Cooking With Agents: Designing Context-aware Voice Interaction for Complex Tasks

Razan Jaber
razan@dsv.su.se
Stockholm University
Stockholm, Sweden

Sabrina Zhong
sabrinazhongzl@gmail.com
UCL Interaction Centre
University College London
London, United Kingdom

Sanna Kuoppamäki
sannaku@kth.su.se
Department of Biomedical
Engineering and Health Systems
KTH Royal Institute of Technology
Stockholm, Sweden

Aida Hosseini
idaho@kth.se
Department of Biomedical
Engineering and Health Systems
KTH Royal Institute of Technology
Stockholm, Sweden

Iona Gessinger
iona.gessinger@ucd.ie
University College Dublin
Dublin, Ireland

Duncan P Brumby
d.brumby@ucl.ac.uk
UCL Interaction Centre
University College London
London, United Kingdom

Benjamin R. Cowan
benjamin.cowan@ucd.ie
University College Dublin
Dublin, Ireland

Donald McMillan
donald.mcmillan@dsv.su.se
Stockholm University
Stockholm, Sweden

ABSTRACT

Voice Agents (VAs) are touted as being able to help users in complex tasks such as cooking and interacting as a conversational partner to provide information and advice while the task is ongoing. Through conversation analysis of 7 cooking sessions with a commercial VA, we identify challenges caused by a lack of contextual awareness leading to irrelevant responses, misinterpretation of requests, and information overload. Informed by this, we evaluated 16 cooking sessions with a wizard-led context-aware VA. We observed more fluent interaction between humans and agents, including more complex requests, explicit grounding within utterances, and complex social responses. We discuss reasons for this, the potential for personalisation, and the division of labour in VA communication and proactivity. Then, we discuss the recent advances in generative models and the VAs interaction challenges. We propose limited context awareness in VAs as a step toward explainable, explorable conversational interfaces.

CCS CONCEPTS

• **Human-centered computing → Interaction design**; **Empirical studies in HCI**.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642183

KEYWORDS

voice interfaces, conversational user interfaces, cooking, conversation analysis

ACM Reference Format:
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 for Complex Tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3613904.3642183

1 INTRODUCTION

Conversational Agents (CAs), or Voice Agents (VAs) [1] are increasingly used as conversational collaborators, with users conversing with them to access news and information, play music, get help with cooking, and control smart-home products. These interfaces are also touted as being able to help users to 'get things done' [56, 58] through conversation.

Despite the prevalence of these devices and the considerable investment in research and development devoted to them, these conversations are still not particularly natural, especially for extended interactions or open-ended conversations. In part, this is because the agents are still, mostly, unable to have multi-turn conversations with users and fail to adapt to the social behaviours of their human interlocutors. Many of the common problems associated with conversational interaction, such as irrelevant responses, misinterpretation of requests, and information overload, can be traced back to a lack of awareness of the conversational context. This is especially problematic when it comes to positioning the VA

[1] We use Voice Agents (VAs) rather than Conversational Agents (CAs) in this paper primarily to avoid confusion with our application of Conversation Analysis (CA)

as a conversational collaborator able to support people in progressing through complex tasks. Although previous research has covered interaction challenges for VAs in complex tasks [35, 71] in general, there is little research that investigates how context-awareness of VAs influences the interaction patterns and challenges in a specific domestic task that requires temporal and spatial synchronisation between objects and the task at hand. Current VA interactions are typically context-blind, limiting the ways that users can interact with them. Further, little is known about how context-aware VAs should be designed to support users in completing their tasks, especially in scenarios where information is urgent or time-sensitive. To support more complex tasks, VAs must be able to understand the context of the task being undertaken, and the context of the ongoing conversation.

This paper contributes to existing research on VA interaction by addressing the mechanisms for developing context-aware voice interaction for complex tasks, including synchronisation between tasks and objects in the shared context. Our research explores interactions with a VA imbued with contextual awareness, providing an understanding of a specific task that is more complex than current prevalent use cases. This work provides much-needed insight into the interaction patterns and challenges faced by users when using VAs to support complex, multi-stage, context-dependent tasks. The task of cooking in a home kitchen was chosen as a common yet suitably complex example of such tasks [75].

We conducted an initial study to identify the current interaction challenges of commercial VAs, namely Google Assistant (GA), in a cooking context. This allowed us to identify key issues that impede the conversational support of a complex task that such systems are supposed to be able to provide. Many of the issues identified in the interaction were caused by contextual misalignment, and the limited capabilities of current VAs to keep track of the knowledge required at a particular moment in the flow of the task or conversation. Informed by the initial study's results, we conducted a Wizard-of-Oz study to investigate user interaction when cooking using a context-aware VA. 16 cooking sessions (a total of 8 hours and 10 minutes of data) were conducted with a VA controlled by a human wizard able to simulate a system imbued with contextual awareness of the cooking task, providing an understanding of the recipe and the objects involved in the task (such as ingredients and cooking implements mentioned in the recipe). This allowed us to examine the challenges and benefits of supporting conversational gambits to maintain and use shared context to ground the state of the task – enabling the system to go beyond the simplistic interactions currently offered.

Our analysis compares the resulting interactions, highlighting the increased complexity of commands and follow-up requests with a context-aware VA and how users could quickly extract more information with less conversational effort. They achieved this by formulating instructions to the system that used the ingredients, task steps, and cooking actions to pinpoint the information or action they required of the VA. They also included information to explicitly ground the state of the task with the VA, providing conversational resources to the VA to align its contextual understanding of the situation and check if the system's understanding matched their own. Our findings highlight the need to design the bounds of context awareness for VA, how it can be interacted with, queried and challenged, and its impact on user-machine dialogue. We discuss the benefits of limited and achievable context awareness in supporting a more fluent and fluid interaction with artificial conversational collaborators, how proactive VAs can balance agency with the user, and how this work connects with generative, multi-modal machine learning-based approaches to conversational interaction.

2 RELATED WORK

2.1 Command Construction with Current Speech-based Voice Agents

VAs are currently used to execute simple user-led tasks such as information search, playing music, and setting alarms and timers [2, 47]. VAs are also thought to be useful in supporting multitasking, offering a way to interact with interfaces in hands-busy, eyes-busy situations, by allowing users' hands to remain on the task [4, 23]. Interactions with VAs tend to take the form of simple adjacency pair dialogues [58]. When constructing these commands, users tend to use simple terms [48, 58], short syntactic structures, and keyword-like utterances [8, 37, 58]. When these interactions breakdown, users tend to adapt their language, using strategies such as hyper-articulating, changing the command structure, or altering their accent [48, 53, 58] to raise the chance that the VA will recognise the user's intended input [60]. This adaptation, as well as the selection of more simplistic utterances, is thought to be due to users seeing VAs as being at risk of communicative failure [14, 53]. Similar to the concepts of recipient or audience design [7, 50], this perception leads users to adapt their speech based on these perceived limitations so as to be more likely to communicate successfully with the system [14, 22].

2.2 Cooking With Voice Agents

Recent marketing of VAs has focused on their ability to support everyday domestic complex tasks such as cooking [56, 71]. Cooking is an example of a complex task that consists of a number of sequentially interdependent steps (e.g., combining, mixing, processing, and handling ingredients using a variety of tools). When cooking, the user must make continuous decisions on when to prioritise one task over another [42]. When supporting cooking through instructional videos, users tend to attend to audio as opposed to video output while they are engaged in cooking [16]. Recent work has shown that when using how-to videos, users tend to look to skip familiar content, or jump to later parts to see the result in order to prepare for future steps [16, 75].

Cooking requires working memory to put in the correct amount of ingredients in a certain order, tacit knowledge of how to manipulate the ingredients into certain states (e.g., kneading and cutting), domain knowledge to solve problems, and multi-tasking skills to coordinate the preparation of different parts to be completed at the same time. As such, when following a recipe using VAs, users often engage in a navigation behavior, often going back and forth to understand the content [32], check the status, control the pace, navigate to a given point of interest [73], and potentially search for additional information. Recent work has shown the impact of smart speaker failures when cooking on users' intention to frequently interact with the device [40]. Hwang et al. [35] presented 9 challenges the users faced when following a recipe with Alexa:

missing the big picture, information overload, fragmentation, time insensitivity, missing details, discarded context, failure to listen, uncommunicated affordances, and limitations of audio . Our paper is also built around the challenges experienced when cooking with a voice assistant, and focuses on the opportunities that context-aware interaction for complex task guidance can present.

2.3 Context Awareness in Voice Agents

In complex scenarios, speech systems need to maintain an understanding of context over multiple turns of interaction [12], and ask appropriate clarifying questions to guide the user The transcriptions used notation derived from[44]. While agents' ability to engage in dialogue has been studied quite extensively [10, 69], the conversational style of these agents has received less attention. However, it has been shown that people's perceptions of conversational agents are influenced by the interaction style of the agent [55]. Many of the interaction issues with current VAs, could be attributed to contextual misalignment, in that the VA lacked a common ground that could be queried, corrected, and drawn upon to further the interactions. This has led to irrelevant or inaccurate responses, with users having to use significant conversational effort to repair specific misunderstandings.

Imbuing VAs with the ability to anticipate a user's needs when delivering a suggestion or command has been noted as something users want [70]. Current VAs are context-blind, and only capable of simple reactive conversations. Developments in speech technology have gone some way to take context into account in the way speech systems operate, recognising the potential performance benefits of such an approach [45]. Recent work suggests developing more context-aware speech recognition systems, using dialog-level, situational, and temporal information to improve recognition results [38, 41, 52]. Yet little is known about how imbuing a VA with contextual awareness can influence user interaction, and how conversational design that leverages this is interacted with by users. Recent developments in generative AI models, notably Large Language Models (LLMs), provide context and memory capabilities, which could result in more natural and interactive conversations. Generative AI models employ various techniques that allow them to find patterns and relationships in the data, without being explicitly told what to look for. The model can generate new examples similar to the training data once it has learned the patterns [33]. LLMs and generative AI can be harnessed to create content and stay in context, which might change the way we design speech agents.

In the study of the organisation of summons-answer Schegloff [65] proposes that the occurrence of a first item in a sequence, such as a summons establishes the relevance of the next item. Thus, the absence of an answer to a summons might be noted by the repetition of the summons, until an answer is obtained, which then allows the summons to move on to further talk. Clark and Brennan [18] detailed the coordination process in communication in human-human conversation and the need to update common ground moment by moment. This grounding process is essential to ensuring mutual understanding [18]. Understanding human interaction with VAs is a crucial step in improving the design of such systems. A significant body of work has shown how people interact with robots, and VAs based on their expectations of human-human interactions [6, 9]. As in human communication, the interaction between users and VA needs to coordinate the content and process of their actions [18, 19]. This type of feature is especially useful in complex tasks such as cooking.

This paper examines the use of context-aware VA when cooking, where a VA could assess the task state and the user's needs to inform VA utterances, provide suggestions, and disambiguate user requests. Such capabilities can lead to VAs being proactive to plan and act in advance of a situation occurring [54]. This could support more mixed-initiative dialogues, with VAs providing useful information at the time it is required by the user to achieve user needs, even if not explicitly requested [5].

3 EXPERIMENTS

As a first step in developing a context-aware speech agent, we conducted a pre-study aimed at a thorough understanding of the user context of assisted cooking. Informed by the results of the pre-study, we designed a context-aware speech agent to investigate how the addition of an understanding of the context of the recipe and the cooking task on the part of the VA influences user interaction.

3.1 Cooking Study 1: Commercial Voice Agent

3.1.1 Study Design and Procedure. In this study, carried out in late 2021, we wanted to explore how well VAs currently assist people in cooking a recipe, and what issues they might experience. We asked participants to video record themselves while cooking with a VA (namely Google Assistant – referred to henceforth as GA) in their kitchens at times of their choosing. Each participant was provided with a smart speaker, which was configured with the participant's own Google account. We selected six recipes, supported by the GA service without the need to install additional skills, as options for the participants to choose from when cooking. The recipes varied in complexity, familiarity, and cooking methods to add diversity to the gathered data. The chosen recipes were (1) chocolate chip cookies (from Allrecipes), (2) chocolate marble cake (from BBC Good Food), (3) seafood pasta (from Food Network), (4) clam chowder (BBC Good Food), (5) Chorizo and pea risotto (BBC Good Food), and (6) spinach and ricotta gnocchi (BBC Good Food). When starting the cooking interaction, GA would specify how the interaction should be structured, giving users instructions on how to interact. Users were commonly informed "When you're ready for the next step, just say 'next step'." or "Would you like to start with ingredients or instructions?" Both these options led to sequential interactions exploring the ingredients or the instructions. GA would also inform the users of the number of steps and the fact that these will be delivered sequentially (e.g., "There are eight steps. I will read them one by one. When you are ready to hear more, you can say 'next step'".

When recruited, we briefed participants on the study procedures and asked them for their consent to participate. The researcher asked the participants to choose three or four of the dishes, without showing them the instructions. Once chosen, the participants were given a shopping list with suggested ingredients and tools to prepare before starting the session. Participants were requested to follow the recipe steps as much as possible. We encouraged them

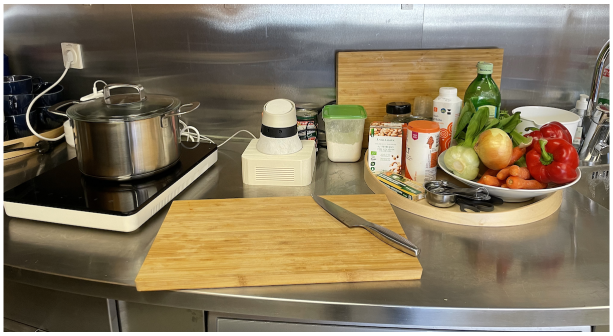


Figure 1: The cooking task area for the context-aware voice agent trial. Cameras were placed on either side of the area shown in the picture. Just out of frame to the right is a sink.

to ask GA questions and interact with it during the task, but also informed them that they were allowed to do other activities at the same time if necessary. They were told to set up their camera so that they could capture themselves, the stove, and the GA – all participants used a laptop for this purpose – and share the recordings with the researcher after they had finished cooking. When all was prepared, the researcher provided the exact phrases to use to start the selected recipes on their GA devices.

3.1.2 Participants. Participants were recruited through convenience sampling, using email and personal contacts from the University. We recruited three participants (aged 25-30 years, 2 females and 1 male) from different households. Prior to the study, participants reported previous experience using commercial VAs in their daily lives to complete basic tasks, such as playing music and searching for information. However, none of the participants had used voice assistants to cook before. All participants signed a consent form describing the details of the trial procedure, data collection, and analysis before the study. The participants were compensated with a £20 Amazon voucher each.

3.1.3 Data Collection & Analysis. In total, 7 hours and 14 minutes of video and audio data were collected across seven cooking sessions, where one session had two cooked recipes, and one had three cooked recipes, resulting in a total of 10 recorded recipes across the study. Across sessions, all recipes were cooked at least once. Two participants chose three recipes, and one chose to cook four recipes. Dialogues within the video recordings, as well as specific actions related to the speech interaction between GA and the participant, were transcribed and annotated manually [43]. The transcriptions used notation derived from [68], noting all participant commands, system responses, and movements related to the interaction, e.g., whether the user moved closer toward the assistant.

Exemplars of interaction issues within the data (termed fragments) were selected by the author who led the study. These fragments were then discussed with two authors who are experienced in conducting conversation analysis and speech interface and human-computer interaction research. Based on these discussions, a number of fragments were chosen from the transcripts and are used to illustrate the key interaction patterns and challenges, explained in the results section.

3.2 Cooking Study 2: Contextually-Aware Agent

3.2.1 Study Design and Procedure. The initial study revealed some issues attributed to *context misalignment* within the ongoing conversation with GA. These are discussed in more detail in the results below. Informed by these interactional challenges, we conducted a second study in mid 2022 aimed at exploring interaction with simulated context-aware VA to address the need for contextual understanding. For this study, we followed a Wizard-of-Oz method [62] to test the interaction with a context-aware VA. The wizard controlled a custom, physical smart speaker running the open-source Mycroft.ai conversational agent software [66], that produced utterances using Google's text-to-speech service. The system was extended with a browser-based Wizard-of-OZ control panel, allowing a researcher to define and control utterances to be spoken by the text-to-speech system. The wizard interface was used to control the device's actions. To ensure that the experiment was not confounded by speech recognition, we used a human wizard to control the output of the system [62]. The wizard sat at the other end of the same room, watching the task through a camera with the view of the participant and the system, and controlled the system's speech output.

We sketched an interaction paradigm that would go beyond the limitations of current VAs in the kitchen. The following three additions to the standard smart speaker VA paradigm were encoded in the interaction rules given to the wizard-operator of the VA: task vocabulary, recipe progression, and proactive suggestions. First, we developed *a shared vocabulary* within the bounds of the recipe interaction, drawing directly from the textual description of the recipe. The wizard interface was populated with both the steps of the recipe and the list of ingredients with their quantities to be triggered with a single click. To simulate a shared context between the user and the agent with respect to the recipe, the wizard replied to "How many…/How much…" questions by instructing the VA to read out the related ingredient's list item, and to other queries regarding the ingredients by triggering the text-to-speech response of the closest recipe instruction. Beyond the list of ingredients, the list of instructions was also considered a resource for shared linguistic context, allowing user utterances to be keyword-matched to instructions that would then be triggered to be read to the user. Instruction list items that included specific lengths of time generated a pre-configured timer next to that instructions. These were also seen as objects in the context of the ongoing task to be queried by the user.

The second enacted addition included the concept of *recipe progression.* The VA demonstrated the ability to track the user's progress through the recipe sequence. For the wizard, this was conceptualised as the VA being able to recognise the ingredients in the list when they were either on the chopping board or had been added to the pot. This enabled the illusion that the VA could respond to direct questions about the current state of the ongoing cooking action, which was central to the common ground of the conversation.

The third enacted change to the standard interaction paradigm was *the inclusion of proactive suggestions* and standardised reactive instructions based on the common ground of the recipe progression.

The suggestions presented to the users were based on *Organisational Assistance* by Kuoppamäki et al. [42]. This type of assistance is provided through proactively offering *advice* such as optional alternatives or additions to the current recipe step to adjust taste or texture to the users' preferences, as well as advice related to the impact that substitutions in the ingredient lists would have on the task. The VA also proactively suggested the use of *timer functionality* within the interaction, where deemed appropriate by the wizard. This tended to be realised as a follow-up question to an instruction that involved a cooking stage that lasted a specific duration (e.g., an instruction including "simmer for 3 minutes" would result in a follow-up question of "should I start a timer?").

This additional contextual understanding enabled the wizard, through the VA, to respond to questions about ordering actions and using ingredients and tools and proactively offer support by way of timers and alternative courses of action based on the available ingredients. The user could also use simple natural language commands to navigate through the recipe by asking the system to move backward, forward, or jump to a specific step. The wizard ignored any commands not directly about navigating the cooking task and the objects related to it.

The study was conducted in a kitchen space at a university, and all ingredients, utensils, and appliances were provided. We deliberately chose a challenging recipe to uncover information and investigate potential assistance needs: "vegetable and white-bean soup". For this study, we skipped the recipe selection stage, which allowed for more effective monitoring of the task and more easily identified times to deliver the prompts for the wizard. This recipe requires intermediate skills and high attention, e.g., several interdependent steps and lots of ingredients. The recipe consisted of 12 recipe steps (not including suggested additions) involving 14 ingredients, where preparation of the ingredients for subsequent steps was expected, at times, to be done in parallel to monitoring and stirring those cooking on the stove. We handed out the recipe about the process or ingredients prior to the actual observation. This pre-exposure to the recipe served to simulate the everyday experience of cooking a somewhat familiar recipe with somewhat familiar ingredients, rather than focusing on the acquisition of new cooking skills or learning new recipes [13, 29]. Prior to cooking, we briefly introduced the kitchen appliances to familiarize them with the lab kitchen. Individual participants followed a recipe with the support of the wizarded VA on the countertop, as shown in Figure1.

The participants were introduced to the interaction paradigm and shown a printout of the recipe they were about to cook. The printed instructions were then taken away, and the participants were told that they would be able to interact with the VA for guidance, clarifications, or reminders of the instructions they were expected to follow while cooking. The participants were informed that they were expected to follow the recipe, but could adjust the ingredients based on their dietary requirements. Two GoPro Hero 7 cameras captured the experiment to the left and right of the users to capture their interactions and provide an overview. The participants were also in the frame of a third camera, providing the video to the wizard.

3.2.2 Participants. A total of 16 participants (aged 25-35 years; 7 female, 9 male) were recruited from the students of the University. The majority (n=13) had experience of interacting with a voice agent. While three participants reported owning a voice assistant, only one used this device daily, and none of the participants had ever used a VA to support cooking. The participants self-ranked their cooking experience into five groups, namely *beginner* (n=5), *intermediate* (n=5), *good* (n=3), *very good* (n=2), and *excellent* (n=1). Three reported that they cook several times a week, with six participants cooking sometimes and one participant only rarely. All participants reported using technology while cooking, such as looking for recipes using their smartphone or tablet. All participants gave informed consent to participate in the study, which was approved under the ethics application of the larger project in which this work is situated.

3.2.3 Data Collection & Analysis. In total, the 16 cooking sessions took 8 hours and 10 minutes (max 41.09 minutes, min 18.16 minutes, mean 30.34 minutes, median 28.42) to complete. Each video was initially cut into individual interactions, resulting in 441 interaction clips between 4 and 125 seconds long. An interaction was defined as complete when the user disengaged with the assistant, or changed the topic of the interaction by asking an unrelated query. As such, 224 interactions consisted of single command and response adjacency pairs, with 95 including follow-up questions, repairs, and multi-stage queries. After the clips were created, three authors independently coded each interaction for the topic, parts of speech, and use of contextual information. Following the coding, the interaction styles were discussed through repeated, shared viewing of similarly tagged clips. This resulted in a selection of sample clips (fragments) that were transcribed. The initial fragments were selected by one of the authors. The selection was reviewed and discussed with two further authors with experience in multimodal human-computer interaction, conversational user interfaces, and conversation analysis.

Our work takes a Conversation Analysis (CA) [64] approach to analyse user interactions. CA has recently gained prominence as a tool to explore user interactions with VAs [1, 58]. This method moves away from assessing the user language based on frequency, focusing instead on an in-depth analysis of the situational components that may lead to specific utterances in interaction. The technique revolves around selecting and assessing key fragments of interest in the elicited dialogues, which illustrate trends or important linguistic effects present within the data. Phenomena commonly observed in CA are now used to scaffold key concepts in conversational user experience design [51] such as *recipient design* [3], *progressivity* [30], and *repair*. Research has used CA as a tool to help critique existing conversational design features, including wake-words [1]. To inform conceptualisation of VA, as opposed to human-human, based 'conversation' [57, 58, 60] as well as explore difficulties experienced by speakers when interacting with VAs [72].

4 RESULTS

This section presents the similarities and differences between interactions with the context-aware VA and the GA to highlight the interactional nature of conversational grounding. The detailed analysis of interactions from both studies shows interaction patterns and challenges. We demonstrate how users interacted with the contextual-aware VA, when the system worked, and when it did not.

The results of the initial study show that the interaction with GA whilst cooking was sequential, with users frequently using simple commands such as "next step" or "next ingredient" to navigate through the recipe, using "what's the step again?" or "repeat previous step" to repeat instructions. Similarly, users of the context-aware VA navigated through the recipe using commands the same as those seen when interacting with GA. These included specifying the numerical position of the step in the sequence (e.g., "What's the second step?"), or using an expression of relative reference (e.g., "next step"). When using contextual VA, participants also made use of repetition commands to parse and manage complex instructions. However, the system's apparent ability to understand contextually dependent queries allowed them to approach the challenges of understanding and confirming their adherence to complex steps in different ways.

4.1 Objects in Context

```
← PA:  Hey Google, how much parmesan do I need?

  GA:  On the website Insider.com,
       [they say]  authentic   [Parmigiano-Reggiano]->
  PA:  [Hey Google]
  PA:  ((moving closer to GA)) [Hey Google]
  GA:        or                [parmesan]->
← PA:  ((moving closer to GA)) [Hey Google]
  GA:  [hails]       from --
  PA:  [Hey Google], how much parmesan do I need?

  GA:  Sorry, I don't have any information about that.
```

Fragment 1: Participant A receives an irrelevant response from Google Assistant to a request for the amount of Parmesan cheese required by the current recipe. *Commercial VA*

In the initial study, we found that many participants repeatedly ran into problems when interacting with GA. Although there were clear issues with speech recognition during the dialogues, many issues can also be attributed to *misaligned context* within the ongoing conversation. One result of this misalignment was the delivery of *irrelevant information*, where the lack of contextual awareness on the part of the VA resulted in responses that were of no use to the participant at that time. This was most common when requesting information about specific ingredients or procedures. Fragment 1 shows clear examples of user *barge in* whereby the user tries to stop GA from completing an irrelevant utterance during cooking. The issue here stems from the contextual misalignment between the expected bounds of the conversation.

In response to such errors, the contextually aware interactions were primed to respond only with *relevant text from the recipe being cooked* rather than the result of an internet search unless explicitly stated by the user. Such a narrow focus of the conversation may be unnatural in a human interlocutor, but for a conversational system, this can result in more fluid – if less conversational – interactions. This reduced the instances where the user had to expend *significant conversational effort* in order to complete their query.

```
P5: Hey <System>, how much carrot should I put
    (.) in it?
              (2.9)

VA: Two small chopped carrots.
```

Fragment 2: Participant 5 triggering the relevant ingredient line to be vocalised by the agent, checking the amount of carrots in the recipe. *Context-Aware VA*

Fragment 3 demonstrates a successful ingredient-based request, whilst also emphasising the increased communicative effort expended to get the required information when issues arose. The user asks GA about the amount of an ingredient needed for the recipe. GA first fails to respond to the query correctly, misrecognising the term *butter* for *battery*. The user attempts a number of repair strategies, including verbatim repetition (Line 7), adding stress to the ingredient (Line 11), speaking the command more slowly (Line 9), and moving closer to the device when interacting with GA (Lines 7,9). Yet, as shown in Fragment 2, focusing on the contextual scope of the conversation provides the same information in a timely and less effortful manner. This clearly shows that the addition of an understanding of the current task context resulted in a reduction in the amount of conversational effort the user needed to do to keep the system replying with relevant information to the task at hand.

The results of this misalignment caused the user to get *irrelevant information*, resulting in increased *communicative effort* expended to get the required information when issues arose. It was common to see users *abandon their primary task* to direct their attention to the GA. The misalignment was made obvious in the users' construction of queries around objects that could be reasonably expected to be part of the ongoing context of the task – the ingredients of the recipe the voice agent was instructing them on. This shows the value of incorporating relevant, shared objects around which to situate conversational interactions when task progression is important.

```
  PA: Hey Google, how much butter do I need?
  GA: Your speaker or smart display doesn't have
      a battery. So if I'm speaking to you,
      you're plugged in and powered up. To see
      the battery level on your computer or
      [mobile device] find the ->
  PA: [Hey Google]
  GA: battery icon on the screen.
←PA: ((moving closer to the speaker)) Hey
      Google, how much butter do I need?
  GA: Your speaker or smart--
←PA: ((moving closer to the speaker)) Hey
      Google, <slowly>how much butter>?
  GA: Sorry, I couldn't find how much of that
      ingredient you need.
  PA: Hey Google, how much BUTTER?
  GA: Sorry, I couldn't find how much of that--
  PA: Hey Google, how much SALTED butter?
  GA: 2 tablespoons salted butter.
```

Fragment 3: Participant B is working through repair strategies after a speech-recognition error to retrieve the amount of butter in the current recipe. *Commercial VA*

```
PA: Hey Google, what’s next?

                   (2.1)

GA: Next up, you have a project supervision
    meeting on Thursday at 11:30 am.
```

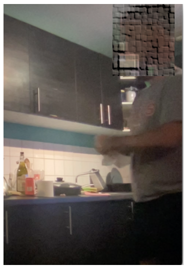
Fragment 4: Participant B receiving an ‘out of context’ response to a request for the next recipe step. *Commercial VA*

4.2 The Task in Context

Moving through the steps of the cooking task was another way in which the lack of a shared context for state and vocabulary increased the conversational effort required of the participants.

In some cases, as can be seen in Fragment 4, the commercial voice agent provided answers far outside the expected context, resulting in delays and frustration for the user. More commonly, however, the navigation interactions with the commercial voice agent were made more complex and less efficient by the lack of relative markers within the task available to the user to ground their query.

As shown in Fragment 5, using GA to navigate through the list of instructions with ‘next’ and ‘previous’ commands was problematic in some cases. While users were also able to request to jump directly to a numbered command (i.e. “Go to step 6”), the steps were not presented with their accompanying numbers by GA, meaning this portion of the shared information had to be extrapolated, and remembered by the user. However, in the navigation examples with the contextually aware VA, users often followed the same pattern of ‘next’ and ‘previous’ step-wise navigation. However, to reduce the communicative effort in moving through the recipe, they constructed more complex navigation commands relying on a shared context around the recipe instructions, tools, and ingredients. In order to ensure that these were successful, participants would often explicitly ground the state of the task and their understanding of the shared context and then employ that to produce shared points of reference to navigation through the recipe. For example, Fragment 6 shows the “next step” command being made implicit and contextualised within the ongoing cooking process, whereby the user also lists the ingredients they have already added to the pot. Compared to the simple use of the next step command seen in the GA interactions, the user, in this case, is listing the ingredients so as to explicitly ground the state of the task. This occurred even though the context-aware VA demonstrated knowledge of the state of the task during the interaction. This effect is seen across a number of fragments within interactions and is discussed in more detail in the following section. As also highlighted in Fragment 6, users would commonly conduct conversational work to explicitly ensure that the context-aware VA shared the same perspective as to the state of the task as they did.



```
PA: Hey Google, next step.
GA: Continue doing this for about 20 minutes until
    most of the stock has been absorbed and the rice
    has swollen but still has a slight bite.
PA: Hey Google, what’s next step?
GA: Tip in the peas, parmesan, and the remaining
    stock. Give everything a good stir.
    Once the cheese has melted, season with black
    pepper, then divide between bowls, topping each
    portion with the remaining crisp chorizo and
    extra grated parmesan.
    That was the last step. Enjoy.
PA: Hey Google, what’s last step?
GA: Sure, tip in the peas, parmesan, and the
    remaining stock--
PA: Hey Google, what’s the step before?
GA: Sure, tip in the peas, parmesan, and the
    remaining stock—
PA: Stop!
```

Fragment 5: Participant C attempting to navigate back in the list of recipe instructions. *Commercial VA*

```
←P6: <System>, what comes after adding the
     celery and carrots (.) and garlic?
                 (2.1)
 VA: Chop and add one garlic clove, then
     peel and chop two carrots in small
     pieces, one celery stick and half a
     zucchini and add them to the pot.
                 (1.5)
 P6: and after that?
                 (1.8)
 VA: Season with some more salt and pepper
     and let simmer for 6 minutes, stirring
     frequently.
                 (1.1)
←P6: [and after]
 VA: [Should I] start a timer?
 P6: Yes.
```

Fragment 6: Participant 6 navigating through the recipe by specifically referring to ingredients to show their current progression. *Context-Aware VA*

Developing this shared orientation toward the ongoing task allowed the users to use that shared context as a reference point when navigating the instructions without the explicit commands seen above. For instance, in Fragment 7, we see the user trigger the system to utter the current step, involving the tomatoes, through reference to their shared understanding of the context of the ongoing task via asking for confirmation on the ‘next step.’ This request for the ‘next’ step was not in relation to the ordering of the steps as they had been read out by the VA, as would be the case in the initial study, but rather in relation to the ingredients (i.e., the tomatoes).

```
 VA: It has been cooking for five minutes.
 P8: Okay (1) <System>, the next step is to add
     the tomato, right?
                 (2.8)
 VA: Add one chopped bell pepper and the
     tomatoes to the pot and let simmer for
     three to four minutes.
←P8: Sorry, can you repeat that?
                 (1.8)
 VA: Yes, add one chopped bell pepper and the
     tomatoes to the pot and let simmer for
     three to four minutes.
                 (3.1)
 P8: °one bell pepper°
                 (3.9)
←P8: Hey <System>, can you (1.1) add a timer
     for (1.8) three minutes?
                 (1.8)
 VA: Sure
 P8:  [and a half]
```

Fragment 7: Participant 8 confirms the shared context of the recipe progression to trigger the next step of the recipe to be vocalised. *Context-Aware VA*

```
P1: So <System>, after putting the carrots
    (.) the celery (.) the zucchini (1.1)
    what’s the next step?
                 (2.8)
VA: Season with some more salt and pepper
    and let simmer for six minutes,
    stirring frequently.
```

Fragment 8: Participant 1 using a contextual query to confirm the next recipe step. *Context-Aware VA*

This type of complex conversational work was not seen in the GA interactions. Such an utterance serves the purpose of explicitly ensuring that both the user and the VA are aware of the state of the task before requesting information relative to that state. This may reflect the nature of voice interface interaction in that, even with a context-aware VA, users still feel that they need to dedicate significant conversational effort to ensuring a shared perspective to ensure they are understood.

We have found that when there was a long wait until executing the next step, some users tried to explore whether the contextual agent could conduct other tasks. These tasks were those that were akin to those expected of a VA, such as asking for the weather or playing music. As the contextual agent did not have this functionality, many users then just waited for the timer or requested updates as to the timer's progress. While, interactionally, there would have been an option to pass-through the query to the commercial voice agent to provide out-of-context responses, this was not added to the WoZ protocol. This raises a question about the balance between keeping a tight interactional focus for the task and keeping the system as a multi-function device. For example, if the participant started playing music (one of the most common VA tasks [2]), the ambiguity of relative commands such as 'next' shown above in Fragment 4 re-emerges.

4.3 Querying the Shared Context

We found that the participants felt able to rely on the shared contextual understanding built up over the course of the cooking session to ask a direct, closed question to retrieve only the information they needed from the recipe. Fragment 12-left shows the participant asking the VA by referring to one of the ingredients in the currently active step, instead of having the agent repeat the recipe step, which included the length of time that the onions should be cooked.

Given the more complex contextual awareness shown by the wizarded agent, we have seen examples where participants relied on shared contextual understanding to resolve uncertainty, such as in Fragment 8. In this example, the participant explicitly requested the VA to provide him with any steps he might forget. P1 produced more complex confirmation checks and showed even greater reliance on the shared context build-up over the cooking session.

In some examples, we have seen that users were confident that the contextual VA preserved enough of the context of the interaction to understand both that the interaction was continuing, and that the shared context of cooking actions and objects would allow relational referencing as it went forward. Participants also used follow-up requests, frequently these were not initiated using the wake word, as can be seen in Fragment 9. While something similar is available in GA through the 'Continued Conversation' functionality, follow-up questions were comparatively rare.

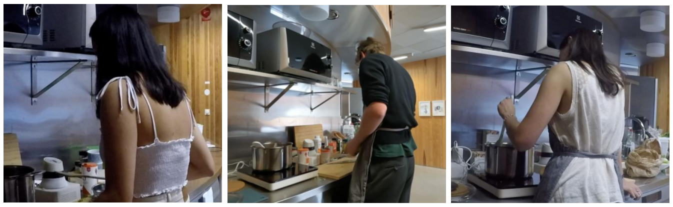
The examples above show a clear willingness of participants to directly query specific aspects of the instructions, relying on the assumption that the knowledge shared between the VA and the user can be used to gather the required information. The use of shared knowledge in this way shows users embracing a more natural form of speech interaction, in comparison to more rigid requesting the full repetition of a previous step when only part of the information from the step is required.

4.4 Proactive Contextual Agents

The contextual VA was able to act upon the shared context and proactively provide selected pieces of advice to the users. This was so as to explore the use of this specific form of interaction in supporting complex tasks like cooking. It led to a mixed-initiative interaction, whereby users and the contextual agent would take the lead in commencing conversation. This proactivity from the contextual VA perspective was conducted with mixed success. To deliver suggestions, the contextual agent would use a sound as an *access ritual* [24, 27], waiting for the users' attention before delivering the utterance. Using a sound rather than an explicit greeting such as "hey" or "excuse me" meant that sometimes the

```
←P2: So <System>, what’s the first step?
                (1.4)
 VA: Heat the butter in a large pot over medium
     heat
                (1.1)
 P2: I don’t have butter, what can I use
     instead?
                (1.2)
 VA: If you don’t have butter you can use olive
     oil.
 P2: How much olive oil?
                (2.8)
 P2: [<System>, how mu]
 VA: [two tablespoons]
```

Fragment 9: Participant 2 using a follow-up question to query the amount of oil to add in lieu of butter. *Context-Aware VA*



```
VA: Hm::mm↓

  ((P5 proceeds to stir
    the pot, ignoring the
    VA))
```

```
VA : Hm::mm↓
P10: Yes
         (2.1)
VA : If you want a
     sweeter soup, you
     can add another
     carrot.
```

```
VA: Hm::mm↓
P9: ye::s
       (2.1)
VA: If you want a spicy
    soup you can add
    some chili flakes.
       (2.5)
P9: °uh::° No, thank you
```

Fragment 10: *Left:* Participant 5 ignoring the voice agent. *Centre:* Participant 10 replying with emphasis to on a suggestion. *Right:* Participant 9 declining the suggestion. *Context-Aware VA*

```
VA: Hm::mm↓
P8: Yes:?
                (2.1)
VA: If you want a sweeter soup, you could add
    another carrot.
                (2.1)
P8: Ok (.) thank_you↓
                (1.3)
VA: You're welcome.
                (3.6)
←P8: <System> do you have any other suggestions?
                (2.7)
VA: No, not for [now
P8:              [° great °]
```

Fragment 11: Participant 8 responding to a suggestion and following up by querying if there were more suggestions. *Context-Aware VA*

users missed these. Fragment 10 shows when this was the case. Within this fragment, the user clearly does not realise that the contextual agent is attempting to gain their attention. This results in the proactive turns not being delivered to the user.

When the access ritual was recognised, users had a mixed response to the contextual VA intervention. For instance, some seemed frustrated, responding with a highly emphasised "what is it?" or "yes?" when hearing the noise (see Fragment 10, right). Within this fragment, as in others, users tended not to act on these suggestions, unless they were seen as critical to the success of the task or included minimal extra effort (e.g. turning down the heat).

Sometimes we saw users trying to explicitly stop, or preempt these suggestions. For instance, in Fragment 11, we see a user asking the contextual VA "Do you have any other suggestions?" In comparison, as in Fragment 12, users tended to engage more positively when the contextual agent volunteered context-relevant functionality through a follow-up question. These suggestions were context-relevant to the preceding instruction, demonstrating the type of functionality that the contextual agent had to support it. This kind of behaviour from the VA is beneficial in that it supports the discoverability [39] of the functions available to the user during interaction.

```
P3: <System>, how long should the
    onions be cooked?
                (2.3)
VA: Let simmer for three minutes.
    Should I start a timer?
P3: Yes.
```

```
P7: <System>, what's the next step?
VA: Season with some more salt and
    pepper and let simmer for 6
    minutes, stirring frequently.
    Should I start a timer?
                (2.1)
P7: No (.) not yet.
```

Fragment 12: *Left:* Participant 3 setting a timer by agreeing to a follow-up question from the agent. *Right:* Participant 7 declining a timer. *Context-Aware VA*

```
   P8: <System>, I've added the onion, eh, the
←      garlic, the carrots, the zucchini (.) and
       the celery, could you please time?
                (1.8)
   VA: Sure.
                (2.1)
   P8: Should I sea, <System>, should I season with
       salt and pepper as well?
                (1.1)
   VA: Yes.
                (3.1)
   VA: How long do you want the timer?
                (1.5)
←  P8: <System>, ask again.
                (1.8)
   VA: How long do you want the timer?
                (1.3)
   P8: So, after <System> So, after I've added the
       onion, eh, the garlic, zucchini and the
       carrots how long time should it be on the
       heat?
                (2.0)
   VA: Let simmer for six minutes, stirring
       frequently.
←  P8: Yup, Ok, so [<System> could]
   VA:             [Should I start a timer?]
   P8: Yes, please start a timer for five minutes
```

Fragment 13: Participant 8 explicitly grounding the state of progression and requesting a timer. *Context-Aware VA*

However, there were times when users rejected this specific functional interruption. For instance, Fragment 13 shows the user making the contextual agent aware that they did not want to use the functionality at that moment in time. In the following utterances, the user then requests the timer at the point at which they feel they are ready to use it in the task. When the contextual VA used these types of requests, they were sometimes missed by the user. Users then looked to get the VA to repeat their utterances. Yet rather than using "can you repeat that" or "Repeat previous step" common in the GA dialogues, users seemed to be more specific about the type of request being made. For instance, when one of the users missed the system's request for how long the system should set a timer, the user responded with "*<system>* ask again?".

They then respond with a highly complex request, for information as to how long the cooking task needs to progress after including specific ingredients. Following this, they shorten the timer length they request based on an estimate of how long these ingredients have been cooking and then ask the contextual VA to add the timer. This not only shows the way that some users may engage with such requests if they are missed, but also that the responses may be highly complex in trying to find information so as to accurately use the functionality after being prompted.

5 DISCUSSION

Our findings highlight the current VA limitations in supporting complex tasks like cooking and demonstrate the conversational improvements from a more context-aware VA. We discuss the potential complexity from added context awareness and its consideration for wider adaptation to personal preferences, how even with increased complexity, VAs are still potentially perceived as constrained dialogue partners, the challenge of balancing proactivity and agency, and how this work connects to LLM-powered dialogue systems.

5.1 Contextual Awareness and Naturalistic, Complex Dialogues

Our work indicates that users often employ complex utterances to ground shared knowledge with a context-aware VA. They engaged in follow-up requests leveraging this shared context, clarified aspects of steps, requested information for subsequent commands, and interacted socially with the VA. This increase in linguistic complexity likely stems from a context-aware VA's perceived advanced conversational abilities, prompting users to use more intricate language. Such interactions facilitated conversational 'grounding', allowing users to confirm shared knowledge crucial for successful communication, particularly when speech recognition errors are minimal, as in our wizard-based evaluation.

Despite advancements in conversational agents, communicating with technology still differs fundamentally from human-to-human dialogue [48, 59]. Users often resort to keyword-based queries, as longer utterances may lead to more recognition errors [36, 46]. Even with our context-aware VA, users bear the burden of ensuring successful communication, with limited assistance from the VA [26, 59].

Our study underscores the potential to improve VA conversation by integrating insights from human conversational mechanics, like turn-taking, sequence organization, grounding, and repair strategies. Grounding in human interaction as a collection of "mutual knowledge, mutual beliefs, and mutual assumptions", which is normally lacking when interacting with VAs, has been shown here to be one such opportunity to improve interaction.

While it is relatively trivial for a pair of human interlocutors to switch conversations between a shared task and any number of topics, this is something that poses a significant challenge for a conversational system. Humans have multiple modalities, backchannels, and prosodic tools to ground an utterance in relation to a particular context, that are both unavailable and, mostly, unparsable by a system.

One approach would be to have explicit context changes, allowing commands such as 'next' to work with an ongoing recipe and an ongoing music playlist. While this increases the conversational burden on the user, it does at least only increase complexity where the user *initiates* more complex interactions. In single-context interaction sessions, the user would not need to attend to manual context switching.

Strategies to help users disambiguate their current conversational context would be necessary. Taking advantage of different prosodic features of text-to-speech modules could imbue each context with a distinctive vocal character, aiding users in distinguishing between ongoing contexts and related tasks. Previous work has noted that such contextual personalisation is desirable [70]. Yet similar to current VAs [23], there may also be privacy and data gathering concerns that need to be addressed when imbuing such systems with wider contextual awareness and personalisations.

Shneiderman [67] argues that spoken language often has severe limitations in human-computer interaction. "By appreciating the differences between human-human interaction and human-computer interaction, designers may then be able to choose appropriate applications for human use of speech with computers [67]." The complex nature of human communication requires careful and thoughtful abstractions for VA design, and does not translate directly to human-computer interaction. That said, our aim is not to simulate human communicative grounding but to investigate how it's employed by users to query and collaborate with a context-aware VA in a constrained context.

5.2 Voice Interfaces, At Risk Partners & Division of Conversational Labour

Following on from Shneiderman, recent work on the nature of conversational interaction with speech agents suggests that, even though emulating human capabilities, such as context awareness, a dialogue with a system is a different genre of dialogue with its own norms and rules [20, 58, 60], with computers being seen as stereotypically more inflexible in their capabilities compared to human interlocutors [14, 21]. Previous work on user partner models of speech interfaces [25], suggests that the perceived flexibility of a system, along with its human likeness and its perceived knowledge scaffold user perceptions of a VA's competence as a conversational partner. Although our findings suggest that users may assume that the context-aware VA is more capable as a communicative partner, some of the effects seen still support the notion that users still may perceive VAs at some level as at-risk dialogue partners [53], adapting aspects of their speech to accommodate their perceived limitations [22, 50, 63]. More work needs to be done to assess how context awareness impacts partner models, whether this leads people to assume that the VA is more able to deal with more complex commands, and how encountering errors may impact this perception.

These perceptions may also influence the perceived conversational effort a user feels they need to invest to ensure communicative success. Recent work in human-human dialogue has concentrated on how expectations as to a partner's exertion of effort affect communication [34], whereby with an uncollaborative partner, the interlocutor has to exert more effort so as to ensure communicative success. This effort tends to be negotiated between the partners [49], yet even with contextual awareness, machine partners may be seen by users as incapable of clearly and easily negotiating the division of labour in conversation, being seen as more fixed in their ability to share conversational effort. Even though the addition of advanced capabilities like context awareness may improve perceived collaborative capabilities of the VA, aspects of our data suggests that some users still look to dedicate conversational effort to explicitly ground specific information as they perceive that they need to take on more of this collaborative effort to be fully certain of accurate updating of mutual knowledge. Future work should look to assess the complex interplay between the design of voice agents, their capabilities and how this influences the division of conversational labour in VA interaction. This type of work would give us a firm grounding as to how this may be contributing to effects seen in more complex voice interface interactions such as ours.

5.3 Balancing Proactivity

Our work investigates how introducing proactivity with a VA and creating a mixed-initiative paradigm impacts user interactions in complex tasks. The results were mixed. When users noticed the system suggestions, many engaged with them *when the suggestions were context-relevant*. Previous work has shown that participants think that they want this feature [70], specifically the ability to anticipate user needs when suggesting or commanding. However, this is likely to depend on user busyness [15] and the urgency or benefit of the interruption to the current task [15, 28, 61, 74]. Studies have also found that users can find such functionality invasive [74], as agents are perceived primarily as tools rather than conversational equals [20]. Generally, the use of proactivity is seen as appropriate when the intervention is seen as critical, relevant, and sensitive to the social context and environment. Users should be given agency in configuring the level of proactivity, while taking into account individual factors such as user mood.

The balance of user agency and the proactivity of the agent presents a challenge for conversational interaction design. This could be done as an iterative process to leverage how pairs of human interlocuters build an understanding of their ongoing interactions [11]. Yet to do this simply from a technical point of view – akin to a throttling algorithm in computing, where negative responses from the user result in fewer agent intrusions – would ignore a lot of the important organisational and contextual queues used by human interlocuters to make their interruptions successful for both parties. The levels of contextual understanding needed to make an interruption successful, in terms of both being acted upon and not unduly irritating the receiver, even in constrained contexts of use, is currently challenging for conversational agent design.

The design of the interruption has also been the focus of recent work. Non-verbal cues, specific verbal cues, and direct interventions have been proposed as ways to design VA-based proactive interruptions [74]. However, our work suggests that non-verbal cues may not always be obvious to users, especially in noisy and busy environments, or when users are engaged in complex tasks. The design of prompts may also depend on the type of proactive functionality being offered. The variability of prompt types could serve as cues for users to identify and decide whether to engage with or ignore an interruption. Further research is needed into new forms of proactive prompts, considering contextual and task-based factors, building on existing work on timing and linguistic content of proactive interruptions [27]. The goal is to establish a user-controllable and context-dependent balance between proactivity and passivity.

5.4 Multi-modal Inference and Contextual Conversational Interaction

The challenges of understanding context from a series of images of the users' actions, and producing dialogue that accurately reflects the changes in the state of that context – including the users' stated and implied goals and plans – do seem to be becoming tractable [17]. The imminent wider release of multi-modal large language models from the likes of Google (Gemini) and OpenAi (GPT-4/Vision) promises the ability to not only generate textual or spoken output in response to transcribed queries, but also to be able to provide reasoned output on the state of ongoing tasks [31]. However, this technical advance does not eclipse the interactional challenges that come with conversational agents. In use cases such as the ones presented in this paper, such systems will still be limited in the modalities available to provide output to the user, and the sensing capabilities to understand the ongoing task outwith user utterances. Problems of discoverability, disambiguation of objects or concepts, and determining a user's intended audience for an utterance will still cause breakdowns in the interactions. It is notable that, as impressive as the results shown by GPT-4 and Gemini are [31], the distribution of the burden of contextualization and grounding work is clearly falling more heavily on the user than in the wizarded examples we have presented above (even if this is somewhat obfuscated in the edited, promotional videos). Current approaches to integrating LLMs in conversational interaction include combinations of rule-based and heuristic methods for determining the correct next utterance [33], where the results of this paper can be directly applied to guide the VA in both informing the user of the current state and in how they can approach querying and correcting it through patterns in their ongoing interaction.

6 CONCLUSION

In conclusion, we have shown the impact of constrained contextual awareness in VAs to support a complex task like cooking. First, we identified interaction challenges of commercial VAs to support complex tasks caused mainly by contextual misalignment. Informed by this, we conducted 16 cooking sessions with a wizarded, context-aware VA. Based on the video analysis of the interactions, we show that, even with the limited interactional capabilities of a scripted text-to-speech based system, context awareness leads to more fluent, less constrained, but more linguistically complex interactions. By providing a constrained context of interaction during a task, a VA is able to reduce the conversational effort expected by users while increasing the relevance and accuracy of responses.

From a design point of view, our findings highlight the need to work more on the design of context-aware and proactive interactions so as to support the cooking experience, whilst ensuring that any future inclusion of context awareness is sensitive to user motivations, the environment, and the level of user preferences. From a theoretical perspective, these results suggest that even with such advanced functionality in place, the limitations of such interaction may lead people to see VAs as at-risk partners. In going forward, we have shown that limited – and therefore technically feasible – accommodation of context-awareness in interaction can provide fluency of interaction while providing an understandable interaction framework to bound and negotiate system capabilities in a way that current VAs do not.

ACKNOWLEDGMENTS

This work was in part supported by the Digital Futures Scholar-in-Residence programme and Digital Futures AAIS project. This work has also been supported, in part, by Science Foundation Ireland (SFI) ADAPT Centre. The ADAPT Centre is funded by SFI and is co-funded under the European Regional Development Fund through Grant 13/RC/2106_P2.

REFERENCES

[1] Saul Albert and Magnus Hamann. 2021. Putting wake words to bed: We speak wake words with systematically varied prosody, but CUIs don't listen. In *CUI 2021-3rd Conference on Conversational User Interfaces*. 1–5. https://doi.org/10.1145/3469595.3469608

[2] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on*

Computer-Human Interaction 26, 3 (April 2019), 17:1–17:28. https://doi.org/10.1145/3311956

[3] Sungeun An, Robert Moore, Eric Young Liu, and Guang-Jie Ren. 2021. Recipient Design for Conversational Agents: Tailoring Agent's Utterance to User's Knowledge. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–5. https://doi.org/10.1145/3469595.3469625

[4] Matthew P. Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. 2014. None of a CHInd: relationship counselling for HCI and speech technology. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 749–760. https://doi.org/10.1145/2559206.2578868

[5] Vevake Balaraman and Bernardo Magnini. 2020. Proactive systems and influenceable users: Simulating proactivity in task-oriented dialogues. In *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue-Full Papers, Virually at Brandeis, Waltham, New Jersey, July. SEMDIAL*. http://semdial.org/anthology/Z20-Balaraman_semdial_0007.pdf

[6] Nikolaus Bee, Elisabeth André, and Susanne Tober. 2009. Breaking the Ice in Human-Agent Communication: Eye-Gaze Based Initiation of Contact with an Embodied Conversational Agent. In *Intelligent Virtual Agents*, Zsófia Ruttkay, Michael Kipp, Anton Nijholt, and Hannes Högni Vilhjálmsson (Eds.). Springer, Berlin, Heidelberg, 229–242. https://doi.org/10.1007/978-3-642-04380-2_26

[7] Allan Bell. 1984. Language style as audience design*. *Language in Society* 13, 2 (June 1984), 145–204. https://doi.org/10.1017/S004740450001037X

[8] Linda Bell and Joakim Gustafson. 1999. Interaction with an animated agent in a spoken dialogue system. In *Sixth European Conference on Speech Communication and Technology*. https://doi.org/10.21437/eurospeech.1999-266

[9] Timothy Bickmore. 2002. Towards the design of multimodal interfaces for handheld conversational characters. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. Association for Computing Machinery, New York, NY, USA, 788–789. https://doi.org/10.1145/506443.506598

[10] Timothy Bickmore and Justine Cassell. 2000. How about this weather?" social dialogue with embodied conversational agents. In *Proc. AAAI Fall Symposium on Socially Intelligent Agents*. https://api.semanticscholar.org/CorpusID:13190315

[11] Jack Bilmes. 1997. Being interrupted. *Language in Society* 26, 4 (Dec. 1997), 507–531. https://doi.org/10.1017/S0047404500021035

[12] Dan Bohus and Alexander I Rudnicky. 2003. RavenClaw: Dialog Management Using Hierarchical Task Decomposition and an Expectation Agenda. (2003), 4. https://doi.org/10.21437/eurospeech.2003-255

[13] Sarah Bowen, Sinikka Elliott, and Joslyn Brenton. 2014. The joy of cooking? *Contexts* 13, 3 (2014), 20–25. https://doi.org/10.1038/scientificamerican0692-116

[14] Holly P. Branigan, Martin J. Pickering, Jamie Pearson, Janet F. McLean, and Ash Brown. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121, 1 (Oct. 2011), 41–57. https://doi.org/10.1016/j.cognition.2011.05.011

[15] Narae Cha, Auk Kim, Cheul Young Park, Soowon Kang, Mingyu Park, Jae-Gil Lee, Sangsu Lee, and Uichin Lee. 2020. Hello There! Is Now a Good Time to Talk? Opportune Moments for Proactive Interactions with Smart Speakers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (Sept. 2020), 74:1–74:28. https://doi.org/10.1145/3411810

[16] 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*. 1–14. https://doi.org/10.1145/3411764.3445131

[17] Alexander Chen. 2023. How it's Made: Interacting with Gemini through multimodal prompting. https://developers.googleblog.com/2023/12/how-its-made-gemini-multimodal-prompting.html

[18] H. H. Clark. 2006. Context and Common Ground. In *Encyclopedia of Language & Linguistics (Second Edition)*, Keith Brown (Ed.). Elsevier, Oxford, 105–108. https://doi.org/10.1016/B0-08-044854-2/01088-9

[19] Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13, 2 (April 1989), 259–294. https://doi.org/10.1016/0364-0213(89)90008-6

[20] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300705

[21] Benjamin R. Cowan and Holly P. Branigan. 2015. Does voice anthropomorphism affect lexical alignment in speech-based human-computer dialogue? 155–159. https://doi.org/10.21437/Interspeech.2015-75

[22] Benjamin R. Cowan, Philip Doyle, Justin Edwards, Diego Garaialde, Ali Hayes-Brady, Holly P. Branigan, João Cabral, and Leigh Clark. 2019. What's in an accent?: the impact of accented synthetic speech on lexical choice in human-machine dialogue. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. ACM, Dublin Ireland, 1–8. https://doi.org/10.1145/3342775.3342786

[23] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What can i help you with?": infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17*. ACM Press, Vienna, Austria, 1–12. https://doi.org/10.1145/3098279.3098539

[24] Michael Dellwing. 2022. Relations in Public. Microstudies of the Public Order. In *Goffman-Handbuch: Leben – Werk – Wirkung*, Karl Lenz and Robert Hettlage (Eds.). J.B. Metzler, Stuttgart, 323–329. https://doi.org/10.1007/978-3-476-05871-3_44

[25] Philip R Doyle, Leigh Clark, and Benjamin R. Cowan. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445206

[26] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is It OK if I Eat You?": Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children (IDC '17)*. ACM, New York, NY, USA, 595–600. https://doi.org/10.1145/3078072.3084330

[27] Justin Edwards, Christian Janssen, Sandy Gould, and Benjamin R. Cowan. 2021. Eliciting Spoken Interruptions to Inform Proactive Speech Agent Design. In *Proceedings of the 3rd Conference on Conversational User Interfaces (CUI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3469595.3469618

[28] Justin Edwards, He Liu, Tianyu Zhou, Sandy J. J. Gould, Leigh Clark, Philip Doyle, and Benjamin R. Cowan. 2019. Multitasking with Alexa: how using intelligent personal assistants impacts language-based primary task performance. In *Proceedings of the 1st International Conference on Conversational User Interfaces - CUI '19*. ACM Press, Dublin, Ireland, 1–7. https://doi.org/10.1145/3342775.3342785

[29] Rachel Engler-Stringer. 2010. The Domestic Foodscapes of Young Low-Income Women in Montreal: Cooking Practices in the Context of an Increasingly Processed Food Supply. *Health Education & Behavior* 37, 2 (April 2010), 211–226. https://doi.org/10.1177/1090198109339453

[30] Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for voice interface design. In *Proceedings of the 1st International Conference on Conversational User Interfaces - CUI '19*. ACM Press, Dublin, Ireland, 1–8. https://doi.org/10.1145/3342775.3342788

[31] GeminiTeam Google, Google. 2023. Gemini: A Family of Highly Capable Multimodal Models. https://paperswithcode.com/paper/gemini-a-family-of-highly-capable-multimodal

[32] 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, New York, NY, USA, 21–30. https://doi.org/10.1145/2556325.2566247

[33] Philipp Hacker, Andreas Engel, and Marco Mauer. 2023. Regulating ChatGPT and other Large Generative AI Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1112–1123. https://doi.org/10.1145/3593013.3594067

[34] Robert Hawkins, Hyowon Gweon, and Noah Goodman. 2021. The Division of Labor in Communication: Speakers Help Listeners Account for Asymmetries in Visual Perspective. *Cognitive Science* 45 (March 2021). https://doi.org/10.1111/cogs.12926

[35] Alyssa Hwang, Natasha Oza, Chris Callison-Burch, and Andrew Head. 2023. Rewriting the Script: Adapting Text Instructions for Voice Interaction. http://arxiv.org/abs/2306.09992

[36] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How do users respond to voice input errors?: lexical and phonetic query reformulation in voice search. (2013), 10. https://doi.org/10.1145/2484028.2484092

[37] Alan Kennedy, Alan Wilkes, Leona Elder, and Wayne S. Murray. 1988. Dialogue with machines. *Cognition* 30, 1 (Oct. 1988), 37–72. https://doi.org/10.1016/0010-0277(88)90003-0

[38] Suyoun Kim and Florian Metze. 2018. Dialog-Context Aware end-to-end Speech Recognition. In *2018 IEEE Spoken Language Technology Workshop (SLT)*. 434–440. https://doi.org/10.1109/SLT.2018.8639044

[39] Philipp Kirschthaler, Martin Porcheron, and Joel E. Fischer. 2020. What Can I Say? Effects of Discoverability in VUIs on Task Performance and User Experience. In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3405755.3406119

[40] Dimosthenis Kontogiorgos, Sanne van Waveren, Olle Wallberg, Andre Pereira, Iolanda Leite, and Joakim Gustafson. 2020. Embodiment Effects in Interactions with Failing Robots. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376372

[41] Geert-Jan M. Kruijff, Pierre Lison, Trevor Benjamin, Henrik Jacobsson, Hendrik Zender, Ivana Kruijff-Korbayová, and Nick Hawes. 2010. Situated Dialogue

Processing for Human-Robot Interaction. In *Cognitive Systems*, Henrik Iskov Christensen, Geert-Jan M. Kruijff, and Jeremy L. Wyatt (Eds.). Springer, Berlin, Heidelberg, 311–364. https://doi.org/10.1007/978-3-642-11694-0_8

[42] Sanna Kuoppamäki, Sylvaine Tuncer, Sara Eriksson, and Donald McMillan. 2021. Designing Kitchen Technologies for Ageing in Place: A Video Study of Older Adults’ Cooking at Home. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (June 2021), 1–19. https://doi.org/10.1145/3463516

[43] Hedda Lausberg and Han Sloetjes. 2009. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods* 41, 3 (Aug. 2009), 841–849. https://doi.org/10.3758/BRM.41.3.841

[44] Lin-shan Lee, James Glass, Hung-yi Lee, and Chun-an Chan. 2015. Spoken Content Retrieval—Beyond Cascading Speech Recognition with Text Retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 9 (Sept. 2015), 1389–1420. https://doi.org/10.1109/TASLP.2015.2438543

[45] Lee Hoi Leong, Shinsuke Kobayashi, Noboru Koshizuka, and Ken Sakamura. 2005. CASIS: a context-aware speech interface system. In *Proceedings of the 10th international conference on Intelligent user interfaces (IUI ’05)*. Association for Computing Machinery, New York, NY, USA, 231–238. https://doi.org/10.1145/1040830.1040880

[46] Manja Lohse, Katharina J. Rohlfing, Britta Wrede, and Gerhard Sagerer. 2008. “Try something else!” — When users change their discursive behavior in human-robot interaction. In *2008 IEEE International Conference on Robotics and Automation*. 3481–3486. https://doi.org/10.1109/ROBOT.2008.4543743

[47] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2019. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science* 51, 4 (Dec. 2019), 984–997. https://doi.org/10.1177/0961000618759414

[48] 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)*. ACM, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[49] Jacob L. Mey. 2010. Reference and the pragmeme. *Journal of Pragmatics* 42, 11 (Nov. 2010), 2882–2888. https://doi.org/10.1016/j.pragma.2010.06.009

[50] Robert J. Moore, Sungeun An, and Guang-Jie Ren. 2022. The IBM natural conversation framework: a new paradigm for conversational UX design. *Human–Computer Interaction* 0, 0 (June 2022), 1–26. https://doi.org/10.1080/07370024.2022.2081571

[51] Robert J. Moore and Raphael Arar. 2018. Conversational UX Design: An Introduction. In *Studies in Conversational UX Design*, Robert J. Moore, Margaret H. Szymanski, Raphael Arar, and Guang-Jie Ren (Eds.). Springer International Publishing, Cham, 1–16. https://doi.org/10.1007/978-3-319-95579-7_1

[52] Youssef Oualil, Dietrich Klakow, Gyorgy Szaszák, Ajay Srinivasamurthy, Hartmut Helmke, and Petr Motlicek. 2017. A context-aware speech recognition and understanding system for air traffic control domain. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. 404–408. https://doi.org/10.1109/ASRU.2017.8268964

[53] Sharon Oviatt, Jon Bernard, and Gina-Anne Levow. 1998. Linguistic Adaptations During Spoken and Multimodal Error Resolution. *Language and Speech* 41, 3-4 (July 1998), 419–442. https://doi.org/10.1177/002383099804100409

[54] Sharon K. Parker and Catherine G. Collins. 2010. Taking stock: Integrating and differentiating multiple proactive behaviors. *Journal of management* 36, 3 (2010), 633–662. https://doi.org/10.1177/0149206308321554

[55] Florian Pecune, Jingya Chen, Yoichi Matsuyama, and Justine Cassell. 2018. Field Trial Analysis of Socially Aware Robot Assistant. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS ’18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1241–1249.

[56] Emma Persky. 2022. Now we’re cooking – the assistant on {Google Home} is your secret ingredient. https://www.blog.google/products/assistant/cooking-with-the-assistant-google-home-your-secret-ingredient/

[57] Martin Porcheron. 2021. What’s in a name and does CUI matter?. In *CUI 2021-3rd Conference on Conversational User Interfaces*. 1–3. https://doi.org/10.1145/3469595.3469619

[58] 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)*. ACM, New York, NY, USA, 640:1–640:12. https://doi.org/10.1145/3173574.3174214

[59] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW ’17)*. ACM, Portland, Oregon, USA, 207–219. https://doi.org/10.1145/2998181.2998298

[60] Stuart Reeves and Martin Porcheron. 2022. Conversational AI: Respecifying participation as regulation. (Dec. 2022). https://doi.org/10.4135/9781529783193.n32 Publisher: SAGE Publications.

[61] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I Interrupt? Diverging Opinions on Proactive Smart Speakers. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–10. https://doi.org/10.1145/3469595.3469629

[62] Laurel Riek. 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction* (Aug. 2012), 119–136. https://doi.org/10.5898/JHRI.1.1.Riek

[63] Clayton Rothwell, Valerie Shalin, and Griffin Romigh. 2021. Comparison of Common Ground Models for Human–Computer Dialogue: Evidence for Audience Design. *ACM Transactions on Computer-Human Interaction* 28 (April 2021), 1–35. https://doi.org/10.1145/3410876

[64] Harvey Sacks, manuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn Taking for Conversation. *Language* 50 (1974), 696–735. https://doi.org/10.2307/412243

[65] Emanuel A. Schegloff. 1968. Sequencing in Conversational Openings1. *American Anthropologist* 70, 6 (Dec. 1968), 1075–1095. https://doi.org/10.1525/aa.1968.70.6.02a00030

[66] Derick Schweppe. 2022. Mycroft – The Open Source Privacy-Focused Voice Assistant. https://mycroft.ai/

[67] Ben Shneiderman. 2000. The limits of speech recognition. *Commun. ACM* 43, 9 (Sept. 2000), 63–65. https://doi.org/10.1145/348941.348990

[68] Deborah Tannen, Shari Kendall, and Cynthia Gordon. 2007. *Family Talk: Discourse and Identity in Four American Families*. Oxford University Press, USA. http://dx.doi.org/10.1093/acprof:oso/9780195313895.001.0001 Google-Books-ID: V2ISDAAAQBAJ.

[69] David Traum and Jeff Rickel. 2002. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems part 2 - AAMAS ’02*. ACM Press, Bologna, Italy, 766. https://doi.org/10.1145/544862.544922

[70] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users’ Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. https://doi.org/10.1145/3411764.3445536

[71] 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*. ACM, Pittsburgh PA USA, 1128–1144. https://doi.org/10.1145/3563657.3595975

[72] Yunhan Wu, Martin Porcheron, Philip Doyle, Justin Edwards, Daniel Rough, Orla Cooney, Anna Bleakley, Leigh Clark, and Benjamin Cowan. 2022. Comparing Command Construction in Native and Non-Native Speaker IPA Interaction through Conversation Analysis. In *Proceedings of the 4th Conference on Conversational User Interfaces*. ACM, Glasgow United Kingdom, 1–12. https://doi.org/10.1145/3543829.3543839

[73] 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*. ACM, Atlanta Georgia USA, 333–344. https://doi.org/10.1145/2678025.2701408

[74] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *4th Conference on Conversational User Interfaces (CUI 2022)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3543829.3543834

[75] Yaxi Zhao, Razan Jaber, Donald McMillan, and Cosmin Munteanu. 2022. “Rewind to the Jiggling Meat Part”: Understanding Voice Control of Instructional Videos in Everyday Tasks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI ’22)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3491102.3502036