *Conversational Web Browsing* is to enable users to navigate the content and services accessible on the Web by "talking to websites": instead of operating graphical UIs using keyboard, mouse, or a traditional screen reader, the users are enabled to express their goals in natural language and access the websites through dialog-based interactions with a conversational agent (e.g., a smart speaker). Before diving into the main contributions of the paper, we here provide an overview of the Conversational Web paradigm through an illustrative scenario of a BVI user browsing Wikipedia pages. Let us assume, for simplicity, that the user is starting from the Wikipedia home page (https://en.wikipedia.org/wiki/Main_Page), even if the navigation could also start from a search engine (e.g., Google), where the user can specify more precisely the information needed before landing on a specific Wikipedia article.

As depicted in Figure 1, starting from the home page, the conversational agent (CA) provides the user (U) with a short description along with the main organization of the website. The user could also at any point get oriented by inquiring about the options available in a given context (e.g., *"I want to know more about topics"*). Informed by these options, the user can navigate the website by following up some selections of interest (e.g., *"I'll read today's article"*). These requests can trigger navigation across or within pages in the website (e.g., to navigate from the home to the article page, or to move to a content section within the article page). Ultimately, when the needed information node is reached, the user can ask the CA to read the found content.

The user requests are intercepted by a conversational AI middleware (*ConWeb middleware* in Figure 1), which is in charge of: *i)* interpreting the natural-language utterances to extract meaningful intents and entities, *ii)* transforming them into automatic navigation commands, and *iii)* building the natural-language response to be sent back to the user.

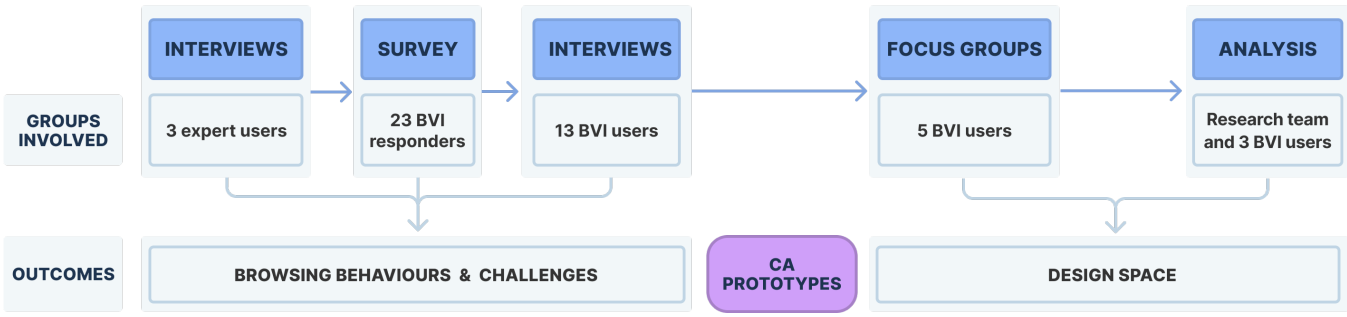

Figure 2: The human-centered approach for the identification of challenges and design dimensions for Conversational Web Browsing.

2.1 Related Works

The Conversational Web browsing paradigm is one of the emerging approaches exploring the integration of conversational capabilities into the Web [7, 11, 16]. It poses technical challenges in terms of how to derive domain knowledge extracted from the Web page content, how to enable natural language interactions, and how to automate browsing actions on a website in order to fulfill a user request expressed in natural language. Previous work already explored these challenges and proved the technical feasibility of the Conversational Web [4, 9]. Researchers have shown that is possible to generate informational bots out of website content [4, 7] and support users in the creation of their own voice-based automation [11, 16]. The proposed technologies, even if preliminary, highlights the potential to make the Conversational Web truly for everyone. Yet, challenges in terms of what type of browsing tasks, conversational patterns and design solutions can improve the UX of this new interaction paradigm still remain unexplored.

The literature, indeed, highlights the need for understanding how the users want to communicate with conversational agents and learn new capabilities [17], for exploring effective conversational patterns [6] and, more in general, for defining guidelines to create accessible Human-AI interactions [3, 14]. The identification of challenges and the definition of guidelines is even more compelling when considering BVI people: despite the evident potential of the conversational paradigm, several studies have highlighted limitations of virtual assistants in the interaction with BVI users, with respect to the input of the information, the control on the information presented, the interaction modalities and even the privacy of users listening to voice personal assistants [1, 2, 10]. In general, creating trust between BVI users and conversational assistants emerges as a paramount factor [1].

The design space that will be discussed in this paper acknowledges the need for further exploring the design of conversational user interfaces, and tries to give a contribution for the specific context of the Conversational Web. Although this paper focuses on challenges highlighted with and for BVI users, the ultimate goal of our research is to promote the notion of Conversational Web to increase the accessibility of the Web for everyone.

3 METHODOLOGY

Figure 2 illustrates the human-centered process [13] that is guiding us in the definition of the new paradigm for Conversational Web Browsing. The goal of this process is to identify and validate conversational patterns and integrate them into a software platform sustaining the notion of Conversational Web Browsing. In this paper we discuss the design space defined by some prominent challenges that we identified through formative studies involving in total 26 BVI users reached out through three Italian BVI associations, Unione Italiana Ciechi e Ipovedenti (UICI), Real-Eyes Sport, Associazione Disabili Visivi (ADV). The studies were authorized by the research ethical committee of Politecnico di Milano (Opinion no. 11/2021). We chose to involve exclusively BVI users as we wanted to identify the most stringent requirements for a conversational browsing detached from the visual channel.

Step 1. Remote preliminary interviews with experts of assistive technologies. We started our research in April 2021, with preliminary structured interviews to gain insights into the challenges and barriers faced by BVI people in accessing digital information and services. The interviews were carried out with 3 digital-technology experts, all but one blind, who educate and assist BVI people in learning and using assistive technologies. The involvement of these experts was important not only for their personal perspective, but also as a proxy to understand the challenges faced by the BVI people they support through their services. Each interview lasted about 2 hours and was conducted remotely, through video-conferencing tools, due to COVID-19 restrictions.

Step 2. Online survey. Guided by the insights collected in the initial interviews, we defined a survey including both closed and open-ended questions[1]; in May we distributed it to a broader group of BVI users who responded to an invitation in the newsletters of the involved associations. We received a total of 23 responses from users aged between 18 and 65 years, 10 of them were totally blind. The survey covered four main aspects: current BVI users' practices for accessing digital services, experience with screen readers, challenges with Web navigation, experience with conversational agents.

[1] See supplementary materials at: https://tinyurl.com/ConWebSurvey

Step 3. Online structured interviews. At the beginning of July we contacted and interviewed 13 BVI users (2 females) among those who had responded to the survey at step 2 and expressed their interest in follow-up activities. We selected them as they had declared to be frequently engaged with digital services, even in work activities, and well versed in the use of assistive technologies. They were aged between 17 and 60 years, 11 were totally blind, 2 visually impaired. Each interview lasted about 1 hour and was conducted remotely. These interviews were useful to further characterize the target users who could benefit from a paradigm for conversational Web browsing, and the challenges in accessing digital services mediated by current assistive technologies and conversational agents. We also wanted to explore in more details how BVI users would need (or like) to be supported by a conversational Web browsing paradigm. All these aspects informed the following design activities.

Step 4. In-presence focus groups and co-design sessions. Among the users who took part to the structured interviews at step 3, we selected 5 participants (2 females) who had a good knowledge of assistive technologies and were also familiar with online navigation. They were aged between 18 and 48 years, two of them were totally blind. At the middle of July, we started the design activities with two exploratory focus groups (the first with 3 users, the second with 2 users). In an 1-hour session, we observed and discussed how the users navigated and explored three different websites to access content (e.g., reading comments on a YouTube video or an article on Wikipedia), and perform some operations (e.g., buying a train ticket). The tasks were chosen according to the familiarity of the users with some online services and websites; we also considered the website level of usability (as defined by Lighthouse, on Google Chrome). We concluded the focus-group activities with a 30-minute session to test some initial hypotheses on conversational patterns. This session was driven by prototypes of conversational agents that we initially created with the intent of eliciting concrete feedback on how people imagine browsing websites mediated by a conversational agent, thus collecting more concrete formative feedback on our design hypotheses.

After one week, a 2-hour co-design session was organized to prototype and test, together with the same users, some scenarios for browsing the Web through conversation. The researchers moderated the development of conversational structures that were rapidly built through DialogFlow[2], deployed on Google Assistant and run on a mobile phone. For selecting meaningful scenarios, we took as a starting point tasks already performed (and challenges already observed) within the focus groups. By allowing participants to actively express themselves in the development of voice chatbots, we were able to materialize the conversational Web scenarios and obtain formative feedback on our initial hypotheses.

Step 5. Analysis and triangulation of results with expert users. Feedback from previous activities were revisited with a further restricted group of 3 participants (1 female), who in the previous steps expressed their extreme interest in being actively involved in our research. In a 2-hour session, we analyzed with them two conversational agent prototypes. The main goal was observing the users' interaction with the conversational agents, and cross-validate with them our design hypotheses.

[2]https://dialogflow.cloud.google.com/

4 DESIGN SPACE

A qualitative analysis of the data gathered in the user study led us to the identification of some concerns and obstacles in the users' current practices for accessing the Web through voice-based assistive technologies. The emerged aspects suggest some design dimensions that we describe in the following and that can drive the definition of design patterns for the Conversational Web in our following research activities.

Shaping-up the map of the navigable space. When describing their experience to access a website through voice-based technologies, the users highlighted that learning the structure of the website is a necessary first step, and also the most cognitive-demanding (P4: *"Sometimes I open a site and it takes me 15 mins or more to understand what it is, where I am and where I can navigate to"*). They reported a number of desiderata that would help in this respect. The most frequent asked for mechanisms to keep track of the **navigational context** during the exploration of the website (P21: *"Sometimes I get lost because I cannot remember the page where I came from. In this way I cannot even return to the home page because I don't recall the path!"*), even when they land on the website coming from external links (P25: *"It would be super-helpful if the CA could provide us with some hints about my previous research on Google! For example: you were searching this, you can find it here on this page"*). Moreover, **high-level navigation mechanisms** are important to convey the website structure in-the-large and the main available functions (P13: *"A well-designed header is undoubtedly one of the standards that a website must meet. I always begin my exploration from there [...]"*). The users also referred to **link predictability** (P20: *"Sometimes I click on a link and I expect to find something that I am interested in, but then the result is not related at all with what I was reading before"*) and **consistency across different websites** (P22: *"Despite the fact that I have never visited this website, I believe it is highly user-friendly because it is very similar to Wikipedia"*).

The users frequently remarked the importance of a **consistent structuring of pages**. P5 said: *"Even if I have a clear idea of what I can or will reach, then it is difficult to locate the content in some pages given their different organization. The pages on Wikipedia all have a similar organization, and so are the paths to explore them"*. The adoption of consistent structures can speed-up the familiarization of users with the website, allowing them to form their model after few exploration of the website (P24: *"It is difficult to get oriented because I can not reuse the same navigation path. Different pieces of similar information are not always at the same importance/navigation level, even if they share similar characteristics"*).

Navigating through intelligible and quick mechanisms. Participants adopted or expressed the need for different navigational strategies depending on their tasks (or goals). They navigated by **in-depth exploration** of the website tree, exploring the different website areas, moving smoothly through different layers by narrowing down navigation options until they reached a node of interest. The most frequently reported need referred to more transparency for the logical connection between nodes (P9: *"Sometimes I don't* *clearly know what I am searching for in a Website and I hope that my navigation would be somehow guided by meaningful links"*). **Navigation by Q&A** emerged as a mechanism to formulate punctual, fast-served requests for specific content, from anywhere and at anytime. "Conversing" with the page to ask about precise information can help users locate the content they need (P16: *"As soon as I open a page I would like to know if there is what I am looking for, without necessarily having to scroll through everything"*). Users have indeed become used to virtual assistants and expect a conversational experience to be able to answer punctual questions (P23: *"It would be really handy for me to be able to navigate from one section to another one directly [...], just like asking Alexa for information, but with the additional advantage of being able to further explore contents if I like them"*). **Navigation by checkpoints** was also highlighted as a means for improving orientation in high-density information websites, where content topics are distributed among different nodes. For these sites, users raised the necessity of surfing through well-known, safe nodes "flagged" by the users themselves (P3: *"It would be interesting to be able to mark elements during my navigation, in order to find them more easily and get oriented even in following browsing sessions!"*).

Summarizing and segmenting the page content. The study clearly highlighted that conversational paradigms cannot consist in a simple transposition of text into voice: this is what happens with screen readers and similar technologies, resulting in poor experiences and high cognitive effort. This aspect became evident when the users asked for **skimming mechanisms**, to summarize page content and prevent unwanted, or unneeded, explorations of target nodes (P6: *"Sometimes reading a page is exhausting because it is too long. I would like something like the preview in Google search for every significant page section!"*). Ideas elaborated during the focus groups and the co-design sessions related to the adoption of **conversational tag clouds** for conveying key concepts, to quickly assess the relevance of the node they are accessing (P12: *"I'd like to know right away if what I'm reading isn't what I'm looking for"*; P17: *"It always takes me too long to figure out why this content is the result of my research. Highlighting the main concepts would help me a lot!"*). Several participants then observed that *an artificial content summarization can be misleading in some contexts, such as in Wikipedia articles.* Therefore, in the co-design sessions users identified the advantage of subdividing content into **labeled content segments** (P4: *"Digging paragraphs makes me annoyed. Perhaps an organized list with the main points and fewer paragraphs would work better"*).

Providing access to conversation-scaffolding intents. During the co-design sessions, users frequently expressed the need for **scaffolding intents** to help them identify the actions that the conversational agent can perform at the page level (*"What can I do here?"*) and within the navigational space (*"Where can I go from here?"*). When accessing a high-density information page, participants would also benefit from re-prompting the reading of specific segments, without necessarily listening to the whole page content (P2: *"Sometimes the information in a page is too dense or articulated and I'm forced to re-listen to it from the beginning"*). When interacting with the conversational prototypes prepared for the focus groups, participants also underlined the benefits of adhering to well-known usability heuristics, for example the provision of **feedback on the system status** to quickly grasp what the conversational agent had done or was going to do (e.g., displacement from a node, navigation to another node, or page loading). The users in particular appreciated the conversational agent "asking before acting" (P24: *"Asking me if I really want to exit a page or telling me that I am going to confirm something would help me know what I have just done and identify what follows up"*).

Controlling and personalizing the Conversational Web. Given the multitude and heterogeneity of websites, it can be difficult to have conversational browsing mechanisms consistent across the whole Web. This was however one frequent need that came out from the users. On the one hand, this leads us to reflect on Web designers' responsibility in enhancing the usability and accessibility of the Web by means of adequate information structures that in turn can ease the creation of a consistent Conversational Web. On the other hand, it is also important to allow the users to control and personalize their conversational experience, to make it coherent with their mental models, and to create trust-building mechanisms. The users must be enabled to control the content manipulation happening during the shift from text to voice (P1: *"I don't want a technology to make navigational decisions for me, not showing me pieces of content because it thinks I could not be interested"*). Additionally, the users frequently expressed the need to cluster information nodes within macro-areas that were more meaningful to them and could help recall a navigation context, e.g., the latest news (P7: *"The biggest obstacle beyond the retrieval of some information items is the difficulty of just recovering them in a simple way, in short bits, coherently with my mental model"*). They also expressed the need to personalize the skimming and summarization mechanisms, to strengthen their control and avoid unwanted side-effects of content filtering or nudging.

Detaching the Conversational Web from the visual Web. A conversational paradigm detached from the visual organization of pages is a strong need for BVI people. In addition to the interpretation of visual features coding information (e.g., colors or text styles), severe problems they often encounter when reading the visual Web are related to page components that require visual abilities. For example, when interacting with a train-reservation website, they had difficulties selecting the departure date on a calendar component: P22: *"These letters spoken by the screen reader I believe are the days' initials; but it took me a bit to figure out"*; P26: *"We had a lot of troubles inserting the departure time since we had to select it from a calendar visualization".*

5 PROTOTYPE

Figure 3 describes the conceptual architecture of the current prototype of *ConWeb*, the platform that we are currently consolidating to enable the conversational Web browsing. The *ConWeb voice client* manages the interaction with the conversational agent, also handling the transformation of the users' voice requests into text, and of the server responses into voice. In the current implementation it is a plugin for the Firefox Web browser; we have already planned additional deployments through virtual assistants (e.g., Alexa), or on dedicated smart objects.

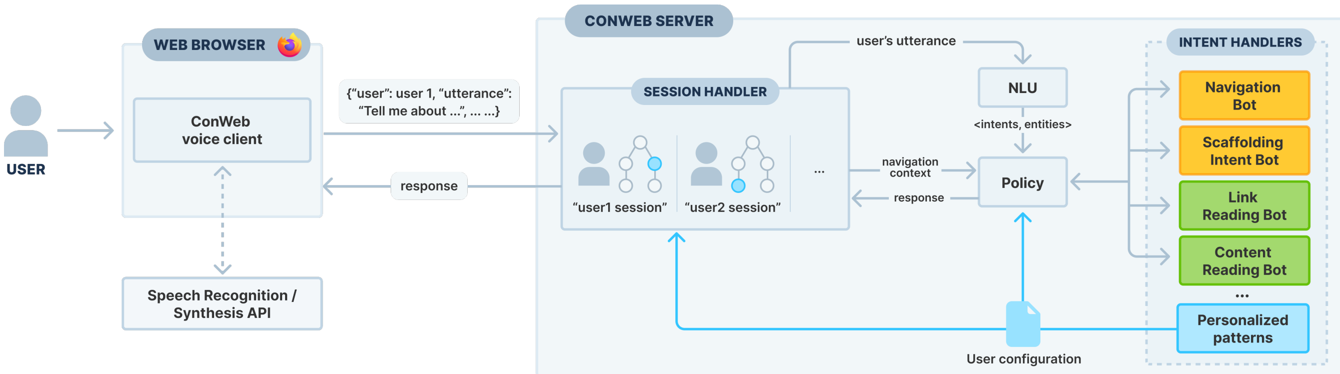

Figure 3: Conceptual architecture of the ConWeb platform. The Intent Handlers support the flexible introduction of conversational patterns and personalization mechanisms.

At the server side, the user's utterances are interpreted by an *NLU engine* (RASA[3] in the current implementation). The *Policy module* further elaborates the extracted intents and entities by contextualizing them with respect to the user's navigation tracked by a *Session Handler*. At the first access to a Web page, the Session Handler builds a "domain knowledge" by automatically extracting from the HTML code some features of the website content and functionality. Also based on this domain knowledge, the Policy module can trigger the *Intent Handlers* serving the user's request in a given navigation state. The Intent Handlers manage the conversation and perform the appropriate Web browsing actions on behalf of the user. Based on their output, the Policy module then builds the response to be sent to the client.

The technical feasibility and performance of the mechanisms for interpreting the user's utterances and translating them into browsing commands have been already discussed in previous publications [4, 9]. Our current work is devoted to configuring the different intent handlers to support the browsing patterns discussed in this paper. The organization of these modules guarantees flexibility in the introduction of conversational patterns. Some handlers manage fixed patterns, for example those for scaffolding commands. Some others can adapt their outcome based on preferences expressed by the users, for example those related to text reading. This last feature responds to the need for personalization recurrently remarked during the focus groups and the co-design activities.

Major efforts are being devoted exactly to enable the users to personalize the conversational patterns. Within the conceptual architecture reported in Figure 3, the user's preferences are represented within a configuration file that is generated by a *Personalized Patterns* handler, in charge of interpreting and managing personalization utterances. This specification feeds the Session Handler and the Policy modules with modifiers acting on the default conversational patterns. Specific attention, however, must be posed on the way users can express their preferences through conversation, which implies the need for adequate conversational patterns. Together with the definition and validation of the navigation and content reading patterns, the user-based specification of personalized patterns will be the object of our future work. These aspects will require extensive users studies conducted on a consolidated implementation of the ConWeb platform.

[3]https://rasa.com/

6 CONCLUSIONS

This paper has discussed some design challenges for a new paradigm for the Conversational Web, as emerged from a user-centered process conducted with BVI users. Evaluation is fundamental to verify the value of our approach. However, the opinions of the users involved in the formative evaluation make us glimpse the positive impact that the approach can have on them. During one interview, P4 said: *"One of the biggest limitations of current conversational agents is that they stop just after opening a website. Siri can lead me until I find something interesting, but after that she is not useful anymore."*

Our current and future work is devoted to the identification and specification of conversational patterns for Web browsing, their integration within our current software prototype, and the validation of the resulting conversational paradigm through extensive user studies. This last activity will also address sighted users, to understand to which extent the assumptions derived with and for BVI users can be extended to any class of users.

In addition to consolidating the ConWeb platform, our activity of pattern definition aims to fill the current gap in the literature and contribute with guidelines that can be reused across different projects for the design of conversational agents for the Web. More in general, in line with some standardization activities already underway (for example the Speakable schema proposed by Schema.org[4]), our ultimate goal is to promote the notion of Conversational Web by means of Web technology extensions that can natively support conversational access.

ACKNOWLEDGMENTS

We are grateful to the associations UICI (Unione Italiana Ciechi), ADV (Associazione Disabili Visivi), Real Eyes Sport, and to all the users who contributed to our research: they helped us understand how inclusiveness can be a value for everyone. This paper is also dedicated to Prof. Florian Daniel, who suddenly passed away in April 2020. He first had identified this value.

[4]See https://schema.org/SpeakableSpecification

REFERENCES

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. 2018. " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 249–258.

[2] Ali Abdolrahmani, Kevin M Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M Branham. 2020. Blind leading the sighted: drawing design insights from blind users towards more productivity-oriented voice interfaces. *ACM Transactions on Accessible Computing (TACCESS)* 12, 4 (2020), 1–35.

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[4] Marcos Baez, Florian Daniel, and Fabio Casati. 2020. Conversational Web Interaction: Proposal of a Dialog-Based Natural Language Interaction Paradigm for the Web. In *Chatbot Research and Design*. Springer, 94–110.

[5] Moshe Chai Barukh, Shayan Zamanirad, Marcos Baez, Amin Beheshti, Boualem Benatallah, Fabio Casati, Lina Yao, Quan Z Sheng, and Francesco Schiliro. 2021. Cognitive augmentation in processes. In *Next-Gen Digital Services. A Retrospective and Roadmap for Service Computing of the Future*. Springer, 123–137.

[6] Sara Bouguelia, Hayet Brabra, Shayan Zamanirad, Boualem Benatallah, Marcos Baez, and Hamamache Kheddouci. 2021. Reusable Abstractions and Patterns for Recognising Compositional Conversational Flows. In *International Conference on Advanced Information Systems Engineering*. Springer, 161–176.

[7] Julia Cambre, Alex C Williams, Afsaneh Razi, Ian Bicking, Abraham Wallin, Janice Tsai, Chinmay Kulkarni, and Jofish Kaye. 2021. Firefox Voice: An Open and Extensible Voice Assistant Built Upon the Web. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[8] Yung-Ju Chang, Chu-Yuan Yang, Ying-Hsuan Kuo, Wen-Hao Cheng, Chun-Liang Yang, Fang-Yu Lin, I-Hui Yeh, Chih-Kuan Hsieh, Ching-Yu Hsieh, and Yu-Shuen Wang. 2019. Tourgether: Exploring Tourists' Real-time Sharing of Experiences as a Means of Encouraging Point-of-Interest Exploration. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4 (2019), 128:1–128:25. https://doi.org/10.1145/3369832

[9] Pietro Chittò, Marcos Baez, Florian Daniel, and Boualem Benatallah. 2020. Automatic Generation of Chatbots for Conversational Web Browsing. In *International Conference on Conceptual Modeling*. Springer, 239–249.

[10] Dasom Choi, Daehyun Kwak, Minji Cho, and Sangsu Lee. 2020. " Nobody Speaks that Fast!" An Empirical Study of Speech Rate in Conversational Agents for People with Vision Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[11] Michael H Fischer, Giovanni Campagna, Euirim Choi, and Monica S Lam. 2021. DIY assistant: a multi-modal end-user programmable virtual assistant. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*. 312–327.

[12] Asbjørn Følstad, Theo Araujo, Effie Lai-Chong Law, Petter Bae Brandtzaeg, Symeon Papadopoulos, Lea Reis, Marcos Baez, Guy Laban, Patrick McAllister, Carolin Ischen, et al. 2021. Future directions for chatbot research: an interdisciplinary research agenda. *Computing* (2021), 1–28.

[13] Joseph Giacomin. 2014. What is human centred design? *The Design Journal* 17, 4 (2014), 606–623.

[14] 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*. 1–11.

[15] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. " Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *CHI 2018*. 1–13.

[16] Gonzalo Ripa, Manuel Torre, Sergio Firmenich, and Gustavo Rossi. 2019. End-User Development of Voice User Interfaces Based on Web Content. In *IS-EUD 2019*. Springer, 34–50.

[17] Francisco AM Valério, Tatiane G Guimarães, Raquel O Prates, and Heloisa Candello. 2017. Here's what I can do: Chatbots' strategies to convey their features to users. In *Proceedings of the xvi brazilian symposium on human factors in computing systems*. 1–10.

[18] WebAIM. 2017. *The WebAIM Million: An accessibility analysis of the top 1,000,000 home pages*. https://webaim.org/projects/million/

[19] Shunguo Yan and PG Ramachandran. 2019. The current status of accessibility in mobile apps. *ACM Transactions on Accessible Computing (TACCESS)* 12, 1 (2019), 1–31.