Digital Voice Assistant vs. Documentation

Study Report
HCI HS 2019
Group 3

Jonas Bokstaller, Nicolas Carmody, Mina Huh, Vu Nguyen, Kunaal Sikka

Table of Contents

Study Goal

Introduction

Hololens, a Mixed Reality smart glasses developed and manufactured by Microsoft uses Voice and Gaze as well as Motion as input. This enables users to work more efficiently in different situations. However, there is an issue that the users are not familiar with using Voice interaction. To solve this problem, we came up with two solutions. First, we introduce Virtual Assistant as an alternative of documentation. Second, users play a simple game to get familiar with Voice Interaction.

Study Goal

In our user study, we want to find out how effective these solutions are. Specifically, we have the following research questions :

Q1: What is the effect of using documentation vs being guided by a Virtual Assistant on the percentage of voice commands?

Q2: What is the effect of using documentation vs being guided by a Virtual Assistant on task completion?

Hypotheses

We formulated hypotheses concerning the features that we think of as important.

i. Users voice usage will increase when helped by a Virtual Assistant
ii. Users will feel more supported with the existence of Virtual Assistant vs documentation
iii. Playing the game will help users get familiar with voice interaction and therefore is helpful for upcoming tasks

Method

Our group was interested in how good voice input would perform when using the HoloLens. The experiment was split into two tests to address specific aspects of voice performance we were interested in. The performance of voice input was first measured in terms of how well a user might learn a task if the information was given through a voice assistant vs. if the same information was given via a description. In a second experiment the question of how comfortable a user might feel by using a voice assistant or a written description to help in a simple task as well as how comfortable the user is using voice input.

Experiment protocol

Set up

To simulate the HoloLens environment, we use a classroom. All the HoloLens overlay (simulated with Windows) is projected to parts of the room (i.e. the wall). The participant can move freely within the room. The computer is operated by a person that watches the hand clicks of the participant (simulated with a laser pointer) and listens to the voice commands. The operator ignores invalid inputs. Another person simulates the digital assistant, which is activated by the participant with eye-contact or by the keyword “Hey Cortana”.

The procedure

The experiment consists of two tasks. For both tasks, we decide randomly whether the user gets introduced to the task by a digital assistant or a pop-up textual help menu/user guide. We do make sure that there is a balanced number of participants using either method.

Prior to starting the experiment, we explain to the participant briefly that they are in a HoloLens environment, and only show them how to open/close the help menu, or how to activate the digital assistant.

The participant is then “left alone”, meaning the task begins and the participant can only interact within the limits of the system (valid inputs, help menu/digital assistant).

In task 1, the participant plays a simple game with voice commands only. Here, we measure the performance of the participant (Number of invalid commands, number of referrals to the documentation, time to complete).

In task 2, the participant is tasked to send an email using the Windows Mail app. The participant may use voice commands or hand input. Here, we measure the performance (as in task 1), plus the number of gaze/hand inputs versus voice.

After the completion of both tasks, the participant fills in a questionnaire that records demographic information, and both graded and textual feedback on the various input and teaching methods.

The digital assistant

Prior to starting the task, the digital assistant explains the objective of the task and shows the user how to operate the application. The digital assistant is a visible person and uses human-like language and gestures. It can also operate on the app for demonstration purposes. The digital assistant guides the participant like a human would teach another human, i.e. it doesn't just read a list of commands (this applies only to the first-use introduction).

While solving the task, the participant can activate the digital assistant and ask questions. As opposed to voice command input of the apps, the digital assistant can understand natural language and react to questions in a natural way. Nevertheless, the digital assistant possesses the same knowledge as the help menu/user guide.

Help menu/user guide

At the beginning of the task, the help menu is shown to the participant. We explain to the participant that he can use the voice commands "open help" or "close help" to close/open the help menu at any time. After closing the menu for the first time, the task begins.

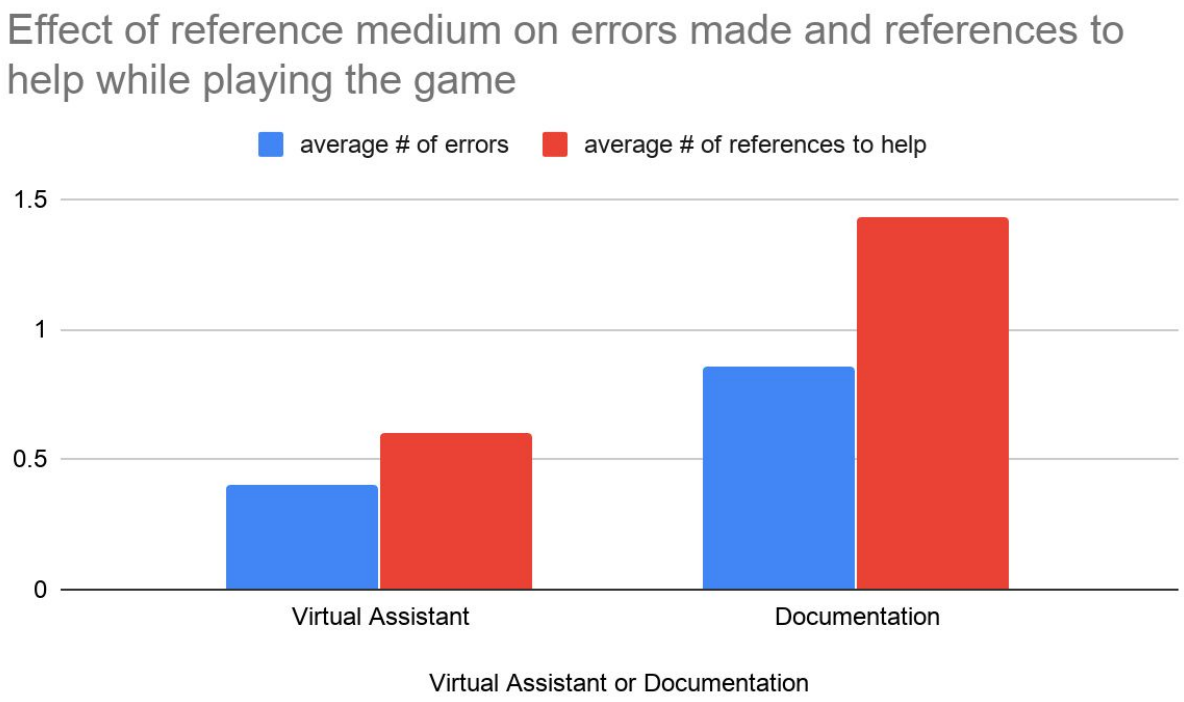
The user guide contains the objective of the task and the voice commands. In task 2, the user guide also contains instructions for hand input.

Results

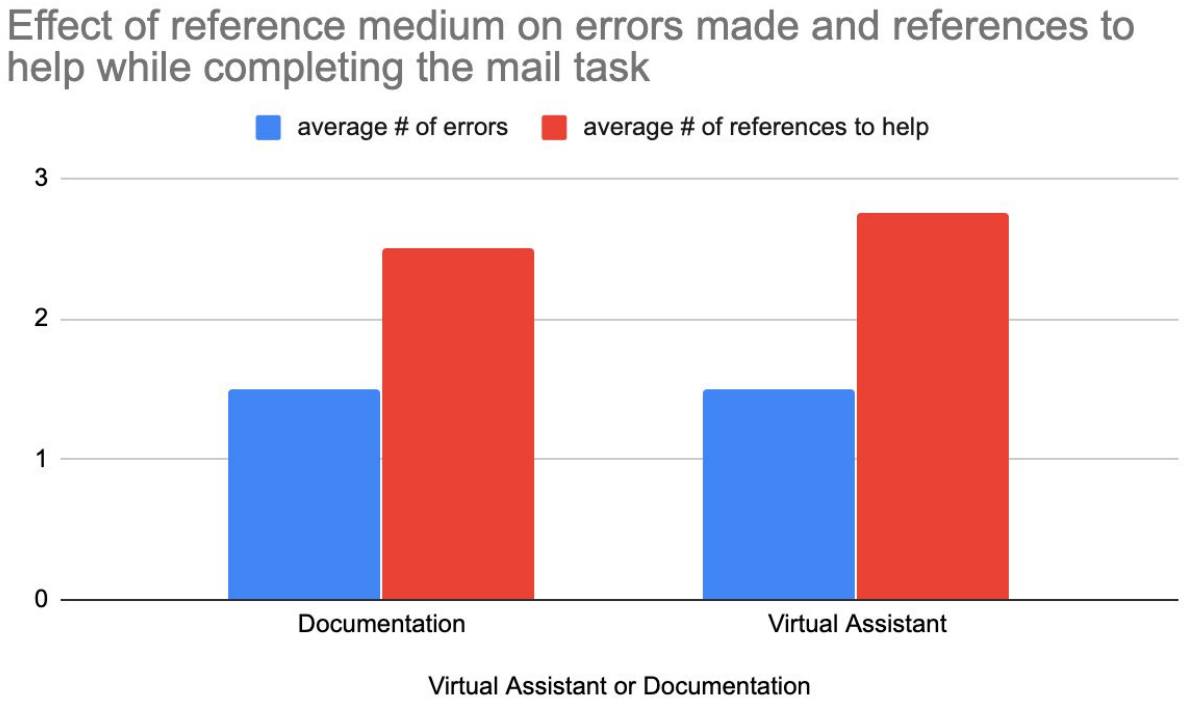
The results of our experiments supported our hypothesis that having a virtual assistant enables the user to learn how to use new applications more effectively, and encourages increased voice-command usage.

We conducted 24 trials in total, 12 where users played the game and 12 where users completed the task we gave them in the Windows 10 mail app. For the game, we had 5 trials with the virtual assistant and 7 trials with the documentation. For the mail task, we had 4 trials with the virtual assistant and 8 trials with the documentation. Ideally, we would have liked to have an even split of virtual assistant and documentation trials for both the game and mail task, but the short time span and our limited resources made it difficult for us to coordinate this. This is something that we would like to change if we were to repeat the experiment.

First, we look at the results of the game. We noticed that, on average, users made **fewer mistakes** and **referenced the help menu** fewer times after being taught by the virtual assistant compared to learning from the documentation. See the chart below for the comparison.

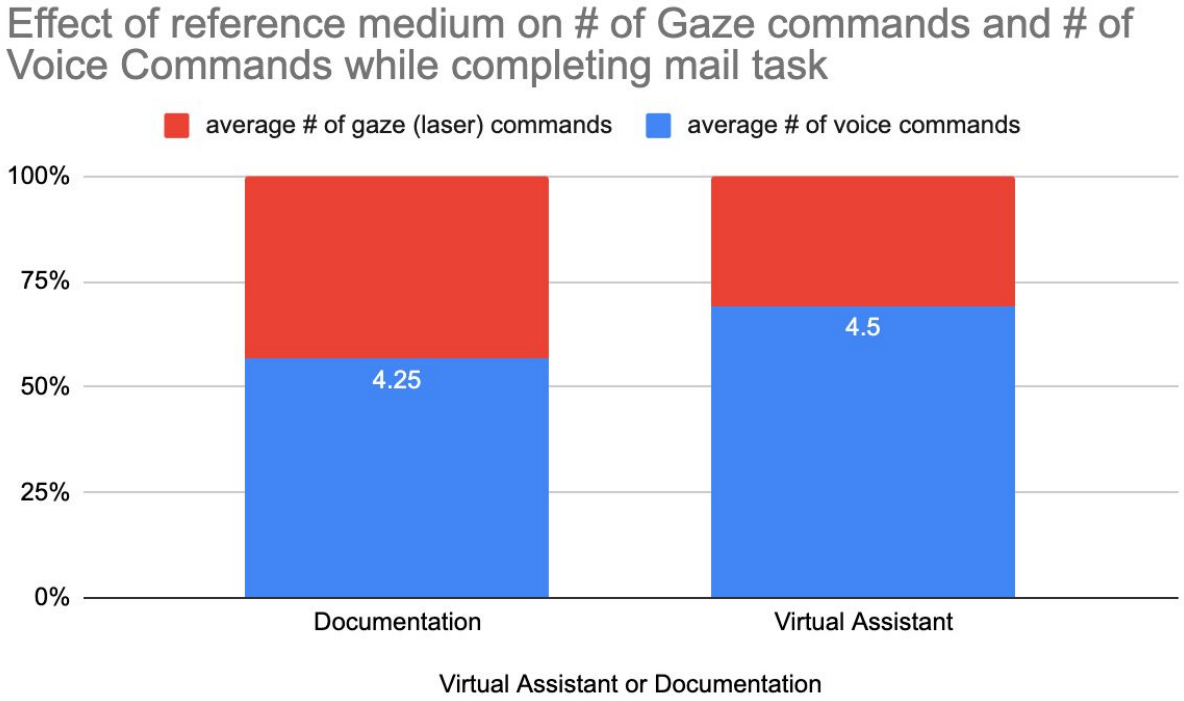


Second, we look at the results of the mail task. Surprisingly, we saw that the number of errors and the number of references to help were quite similar, as depicted in the graph below.

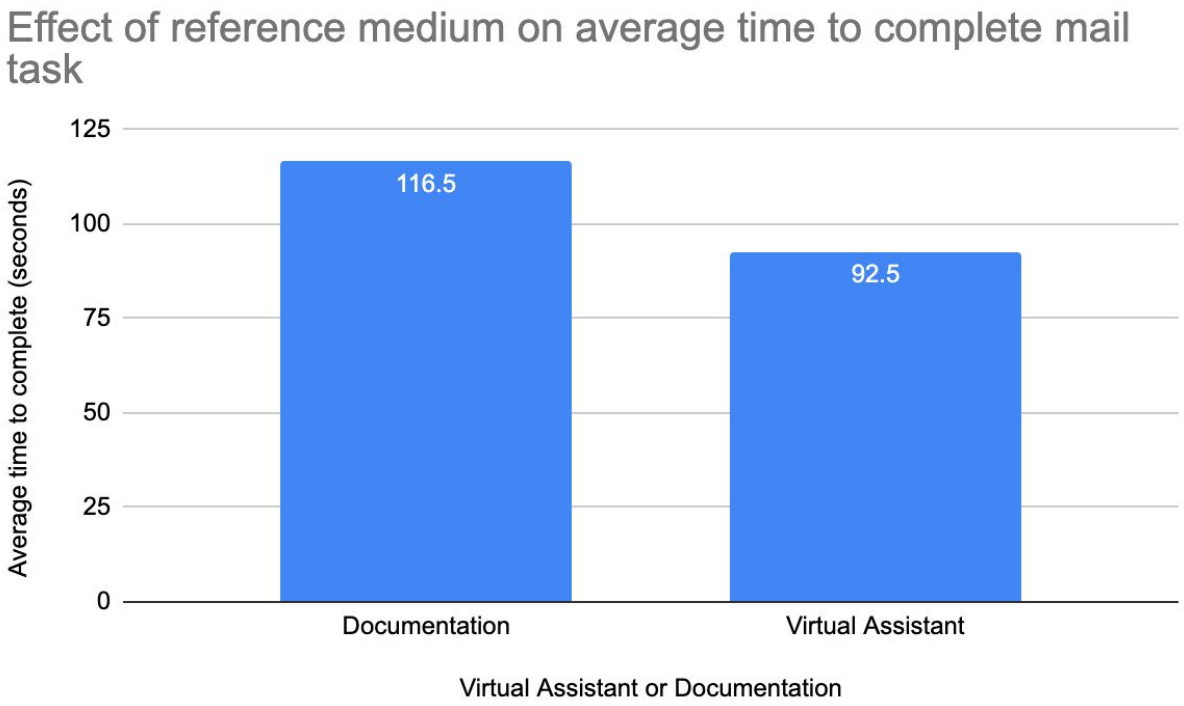


In fact, users who used the virtual assistant had slightly more references to help than users using documentation. The difference is small enough here that we say that the virtual assistant and documentation are on-par with each other.

However, users who used the virtual assistant while completing the mail task were more inclined to use voice commands than those using documentation. We saw the **percentage of voice commands increase from 59% to 72%**. This is a great sign because it signals that one potential avenue that developers of the Hololens could use to increase voice command usage is to introduce virtual assistants to help onboard users to their application.

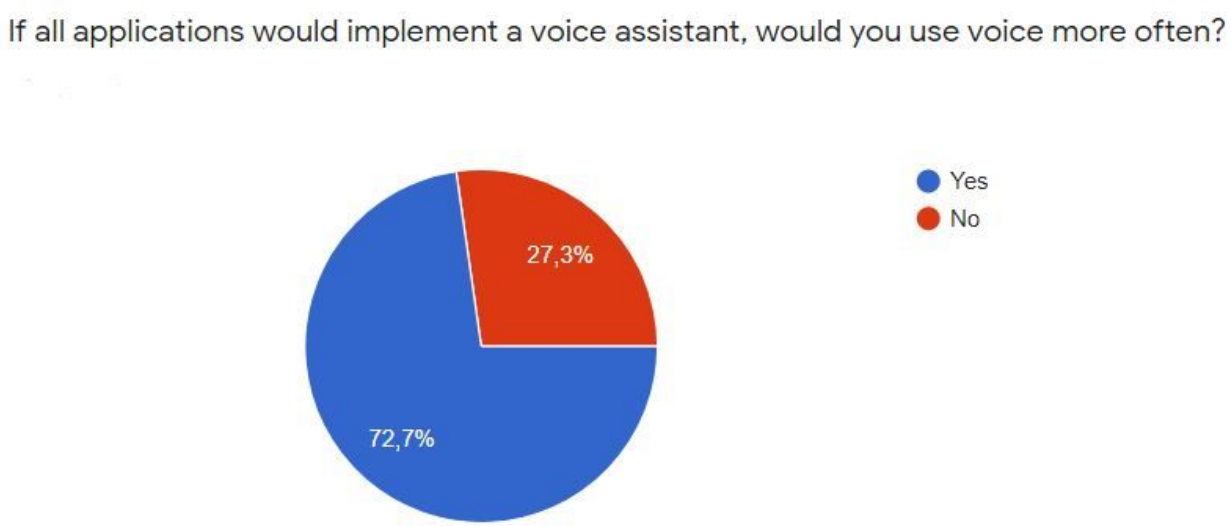

Effect of reference medium on # of Gaze commands and # of Voice Commands while completing mail task
average # of gaze (laser) commands
average # of voice commands
100%
75%
50%
25%
0%
4.25
4.5
Documentation
Virtual Assistant
Virtual Assistant or Documentation

In addition, the average amount of time to complete a task was lower with the virtual assistant. See the chart below.

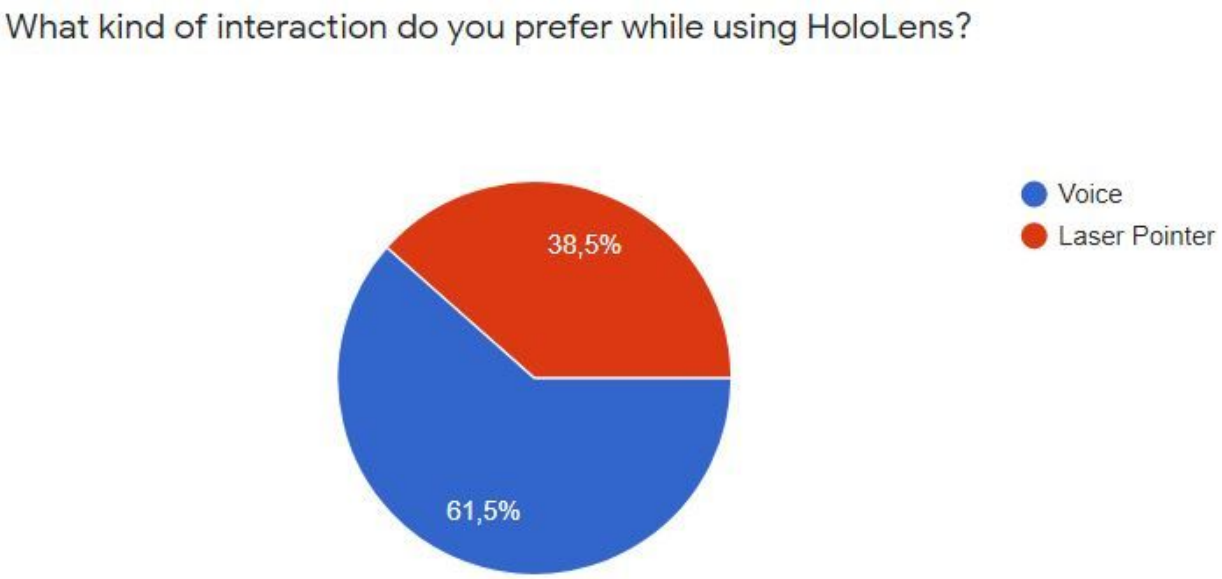


It is important to note that Cortana typically spoke longer than the user read the documentation for, and in future experiments, we would like to observe the relationship between how long the user is spoken to or how long they read the documentation on how quickly or how well they can complete a task.

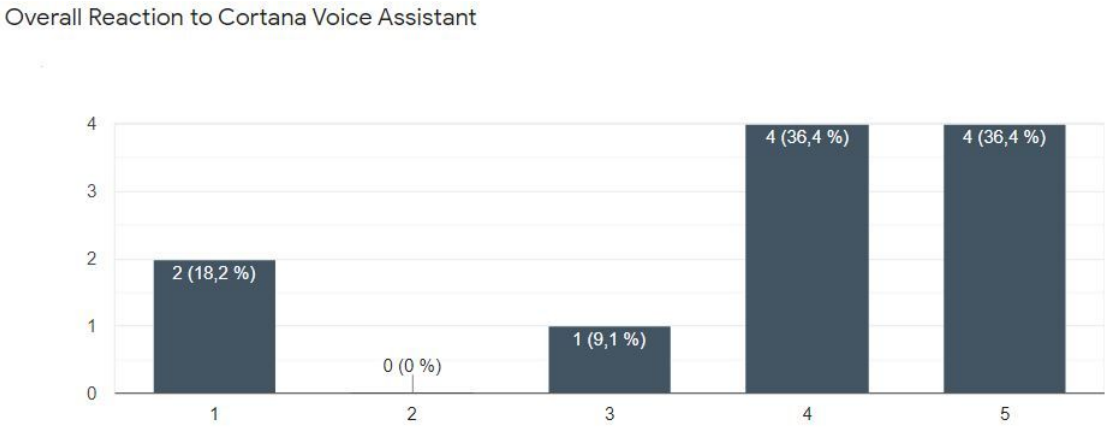
We also examined users' feelings toward the voice-enablement of applications. In the survey that we gave to test subjects we found that most **would use voice assistants more often** if all applications offered them:

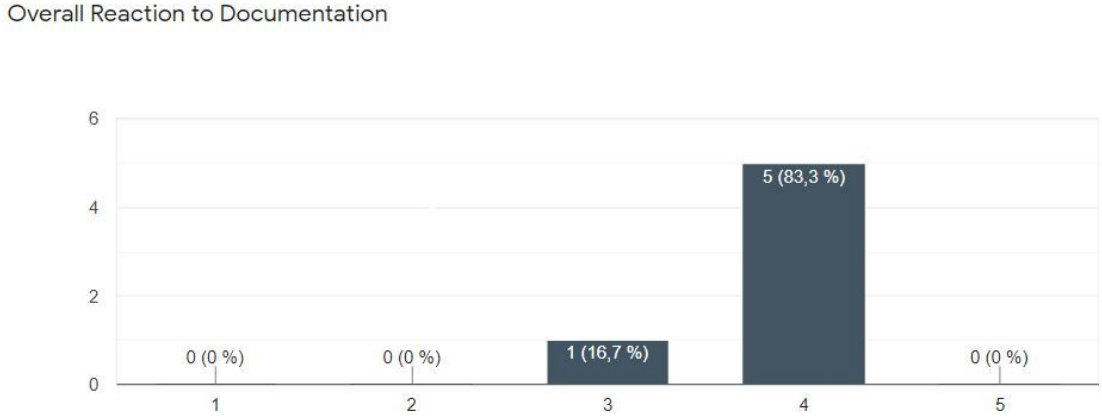


We also found that most users preferred using voice over pointing technology.



Users found the Cortana virtual assistant more interactive too, with most users rating interactivity with Cortana 5 / 5 in our survey, compared with 3.5 / 5 for documentation. Overall reactions to the Cortana voice assistant were more mixed than the overall reaction to documentation, as one can see pictured below.





Implications

Based on the results from our experiment, where we compared the learning curve of users playing a game and writing an email with voice and gaze on HoloLens, we concluded that with a virtual assistant like Cortana, it is more intuitive for people to learn how to use a new application with mostly voice interaction.

Most people don't use a virtual personal assistant and therefore it is an important task to first teach, how to properly interact and use the virtual assistant with the most benefit. Without the right voice commands for the virtual assistant, most people won't interact with it in the most efficient way or don't use it at all. There should be an introduction to the Cortana voice assistant at the beginning and after launching an application for the first time with HoloLens, an own introduction to that specific application, which can be reopened at any time.

For the future works, it was advised to implement synonyms for existing commands because one of the main problems was to remember the exact phrase to execute a specific command. As a consequence, users were seeking help; the testers of our experiment wanted more variations to call the voice assistant for help, for example, one could use the phrase "I have a question" in addition to "Hey Cortana". Implementations of voice input should, therefore, focus on the ability to use synonyms. This will increase acceptance of voice input as users respond with frustration when the system does not respond naturally.

Another feature to add is the modulation of the Virtual Assistant. Users complained about having to listen to the whole instruction of the assistant whenever they asked for help. Even when users are asking just for the keyword, Virtual Assistant would explain all the related things in detail. The fixed narrative style of the assistant may lead to users switching to documentation because they can spot the missing command faster. This feedback leaves us another research question: *"How to modulate the style of Virtual Assistant to make it informative and time-efficient at the same time?"*

In conclusion, our human-computer interaction experiment found that it was more intuitive to use voice input in general. It is more natural to just ask a human-like assistant for help than reading whole documentation just to find one simple command that could be found by the assistant in seconds.

Our experiment mostly reflects the things in our hypothesis and therefore it is a good idea to make use of a virtual assistant in HoloLens to teach people using an application with their most natural way of expressing themselves: voice.