# Qualitative evaluation:

We used a questionnaire based on a template from Gary Perlman[[1]](#footnote-1). According to our general research goal we wanted to compare a multimodal interface with a traditional controlled interface. Therefore we asked the user after he interacted with both of these interfaces to rate both of them. The questionnaire was divided into four topics. “Usefulness”, “Ease of Use”, “Ease of Learning” and “Satisfaction”. The first part should review the general effectiveness of the two interfaces according to the judgment of the user. In the second part the user should describe how easy it was to interact with the system. In the third part the user should rate and compare the two ways of interaction, regarding to how quick he got accustomed with the possibilities of them. And the final part should reflect the general satisfaction of the user.

The user was able to assign a scalar score from 1 to 7 (strongly disagree to strongly agree) to a bunch of closed questions. At the end a last question asked the user about his general preference for one or the other interface. To give the user the possibility to add some individual remarks about what he liked/disliked about each interface, we asked him to write down three positive and three negative aspects for each interface. Furthermore a general comment page allowed him to give even more feedback and his own opinion.

# Quantitative Evaluation

## Evaluation setup

As stated our set of users was rather small. Therefore we decided to do an evaluation within-group. Because our users where all within a range of age and intellect, where they already interacted a lot with mouse and keyboard, we first asked them to interact with the system by voice and gesture for two minutes without a specific task, to accustom them with voice and gesture interaction. Additionally we provided a set of instructions, which explained the functions of the application. We gave those instructions, because we wanted the application itself to be an independent variable. To diminish the learning effect, that occurs with a within-group setup, and the fact that the users were already accustomed to mouse and keyboard, we let them first interact with mouse and keyboard and then with gesture and voice.

Each user received the following set of tasks, he or she should accomplish with both interfaces:

1. Open the book “A History of Mathematics”.
2. In the book, search for the word “Gutenberg”.
3. Go to page 10.
4. Bookmark this book.
5. Bookmark the book “First Six Books of the Elements of Euclid”
6. Show all bookmarks

For later evaluation of the dependent variables we video-captured the screen and the user himself.

**Variables**

Independent variables:

* Application
* Computer setup
* Subject type (average skilled computer user)

Dependent variables:

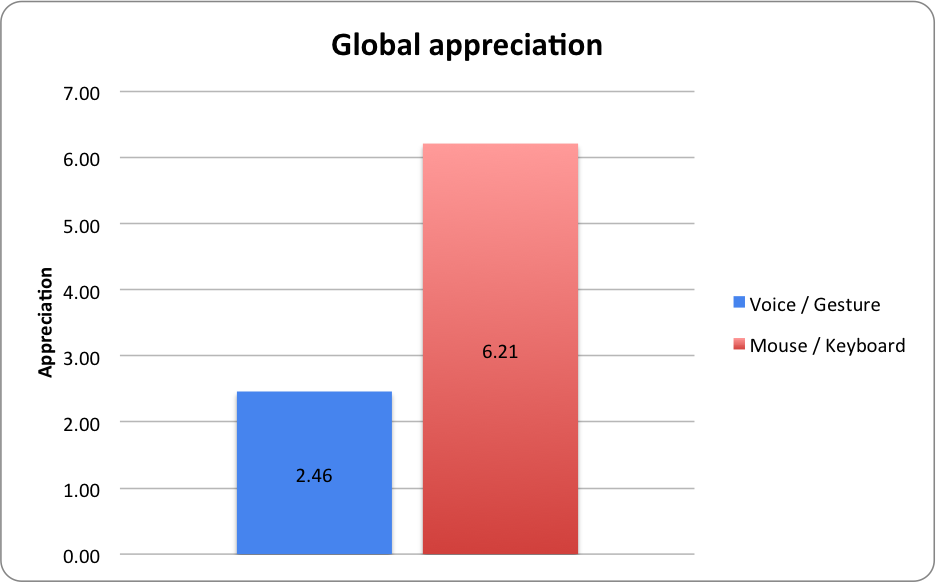
* Time to complete all tasks
* Number of steps (Number of mouse-clicks, number of mouse-drags, number of connected keys pressed[[2]](#footnote-2), number of words spoken, number of gestures, number of arm movements).
* Number of recognition errors[[3]](#footnote-3)
* Number of user related errors

# Evaluation’s results

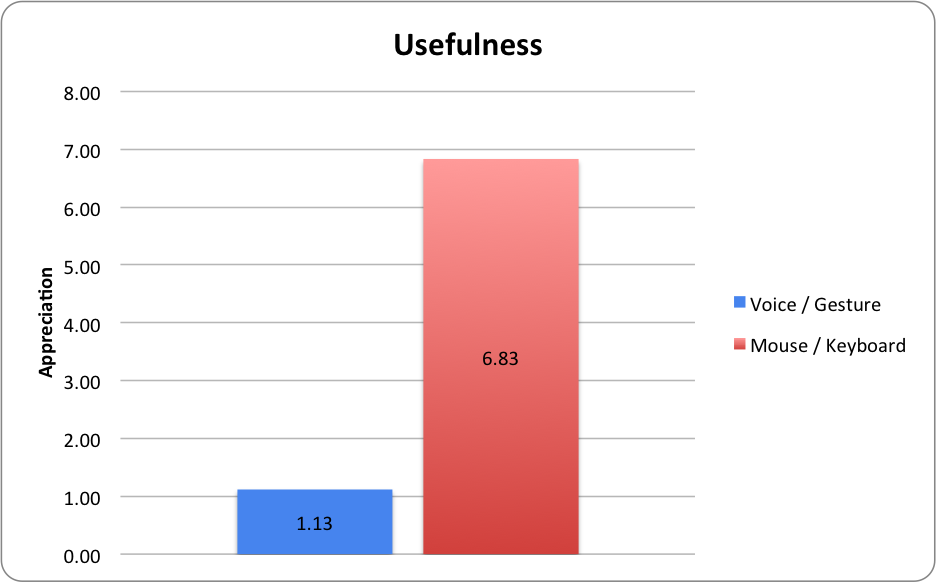
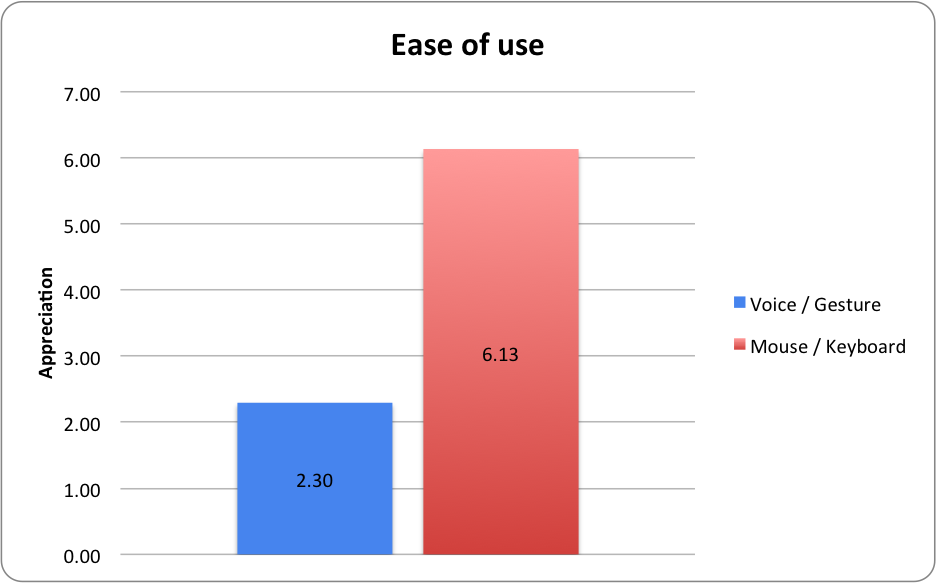
The evaluation was run on a group of six users, all within an age range of 20 to 60 years old. Before the evaluation they all knew how to use a computer and never had really used a Kinect or any gesture-and-voice interface previously. Our hypothesis was that keyboard and mouse interaction is more efficient than voice and gesture for the virtual library application.

## Qualitative results

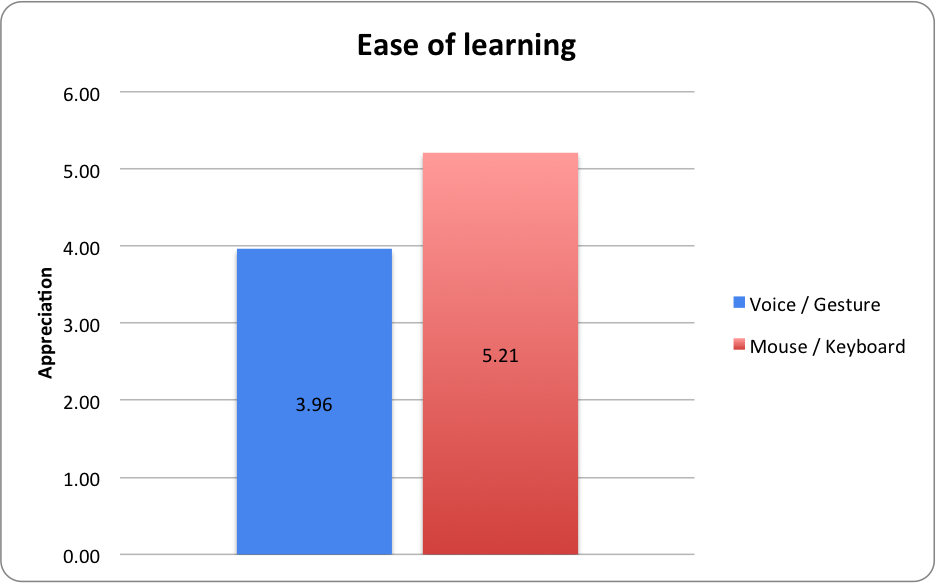
As we can see in [reference to „Global Appreciation“ graph], the classical set of modalities mouse/keyboard was widely preferred by the users. This result is not surprising, as we will see it in the following of the results.



The two next graphs [reference to „Usefulness“ and „Ease of use“ graphs] show us that the users found the mouse/gesture a lot more effective to work with, and easier to interact with. We can attribute this result to the fact that users were more comfortable with an interface they already knew well and use everyday.

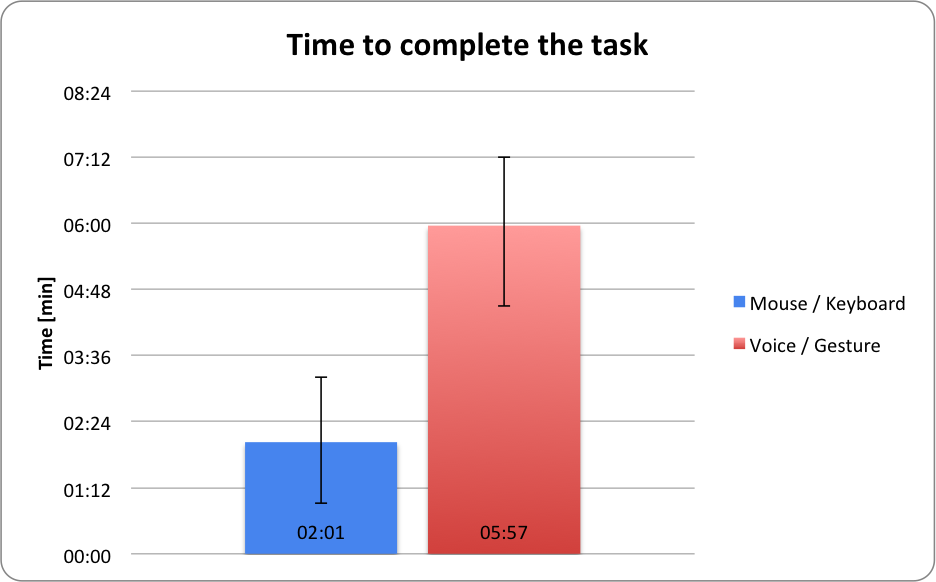
The two graphs below show an interesting contrast with the other qualitative results. While users preferred the mouse and keyboard interaction, they were confident in the fact that they would be able to learn to use the Virtual Library nearly as quickly with both interfaces. This result is encouraging, as it means that users do not feel too much difficulty in learning new modalities.



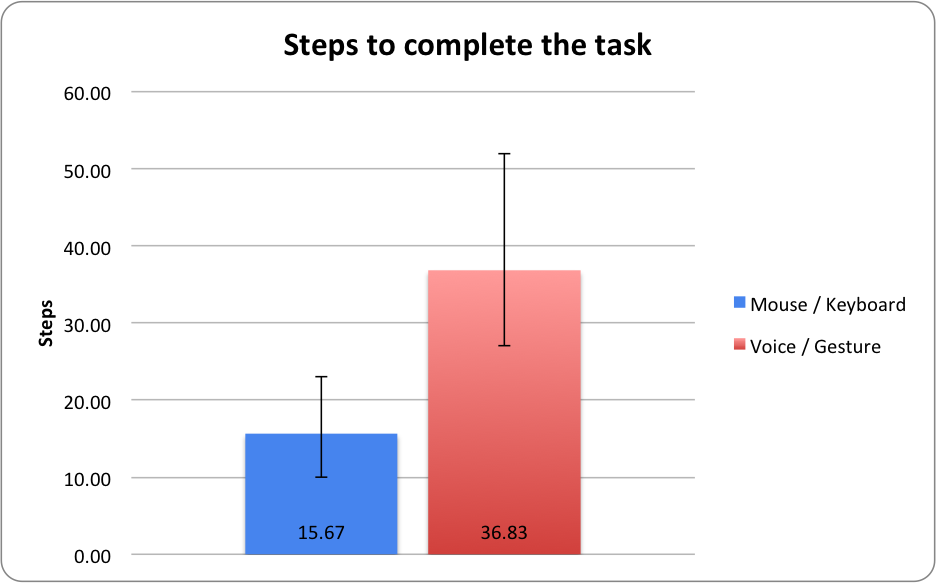
## Quantitative results

Our hypothesis was that since users were already very familiar to the use of a mouse and a keyboard, they would be more efficient in the task they would be given. The results we obtained, visible in [graph reference], confirmed our prediction: users completed the given task around three times faster with a mouse and a keyboard than with the voice and gesture commands.

With the qualitative results in mind, we think that by giving the users more time to learn how to use the voice and gesture interface, the time difference between both interfaces would decrease significantly.



The steps required by the user to complete the given task were also measured. In the following graph [reference to it] we can see that the voice and gesture interface required more steps.



According to us, this is mainly due to the recognition errors that occurred with the speech and gesture recognition. As we can see in the table below [cf. Reference to the table], in average 20.5 errors occurred for each user while using voice and gesture, and only 1 occurred with the mouse and keyboard interface.

Moreover, among these 20.5 errors, about 60% were due to the system itself (e.g. not understanding a word, missing track of the hand). If the ergonomics of the virtual library were improved, most of these errors would probably not occur. The remaining 8 errors can be imputed to the fact that the users were not familiar with voice and gesture systems, and did not know very well how to behave with it.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of interface** | Users errors (average) | Recognition errors (average) | Total (average) |
| Mouse and keyboard | 1 | 0 | 1 |
| Voice and gesture | 8 | 12.5 | 20.5 |

In conclusion of these results, we can say that our hypotheses revealed to be true: keyboard and mouse interaction is more efficient that voice and gesture for a virtual library application. One of the possible explanations is that the mouse and keyboard are widely used since decades, as voice and gesture systems just start to be adopted by consumers (e.g. Google Now, Apple’s Siri, Samsung TVs). The overall efficiency of gesture and voice commands would certainly increase if its ergonomics were improved accordingly to further user centered design evaluations (by giving the user more feedback, e.g.), and if more time was given to the users to get familiar with the voice and gesture system.

# Conclusion

According to the results of our evaluation, we should say that mouse and keyboard interaction is more efficient, useful, satisfying and precise than a multimodal way of interaction, as represented by voice and gesture interaction in our system.

But the low scoring of the voice/gesture-input can be traced back to several circumstances. First of all: the efficiency and usefulness of a modality is strongly dependent on the task it is used for, the feedback, which is given to the user and the environment it is used in.

Our main mistake was to try to use a voice-gesture-interaction with a classical GUI-interface. The user had to stand a certain distance to get recognized by the Kinect. It was impossible to read the small GUI-Elements from such a distance, therefore he had to get nearer and therefore his gestures could no longer be interpreted by the system. It would have been more successful to use also more channels for the output (e.g. synthetic speech).

Further problems were technical challenges. The visual tracking of the Kinect did not work if the user moved his limbs too quickly. Therefore he had to move slowly and because of that the efficiency of the system was reduced a lot.

A positive aspect of our gesture-voice-interface, which we could observe during evaluation, was the flexibility of redundant or equivalent modalities. User often switched from gesture to voice and vice versa, if the Kinect didn’t interpret the input correctly. Redundant or equivalent modalities are therefore a really useful instrument for error correction. We could also observe how users changed modalities if the situation asked for it. E. g. a user was interacting with the system through gesture, if his hands were occupied because he had to read the instructions, he immediately changed to voice-interaction.

As a final conclusion we can state, that multimodal interfaces can be more efficient, faster, flexible and less erroneous, but only if you implement them thoughtfully, with according feedback and with enough sensor-precision.

1. http://garyperlman.com/quest/ [↑](#footnote-ref-1)
2. E.g. You need 4 keys to write the word „five“, we counted those pressed keys as 1. [↑](#footnote-ref-2)
3. E.g. The user said the correct word or made the correct gesture, but the system didn’t recognised them correctly. [↑](#footnote-ref-3)