FindAssist: Assessing Contextual Keyword Search in Voice Agents

Danqi Li[*]
lidanq01@student.ubc.ca
University of British Columbia
Canada

Du Pyo Park[*]
dparkdp@student.ubc.ca
University of British Columbia
Canada

Karel Harjono[*]
harjono@student.ubc.ca
University of British Columbia
Canada

Abstract

Voice Agents (VAs) often struggle with rigidity and limited contextual awareness, especially during hands-free and focus-intensive activities. We propose a technique that enhances document search within voice-based systems by enabling context-aware keyword searches through voice commands. This approach combines semantic understanding with the conversational capabilities of VAs, allowing for more precise and relevant information retrieval. By seamlessly integrating advanced search algorithms with voice interaction, our technology improves user efficiency and satisfaction during focus-intensive tasks. We then discuss the implementation and evaluation of this technique, highlighting its positive impact on usability and task performance, and demonstrating its potential to transform voice-based document exploration.

ACM Reference Format:
Danqi Li, Du Pyo Park, and Karel Harjono. 2024. FindAssist: Assessing Contextual Keyword Search in Voice Agents. In . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Voice Agents (VAs) have become an integral part of modern technology, facilitating hands-free interaction and providing users with convenience in various contexts. Despite their growing popularity, VAs often exhibit limitations in understanding context and providing relevant information, especially during complex tasks such as cooking[5]. These shortcomings highlight the need for more sophisticated, context-aware systems that can enhance user experience and efficiency. In recent years, significant studies have been made in developing multimodal interaction techniques to aid users in various activities. For example, Yadav et al. (2015)[16] designed a system to enable non-linear navigation in educational videos using a combination of audio and visual content. Their approach employed dynamic word-clouds and 2-D timelines to facilitate quick navigation to points of interest, demonstrating substantial improvements in user efficiency compared to traditional transcription-based interfaces[16]. Similarly, systems like RubySlippers have been developed to support content-based voice navigation for how-to videos, allowing users to issue keyword-based queries instead of temporal commands[2]. Building on these advancements, our research focuses on making VAs more contextually aware and responsive. Previous studies show that adding contextual information can greatly improve interaction quality and user satisfaction, particularly in tasks like cooking, leading to smoother and more relevant responses [5]. However, it was highlighted that User tends to lose trust in VAs once an error has occurred[9]. This forces user to use simple words, short phrases, and keyword-like expressions when forming commands [8, 9]. When these interactions fail, they often adjust their language by hyper-articulating, restructuring commands, or changing their accent to increase the likelihood of the VA understanding them [8, 9, 10, 11]. In an effort to explore ways to increase trust in VA systems, we introduce FindAssist, a novel technique that enhances document search within voice-based systems by enabling context-aware keyword searches through voice commands. Our approach combines semantic understanding with the interactivity of a voice interface, allowing for more precise and relevant information retrieval. This technique is particularly useful for tasks such as cooking, where users need quick access to specific information without manually navigating through documents or videos. By incorporating our method, we aim to address the limitations of previous work by facilitating quick navigation through documents, thereby complementing VA interactions. We evaluate the performance of our approach through comprehensive user studies, demonstrating its impact on user satisfaction and task efficiency. The results indicate a significant improvement in the usability of voice-based document searches, suggesting that our technique not only addresses the current limitations of VAs but also sets a new standard for voice-activated content exploration.

1.1 Contributions

- **Context-aware Keyword Searches:** We introduce a method that integrates semantic understanding and contextual awareness into VA keyword searches, enhancing precision and relevance.
- **User Study and Evaluation:** Our user studies reveal a substantial improvement in task efficiency and user satisfaction when using our context-aware VA system.
- **Enhanced Usability:** The findings suggest that our approach significantly improves the usability of voice-based document searches, particularly in hands-free, focus-intensive scenarios.
- **Accessibility Improvements:** By focusing on hands-free and context-aware interactions, our system offers substantial improvements in accessibility for users with vision impairments, particularly in complex tasks such as cooking.

Through this study, we aim to push the boundaries of VA capabilities, ensuring that they can effectively support users in complex

[*] All authors contributed equally to this research.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

and dynamic environments, thereby enhancing productivity and user experience.

2 Main Topics

The number of people who use VAs, which is a sophisticated autonomous system that can communicate and interact with the user. to help with their cooking are increasing. But their usage of the VAs are not yet completely satisfactory. When using VAs without "voice-to-search" it was found, users move up and down a text to search for what they are looking for, showing difficulty[3]. With no functionality to search through voice, users sought to manually search up what they heard[16].

Weber et al. (2023) states that participants of their study responded positively to the idea of an entity that assists you based on what actions you make in the kitchen[15]. This suggests that there are more people who may positively receive the idea of a VA that allows one to search through voice. Moreover, they state that it is important for the user to have a feeling of control while cooking. They state that users were more confident in the assistant they are using when it is transparent and communicative in what action it takes before it makes any conclusive action. Our implementation gives the user complete control over what they can ask the user and does not take any action before warning, therefore our implementation of searching through voice should be well received by the general public.

2.1 Improvements of Previous Works

Our study builds on Jaber et al. (2024)'s study by giving a method to the user that quickly finds and reads a keyword from the original recipe text[5]. Their study assumes the participant does not wish to manually search up the context of the recipe article the VA gets its information from. Their study uses the wizard-of-oz technique, which means they did not give any implementation of practical code. However, our study uses an actual VA powered by the OpenAI engine for semantic searching. The semantic search is aware of the context of what the user asks, making it context-aware in the scope of the text of the recipe only.

Li et al. (2024) states that the impaired required reducing the need for precise physical movement for tasks such as using the knife and throwing things in the garbage[7]. In order to help the impaired have the reduced need of precise physical movement, our system completely relies on vocal cues to start the "voice-to-search" technology. Furthermore, Li et al. (2024) suggests that users prefer lower verbosity levels, so our project has a low-verbosity level[7]. Li et al. (2024) did not consider the visually impaired people who had no cooking experience (part 8). They only considered the people who had some cooking experience. Our semantic search engine provides even the visually impaired person who has no cooking experience an opportunity to cook[7].

Schneider et al. (2023) 's study states that VAs also help the elderly when they search up news. Our project allows a user to cook without reading the recipe text at all[13]. If our project does not require one to know how to navigate through new technology, then the elderly should be able to have an easier time using our app. Our project only requires the elderly person to speak to the phone and ask it questions concerning the recipe text. Thus, our project makes the additional contribution of providing the elderly a new way of cooking, much like how our our project helps the visually impaired.

Reicherts et al. (2022)'s study states that the VA provides the user a faster data analysis and that a conversational interactive VA, caused the user to interact with the agent more and had the potential to be seamlessly blended into our social interactions[12]. Through quantitative results in the efficiency and accuracy of searching through voice, our study provides further concrete proof that using the VA to search is a much faster way to analyze the data than manually searching. Furthermore they state that VAs were the preferred method of communicating to the computer rather than the method of touch screens. Our quantitative results in asking the user questions provides us further proof that the user prefers to use the context-aware voice over touching the screen to search something up.

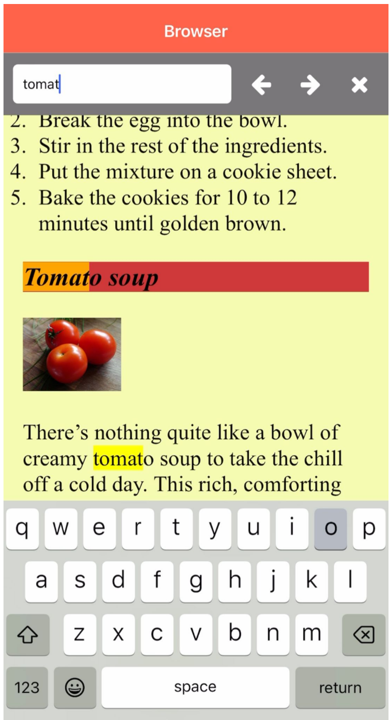
According to Seaborn et al. (2021), voices in vHAI (virtual-human-AI) have the added potential of influencing our thoughts and expressing its own opinions[14]. Seaborn et al. states that there is a special need to evaluate more on the impact of voices in computer technology[14]. Aggarawal et al. (2023) discovered that users are worried VAs with excessive empathy will negatively affect our lives, including our interaction with other people and our general emotional mood[1]. Kim et al (2022)'s study also makes a similar conclusion: participants wished for VAs to work more as assistants rather than as human substitutes[6]. They recommended future VAs in production to be more wary of the how one can perceive the difference between a computer-assistant and a human being [6]. Therefore, our VA incorporates little to no empathy on the part of the assistant and also allows the user to respond to the VA in a more primitive way - by using the keyboard. Our VA expresses little to no opinion due to the fact that, when the VA speaks out loud, it only has the ability to repeat the sentence a keyword is found in and requests for confirmation of a completed query. The low verbosity of the VA also makes our VA to be less human-like.

Therefore, our study will improve context-based searching through the voice that helps the elderly and the visually impaired. It will provide faster data analysis, better proof that the user prefers the voice command over the manual search, and better assistance to visually impaired people who have no cooking experience. Our implementation gives the user control over the kitchen, does not clutter their head with high verbosity, and does not make them anxious that a robot will have too much impact on their emotions and autonomy.

3 Experiment Design

3.1 Goals

In order to investigate the effects of different input methods on the efficiency, accuracy and user experience of people looking up specific keywords in digital recipes, we conducted a within-subject design where participants looked up specified words in two different ways while browsing two digital recipes. With this study, we aimed to determine whether voice agent technology provides a more efficient and user-friendly approach to this task than traditional cell phone keyboard. This study differs from previous experiments in that it focuses specifically on the context of recipe lookup, a



2. Break the egg into the bowl.
3. Stir in the rest of the ingredients.
4. Put the mixture on a cookie sheet.
5. Bake the cookies for 10 to 12 minutes until golden brown.

Tomato soup

There's nothing quite like a bowl of creamy tomato soup to take the chill off a cold day. This rich, comforting

Figure 1: Screenshot taken from one of the participants performing the experiment with Keyboard Interface in [4]

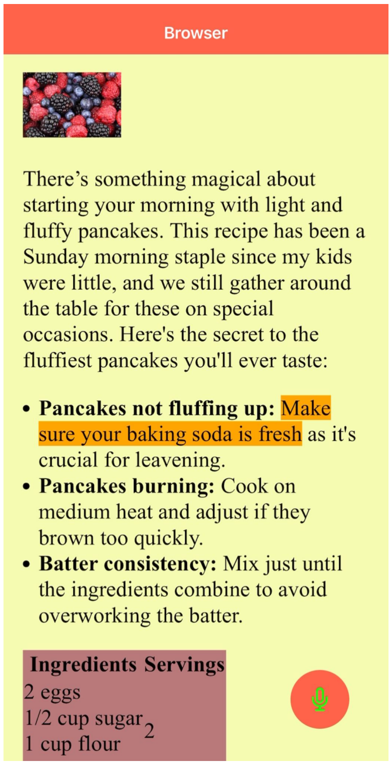
very common setting in everyday life but was not fully developed by past works. The study aims to improve one of the challenges that speech agents often face - insufficient context awareness. In addition, the study utilized a within-subjects design, which controls for possible individual differences by providing a direct comparison between voice input and cell phone keyboard.

3.2 Participants

Four male participants, aged between 18 and over 60 years, were recruited for the study. The participants represented diverse backgrounds. At the time of the study, two participants were residents of Canada, while the remaining two were from Indonesia and China, respectively.

3.3 Apparatus

The experimental setup consisted of two types of hardware and two types of software, where the hardware includes smartphone with built-in microphone for voice input and standard keyboard (QWERTY) support typing. The software consisted of a web browser application, similar to *Safari*, that supported both voice recognition2 and keyboard input1 capability as well as a timing software that was required to record the time taken to complete each task. There are potential features that may affect the results of the experiment, such as the built-in microphone that may have voice recognition delays.



There's something magical about starting your morning with light and fluffy pancakes. This recipe has been a Sunday morning staple since my kids were little, and we still gather around the table for these on special occasions. Here's the secret to the fluffiest pancakes you'll ever taste:

- **Pancakes not fluffing up:** Make sure your baking soda is fresh as it's crucial for leavening.
- **Pancakes burning:** Cook on medium heat and adjust if they brown too quickly.
- **Batter consistency:** Mix just until the ingredients combine to avoid overworking the batter.

Ingredients	Servings
2 eggs	
1/2 cup sugar	2
1 cup flour	

Figure 2: Screenshot taken from one of the participants performing the experiment with Voice Interface in [4]

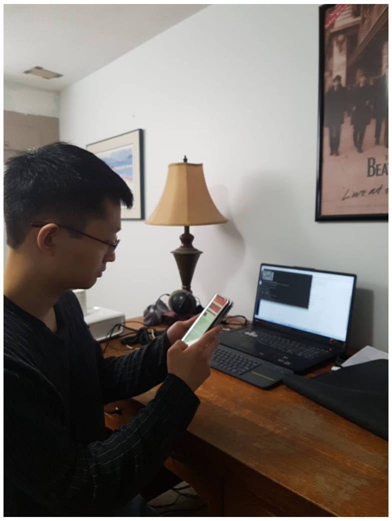


Figure 3: Picture of participant doing the experiment

3.4 Procedure

3.4.1 Task. To explore the effects of input method, we designed a within-subject design study with input mode as a within-subject factor (voice vs. cellphone keyboard)—all participants performed keyword in each condition with different recipes (i.e., Recipe A and

Recipe B) retrieval were exposed to both conditions of input mode. Across the sample, we counterbalanced the order of input methods (conditions) between participants using Latin squares. This resulted in 4 participants being randomly assigned to either group 1 or group 2, with 2 participants in each group. Group 1 participants started with voice input while group 2 participants started with keyboard. In each case, participants were asked to use either voice input or cell phone keyboard to find 6 specific words in that Recipe. After the task was completed, each participant was asked to answer a series of questions[4].

3.4.2 Participant Procedures. Participants were first provided with a brief description of the purpose of the study. The experiment was then initiated by collecting verbal consent from each participant, and participants were informed of their right to withdraw at any time throughout the experiment. During the experiment, we first gave participants a brief tutorial on using the speech technique and the cell phone keyboard, and gave each participant the opportunity to practice both methods to familiarize themselves with the input method. Next, participants were randomly assigned to either Group 1 or Group 2. Regardless of group, each participant retrieved the keywords using both input methods and searched separately in different recipes. For example, if in the first recipe condition, participants used voice input for keyword retrieval, then in the second recipe condition, the same participant would use cell phone keyboard typing for keyword retrieval. For each input method, participants were required to perform a retrieval for six specific words. It is worth noting that we will give 6 specified questions for each recipe and the participant has to think of a word to answer that question. When they find the answer, they move on to the next question until they have completed the full keyword search. For each participant, we measured and recorded the time it took to complete each task as well as recording the number of errors in each task. Also, in order to simulate a real-life cooking situation when a user tries to complete a keyword search on the phone's keyboard, we had to assume that the user had washed his/her hands before touching the phone, so we added a fixed amount of time (5 seconds) to each trial for the keyboard task. In addition, after completing all tasks, participants are required to fill out a short questionnaire that includes rate their experience on ease of use and satisfaction of each input methods. each experiment will last approximately 15 minutes, including training, task completion, and post-task scoring.

3.5 Design

We used a within-subject repeated measures design. For each input mode, participants had to perform keyword searches in one recipe, with a total of two recipes using two input modes. Each recipe consisted of roughly 20 simple sentences to balance the length of the recipe. The independent variable is input type. It has two levels, voice and cell phone keyboard). The dependent variable is the completion time, accuracy and the user satisfaction. the completion time refers to the time taken for one participant to find the specific word (measured in seconds). The accuracy is the number of errors each participant made during the retrieval process. The user satisfaction is the participants' ratings of their satisfaction with the input method (measured using a Likert scale from 1 to 5). Each participant will complete two tasks, one using voice technology and the other using ell phone keyboard. The order of the tasks will be counterbalanced to control for order effects. As well as each task will be repeated six times to ensure unbiased data. Each participant will provide data for two input methods for three different dependent variables, for a total of six data points per person:

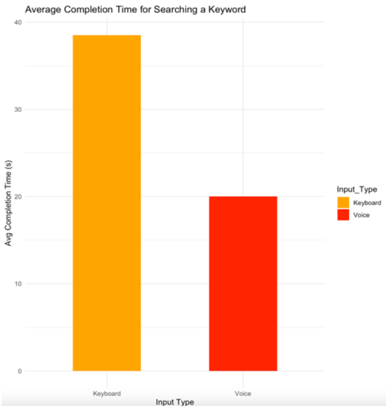


Figure 4: Bar Graph Demonstrating the Difference in Average Completion Time for Searching a Keyword in a recipe between the Voice and Keyboard Input Methods.

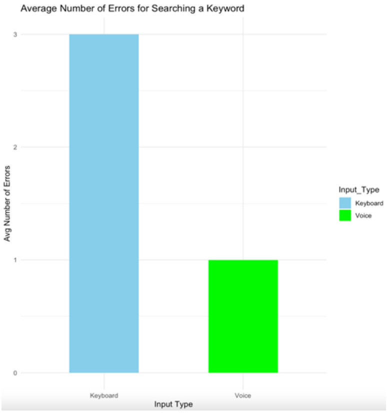


Figure 5: Bar Graph Demonstrating the Difference in Average Number of Errors for Searching a Keyword in a recipe between the Voice and Keyboard Input Methods.

(1) Completion Time for Voice Input
(2) Completion Time for cell phone keyboard
(3) Accuracy for Voice Input
(4) Accuracy for cell phone keyboard
(5) User Satisfaction for Voice Input
(6) User Satisfaction for cell phone keyboard

This design ensures that each participant serves as their own control, reducing the influence of individual differences and increasing the internal validity of the study. The data collected will allow for a comprehensive analysis of the efficiency, accuracy and user experience of the two input methods.

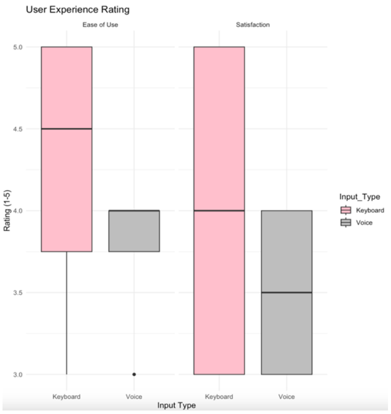


Figure 6: Box Plot Demonstrating the Difference in User Experience Rating between the Voice and Keyboard Input Methods.

4 Result

We first report the effect of input method on the efficiency of users in finding specific keywords in recipes. Then, we report an analysis of lookup accuracy when different input methods are used. Finally, we investigate the impact of different input methods on the user experience. Across all three measures, no outlier was found.

- **Completion Time**: To investigate whether input method affected participants' search efficiency, we analyzed the data using paired t-tests and reported the effect sizes using *Cohen'd*. Figure 4 shows the average time participants took to find a specific word in a recipe using each of the two input methods. A paired-samples t-test was conducted and revealed that there was a significant difference in the completion times for voice input (*M* = 14.71, *SD* = 3.68) and keyboard (*M* = 29.54, *SD* = 9.21); *t*(3) = -3.35, *p* = .04, 95% *CI* [-28.92, -0.75]. These results suggest that on average, the voice input method was faster than keyboard method, the difference was statistically significant.The effect size (*Cohen's d* = 1.68) also revealed a large effect, indicating the difference in completion times is practically substantial.
- **Accuracy**: We measured error rates in our experiments to investigate how different input methods affected participants' search accuracy. Figure 5 shows the average number of errors made by participants in finding specific keywords in a recipe using each of the two input methods. A paired-samples t-test was also conducted to compare the error rate of voice and cell phone keyboard input methods. There was not a significant difference in the scores for voice input (*M* = 0.25, *SD* = 0.50) and keyboard (*M* = 1.25, *SD* = 1.50); *t*(3) = -1.41, *p* = .25, 95% *CI* [-3.25, 1.25]. This suggests that the difference in accuracy between the two input methods was not statistically significant. In addition, the effect size (*Cohen's d* = .63) revealed a medium effect size, suggesting a moderate difference in the error rate between the two input methods.
- **User Satisfaction**: In the post-task rating, we asked participants to rate the ease of use and satisfaction (scale of 1-5) for each input type. Figure 6 shows the participants' experience ratings of the two input methods for keyword finding.A Friedman test was conducted to evaluate differences in ease of use between voice and keyboard input methods. the results indicated that the difference was not statistically significant, ^{2}*(1, N = 4)* = 3.00, *p* = .08. This suggests that participants did not perceive a particular input method to be significantly easier to use. Similarly, a Friedman test was also conducted to evaluate differences in satisfaction between voice and keyboard input methods. The results revealed a statistically significant difference in satisfaction, ^{2}*(1, N = 4)* = 4.00, *p* = .046. This suggests that participants were more satisfied with voice input compared to keyboard input, with the difference being statistically significant.

5 Discussion

The results of this study reveal important differences between voice input and keyboard input on three key metrics: completion time, accuracy (number of errors), and user experience (ease of use and satisfaction).

For the measure of completion time, the results confirmed our hypothesis that participants using voice input were able to complete the context-based keyword finding task faster compared to the cell phone keypad method. The reduced time can be attributed to the context-awareness feature of our VA system. It allowed participants to express exactly what they were looking for and give the most contextualized answer. In fact, when performing a keyword search task on a recipe, the user has to think of and search a word that answer the question we gave based on the recipe. Since our voice app is based on semantic understanding specifically, our app performs a context-based semantic search based on the participant's voice commands to highlight the correct or closest to correct keywords for them. This greatly reduces the time required for searching compared to doing it manually. This increased efficiency further supports the usability and effectiveness of VA in context-based scenarios. The large effect size emphasizes the practical significance of this finding.

For the measure of accuracy, the effect of the difference in input method on the participants' accuracy in performing the keyword search task on the recipes was not statistically significant, suggesting that both input methods were fairly accurate. It is worth noting, however, that on average, the number of errors for voice input was slightly lower, a slight trend that may require further investigation with larger sample sizes.

For the measure of user satisfaction, in partial agreement with our hypotheses, the results showed that while there was no significant difference in ease of use between the two input methods, participants were generally more satisfied with using voice input. This finding suggests that while participants found both methods easy to use, they were more satisfied with voice input, which may be related to the efficiency it provided in completing tasks more quickly.

Interestingly, the analysis showed that participants who were more familiar with the VA technique showed greater improvement in completion time, suggesting that familiarity with the technique may improve its effectiveness.

6 Limitations

We recognize several limitations of this study. First, the small sample size (N = 4) reduces the statistical power of this study and limits the generalizability of the findings. While the within-subjects design effectively controlled for individual differences, a larger sample size would have provided more reliable and robust results. Second, we artificially added a fixed number (5s) for task completion time using the keyboard, designed to mimic the step of hand washing that would occur before looking up a recipe when cooking in real life, but again, this introduces potential confounding variables. Then, participants' potential unfamiliarity with the voice input technology could then affect their performance. In the experiment, two participants required external prompts, which could have introduced variability not entirely caused by the input method itself. These factors may have affected the results, especially in terms of ease of use and satisfaction. Fourth, in this study, participants conducted the experiment in a controlled environment, so the external validity we found is limited. Finally, our study is limited to the cooking setting, and is not necessarily generalizable to other situations.

7 Implications for Design

Considering the significant improvements in speed and user satisfaction, this research has clear implications for the design of VA context-aware systems. Especially for systems used in complex and context-based tasks which users need to perform fast document searches, such as cooking applications and educational software.

8 Conclusion

In conclusion, the results of this study show that there was no significant difference in accuracy and ease of use ratings between voice and keyboard input methods. However, there was a significant difference in efficiency and satisfaction ratings, with the voice input method being significantly faster and more satisfying when searching a specific keyword in a digital recipe. These findings highlight the potential of a more sophisticated, context-aware systems that can enhance user experience.

In future studies, it is necessary to recruit more participants for the experiment to ensure the reliability of the results. For the step of simulating handwashing during the experiment, future research could attempt to have participants go through a realistic simulation of the cooking process thereby increasing the internal validity of the experiment. Also, given the technical challenges faced by participants, future studies may include more tailored training sessions to ensure that participants are familiar with the technique before starting the experiment. The external validity of our findings is also limited, therefore, future studies are encouraged to replicate our study in a more realistic setting. The robustness of this study can be verified by using different recipes and study populations. Furthermore, in the study, Our FindAssist app is a semantic searching methodology for the cooking context, and in future studies, we hope to explore FindAssist effectiveness on other complex tasks, thus providing broader insights into its usability and effectiveness under different domains. As AI technology continues to evolve, harnessing the power of context-aware VA based may revolutionize the way users interact with technology, improving the efficiency of everyday tasks and user satisfaction.

References

[1] Tanuj Aggarwal and Jorge Goncalves. 2024. Towards empathetically responsive voice assistants. In *Proceedings of the 35th Australian Computer-Human Interaction Conference* (OzCHI '23). Association for Computing Machinery, Wellington, New Zealand, 669–678. ISBN: 9798400717079. DOI: 10.1145/3638380.3638398.

[2] Minsuk Chang, Mina Huh, and Juho Kim. 2021. Rubyslippers: supporting content-based voice navigation for how-to videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21) Article 97. Association for Computing Machinery, Yokohama, Japan, 14 pages. ISBN: 9781450380966. DOI: 10.1145/3411764.3445131.

[3] Philip J. Guo and Katharina Reinecke. 2014. Demographic differences in how students navigate through moocs. In *Proceedings of the First ACM Conference on Learning @ Scale Conference* (L@S '14). Association for Computing Machinery, Atlanta, Georgia, USA, 21–30. ISBN: 9781450326698. DOI: 10.1145/2556325.2566247.

[4] Karel Harjono. [n. d.] Voice interface experiment video. https://drive.google.com/file/d/15y-5-OT-SBFL0aC2fSw6KBt4fxkFZ3Zy/view?usp=sharing. Accessed: 08-12-2024. ().

[5] Razan Jaber, Sabrina Zhong, Sanna Kuoppamäki, Aida Hosseini, Iona Gessinger, Duncan P Brumby, Benjamin R. Cowan, and Donald Mcmillan. 2024. Cooking with agents: designing context-aware voice interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24) Article 551. Association for Computing Machinery, Honolulu, HI, USA, 13 pages. ISBN: 9798400703300. DOI: 10.1145/3613904.3642183.

[6] Hyeji Kim, Inchan Jung, and Youn-kyung Lim. 2022. Understanding the negative aspects of user experience in human-likeness of voice-based conversational agents. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (DIS '22). Association for Computing Machinery, Virtual Event, Australia, 1418–1427. ISBN: 9781450393584. DOI: 10.1145/3532106.3533528.

[7] Franklin Mingzhe Li, Michael Xieyang Liu, Shaun K. Kane, and Patrick Carrington. 2024. A contextual inquiry of people with vision impairments in cooking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24) Article 38. Association for Computing Machinery, Honolulu, HI, USA, 14 pages. ISBN: 9798400703300. DOI: 10.1145/3613904.3642233.

[8] 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* (CHI '16). Association for Computing Machinery, San Jose, California, USA, 5286–5297. ISBN: 9781450333627. DOI: 10.1145/2858036.2858288.

[9] Sharon Oviatt, Jon Bernard, and Gina-Anne Levow. 1998. Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech*, 41, 3-4, 419–442. PMID: 10746365. eprint: https://doi.org/10.1177/002383099804100409. DOI: 10.1177/002383099804100409.

[10] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, Montreal QC, Canada, 1–12. ISBN: 9781450356206. DOI: 10.1145/3173574.3174214.

[11] Stuart Reeves and Martin Porcheron. 2022. Conversational ai: respecifying participation as regulation. William Housley, Adam Edwards, Roser Beneito-Montagut, and Richard Fitzgerald, (Eds.) hardback isbn. (2022). https://nottingham-repository.worktribe.com/output/8499241.

[12] Leon Reicherts, Yvonne Rogers, Licia Capra, Ethan Wood, Tu Dinh Duong, and Neil Sebire. 2022. It's good to talk: a comparison of using voice versus screen-based interactions for agent-assisted tasks. *ACM Trans. Comput.-Hum. Interact.*, 29, 3, Article 25, (Jan. 2022), 41 pages. DOI: 10.1145/3484221.

[13] Phillip Schneider, Nils Rehtanz, Kristiina Jokinen, and Florian Matthes. 2023. Voice-based conversational agents and knowledge graphs for improving news search in assisted living. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments* (PETRA '23). Association for Computing Machinery, Corfu, Greece, 645–651. ISBN: 9798400700699. DOI: 10.1145/3594806.3596534.

[14] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. 2021. Voice in human–agent interaction: a survey. *ACM Comput. Surv.*, 54, 4, Article 81, (May 2021), 43 pages. DOI: 10.1145/3386867.

[15] Johanna Weber, Margarita Esau-Held, Marvin Schiller, Eike Martin Thaden, Dietrich Manstetten, and Gunnar Stevens. 2023. Designing an interaction concept for assisted cooking in smart kitchens: focus on human agency, proactivity, and multimodality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (DIS '23). Association for Computing Machinery, Pittsburgh, PA, USA, 1128–1144. ISBN: 9781450398930. DOI: 10.1145/3563657.3595975.

[16] Kuldeep Yadav, Kundan Shrivastava, S. Mohana Prasad, Harish Arsikere, Sonal Patil, Ranjeet Kumar, and Om Deshmukh. 2015. Content-driven multi-modal

techniques for non-linear video navigation. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (IUI '15). Association for Computing Machinery, Atlanta, Georgia, USA, 333–344. ISBN: 9781450333061. DOI: 10.1145/2678025.2701408.