Developing a Multimodal Entertainment Tool with Intuitive Navigation, Hands-Free Control, and Avatar Features to Increase User Interactivity

A Technical Report submitted to the Department of Engineering Systems and the Environment

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Grace Ko

Spring, 2022

Technical Project Team Members

Nathaniel Barrington

Caton Gayle

Erin Hensien

Megan Lin

Saimanga Palnati

Gregory Gerling

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Gregory Gerling, Department of Engineering Systems and the Environment

Developing a Multimodal Entertainment Tool with Intuitive Navigation, Hands-Free Control, and Avatar Features, to Increase User Interactivity

N. Barrington, C. Gayle, E. Hensien, G. Ko, M. Lin, S. Palnati, G.J. Gerling

***Abstract*—Advances in user interactivity in digital entertainment offer the potential to engage users beyond traditional passive and solitary experiences. Indeed, a high level of interactivity is inherent in tasks involving cooking, auto repair, and home improvement, all of which require users to complete multiple, detailed, and interdependent steps. Such tasks also require access to visual and audio instructions while a user's hands are engaged in a primary, physical task, and are often conducted in unique locations, e.g., kitchen or garage. This effort describes the design of an interactive, multimodal digital entertainment user experience for an 'edutainment' cooking show. The prototype wireframes incorporate three novel features, identified through requirements gathering and iterative design, of an interactive recipe map for hierarchical content navigation, voice command hands-free control, and avatars to further engage users. The overall design and three features were evaluated via usability testing in real kitchen settings in the conduct of actual cooking with real ingredients with a diverse range of seven participants. The results illustrate that the hierarchical navigational feature, alongside interactive voice communication, were effective at reducing users' cognitive load, and streamlined necessary information in order to support task completion.**

I. Introduction

Digital, in-home entertainment is a roughly $295 billion USD industry [1] that has become mainstream and ubiquitous over the past decade, due in part to its pricing model and the availability of broadband internet streaming. Through various private companies and governmental agencies, individuals can now access, on demand, a wide range of programming options. While now in common usage on mobile platforms, the majority of users tend to engage with such media in a solitary, sedentary fashion and with very little interaction. In particular, consumers are usually seated, and not prompted to act while in the middle of a show, in contrast to the high amount of interaction inherent in a video game. As a result, multiple entertainment companies are working on developing novel, next-generation paradigms.

Recent signals point toward a transition to increased engagement with entertainment programming and virtual reality. For example, Facebook, rebranded as "Meta," has promoted the "metaverse," which is defined as "a virtual world where people can socialize, work, and play" [7], which may change how people integrate their virtual and real lives. In addition, other companies have begun to integrate "Choose Your Own Adventure" content into their platforms, taking inspiration from second-person point of view children's books stemming from the 1970s to the present [3]. For example, Netflix now offers an interactive trivia game called "Trivia Quest," in which users receive a reward for answering questions correctly [8]. Users directly input their answers on a mobile or laptop screen or use a remote control to do so on a standard television. Disney+ offers a similar interactive episode called "Baymax Dreams," where users will use touch input to progress through an episode [2].

Moreover, beyond traditionally passive and sedentary experiences, a high level of interactivity is inherently found in tasks such as cooking, auto repair, and home improvement. Such tasks tend to require users to complete multiple, detailed, and interdependent steps. In addition, such tasks require access to visual and audio instructions while a user's hands are engaged in a primary, physical task, and are often conducted in locations, e.g., kitchen or garage, other than those for which media are traditionally viewed. In these cases, interactive guidance may enable users to efficiently break down difficult concepts and tasks, and become personalized to a user's or situation's specific needs, thereby reaching a wider variety of users, and regardless of disability or device limitation. In specific, those with motor impairments in the form of limited hand dexterity may benefit from hands-free, voice only interaction [4]. Such interactive applications also aid in dealing with pervasive issues such as mental health in the form of health behavior apps [5].

II. Methods

In the work herein, we describe the design of an interactive, digital entertainment experience that utilizes multimedia and multimodal methods to engage users in an active 'edutainment' how-to experience. The specific use case is tied with the task of cooking, chosen due to its hands-on nature and ubiquity, and conducted by people who range widely in age and other demographics. Focusing upon three novel features involving hierarchical content navigation, hands-free voice commands, and digital avatars, the prototype design aims to increase user interactivity and engagement with the content. First, an interactive map of a tree-like format allows the user to navigate the timeline and hierarchically organized content in completing a recipe. Second, a limited subset of hands-free voice commands allows users to control the navigation and video clips. Indeed, user may not be able use their hands to navigate the platform because they are covered in ingredients or residues. Third, two digital avatars support the entertainment aspect, one for comedic relief and one for educational tidbits, delivered in a question-style, multiple-choice format. Five common attributes of usability are taken into account in the design, including efficiency, satisfaction,

learnability, memorability, and a low error rate [6]. Most importantly, the design seeks to emphasize the importance of active engagement in the field as compared with more traditional and passive, sedentary entertainment.

III. Methods: Requirements Gathering

Requirements for building the user interface and experience were derived from interviews with various stakeholders coming from a large demographic range. Both information and functional requirements were iteratively refined during the initial design phases. The major constraints in the information requirements focused upon accessibility concerns with regards to recipes. Functional requirements focused on providing a medium for users to follow a recipe and to navigate the medium so that they can complete the recipe. The details of these requirements are listed below.

A. Information Requirements

Clear and Intuitive Formatting: The user interface must be easily learnable through the means in which information is presented to the user via its navigation, video, or tutorial.

Ingredients and Materials Overview: An overview of ingredients and materials necessary to cook the recipe will be provided before cooking begins. This makes the cooking process more efficient, and less error-prone, as users will not have to stop cooking midway into the recipe to retrieve necessary items.

Group Size and Allergies Information: To curate a personalized experience, basic parameters in the form of a pre-recipe questionnaire, tied to the recipe quantity, number of individuals to be fed, allergies or ingredients to be removed, etc. will be presented before cooking begins.

Recipe Selection Detail: Information at various levels of detail to allow users to initially choose between and select a recipe, including its name, duration of time to complete, if there is a particular celebrity involved, etc.

Recipe Completion Status: Clear guideposts are needed for users to know where they are temporally within the process of completing the recipe.

Suggested Chronology of Tasks: The user should be able to easily discern the suggested chronological order of completing the steps in the recipe. Individual steps throughout the process shall be broken into smaller tasks, each composed of individual ingredients, and include visual indication of completion.

Hands-Free Navigation Instructions: To utilize the hands-free component of this platform, the keywords must fit within lexicons commonly used for other common applications, and users will need a means to access keywords.

B. Functional Requirements

Intuitive Navigational Structure: By following the navigational hierarchical structure both vertically and horizontally, users shall be able to complete the recipe's steps in the most efficient order. It should be made intuitive that the ingredients on the bottom nodes of the recipe map are dependent on the node they are attached to above.

Active Navigation: The user must directly interact with the experience using either touch or voice to navigate through the recipe to reach completion.

Hands-free Controllability: Users must be able to navigate the interface entirely via voice-interaction if necessary for both accessibility concerns and users who are not able to use their hands (i.e., hands are wet/dirty from chopping, do not want to touch their phone).

Full User Autonomy: Users must be able to actively navigate through the hierarchical structure, maintain access to the recipe and tutorials/instructions, and have on-demand control of the flow of the show if they so choose.

Flexible and Decision Driven Environment: Throughout the experience, the user must be able to navigate through each step at their own pace, with the freedom to skip any additional functions that they deem unnecessary to their experience. Additionally, users must be able to return to previous steps if they feel the need to.

Active Engagement Opportunities: The avatars must be fully functional, timely, and add to the overall experience. The avatars must be equipped with the relevant data they need to execute their intended tasks, which are comedic content and trivia questions. The user must be able to easily opt out of such additional content.

IV. Methods: Prototype Development

Over the course of the user experience design, multiple prototype iterations were produced, evolving with feedback from stakeholders, and focusing more precisely on the scope of the final content. The wireframing tool Figma was used to create prototypes. Several model recipes were evaluated to help pinpoint issues, e.g., longer recipes required more complex and detailed navigational structure. To perform usability testing, the application Framer was used, which allowed us to integrate mock cooking videos into the prototype screens created in Figma.

A. Alternatives Generated

The first design concept sought to develop a navigational hierarchy by utilizing a horizontal mapping structure to illustrate the necessary steps to completion, with the ability to break down the cooking experience into separate sessions (Fig. 1A). The horizontal structure sought to optimize readability as the eye naturally travels from left to right across the screen. The user had the ability to select the number of sessions in which they would like to complete the recipe in order to promote flexibility and customization. Color-coded components signified the order of completion. The design lacked progress tracking to help the user orient themselves within the recipe, contained an overwhelming amount of text and colors, and did not maximize usage of the screen space. Therefore, the second iteration aimed to decrease the user's cognitive load while providing additional information (Fig. 1B). This goal was achieved by simplifying the map overview while including more informational depth in subsection pages. The subsection pages include the ingredients and their associated actions, indicating to the user how to best prepare

ahead of viewing the video content. The prior horizontal mapping structure was changed to a vertical format in order to maximize usage of screen space. A duration bar was added at the bottom of the page to offer feedback on completion progress. Separation by session, which allowed a recipe to be conducted over multiple days, was removed and the color scheme was simplified. Overall, we found that the prototype oversimplified the representation of the recipe, in both its iterations. Moving forward, we focused upon identifying and streamlining necessary information in order to maintain a reasonable cognitive load while still supporting task completion.

B. Final Concept and Designs

The final prototype design took feedback from preliminary user testing and stakeholder discussions to incorporate three main features: hierarchical content navigation, hands-free voice commands, and two entertainment focused avatars. Each novel feature is described in detail.

Hierarchical Content Navigation: An interactive map was designed to allow users to vertically and horizontally navigate a recipe at their pace (Fig. 2). First, an overview of the navigation with a tree-like format is displayed, with the main node being "Garlic Bread" and each subsequent sub-node describing a step of the recipe that connects to form its entirety (Fig. 2A). When the user zooms in, a more detailed view with images is shown that provides the user with information such as how much time each step of the recipe will take and quick tips for the recipe (i.e., "Best when used immediately" (Fig. 2B). A scroll button to the left traverses down the content vertically (Fig. 2C). Once a sub-node has been completed, a green overlay and a checkmark indicate completion of that step. The recommended next sub-node of the recipe is then highlighted with a yellow box outline (Fig. 2D). Once a sub-node is interacted with, a timed intermittent screen will pop up with a list of materials needed to complete this step of the recipe so the user can prepare for the video to begin (Fig. 2E). The video will then begin, providing several useful functions including an icon to get back to the main overview recipe map (Fig. 2F).

Hands-Free Voice Commands: Voice command functionality was added to avoid common issues of touching an electronic device with messy hands. If a user is preoccupied with other tasks while cooking, they may not always be able to return to the recipe. By touching the microphone icon (Fig. 3A) or by saying the phrase "Show me voice commands," an overlay of possible voice commands is introduced (Fig. 3B), with "Fast Forward," "Next Ingredient," and "Back to Map" meant to assist the user in navigating through the recipe. This aspect also considers a wider range of users, for example, users who have a disability and do not have use of their hands.

Avatars: The user can see where the different avatars will appear in the video using the transparent icons on the video timeline. The avatars may be turned on and off using the question mark and funny face icons pictured on the right side of the screen (Fig. 4A). The first avatar is named "Albert", who probes users with trivia questions related to the recipe (Fig. 4B). The second avatar is named "Otis", who interjects the recipe with humorous jokes (Fig. 4C).

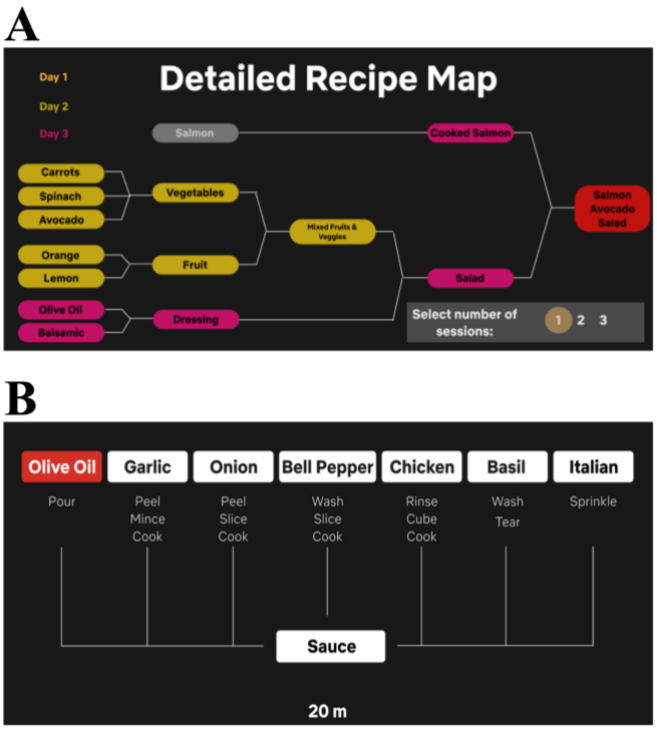

Fig. 1. Alternative maps generated. (A) The first hierarchical content navigation prototype focused on optimizing user flexibility and engagement **(B)** The second navigation prototype focused on increasing provided information and decreasing cognitive load.

V. Methods: Usability Evaluation

Users: Usability testing with volunteers was conducted to evaluate the features, overall flow, and diagnose concerns. User testing was completed with seven individuals from two age groups, the first with an average age of 23, and the second with an average age of 65. Four men and three women participated, across a range of cooking and technical expertise, and experience with voice commands, to reduce bias, given the small sample size.

Procedures: Each evaluation was procured in the same type of environment, i.e., that participant's own kitchen, to reduce the role of external factors in the usability evaluation. All participants cooked in their kitchen with a computer screen in front of them, which simulated the area of a mobile phone. The same recipe was used for everyone, which was making garlic bread with an air-fryer (Fig. 1). All ingredients necessary to make the garlic bread were provided, and tools were those of the individual's kitchen. Through the use of the Framer tool, which aids in screen-to-screen transitions, and Wizard-of-Oz (WoZ) techniques, participants traversed through the recipe. With respect to the voice commands, users spoke those to a facilitator, we then through WoZ manually input the commands into the prototype system in place of a working voice recognition system.

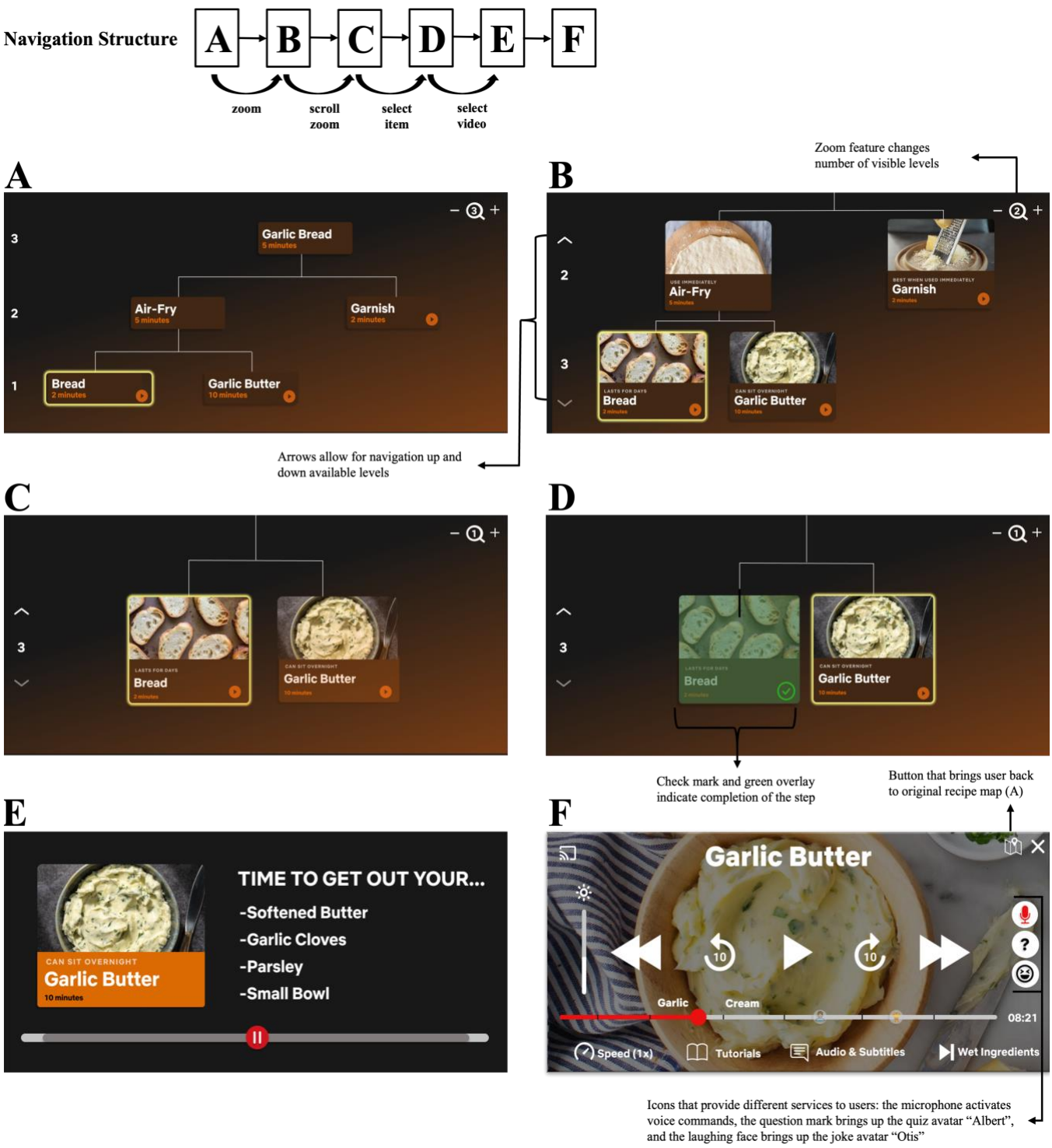

Fig. 2. Overall navigation flow of the multimodal tool. (A) Hierarchical content navigation **(B)** Zoomed-in view of the first two levels of the recipe map in A **(C)** After clicking the plus icon of the zoom arrow, the third level ingredients under the step "Airfry" are revealed **(D)** After the "Bread" step has been completed, a green overlay will cover the box, prompting the user to go to the next step "Garlic Butter" **(E)** Informational overview page appears before the recipe video starts informing the user on the needed ingredients and tools **(F)** Paused screen view of "Garlic Butter" video.

Evaluation Metrics: Multiple metrics were used throughout the evaluation process: whether participants completed the recipe fully, duration of time to complete the recipe, how many of the features they rated positively/negatively, and the number and severity of pain points. Throughout, the participants were highly encouraged to use the "talk-aloud" method to articulate their thought processes for each cognitive choice and selection action, even if their thoughts were not

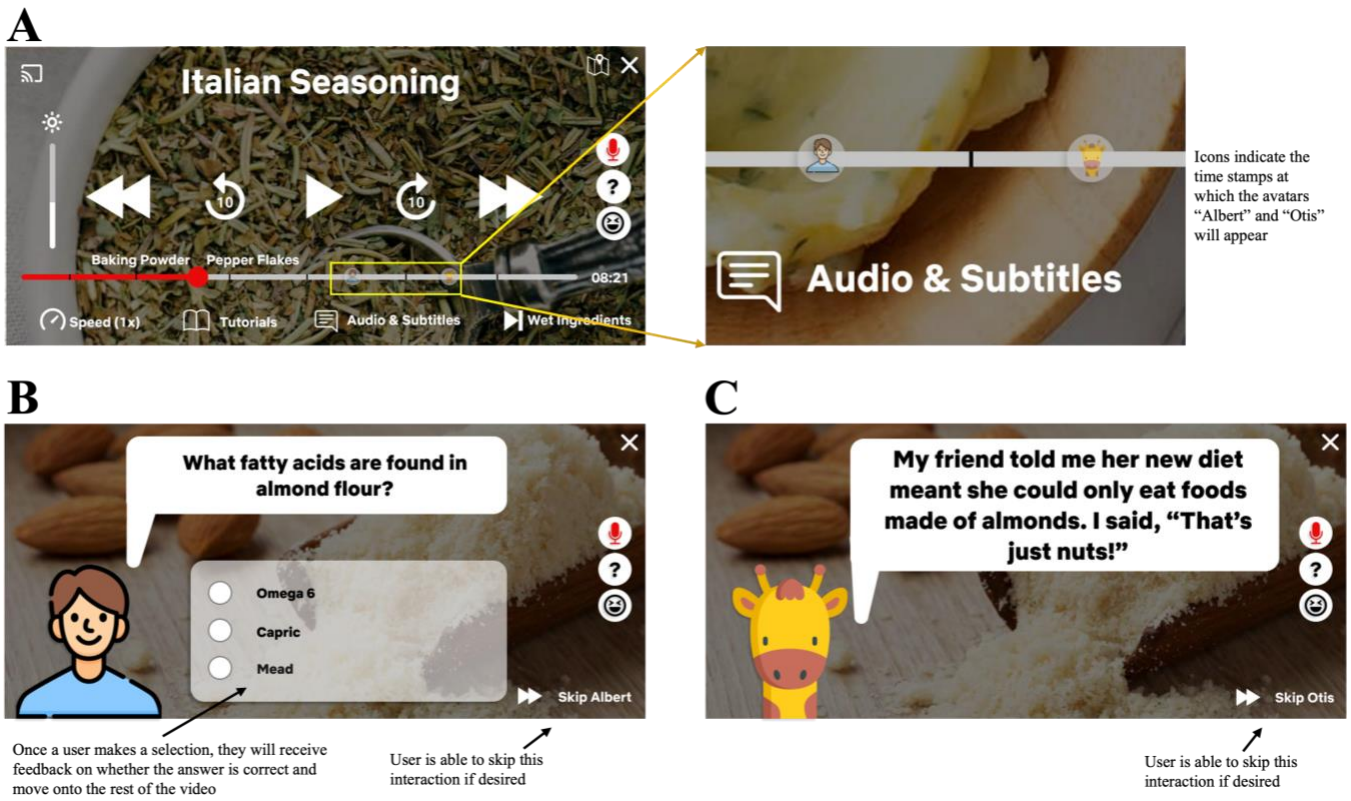

Fig. 3. Video screen and avatars. (A) Video overlay screen with time stamps of upcoming avatar interactions indicated on slider **(B)** Avatar named Albert that asks the user an interactive trivia question related to the recipe **(C)** Avatar named Otis that inputs entertaining jokes to keep user engaged.

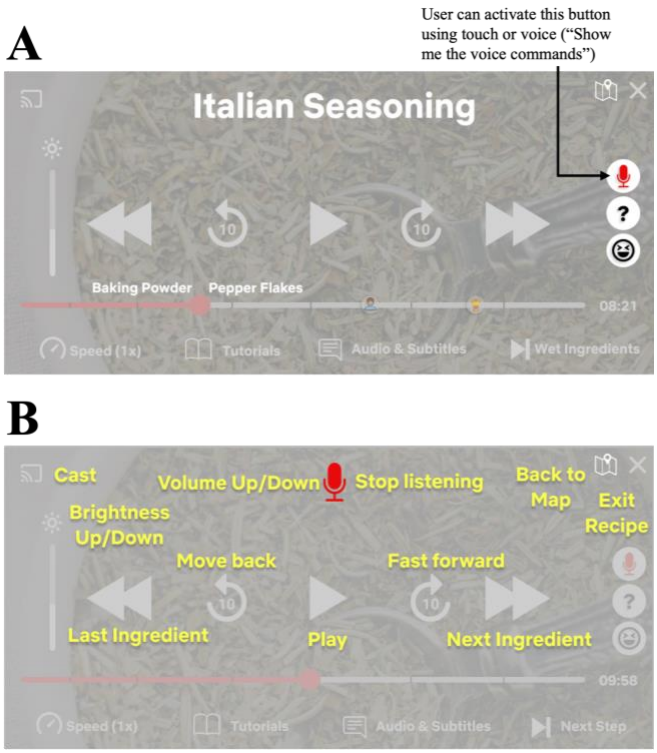

Fig. 4. Voice commands. (A) The hands-free function is displayed by the red microphone button which can be activated by either voice or by touch. **(B)** Once activated, an overlay of common voice commands will fill the screen to inform the user of what actions they can take.

connected directly to a voice command. Users were also instructed to progress through the prototype evaluation under the presumption that the facilitator was not there. However, the participant was able to ask questions to the facilitator in order to clarify any issues or offer suggestions. Additionally, participants were also asked a series of questions about their overall experience with the interface at the conclusion of their functionality test. Examples of these questions include: "What aspect did you feel most engaged with?," "What features did you expect or wish were included that were not available?," and "How likely would you be to use this product again in the future?" These questions provided general feedback on certain aspects of the interface, while the "talk-aloud" articulation throughout the evaluation provided deeper insights into pain points with the prototype interface.

VI. Results

The usability tests, utilizing the streamlined garlic bread recipe, involved seven individual users, all of whom successfully used the service to cook a meal. Each test was completed in approximately 45 minutes, with one participant's duration extending beyond an hour. After completion, each participant was asked an identical series of questions about their experience, with a particular emphasis on the three main features. Overall participant evaluations of both the hierarchical content navigation and avatars were positive, while the voice commands received a more neutral

evaluation. On average, facilitator intervention was required twice during each evaluation. Certain qualitative sentiments were consistently observed.

Hierarchical Content Navigation: Five participants found the interactive nature of the hierarchical content navigation enhanced their enjoyment of the prototype, while two participants felt the additional steps of the navigation prevented them from accessing the cooking portion in a timely manner. A specific issue among participants involved confusion with the zoom-based navigation within the hierarchy. During the test, the participants could adjust the number of visible steps by altering the zoom level. Four users noted that the limited number of visible options at the maximum zoom directed them to the incorrect steps in the process because the appropriate step was not immediately visible. This misunderstanding would likely lead future users to initiate an unintended portion of the recipe and derail the flow of the experience.

Voice Commands: The users were divided into two groups based on their preexisting familiarity with voice command systems. Some users' previous experience with voice command technology such as Amazon's Alexa and Apple's Siri led to seamless and persistent usage of the hands-free voice commands. Those participants with less prior experience were more intimidated and confused with these functions, leading to infrequent usage. This disparity could develop feelings of alienation in a potentially large user-base and discourage their use.

Avatars: General user sentiment on the avatar involvement was positive, however three users raised a common complaint relating to avatar interjections. These users felt that the avatars could be overly intrusive and that they were more of an interruption of the experience rather than an enhancement. All three users suggested improvements, including their integration into the video itself with smaller pop-ups that do not interrupt the video.

VII. Discussion

This project developed an interactive, multimodal entertainment tool that has three main features. The prototype wireframes incorporate three novel features, identified through requirements gathering and iterative design, of an interactive recipe map for hierarchical content navigation, voice command, hands free control, and avatars to further engage users. After conducting usability testing with a diverse set of users and stakeholders, several changes were made, including streamlining the navigation, a more comprehensive tutorial, and more accessible information. All of these changes increased the ease of using this tool.

The user experience design considered herein differs from traditional passive and sedentary experiences because it focuses on actively engaging the users in tasks with a high level of interactivity. Additionally, the design focuses heavily on voice commands as a way to engage a new section of users that are not traditionally considered in the design process, which limits motor capacity.

One limitation to this project was the constraint of time for creating the prototypes and conducting user testing. Some capabilities of our user testing platform were non-functional and this limited the scope of our usability testing results. The novelty and success of this multimodal tool demonstrates the potential for active engagement in entertainment experiences.

Acknowledgment

We would like to thank Victoria Bellotti and Avi Naim for their support and guidance.

References

[1] E. Corbett and A. Weber, "What can I say?," *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Sep. 2016, doi: 10.1145/2935334.2935386.

[2] "Disney's Immersive 'Baymax Dreams' Short Pushes New Boundaries For Interactive Animation | What's On Disney Plus," Jan. 28, 2021. https://whatsondisneyplus.com/disneys-immersive-baymax-dreams-short-pushes-new-boundaries-for-interactive-animation/

[3] "History of CYOA," *Chooseco LLC*. https://www.cyoa.com/pages/history-of-cyoa

[4] E. Corbett and A. Weber, "What can I say?," *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Sep. 2016, doi: 10.1145/2935334.2935386.

[5] L. Dennison, L. Morrison, G. Conway, and L. Yardley, "Opportunities and Challenges for Smartphone Applications in Supporting Health Behavior change: Qualitative Study," *Journal of Medical Internet Research*, vol. 15, no. 4, p. e86, Apr. 2013, doi: 10.2196/jmir.2583.

[6] R. Harrison, D. Flood, and D. Duce, "Usability of Mobile applications: Literature review and rationale for a new usability model. Journal of Interaction Science," *Journal of Interaction Science*, May 07, 2013. https://journalofinteractionscience.springeropen.com/articles/10.1186/2194-0827-1-1

[7] T. A. Press, "Facebook wants to lean into the metaverse. Here's what it is and how it will work," *NPR*, Oct. 28, 2021. [Online]. Available: https://www.npr.org/2021/10/28/1050280500/what-metaverse-is-and-how-it-will-work

[8] "Watch Trivia Quest | Netflix Official Site," *www.netflix.com*. https://www.netflix.com/browse?jbv=81449637