

## USABILITY STUDY & FINDINGS

# A Usability Study on Trip-Planning with Apple's Siri

TEAM 10

SHELLEY XIA

SHRAVYA NEERUGANTI

ALICE MERRICK

## INTRODUCTION

Perhaps unsurprisingly, the Internet has become one of the most effective ways for individuals to seek information about travel. However, frequent travelers claim that they are often overwhelmed by the huge amount of information available online and at times struggle to find what they are looking for, causing trip-planning to be a frustrating experience. [1] Considering the complexity involved with planning a trip online, we were inspired to investigate how a natural language user interface technology such as Apple's Siri can assist with tasks related to trip-planning.

The research team that conducted the usability tests consisted of three MHCID graduate students at the University of Washington. The usability tests were conducted in a quiet office space on November 30 and December 1 using iOS version 10.1.1 of Siri on the iPhone. The findings of the usability test are recorded in this document.

## REFERENCES

1. Donna L. Hoffman, Thomas P. Novak, and Alladi Venkatesh. 2004. Has the Internet become indispensable? *Communications of the ACM* 47, 7 (January 2004), 37–42. DOI:<http://dx.doi.org/10.1145/1005817.1005818>

## RESEARCH GOALS

The study examined the use of Apple's Siri voice assistant for the purpose of trip-planning for frequent long-distance travelers. The goal was to determine what attitudinal and usability obstacles users faced when using Siri for basic tasks associated with planning a trip, in order to discover ways to improve the experience for future users.

## RECRUITING & PARTICIPANTS

Our participants were recruited from a Facebook group called HCDE Peeps (a group for students and alumni of the HCDE department) and from the MHCID program at the University of Washington. We recruited four participants total, two of which were full-time graduate students at the University of Washington. The other two participants were non-student, working individuals who live in Seattle, WA.

## SESSIONS

Each usability test lasted approximately 30 minutes. During the test, the facilitator briefly interviewed each participant about their travel habits, had them complete four tasks, and then closed the interview with questions regarding their impressions of using Siri.

Before and after each task, the facilitator asked the participant to rate the task on a 5-point likert scale with measures ranging from Very Easy (1) to Very Difficult (5), in order to understand the following:

- How difficult the participants anticipated using Siri will be
- How difficult the participant felt using Siri actually was after the task

After each task was completed, the facilitator asked the participant to rate their experience of using Siri for the task on a 5-point likert scale with measures ranging from Very Unsatisfied (1) and Very Satisfied (5) for subjective measures including:

- How confident they were in completing the task
- If the results from the task were useful or relevant
- How many times they had to attempt the task

In addition, the facilitator asked the participants the following after each task:

- Overall impressions of using Siri
- Difficulty in understanding Siri
- Difficult being understood by Siri
- Surprises by what Siri could or could not do
- Changes they would make to Siri
- Using Siri for trip-planning in the future

## DATA COLLECTION

The facilitator used her iPhone as the testing phone, which participants used to engage in the test. The test phone captured participants' text messages sent using Siri while the note-taker documented hand-written notes and quotes from participants. The third researcher took photos of the participants using Siri during the session. The notes recorded were the main source of data collected as there were no audio or video recordings made.

## KEY FINDINGS

- **Interruptions** - In a few cases, Siri interrupted participants as they were speaking, which frustrated participants.
- **Misunderstanding Phrases** - When participants used certain phrases or words, Siri could not understand and gave results the participants were not looking for.
- **Misrepresentation of Visual Feedback** - Siri transcribed what participants were saying but occasionally, the visual feedback given was different than what was spoken. Participants were confused by the visual transcription Siri provided.
- **Ongoing Dialogue** - Participants wanted to keep a conversation with Siri.

## PRIORITY RATINGS

Priority ratings are assigned to usability issues found in the tasks and provide recommendations for the order in which issues should be fixed. For each usability issue we found, we assigned it with a rating based on the following criteria:

- The frequency in which this problem occurs (*Is it common or rare?*)
- The impact the problem has on users (*Is it easy or difficult for users to overcome?*)
- The persistence of the problem (*Is it a one time problem or do users continuously face this problem?*)

### 1 HIGH PRIORITY

This rating indicates a usability catastrophe and it is imperative to fix this before the product is released. Issues in this category are major functionality problems that prevent or severely impede the user from successfully using the voice UI.

### 2 MEDIUM PRIORITY

This rating indicates a major usability problem and is important but not crucial to fix before product release. Issues in this category have some functionality issues that may hinder the users' interaction with the voice UI but does not prevent them from carrying out their tasks.

### 3 LOW PRIORITY

This rating indicates a minor usability problem and fixing this should be given low priority. Issues in this category do not impact a user's ability to use the voice UI.

TASK COMPLETION SUMMARY

This chart gives an overview of the task completion among the four participants.

To learn more about each task, please refer to the following Task Analysis section.

Participant	Task 1	Task 2	Task 3	Task 4
1	✓	✓	✓	✓
2	✓	✓	✓	✗
3	✓	✓	✓	✗
4	✓	✓	✓	✗
Success	4	4	4	1
Completion Rate	100%	100%	100%	25%

## USABILITY STUDY & FINDINGS

# Task Analysis

TASK 1: CHECK CALENDAR // TASK ANALYSIS

You’ve been invited to a conference in San Francisco this Saturday and Sunday. You would like to check if you are free this weekend to travel.

**Success Criteria:** Find calendar and check availability using Siri

FINDINGS SUMMARY

The participants did not know how easy or hard the task would be. All participants were surprised in one manner or another by the results of the task - three participants were surprised by how easy the task actually was, and one participant was surprised by how difficult it actually was. In every case, the participants’ expectations were not met.

USABILITY ISSUES

- 2

In a few cases, Siri interrupted participants as they spoke.

**Example:** When one participant spoke to Siri, Siri abruptly interrupted her in the middle of her sentence. When the participant tried to continue, Siri said, “I didn’t understand what you meant”.
- 2

Siri sometimes gave undesirable or irrelevant results when participants’ used certain wording.

**Example:** One participant had to repeat the task more than once due to the language/speech that the participant had used. She expressed that she had to learn “Siri’s language”. She first tried, “Am I completely open this weekend?” (see Fig. 1) which Siri did not understand. She then rephrased to ask Siri to “check my calendar” (see Fig. 2) and that yielded the desired response from Siri. She expressed that she would need to remember to use those exact words if she wanted to use Siri for such a task again.

TASK COMPLETION RATE

100%

5-POINT LIKERT SCALE (VERY EASY - VERY DIFFICULT)

Participant	1	2	3	4	Summary
Rating before task	3	3	3	1	neutral
Rating after task	1	1	1	4	easier than expected

**Takeaway:** All participants found the task easier than they expected, except for one participant.

TASK 1: CHECK CALENDAR // TASK ANALYSIS CONT'D

SCREENSHOTS

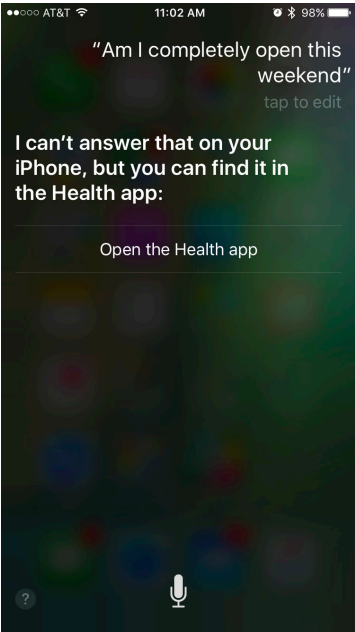


Fig. 1 Some words or phrases will yield irrelevant or undesirable results.

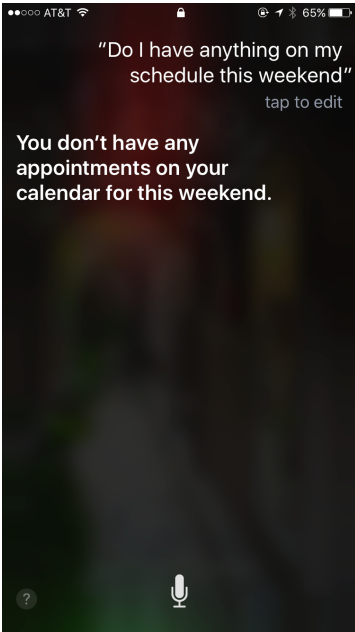


Fig. 2 Using certain key words like "schedule" will prompt a more appropriate answer from Siri.

QUOTES FROM PARTICIPANTS

*"I assume Siri can handle it. It takes 2 seconds on my computer."*

*"Easier than I expected."*

SINGLE WORD DESCRIPTIONS SUMMARY

- **Most participants found the task "straight-forward" to complete.** Some other descriptions of the task included "simple", "easy", and "smooth".
- **One participant mentioned that the task was "annoying", "repetitive", and "borderline frustrating".**

5-POINT LIKERT SCALE (VERY UNSATISFIED - VERY SATISFIED)

Participant	1	2	3	4
Rating after task	4	5	5	3

**Summary:** Participants found the process of completing this task generally satisfying.



TASK 2: CHECK WEATHER FORECAST // TASK ANALYSIS

Now that you’ve established that you are free to go to the conference in San Francisco this weekend, you would like to know if you should pack your rain gear for the trip.

**Success Criteria:** *Using Siri, check weather conditions to see if it is raining during the weekend*

FINDINGS SUMMARY

This was a very successful task. One participant wanted to know the percentage chance of rain because that would give her more information on whether or not she would pack her rain gear. All in all, this was an easy task to complete for all participants.

USABILITY ISSUES

No usability issues were found specifically for this task.

TASK COMPLETION RATE

100%

5-POINT LIKERT SCALE (VERY EASY - VERY DIFFICULT)

Participant	1	2	3	4	Summary
Rating before task	2	1	2	2	generally easy
Rating after task	1	1	1	1	easier than expected

**Takeaway:** All participants found this task easy to complete.

TASK 2: CHECK WEATHER FORECAST // TASK ANALYSIS CONT'D

SCREENSHOT



Fig. 1 Siri will display a visual weekly forecast.

QUOTES FROM PARTICIPANTS

[The weekly forecast display is] *"nice and informative."*

One participant has used Siri to check the weather before and commented, *"It actually works quite well."*

SINGLE WORD DESCRIPTIONS SUMMARY

- **All participants found this task to be "easy" and "clear".** Words such as *"simple"*, *"informative"* and *"straight-forward"* were mentioned more than once by the participants.

5-POINT LIKERT SCALE (VERY UNSATISFIED - VERY SATISFIED)

Participant	1	2	3	4
Rating after task	5	5	5	5

**Summary:** All participants found the process of completing this task very satisfying.

TASK 3: SEARCH MAPS FOR A RESTAURANT // TASK ANALYSIS

While you are in San Francisco, you would like to meet up with your friend Gloria for dinner. Please find a restaurant near the San Francisco airport where you can meet with her.

**Success Criteria:** Find a list of restaurants nearby the San Francisco International Airport

FINDINGS SUMMARY

Many of the participants were not sure whether or not they completed the task. Two participants mentioned that the list of restaurants displayed did not contain enough information (e.g, lack of distance) for them to make a decision. One participant was surprised about how informative and visual the results were.

USABILITY ISSUES

- 1

Participants were unclear if Siri showed restaurants that were near the airport.

**Example:** There were no indications of how far the restaurant is to the airport, which caused the participants to be unsure of which restaurant to select.
- 2

Some participants needed to reword or rephrase their query a few times.

**Example:** When participants asked Siri to show restaurants near San Francisco International Airport, Siri confirms with “Showing restaurants near San Francisco”. However, participants were confused by the wording and weren’t sure if the list displayed restaurants near the airport or restaurants located generally in the city of San Francisco (see Fig 1.)
- 2

Siri does not remember the context of previous queries from the participant.

**Example:** One participant wanted to see a list of restaurants near the San Francisco Airport but Siri did not understand the request. The participant then reiterated, “Give me a list of restaurants near the airport” but Siri displayed a list of restaurants near Seattle-Tacoma International Airport, not SFO. The participant expressed frustration, mentioning “I was just talking about San Francisco, why would Siri give me results for Seattle?”

TASK COMPLETION RATE

100%\*

\* Despite completing the task, none of the participants picked a restaurant from the list, which was our original intention. This was because our task prompt did not specify this clearly to our participants.

5-POINT LIKERT SCALE (VERY EASY - VERY DIFFICULT)

Participant	1	2	3	4	Summary
Rating before task	3	4.5	4	4	generally difficult
Rating after task	1	2	5	3	easier than expected

**Takeaway:** Most participants found the task to be much easier than they expected, except for one participant.

## TASK 3: SEARCH MAPS FOR A RESTAURANT // TASK ANALYSIS CONT'D

## SCREENSHOTS

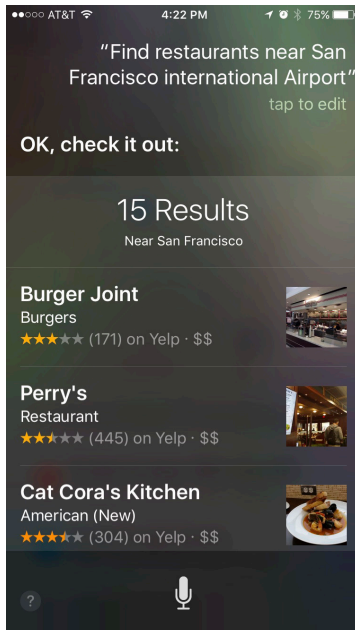


Fig. 1 Siri displayed results "near San Francisco" despite the request for finding restaurants near the San Francisco International Airport.

## QUOTES FROM PARTICIPANTS

*"I guess it's good enough"*

*"Didn't get the info I needed"*

*"How do I know where they (restaurants) are? I would expect it to say 0.2 miles so I can see if some of them are in the airport"*

## SINGLE WORD DESCRIPTIONS SUMMARY

- **Only half of the participant described this task as "simple" and "easy".**
- **The other participants described their frustrations with completing the task by using words such as "intuitive" and "difficult".** Some participants felt like they did not or were not able to complete the task successfully, describing the process of the task as "unsuccessful" and "so-so".

## 5-POINT LIKERT SCALE (VERY UNSATISFIED - VERY SATISFIED)

Participant	1	2	3	4
Rating after task	5	3	2	3

**Summary:** Participants found the process of completing this task generally unsatisfying.

TASK 4: COORDINATE WITH A FRIEND // TASK ANALYSIS

You have now found a place to meet your friend. Using Siri, ask your friend Gloria to meet you at the restaurant on Saturday evening at 8:00pm.

**Success Criteria:** Message, text, or email Gloria asking to meet at chosen location Saturday at 8:00pm

FINDINGS SUMMARY

Only one participant was able to successfully include “Saturday” and “8PM” in their communication. Three of the 4 participants had to construct their messages more than once in order to write a message they wanted to send; a participant had to attempt the task a few times because Siri did not understand the word “friend” before the name of the contact was given. One participant did not include either “Saturday” or “8 PM” in her message because she assumed that Siri did not display all of the text and but didn’t bother to check if the message was constructed as such. See Fig 2 for examples of participants’ messages.

USABILITY ISSUES

- 1

Siri provided inappropriate wording and feedback to participants.

**Example:** When participants asked Siri to send a message to Gloria, the messages were not transcribed correctly. 8PM did not show up in the message. Siri did not understand “my friend, Gloria” and would ask who to send the message to (see Fig 1.).
- 3

Siri provides inconsistent results.

**Example:** Despite the similar or identical wording of the request from the four participants, only one participant was able to successfully include “Saturday” and “8PM” in the text message sent (see Fig 2.).

TASK COMPLETION RATE

25%

5-POINT LIKERT SCALE (VERY EASY - VERY DIFFICULT)

Participant	1	2	3	4	Summary
Rating before task	1	1	4	2	generally easy
Rating after task	2	1	4	3	more difficult than expected

**Summary:** Two participants found the task to be more difficult than expected. The other 2 participants found it to be as difficult or easy as they had predicted.

TASK 4: COORDINATE WITH A FRIEND // TASK ANALYSIS CONT'D

SCREENSHOTS

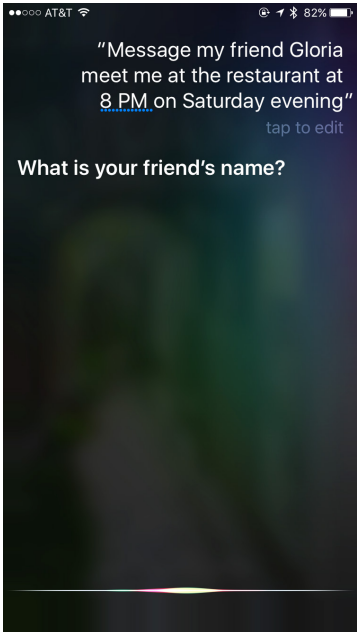


Fig. 1 Siri was not able to recognize Gloria as a contact at first attempt.

QUOTES FROM PARTICIPANTS

One participant wanted Siri to display more information - *"It didn't show me the full text, only ellipses"*



Fig. 2 The various messages participants recorded and sent. Only one message contained both "Saturday" and "8 PM".

SINGLE WORD DESCRIPTIONS SUMMARY

- **Some participants found this task to be "easy" and "satisfying".**
- **Many participants thought the process of completing this task to be a "hassle" and "complicated".** Other participants described the process as "glitchy" and "mildly tedious" because they had to attempt the task a few times before completion.

5-POINT LIKERT SCALE (VERY UNSATISFIED - VERY SATISFIED)

Participant	1	2	3	4
Rating after task	2	4	3	4

**Summary:** Participants found the process of completing this task generally satisfying, with some mixed results.

## RECOMMENDATIONS

Through identifying usability issues and understanding behavioral and attitudinal obstacles that exist for the participants in using Siri, we have developed the following design recommendations:

1. Siri should be context-aware or context-sensitive. For example, if a user asks Siri about the airport in San Francisco and refer to the airport in the next query, Siri would remember that the airport referenced is the San Francisco airport mentioned in the previous request.
2. Siri should provide a visual interpretation of the user's query, so that the user is able to alter the assumptions and interpretations made by the natural language UI.
3. Siri should pick up on the user's natural pauses and filler words (i.e., "umm" or "uh") and not interrupt the user in those scenarios.
4. Siri should also filter out filler words when interpreting the user's query.
5. Siri should provide more in-depth visual affordances depending on the context of the inquiry. For example, when displaying a list of restaurants, the UI should display distance related to the location requested.

## POST-TASK RESULTS

In a post-task questionnaire, we asked the participants about their overall impressions of using Siri, including any difficulties they might have had understanding or being understood by Siri. Generally, most participants expressed that they found Siri to be useful only in certain contexts. One participant commented that he found Siri to be *“rough around the edges”*, while another participant indicated that it could be helpful, but *“is sometimes more trouble than it’s worth”*. The participants found that it was overall easy to understand Siri, but found it more difficult to be understood by Siri using their natural speech.

We also gathered impressions on whether or not our participants would use Siri to help plan a trip in the future. All four participants mentioned that they would likely use Siri for tasks such as scheduling or checking the forecast, but would not try using it for anything beyond these simple tasks. This was because they did not feel that Siri was capable of completing end-to-end trip-planning activities, such as searching and booking flights or hotels.

## CONCLUSION

Siri can be effective in carrying out simple tasks to help an individual with an activity such as planning a trip. However, many of our participants experienced difficulty with using Siri, due to the program’s inability to understand certain phrases, words, or contexts, which led to inappropriate responses that caused user confusion. Since many participants perceived Siri as unable to perform certain tasks, this led to the belief that Siri has many limitations in its capabilities and functions. If these aspects of Siri can be improved, Siri would be a stronger voice assistant.