PICaFACE

Controlling camera features with facial expressions

Anja Bürger, Aynira Wawersig
Winter Semester 2017/18

Prof. Dr. Michael Kipp
Interaction Engineering
Augsburg University of Applied Sciences

Abstract

Taking pictures with the mobile device can get difficult if only one hand is available for the interaction, especially if the device is very large and the user wants to take quick snapshots. PICaFACE, a camera application prototype, introduces new interaction techniques, which integrate the performance of facial expressions to simplify the user's single-finger interaction. The zoom mode is triggered by pursing the lips and widening the mouth or tilting the head forwards and backwards. The release function is fired by raising the eyebrows or opening the mouth. The zoom center can always be defined with one touch onto the screen, which is simulated with a mouse click in PICaFACE. With a small user study the different zoom and release gestures have been analysed and compared to get hints about how efficient, intuitive, comfortable and accepted each technique turns out to be for the users. As a result mouth gestures were in general perceived as entertaining but not as very advantageous to improve the one-handed process.

Motivation

Nowadays it seems that many people exclusively take pictures with their mobile devices as the camera technologies improve permanently within their smartphones. However, in everyday life many situations occur where they can handle their devices with only one hand, e.g. if they have to hold on to the bar while standing in a bus or if they have to carry things around with them. In case they want to take spontaneous pictures in those situations some inconvenient interaction problems can occur *(Figure 1)*, e.g for zooming in and out a second finger is needed and it is often very difficult to reach the release button and hold the device steady at the same time. If the release button can still be reached though, in most cases the pictures will be blurry or out of frame and the focus might differ from how the user actually intended to set it *(Figure 2)*. With PICaFACE the user can easily take unique snapshots for example while holding a big ice cream cone in one and the smartphone in the other hand. With one tap onto the screen he or she can define the zoom center. By performing facial expressions the user can zoom and trigger the release. These functions relieve the hand which holds the device in a way so that the user can focus on an enhanced stability, because the hand can be held relatively still.

Figure 1: Cumbersome picture taking process with one occupied hand

Figure 2: Unintentional change of framing and blurring

Another motivation for developing the prototype was that people often use some kind of facial expression during a photo shoot anyway. This leads to the assumption that facial gestures would possibly be an accepted and useful method to control the camera. Finally the goal was to find out how comfortable, fast and easy it is for the user, to control an application by executing facial gestures compared to conventional interactions involving only touch input and how fluent the interaction flow of different technique combinations is perceived.

Related Work

The work was inspired by the paper Air+Touch: Interweaving Touch & In-Air Gestures[1], which presents a new class of interactions that combine in-air gestures and touch events to simplify single-finger interactions. The authors developed for example the possibility to define the zoom center by tapping onto the screen and then cycle in the air with the thumb to trigger the zoom mode. Although this method appears to be a smart solution for single-finger interaction problems, it could perhaps be a bit annoying for instance in case of using the camera, because the user might accidentally touch the screen by circling with the thumb and also occlude the view. Accordingly the idea came up that connecting touch events with facial gestures could possibly ensure an undisturbed process in situations like these.
Another related project is faceTYPE, which was developed by the students Alice Strunkmann-Meister and Rodrigo Blásquez[2]. With faceTYPE the user can format text in an editor by performing facial expressions and therefore avoids interrupting the writing flow. How the integration of this novel interaction technique could improve the fluidness of completing tasks encouraged the development of PICaFACE.

Technical Realisation and Setup

In our prototype we simulate the situation by using a laptop (MacBook Pro), its built-in webcam and a mouse *(Figure 3)*. Face gestures are captured and processed via an application called FaceOSC, which can communicate with other applications by the Open Sound Control (OSC) Protocol.[3] FaceOSC is based on OpenCV, a library which is often used to detect movement in general.[4] After receiving facial data the FaceOSC application sends it to the software sketchbook Processing[5] (release 3.3.6) through a network connection, which is build via the osp5 library for Processing. The interface for PICaFACE, which is written in Processing, shows different pictures of nice scenic or urban views *(Figure 4)*.
The face-mesh from FaceOSC is shown in an extra window. As soon as a face is recognized PICaFACE receives information about the facial data usually as float values. To realize the concept the values for the width and the height of the mouth, the height of the right and the left eye and eyebrow, as well as the information about the orientation of the head as a x,y,z-vector and more precisely the float value of the x-pose were implemented.

---

[1] Chen, Xiang „Anthony"/Schwarz, Julia/Harrison, Chris/Mankoff, Jennifer/Hudson, Scott E. (2014): Air+Touch: Interweaving Touch & In-air Gestures, in: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology. New York, NY, USA: ACM, S.519–525

[2] http://michaelkipp.de/interaction/projects/?/201516w/FaceType/
[3] https://github.com/kylemcdonald/ofxFaceTracker/releases
[4] https://opencv.org/
[5] https://processing.org/

Within the Processing software the calculation of the threshold values for each gesture is made to specify the correspondent zoom and release function. Mouse interactions resemble touch and are also integrated to give visual feedback and round the interaction flow within PICaFACE in terms of clicking, pressing and dragging the left and right mouse buttons. For the conventional version of the prototype a library called giCentre, which contains utilities to assist creating data visualisation sketches in Processing, is integrated.[6]

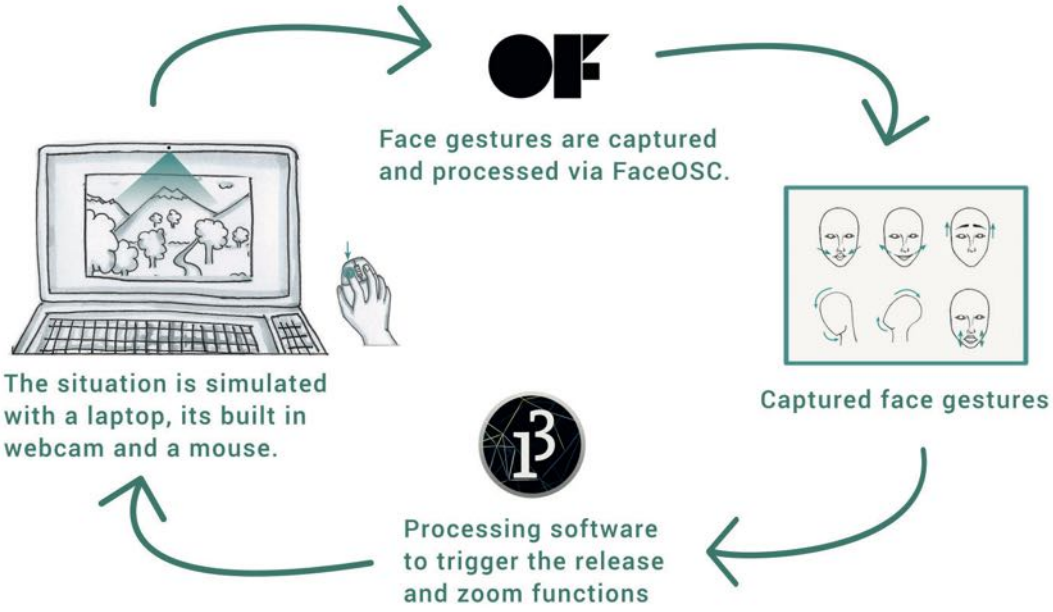

Figure 3: Technical setup of PICaFACE

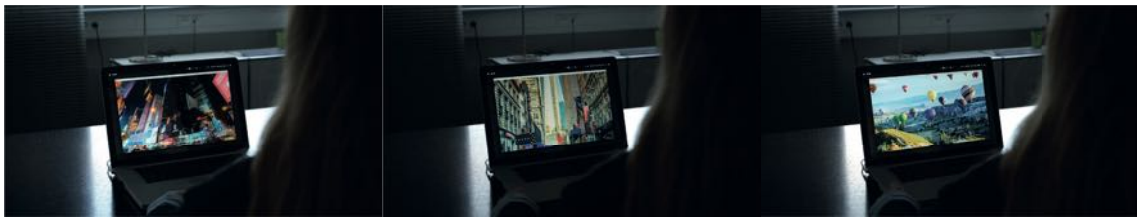

Figure 4: The three prototype interfaces

[6] https://www.gicentre.net/software/#/utils/

Interaction Techniques

To investigate the potential of the concept three interaction scenarios were implemented *(Figure 5)*. In every scenario the zoom center can be defined by clicking or pressing the left mouse button and the picture can be panned by dragging the mouse while the right mouse button is pressed.

functions	control condition: mouse	gesture scenario 1: mouth + eyebrows	gesture scenario 2: head + mouth
defining zoom center	left mouse button pressed	left mouse click	left mouse click
pan picture	right mouse button pressed + drag	right mouse button pressed + drag	right mouse button pressed + drag
zoom in	left mouse button pressed + drag up	small mouth	tilt head forwards
zoom out	left mouse button pressed + drag down	wide mouth	tilt head backwards
release	left mouse click on button	raise eyebrows	open mouth wide

Figure 5: Table of all interaction scenarios and functions

Control Condition

A conventional interaction scenario or so-called control condition *(Figure 6)*, which consists of pure mouse interactions, is designed to compare it to the features of the novel interaction versions, which include the detection of facial gestures. The control condition therefore represents the touch interaction with a mobile device. The user can zoom in and out by dragging the mouse up and down. By clicking a button on the screen the current frame is saved as a jpg-file to the sketch's folder.

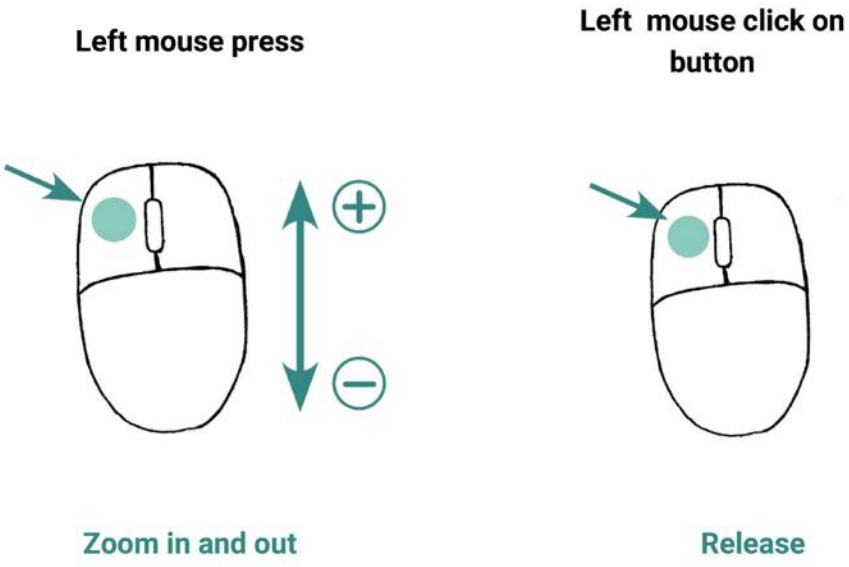

Figure 6: Control Condition: Mouse

Gesture scenario 1: Mouth + Eyebrows

Within the first gesture scenario the user can zoom in by pursing the lips and therefore making the mouth small or narrow and zoom out with a wide mouth. A picture is taken by raising the eyebrows *(Figure 7)*.

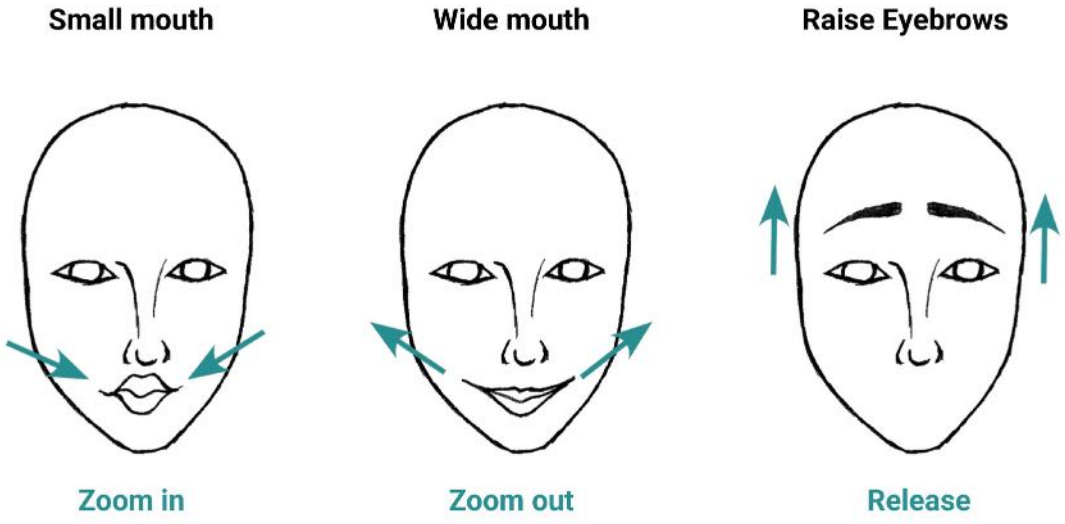

Figure 7: Gesture Scenario Mouth + Eyebrows

Gesture scenario 2: Head + Mouth

The other gesture scenario implies that the user can trigger the zoom function by tilting the head forwards to zoom in and backwards to zoom out. By opening the mouth the release is triggered and the current excerpt of the picture is saved *(Figure 8)*.

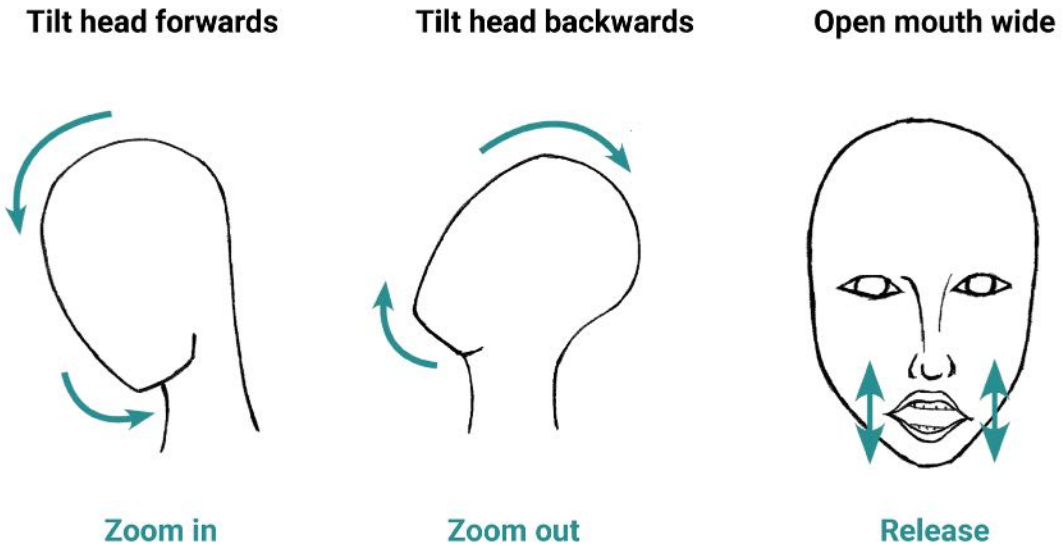

Figure 8: Gesture Scenario Head + Mouth

The choice of these facial expressions arouse from the fact that they could be detected very reliably by the system including a common built-in webcam and also, especially regarding the zoom functions, because they could be perceived as intuitive. Several sets of gestures were implemented to test which expression performs better and to observe how well each is being accepted by the user.

Evaluation

The three prototypes were tested by six test persons, three of them male and three of them female. All of them prefered using their smartphones for everyday snapshots. Every participant had to complete one task for each scenario. After they absolved all three tasks they also answered a questionnaire *(Figure 9)*, which contained open and closed questions in order to find out how intuitive, effective, comprehensible, easy and suitable for everyday life the different interaction techniques were regarded as.

Figure 9: Test scenario

In each prototype version a picture with two boxes was shown *(Figure 10)*. The task was to first zoom in to box 1, take a picture by triggering the individual release function, then zoom to box 2 and trigger the release once more. *Figure 11* Ideal image dimensions of a taken picture shows a taken picture in its best-case form. For each task the screen was recorded together with a video, captured by the webcam so that facial expressions could me observed next to the prototype interface. The pictures, which were taken by the test persons, were saved in individual folders. The order of the tasks was swapped so that every task was at each position two times to counterbalance learning effects and to spread the initial excitement of the participants, which could have influenced their performance within the first task.

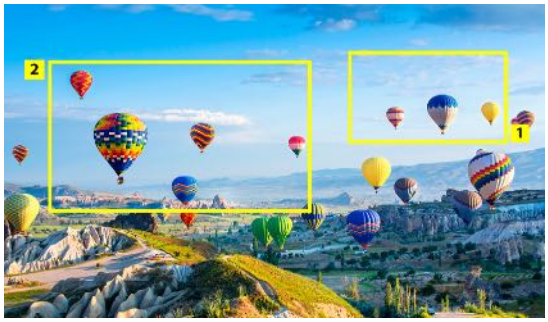

Figure 10: One of the pictures for the tasks

Figure 11: Ideal image dimensions of a taken picture

Metrics

By analysing the screen recordings the task completion times (total time) of each test person and for each task could be measured. It is defined by the time span from starting the program till triggering the final release. Within three categories the error rates for the zoom and the release function were counted. The total error rate therefore is the summation of first, all accidental performed gestures which triggered a function, secondly, the intentionally performed wrong gesture caused by confusion and thirdly, the intentional performance of the right gesture which wasn't detected. The accuracy of each scenario was measured by counting the amount of zoom changes needed to fulfil the task. By comparing the dimensions of the saved images from the test persons with the ideal version of the excerpt inside the box and calculating the amount of false pixels in Photoshop the precision of each interaction technique could furthermore be evaluated. Finally the questionnaire gave precious insights about subjective opinions.

Results

The results collected from six persons can barely be considered as representative but they give precious hints on how PICaFACE could be improved. The measurements being conducted with automated tools show that the average completion times for each interaction scenario don't differ very much, but still the combination of head (zoom) and mouth (release) gestures took a little bit longer than the other two scenarios *(Figure 12)*. The total error rates for the zoom and the release functions reveal that the participants had a higher error rate with mouth gestures in general *(Figure 13)*.

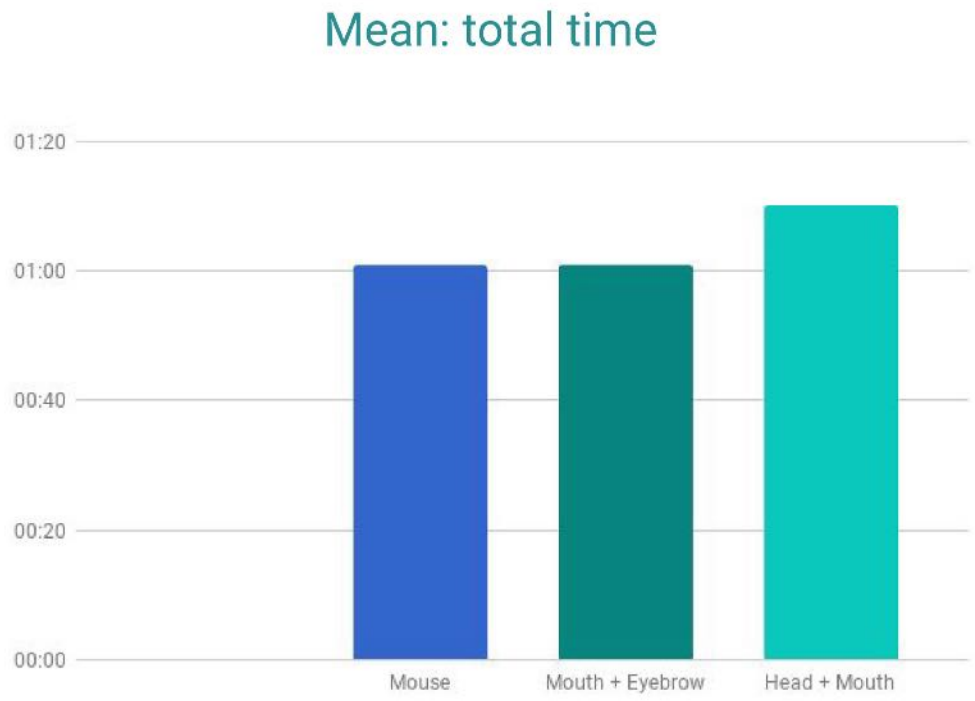

Figure 12: The average completion time for all scenarios in minutes

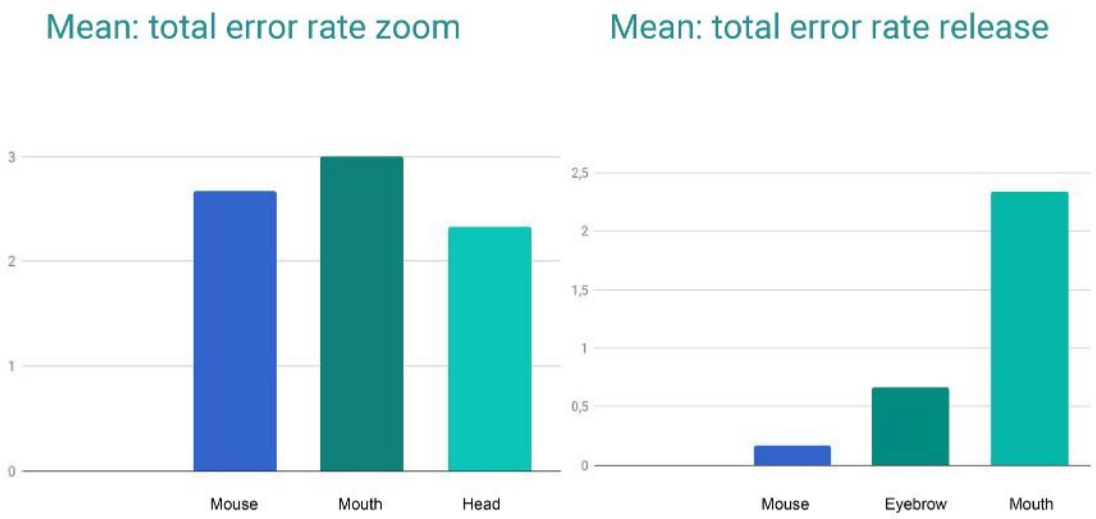

Figure 13: The average of errors for zoom and release

A mean value for the efficiency concerning the precision of each scenario was calculated by dividing the percentages of correct pixels within the test persons' screenshots into the completion times in seconds *(Figure 14)*. As a result it can be observed that there aren't remarkable differences. Evidence to this fact is also the outcome of conducting t-tests with the data sets of all measurements. There's only one significant difference between the total release error rates of clicking a button to take a picture in the control condition and triggering the release with opening the mouth wide in gesture scenario 2. Thus the p-value of 0,048 proves that the click-interaction is significantly more reliable than opening the mouth.

Figure 14: Mean values for the efficiency concerning the precision of each scenario

The results of the questionnaire additionally illustrate that although all mouse interactions are still perceived as the most intuitive *(Figure 15)*, some participants really liked the interaction by raising the eyebrows or tilting the head *(Figure 16)*. If they had the choice, two of the six test persons wouldn't use any of the facial gestures in public but if all participants had to choose among the gestures, they would again prefer raising eyebrows and tilting the head *(Figure 17)*. Every facial expression was rated as being more entertaining compared to conventional mouse interactions.

During the test it was obvious that some participants felt really uncomfortable to open their mouth. Moreover, two persons had great problems with the gesture scenario 1, because the release expression (raising eyebrows) also triggered the zoom function (small and wide mouth). One person had the same problem with gesture scenario 2 (release: open mouth, zoom: tilt head).

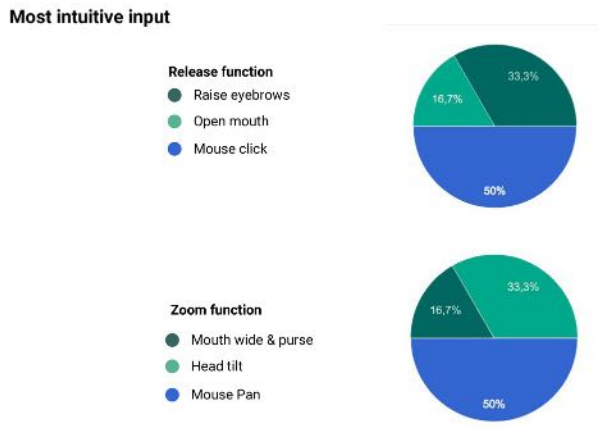

Figure 15: Most intuitive input

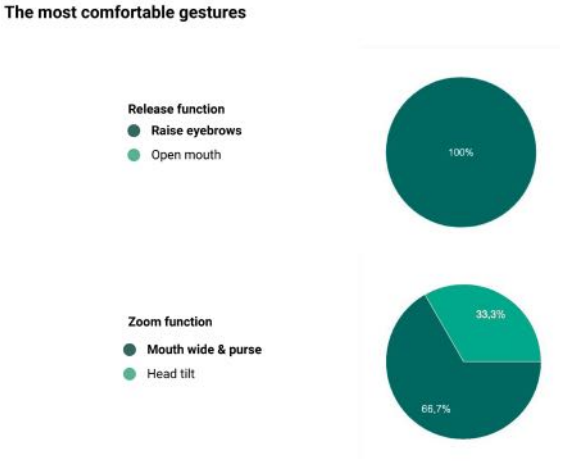

Figure 16: Most comfortable gestures

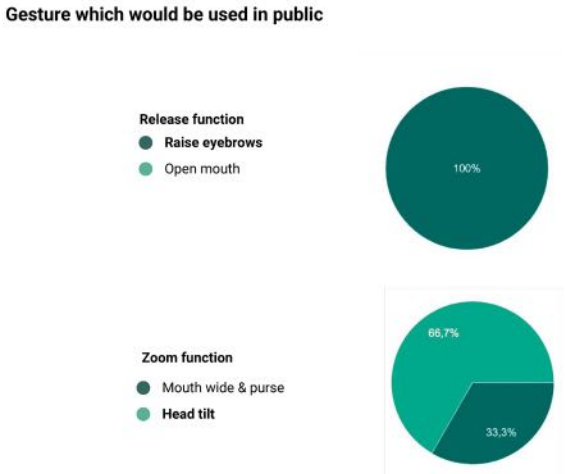

Figure 17: Gestures which would be used in public

Conclusion

Finally the user study revealed that facial expressions by using the mouth might be fun but their usage wouldn't be accepted in public and is not being perceived as very convenient. Whereas raising the eyebrows and tilting the head seem to be interaction techniques the users feel more comfortable with. This fact can also be read from four out of six answers the test persons gave to the open question what they liked best. For example one person mentioned:

> "I liked the function by nodding the head best. Combined with taking the picture by raising the eyebrows and further refinements, perfect!"

In five out of six answers to the question what the participant didn't like at all a mouth gesture was mentioned, for example one person commented the displeasure like this:

> "rip mouth open because of dentist feeling"

After the experience of the mutual influence of the zoom and the release function during the test it might be useful to include a clutch mechanism for either the zoom or the release function to avoid the blending of both mechanisms. Maybe integrating proximity attributes to achieve this improvement could be content for further research. Another helpful feature for the user would be to control the pace of zooming by performing the zoom expression less or more intensely.

Appendix

Questionnaire the participants had to fill in after testing the prototype versions:

Picture capturing with PICaFACE

With the following anonym questionnaire we want to find out how you individually perceived the different interaction techniques.

General informations

1. **Which is your preferential device to take daily pictures.**
   Markieren Sie nur ein Oval.
   - ◯ Digital camera
   - ◯ Smartphone
   - ◯ Tablet
   - ◯ System camera/ single-lens reflex camera
   - ◯ Other

2. **In order to take pictures with the smartphone two hands are needed.**
   Markieren Sie nur ein Oval.

	1	2	3	4	5	
disagree	◯	◯	◯	◯	◯	agree

3. **With a mobile device it's difficult to take sharp pictures with only one hand.**
   Markieren Sie nur ein Oval.

	1	2	3	4	5	
disagree	◯	◯	◯	◯	◯	agree

The different interaction combinations

4. **Completing the task exclusively with the help of mouse operations was easy.**
   Markieren Sie nur ein Oval.

	1	2	3	4	5	
disagree	◯	◯	◯	◯	◯	agree

5. **It's easy to accomplish the activity by zooming with the mouth and triggering the release with the eyebrows.**
   Markieren Sie nur ein Oval.

	1	2	3	4	5	
disagree	◯	◯	◯	◯	◯	agree

6. **The combination of head tilt (zoom) and opening the mouth (release) is easy to perform to finalize the task.**
   Markieren Sie nur ein Oval.

	1	2	3	4	5	
disagree	◯	◯	◯	◯	◯	agree

7. **Which task could be performed the most fluent in your opinion?**
   Markieren Sie nur ein Oval.
   - ◯ Pure mouse interaction
   - ◯ Combination of mouth (zoom) and eyebrow (release) gestures
   - ◯ Combination of head tilt (zoom) and mouth (release) gestures

The Zoom Gestures

8. **Which zoom function is the most intuitive for you?**
Markieren Sie nur ein Oval.

- ◯ Mouse (pressing mouse button while dragging up and down)
- ◯ Mouth (small, narrow mouth and wide mouth)
- ◯ Head (tilting head forwards and backwards)

9. **Which gesture is more comfortable?**
Markieren Sie nur ein Oval.

- ◯ Mouth (small, narrow mouth and wide mouth)
- ◯ Head (tilting head forwards and backwards)

10. **Which action is the most exhausting?**
Markieren Sie nur ein Oval.

- ◯ Mouth (small, narrow mouth and wide mouth)
- ◯ Head (tilting head forwards and backwards)
- ◯ Mouse (pressing mouse button while dragging up and down)

11. **Which modality zooms the most precisely?**
Markieren Sie nur ein Oval.

- ◯ Head (tilting head forwards and backwards)
- ◯ Mouse (pressing mouse button while dragging up and down)
- ◯ Mouth (small, narrow mouth and wide mouth)

12. **Which technique was the most fun?**
Markieren Sie nur ein Oval.

- ◯ Mouse (pressing mouse button while dragging up and down)
- ◯ Head (tilting head forwards and backwards)
- ◯ Mouth (small, narrow mouth and wide mouth)

13. **Which zoom method worked worst?**
Markieren Sie nur ein Oval.

- ◯ Mouse (pressing mouse button while dragging up and down)
- ◯ Mouth (small, narrow mouth and wide mouth)
- ◯ Head (tilting head forwards and backwards)

14. **Which zoom function is the most comprehensible?**
Markieren Sie nur ein Oval.

- ◯ Head (tilting head forwards and backwards)
- ◯ Mouse (pressing mouse button while dragging up and down)
- ◯ Mouth (small, narrow mouth and wide mouth)

15. **Which interaction technique do you prefer all in all?**
Markieren Sie nur ein Oval.

- ◯ Mouth (small, narrow mouth and wide mouth)
- ◯ Mouse (pressing mouse button while dragging up and down)
- ◯ Head (tilting head forwards and backwards)

16. **Which facial expression would you use in public?**
Markieren Sie nur ein Oval.

- ◯ Mouth (small, narrow mouth and wide mouth)
- ◯ Head (tilting head forwards and backwards)
- ◯ Neither

17. **If you had to choose, which facial expression would you use in public?**
Markieren Sie nur ein Oval.

- ◯ Mouth (small, narrow mouth and wide mouth)
- ◯ Head (tilting head forwards and backwards)

The Release Gestures

18. **Which release function is the most fun?**
Markieren Sie nur ein Oval.

- ◯ Click the release button
- ◯ Raise eybrows
- ◯ Open mouth wide

19. **Which method is the most intuitive in your opinion?**
Markieren Sie nur ein Oval.

- ◯ Click the release button
- ◯ Open mouth wide
- ◯ Raise eybrows

20. **Which release modality is the most exhausting?**
Markieren Sie nur ein Oval.

- ◯ Click the release button
- ◯ Open mouth wide
- ◯ Raise eybrows

21. **Which release did work worst?**
Markieren Sie nur ein Oval.

- ◯ Raise eybrows
- ◯ Click the release button
- ◯ Open mouth wide

22. **With which release function can pictures be taken the most precise?**
Markieren Sie nur ein Oval.

- ◯ Open mouth wide
- ◯ Click the release button
- ◯ Raise eybrows

23. **Which technique is the most comprehensible?**
Markieren Sie nur ein Oval.

- ◯ Raise eybrows
- ◯ Click the release button
- ◯ Open mouth wide

24. **Which gesture is more comfortable?**
Markieren Sie nur ein Oval.

- ◯ Raise eybrows
- ◯ Open mouth wide

25. **Which method do you prefer all in all?**
Markieren Sie nur ein Oval.

- ◯ Click the release button
- ◯ Raise eybrows
- ◯ Open mouth wide

26. **Which facial expression would you use in public?**
Markieren Sie nur ein Oval.

- ◯ Raise eybrows
- ◯ Open mouth wide
- ◯ Neither

27. **If you had to choose, which facial expression would you use in public?**
Markieren Sie nur ein Oval.

- ◯ Raise eybrows
- ◯ Open mouth wide

Conclusions

28. **As a whole, what did you like best and why?**

29. **What didn't you like at all and why?**

30. **What could be improved or realized differently?**

About you

31. **I'm absolutely up to date with modern technologies.**
Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	◯	◯	◯	◯	◯	◯	◯	strongly agree

32. **What's your age group?**
Markieren Sie nur ein Oval.

- ◯ 18 and younger
- ◯ 19-24
- ◯ 25-34
- ◯ 35-44
- ◯ 45-54
- ◯ 55 and older

33. **Gender**
Markieren Sie nur ein Oval.

- ◯ male
- ◯ female

Thank you very much for your participation!!