

USABILITY TESTING FOR VOICE INTERACTIONS

by Bill Franks | 10/12/2021

USABILITY TEST PLAN

Introduction

Cooking with Nonna is an Alexa skill that is being developed as a cooking application for home cooks. The skill provides a variety of quick and easy cooking recipes while assisting the user with step-by-step instructions.

Goals

The purpose of this usability test is to assess the usability and value of the skill with first time users. The main areas of focus will be the ability of the user to launch Cooking with Nonna, find recipes, change recipes and cook the chosen recipes by using the provided steps.

Test Objectives

1. Determine if users can successfully launch the skill.
2. Determine if the user can successfully chose a recipe for breakfast and change said recipe to another.
3. Determine if users can successfully navigate the steps of preparing the recipe.

Methodology

The majority of the tests will be conducted as moderated remote tests using video conferencing software (ie. Zoom, Skype, etc.). The remote tests will use the Alexa Developer Console with me signaling the users when to speak their utterances while I activate the mic in the Console. The computers used by the participants will need video and sound capabilities without the use of headphones. The additional tests will be in person using an Echo dot devices or the Alexa web app.

Participants

The study will take place between 10/15/21 - 10/22/21 and test 5 participants obtained through personal networks, friends and family.

Script

Hi and thank you again for taking the time to participate in some user testing. I'm going to give you a brief overview of what we're going to be doing today so you know what to expect. I'm a designer working on a new Alexa voice skill called "Cooking with Nonna".

Basically, I'm testing to see if users can use the skill efficiently, have any problems and if they like using it. The goal of this session is to evaluate the skill and not you. There are no right or wrong answers – I just want to hear your honest thoughts and opinions.

If you don't mind I would like to record this session and it will only be used by me for analysis purposes so I can improve Cooking with Nonna. I'll start by asking you a few basic questions before we start testing with the skill.

Do you have any questions before we start? Is it okay to start recording now?

1. What age range do you fall under?

- 21-25 years
- 26-30
- 31-35
- 36-40
- more than 40 years

2. Have you ever used a voice assistant before? If yes, how often would you say and for what type of tasks?

3. Do you cook? How would you categorize yourself as a cook?

- Beginner
- Intermediate
- Expert

USABILITY TEST PLAN

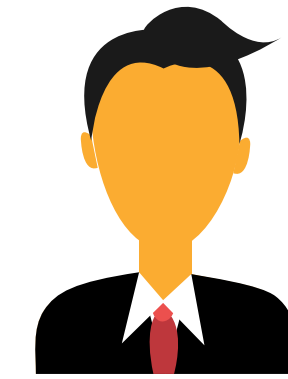
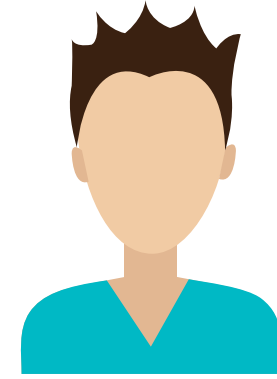
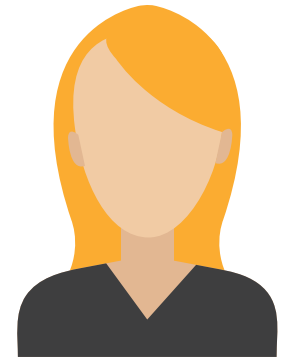
Tasks

1. You just woke up and want to have a quick breakfast but don't know what to make. Use Cooking with Nonna to find a breakfast recipe. (Find a breakfast recipe).
2. Once you start making your breakfast recipe you decide you don't like it and want something else (Change the recipe once you have started).
3. While you are making the breakfast recipe you miss a step and need a step repeated (Repeat a step in the instructions).

Follow-up Questions

1. On a scale of 1-10, how easy or hard was it to accomplish that task (1 being easy, 10 being hard)? Why did you give it that rating?
2. Was there something that you liked?
3. Was there something that you didn't like?
4. Did you feel like you were in charge of the interaction? Why or why not?

REPORT SUMMARY



Name: Kerry, Over 40
Voice Assistant Usage: 1-2 a day
Cooking Skill: Intermediate/Expert

Name: Dalena, 26-30
Voice Assistant Usage: 1-2 per week
Cooking Skill: Intermediate/Expert

Name: Cody, 26-30
Voice Assistant Usage: Rarely
Cooking Skill: Beginner

Name: Michelle, 36-40
Voice Assistant Usage: 1-2 per week
Cooking Skill: Intermediate

Name: Miguel, 26-30
Voice Assistant Usage: 1-2 per week
Cooking Skill: Intermediate

	Kerry	Dalena	Cody	Michelle	Miguel	Success %
Task 1: You just woke up and want to have a quick breakfast but don't know what to make. Use Cooking with Nonna to find a breakfast recipe. (Find a breakfast recipe).						100%
Task 2: Once you start making your breakfast recipe you decide you don't like it and want something else (Change the recipe once you have started).						40%
Task 3: While you are making the breakfast recipe you miss a step and need a step repeated (Repeat a step in the instructions).						80%

TEST RESULTS

MOBILE USABILITY TEST	Kerry	Dalena	Michelle	Cody	Miguel	TOTAL	POSSIBLE SOLUTIONS & NEXT STEPS	NOTES
Errors								
<i>[Error Rating 1] Trouble launching skill</i>						3	Be sure to add and describe the Name of the skill and wake phrase when marketing the skill on Amazon.	Some issues with participants forgetting the name of the skill or not knowing how to launch a skill.
<i>[Error Rating 2] : Interrupting of Alexa's responses caused errors.</i>						4	Maybe make some of the responses shorter since users kept trying to interrupt.	Most users tried to interrupt the response which later caused problems.
<i>[Error Rating 4] : Utterances didn't work all of the time because they were not defined.</i>						5	Add more utterances since errors were usually caused by lack of utterances.	Although not part of the task, some participants had a hard time ending the tasks early.
<i>[Error Rating 1] : Utterances didn't work because of Alexa not picking up utterances over video.</i>						2	Problems were probably magnified because of doing the test over video.	Once utterances were spoken again in a clearer manner the skill worked.
<i>[Error Rating 5] : Skill failed on the second task. "Problem with Skill Response"</i>						2	Problem with Alexa skill. Once I reset the developers console it seemed to work fine.	The skill failed in 2 different tests. The first one we chose lunch instead. The second one I resaved the code and it worked.
Observations								
<i>Participants kept talking over or interrupting Alexa.</i>						3		
<i>The farther along in the test the more comfortable the participants became with using the skill.</i>						3		
<i>Participant mentioned that they did not want any backstory that it was important to just get to the recipe.</i>						1	Keep recipes as short and succinct as possible.	
Negative Quotes								
<i>"It was annoying saying next all of the time."</i>						1	Shorten the skill's responses whenever possible and use tapering for returning users.	Users generally used the same next or repeat command even though they didn't need to.
<i>"I didn't like that I said next and it wouldn't listen and kept going. It wouldn't listen to me."</i>						2	Other than making the response shorter, I'm not sure that I can improve since Alexa doesn't allow burst in.	
<i>"Not sure the parfait was an appropriate recipe for the context since it was more assembly than cooking."</i>						1	Maybe replace this recipe with another one since this was more of a make ahead recipe.	
Positive Quotes								
<i>"It was really easy, I liked the different steps. I liked that it stopped in case you wanted to write it."</i>						2		
<i>"Easy to navigate once I heard the defined commands. "</i>						2		
<i>Made me feel like it was accepting what I was saying. I had the most freedom for that last task."</i>						1		
<i>"It was nice that it was able to change and do what I needed to do after the first error."</i>						1		
<i>"It's easier for me to listen to the instructions than to try to read them, so I really liked that there was a pause between steps."</i>						2		

OBSERVATIONS & REVISIONS

1. The majority of the errors were caused by a lack of utterances. I adjusted the script to incorporate the utterances used by participants.
2. I added the additional lunch, dinner and snack recipes since there were some problems with getting the skill to work correctly for breakfast.
3. I shortened the `first_time_instructions` response and added in some tapering responses for returning users to avoid some frustration.
4. The follow-up questions provided valuable information regarding "positive & negative" quotes and general insights. The difficulty scale of 1-10 did not provide any additional information since most users rated the tasks as easy even though they couldn't complete the task. The ratings seemed to be derived from the ease of the instructions and not whether or not they could finish the task.
5. Most of the users had a hard time not interrupting the skill's responses. This caused some frustration for a majority of users. I think most users were used to voice systems that could easily be interrupted or "barged in" on. Many commented on phone systems and how easy it was to talk over them to get what they wanted.
6. Most users had some sort of problem launching the skill. If it was not at the very beginning it was later when they forgot the name of the skill. I think that the name is unique and easy enough to remember. Repetition and use of the skill should help with users remembering.
7. I'm still not sure why but for some reason I got a "problem with skill response" during two of the tests. This response came well into the tests and seem to happen at random. For one test I instructed the user to choose "lunch" instead of "breakfast". The the other test I re-deployed the code from the development console and that fixed the problem.
8. All of my usability tests except for one, were done via zoom with the user speaking to a echo dot device. This seemed to work really well with the exception of a few times when the user mumbled or could not be heard very well over the video. After the first test was conducted this way I realized that it was beneficial to place the dot within the users view so they could see the light prompts that were provided by the dot.
9. The majority of participants became more comfortable using different utterances and with the device in general by the end of the three tasks. It was interesting to see that some users thought that they were using the defined utterances when in fact they were varying them quite often. An example would be saying "next" to go to the next step versus "continue" or "keep going".