IT UNIVERSITY OF COPENHAGEN

# Hand-Gesture-Based Interaction in Hybrid Meetings


Authors:

**Lucas Frey Torres Hanson – luha@itu.dk**

**Mads Aqqalu Roager – mroa@itu.dk**

Supervisor:

**Fabricio Batista Narcizo – narcizo@itu.dk**



**STADS code:**

BIBAPRO1PE



IT University of Copenhagen

Copenhagen, Denmark



May 15th 2024



## Abstract

Hybrid meetings, which include remote and in-person participants, pose unique challenges in managing virtual meeting controls. Traditional methods requiring physical interaction with the hosting device disrupt the meeting flow. This thesis investigates whether hand-gesture-based interaction can provide a seamless solution, improving efficiency and user satisfaction. We developed a Minimum Viable Product utilizing gesture recognition to control meeting settings like volume and camera toggling in Zoom. Using a gesture elicitation study and A/B testing with 206 participants total and an average statistical power of 0.734, we identified popular gestures for specific tasks. Our findings show, that there are no natural gesture commands for meeting controls among university students based in Denmark. However, the study proposes important points to consider when creating a vocabulary, such as reversible gestures, analogous interfaces, dynamic gestures, and visual feedback.

# Contents

# Chapter 1

# Introduction

## 1.1 Background and Motivation

New challenges arose after the global shift to remote work during the pandemic. Virtual and hybrid meetings have become highly popular in corporate and educational environments. A particularly notable difficulty is during hybrid meeting scenarios, where some participants are physically present while others join remotely. The source of friction emerges when individuals physically present need to adjust the meeting, such as volume control and other system settings on the hosting device. This challenge disrupts the meeting flow, requiring one to either physically approach the device for the adjustments or relay the information to the one managing the device.

In recent years, using hand gestures as a means of interaction has expanded beyond traditional domains. Interacting with drones, robotics, 3D virtual objects, and vehicles or facilitating communication through sign language highlights the versatility and efficiency of gesture-based communication across different domains. Despite these advancements, the virtual and hybrid meeting domain has yet to address problems such as controlling the meeting scenario settings.

Addressing the friction and ensuring a smoother hybrid meeting experience is at the core of this project. A way to reduce this friction is to enable hand-gesture-based interaction.

The project aims to develop a minimum viable product, MVP, that will contribute to solving this problem. This extension will enable hand-gesture-based interaction with the virtual meeting client, Zoom. The MVP will use a pre-existing machine-learning model. The key focus will be to address the friction felt during hybrid meetings. We will strive for a good user experience by considering multiple articles and basing the final dictionary on a gesture elicitation study and an A/B test.

In this project, we will research and document findings in natural gestures, meaning users find the gestures natural given a task description suitable for specific interactions to manipulate the configuration of the virtual meeting gesture-to-task dictionary. We will develop an extension to modify the operating system volume, toggle the camera, mute/unmute the microphone and leave meetings through intuitive hand-gesture-based controls. We will research and collect data about the natural hand gestures for specific actions in virtual meetings. The MVP will be built around the data, as it will specify which gestures are relevant for which actions.

By incorporating natural hand gestures, the project aims to improve the user experience in hybrid meeting scenarios, reducing friction and increasing focus. Additionally, the data analysis will explore the correlation between hand gestures and tasks in hybrid meeting scenarios and explore how to create a gesture-to-task vocabulary.

### Hypothesis

Within the context of hybrid and virtual meeting scenarios, we hypothesize there exists a set of hand gestures intuitively associated with specific meeting controls, such as volume adjustments and camera toggling, by a significant majority of users from diverse backgrounds, indicating a universally recognizable gesture-to-task vocabulary.

### Research Question

Is it possible to make gesture commands for common controls for hybrid and virtual meeting scenarios, using gesture elicitation and an A/B test with a diverse international group of university students, that individually will have a consensus of at least 80%?

## 1.2 Objectives

The research paper will have the following objectives:

1. Create a large data set based on a gesture elicitation study. It will include participants' demographics, elicited gestures, and feedback.
2. Conduct an A/B test to maximize the precision of the proposed gestures. It will include participants' demographics and the gestures they prefer.
3. Create a gesture set that is based on the results from the A/B test that can be used for the MVP to interact with.
4. Create an MVP on the gesture set that will enable users to interact with hand gestures with a virtual meeting client.
5. Conduct a usability test that uses the created MVP. It will include the participants' feedback on using the MVP.

## 1.3 Contributions

The thesis will provide a significant contribution of a large data set, named "GestureVerse: Exploring the Multiverse of Interactive Gestures". The dataset will be curated based on the objectives outlined in Section 1.2, It encompasses insights gathered from a large dataset of video recordings from the elicitation study. Moreover, it incorporates feedback from 102 individuals from the elicitation phase, followed by 104 responses from the A/B test, and finally insights from 5 participants in the usability test.

In addition, the thesis will provide a framework that lays the groundwork for future research in this domain. The framework serves as a valuable stepping stone for further exploration and analysis.

## 1.4 Thesis structure

Chapter 1 introduces the background and motivation for conducting this study. It describes the problem statement in the form of a hypothesis and research question, as well as the objectives and contributions of the study.

Chapter 2 describes work related to the study that explores gesture recognition in different technologies, as well as frameworks for elicitation and evaluation of vocabularies.

Chapter 3 introduces relevant theory to the study and describes important concepts such as Human-Computer Interactions, gestures, and virtual meeting technology.

Chapter 4 describes the methods used for the experiments conducted in the study and how we conducted a literature review, gesture elicitation, A/B test and usability test. It introduces relevant data analysis methods and formulas for evaluating results.

Chapter 5 introduces the Minimal Viable Product produced with the gesture set extracted from the data set. It gives an overview of the program architecture, as well as the machine learning model. Lastly, it describes relevant future extensions.

Chapter 6 presents the findings from the gesture elicitation study, the A/B test, and the usability test together with statistical analysis including agreement rates, correlation matrices, and PCAs.

Chapter 7 discusses the implications of the findings for gesture interactions in hybrid meetings, as well as the research papers' limitations and in-depth suggestions for future work.

Chpater 8 summarizes the key points of the research study. It discusses the work's relevance and offers suggestions for future work. Finally, the gesture-to-task vocabulary based on the data set, "GestureVerse: Exploring the Multiverse of Interactive Gestures", is presented visually.

# Chapter 2

# Related Work

Numerous studies have explored gesture recognition in contexts such as virtual reality, robotics, and vehicles. However, their application in hybrid meeting scenarios remains underexplored. This chapter reviews key technologies and methodologies in gesture recognition, providing a foundation for our research as well as related publications in the field of hand-gesture-based interactions of virtual or hybrid meetings.

In our literature review, we did not find any related work directly covering hand-gesture interaction in hybrid and virtual meetings, however, we found a few different publications interacting with other interesting and related interfaces. He *et al.* [1] proposed a hand-gesture set for controlling a smart kitchen environment. The study covered the procedure of eliciting gestures for the system, which involved two experiments. In the first experiment, they showed the users a task in a smart kitchen and then asked the users to come up with two gestures for said task. They extracted frequently chosen gestures from the first experiment and evaluated them in a second experiment in terms of ease of execution, joy of execution, memorability, and good match on Likert scales. The study presented the final elected user-defined gesture set, which included gestures for turning the range-hood on and off.

Another technology covered by Löcken *et al.* [2] is a hand-gesture set for controlling music playback. They elicited the gestures within a process focusing on constant user feedback. The process was composed of analysing the referents relevant to gesture mapping and then asking users to perform gestures for each referent. These gestures were then formalised and used to define consistent gesture sets. The defined gestures were finally evaluated with Likert scales in terms of simplicity and intuitiveness and improved to compare results.

Pereira *et al.* [3] elicited gestures for laptop interaction. They conducted an elicitation study on 30 participants. This resulted in 1300 gestures generated based on interaction with a laptop. This involved many general interactions with a laptop, with tasks like *shrink*, *enlarge*, *open*, *close* etc. They analysed the results and then reduced the number of gestures to 84. They then rated the vocabulary based on preference, usability, and effort. They proposed a vocabulary containing 13 distinct hand gestures.

Wu *et al.* [4] and Ali *et al.* [5] conducted elicitation studies on the potential of demographic and confirmation bias by priming affecting the natural gestures. Wu *et al.* [4] focused on comparing Chinese and American participants and found a significant difference, and that gesture design is culturally biased for some tasks, and for others not. The domains researched were both set in vehicles, TVs, and VR. Ali *et al.* [5] focused on video-related controls and had three groups: a control group, a creative mindset group, and a sci-fi group. This resulted in gestures that were significantly different compared to the non-primed group.

Wu *et al.* [6] also worked with gestures within the TV domain. They conducted a four-stage study. The first stage was centred on what functionality the users needed to interact with gestures in the TV domain. In the second stage, they had a user elicitation based on the features that were deemed needed from stage one. The third stage was about evaluating the candidate gestures. Lastly, they let the end users design their own gestures for specific tasks for a personalised experience.

Some publications do not elicit gestures in their own research, but instead gather other people's findings and compile them together. Vatavu *et al.* [7] explored 12 existing gesture elicitation studies of gestures used by people with visual impairments. They compiled the collected gestures into a new set for TVs, smartphones and tangible user interfaces. Likewise, Villareal-Narvaez *et al.* [8] also conducted a systematic literature review of gesture elicitation studies. They reviewed a total of 267 studies and presented a set of gestures with many associations with the most associated referents mapped to them. The referents were all generalised and can be used in many different systems. A set of recommendations and guidelines for gesture elicitation studies was then proposed, including the importance of assigning every referent to a single action and collecting one gesture per referent per participant.

Some researchers investigated gesture interaction more concretely connected to specific pieces of software. Koh *et al.* [9] conducted a user elicitation study and the development of a hand gesture recognition system that can map symbolic hand movements to analogous emojis. After the user elicitation study, they asked the participants to rate the different gestures suggested for the emojis. Finally, they asked the participants to link each elicited gesture to an emoji.

Parthiban *et al.* [10] elicited gestures for their own piece of software. They presented a new interface called Large User Interface, LUI, that coordinates gestural and vocal interaction to increase the range of possible dynamic interactions on large displays. After they showed the participants the different predefined gestures, they then asked the participants to conduct specific tasks. They attained an accessible and learnable vocabulary to interact with large displays.

Andolina *et al.* [11] worked with gestures for a more general use case. They presented the results of two experiments regarding the general digitisation of the manufacturing sector. One experiment was about designing one-handed gestures for this environment, the other was about cursor feedback. They experimented with hand gestures with a specifically designed application, with no real-world use. Still, it had general features based on real-world web applications in the manufacturing sector. The experiment asked users to perform one gesture per interaction in the presented application. Afterwards, they formalised the most common gestures into a gesture set and evaluated them in terms of guessability and agreement. Agreement rates ranged between 38% and 90%.

Pulina *et al.* [12] also worked on a mock-up level, but they conducted a user elicitation study to support interactions across multiple device formats. Later, they reviewed the gestures. They did the elicitation online and with a mock-up UI to provide context for the actions. 55 individuals participated. The domain they researched was specifically cross-device applications, meaning that the gestures would initiate a specific communication between two devices. The study resulted in a vocabulary that the participants rated with a mean of 3.1 and a median of 3.5, in how easy it was to learn on a Likert scale of 5.

Some researchers publish work that instead focuses on the process of eliciting gestures. Delamare *el al.* [13] conducted a study exploring memorability and recognition issues with one-to-one gesture-command mappings. They conducted an experiment in which they tested two different gesture sets. The first set was a standard one-to-one gesture-command set, and the second one was a set combining gestures with the use of both hands. The second, combining gesture set enabled the categorisation of gestures as an attempt to improve the memorability of the gestures. They found that combining gestures was preferred in their experiments when single gestures were abstract, and combining gestures improved memorability.

Uva *et al.* [14] instead proposed a user-centered framework for designing midair gesture interfaces. The framework consisted of, firstly, analysing the scenario. Then the gesture elicitation consisted of showing the users actions and asking them to propose a corresponding gesture. The next step was defining the vocabulary based on the results of the previous step. Finally, the vocabulary was validated and evaluated in an experiment on fatigue and memorability. Their agreement rates ranged between 25% and 54%.

In summary, a lot of work has been published in the field of hand gesture interaction with different technological interfaces, but not much of direct relevance to interaction with hybrid or virtual meetings. Instead, work has been published about interaction with TVs, laptops, and smart kitchen environments. But also for analogous emojis and more conceptual pieces of software. Common for most of these papers is, that they elicit gestures with a user-centred framework, where they first freely elicit gestures for given tasks, after which the most popular gestures are evaluated in some way, either a Likert scale or another measurement like memorability, guessability and agreement. Therefore some publications also cover formalising the process of eliciting gestures.

# Chapter 3

# Theoretical Background

This chapter provides a concise overview of relevant theory to understand the concepts discussed in the research study.

## 3.1 Fundamentals of Human-Computer Interaction

The fundamentals of Human-Computer Interaction, HCI, consists of understanding how humans will use an interface and the issues that might be faced with the interface.

HCI is composed of four main components according to Krasovskaya [15].

**The user** plays a central role in HCI. The user interacts with a system, and therefore in HCI, we analyse the user in terms of different parameters, such as the users' emotions and cognitive capabilities.

**The goal-oriented task** represents the fact that a user always has a goal in mind when interacting with a system. When we design systems it is therefore necessary to keep in mind how users achieve their goal. We must therefore consider when we design systems the complexity of the task the user is trying to complete, the knowledge necessary to do so and the time required to carry out the task. This is why system designers carry out usability tests when designing a system, to see if the user can achieve the task they are trying to complete, and how difficult for them it is.

**The interface** is the link between the user and the computer system. This enables the user to interact with a system intuitively and efficiently. When designing an interface one must keep in mind various interface-related aspects, such as screen resolution, display size and interaction type. For example, a website should be responsive and change based on whether it is accessed from a mobile phone or a computer.

**The context** is crucial in an HCI, as we can enhance the user experience based on the context and environment it is accessed through. For example, we could change a user interface based on the lighting from which it is accessed, or how an application will perform with a bad network connection.

HCI is very important to keep in mind when designing applications and systems like the one we made. We must keep in mind the user and the goal the user has when interacting with our application, and also the interface the user is interacting through, and the context the interface is being accessed from. Several proactive methodologies can be used independently or in combination to identify the interactions and issues and ensure a robust analysis. This research paper will be using gesture elicitation and A/B test, which are respectively used by He *et al.* [1] and described by Tomitsch *et al.* [16]. Lastly, usability tests will be conducted as described by Tomitsch *et al.* [16].

## 3.2 Gesture Recognition Principles

A gesture is the orientation of a part of or the whole human physique to convey non-verbal information, according to Konar *et al.* [17]. A hand gesture is therefore the hand's orientation to convey this information. We can further classify these gestures into static and dynamic gestures. A static hand gesture represents a static fixed hand orientation that conveys non-verbal information. A thumb sticking up from the hand is an example of this, as this static orientation of the hand, represents and *ok* or *I like this* information to most people. On the contrary, a dynamic hand gesture represents a movement of the hand, that conveys non-verbal information. A person waving is an example of this, as this movement represents a greeting or a farewell in most cultures.

For gesture recognition to work, several input types can be used. This research paper used a camera as input, and Machine Learning to recognise the gestures. This technique benefits from an already defined gesture set and a training dataset, to make a mapping from a gesture to a gesture classification. This way, we can compute a classification from a hand gesture, and act based on this classification.

## 3.3 Technologies for Gesture Recognition

When recognizing gestures in practice using a machine learning model, The Google Mediapipe Framework offers a simple way to do this. MediaPipe Gesture recognition task, as described by Google [18], can be used to recognise the landmarks of the hand, which are essentially all the joints in the hand. The gesture recognizer is fed with a model that maps a collection of hand landmarks to a gesture classification. Said model can be trained with the Mediapipe Model Maker, given classified static images of gestures as a training dataset. The gesture recogniser uses a model already trained to interpret landmarks and only the gestures have to be trained with a training data size of at least 100 images per gesture, as described by Google [19].

In this research paper, we will employ this exact framework, to make a Minimal Viable Product.

## 3.4 Hybrid Meeting Technologies

A virtual meeting is a meeting where participants are not together physically, but are connected via. video and audio, typically via. the internet. Since the COVID-19 pandemic in 2020, many employees of different companies have started working remotely using this technology to communicate with co-workers and partake in important meetings.

Hybrid meetings are meetings where some participants are physically present while others join remotely. This results in more than one person joining a virtual meeting from a single computer, which results in some of the participants physically present not being able to interact with the computer connected to the meeting.

Popular technologies for hosting these virtual and hybrid meetings include Zoom, Google Meet, and Microsoft Teams among other platforms.

In this research paper, we will build a Minimal Viable Product that extends the functionality of one of these technologies to solve the problem of not being able to reach the computer during hybrid meetings.

# Chapter 4

# Methodology

This section covers the different methodologies used in the research study during its different stages. It concerns both how data collection has been conducted as well as which analysis has been used for the data.

## 4.1 Literature review

To have a background for our work, and to get an overview of what work already has been done, we conducted a literature review. For this review, we constructed a research protocol, defined in Appendix A.3. In the protocol, we define object, scope and domain, to specify exactly what we were looking for. Furthermore, we describe the strategy of our search, and this results in two search strings we used for the sites, IEEE Xplore and ACM as well as a few manually selected papers. The papers in the literature review had to have some hand gesture vocabulary. This means that they need to describe some kind of hand gesture pose, used to interact with a system. This system needed to be based on the core functionality of video processing of said gesture, and not another type of sensor or other technologies. The domain for the papers should preferably be within collaboration systems for hybrid and virtual meetings, but since this domain is very narrow, and has no publications, it was necessary to open the domain to related domains, such as Virtual Reality and general collaborative products. Furthermore, we specified that we needed papers from 2019 to present and that they needed to be in English. This process resulted in two search strings.

The search string used for IEEE Xplore was:

```
(("hand gesture*" OR "gestural" OR "mid-air" OR "gesture-based") AND ("vocabular*" OR "lexic*"
OR "gesture set*" OR "gesture command*" OR "command set"))
```

The search string used for ACM was:

```
AllField:("hand gesture*" OR "gestural" OR "mid-air" OR "gesture-based") AND AllField:
("vocabular*" OR "lexic*" OR "gesture set*" OR "gesture command*" OR "command set")
```

This resulted in 317 results on ACM and 108 results from IEEE. We looked these papers through and narrowed them down to 14 relevant publications based on the criteria of the research protocol.

We then used ChatGPT to organize all the gestures from the papers into a table, which is shown in Appendix A.5. We examined the output to ensure correctness. ChatGPT was used to make the process of gathering the gestures from the 14 papers into a single CSV file more effective.

We gathered and categorized all gestures from these papers, with all gestures used for the same task in the same category. This enabled us to see what gestures were frequently used for what tasks in related papers.

## 4.2 Gesture elicitation

Gesture elicitation is a process in which participants choose gestures freely for a given referent or task. This section will cover how we performed and conducted a gesture elicitation study and which data we focused on gathering and how it was analysed. Gesture elicitation was used as we wanted the input of the participants and it was also a method that was used by He *et al.* [1] and most other papers described in Chapter 2.

### 4.2.1 Participants

Participants N = 102, were all university students They came from different fields of study but mostly IT, as displayed in Appendix A.28. 64 were male, 36 were female and 2 were of other genders. The age distribution in the study was as follows. The female's age was $24 \pm 2.32$ , while the men's age was $25 \pm 3.67$, as defined in Appendix A.30. Most individuals were recruited from the IT University of Copenhagen, while still keeping the diversity of the sample in mind.

We limited the participants to only being students to have a natural project size limit and because we primarily have access to this group. The sample size was set to 102 to achieve a statistical power of 0.8. This was calculated on the webpage ai-therapy.com. The *test-family* was set to *t-test*, *sample-groups* to *independent groups*, *number of tails* to *one*, *effect size* to *0.5*, *significance level*, $\alpha$, to 0.05, and *power* to 0.8.

### 4.2.2 Apparatus

The users were presented with our laptops. To gather preliminary data, they answered a Microsoft form, available in Appendix A.7.

During the experiment phase, we asked the users to use the think-aloud method, which involves participants thinking aloud while selecting gestures, as described by Nielsen [20]. We also showed them a Wizard of Oz prototype, as described by Ramaswamy *et al.* [21]. A Wizard of Oz prototype is a mock interface, controlled by a person. In our case, this was a PowerPoint presentation, showing a Zoom meeting and simulating actions when a user made a gesture.

We also recorded the participants during this phase with the computer webcam. Their faces were obfuscated, to comply with GDPR, as displayed in Figure 4.1. While eliciting gestures, the participants sat roughly 2 meters from the computer, so the recording would pick up all the gestures no matter where they were performed relative to the body.

Figure 4.1: The figure shows an example of one of the writers and how the stored videos were obfuscated for storage

The follow-up questions about general feedback were also gathered in the Microsoft form as shown in Appendix A.7.

### 4.2.3 Procedure

The participants were asked to complete the form until the *Experiment* part of the form, which is available in Appendix A.7. This was to get data on the individuals' demographics and other cultural and environmental factors, that might affect how they would answer our experiment.

During the experiment phase, we would ask them the questions specified in Appendix A.7 to provide a verbal walkthrough of a meeting scenario. We would at this point also focus on not making any gestures, to affect the participants as little as possible. The Wizard of Oz prototype, available in Appendix A.4, was also shown to give visual guidance as well as to ensure that their gestures would not be severely affected by the participants looking at themselves, as virtual meetings clients do not have big indications of their current triggered actions. We then asked them questions related to each task we wanted them to select a gesture for. For example, question 1, available in Table 4.1 relates to the task *Increase* *volume*.

Table 4.1: The table contains the questions asked during the gesture elicitation study

| | |
|---|---|
| **Q1** | You have a hard time hearing what the client connected to the meeting says. You therefore want to increase the volume. How would you do this using only your hands? |
| **Q2** | The volume is now a bit too loud. How would you decrease the volume using your hands? |
| **Q3** | After a bit of discussion back and forth with the client, you want to discuss something with your coworkers without the client being able to hear what you are saying. How would you mute your microphone using your hands? |
| **Q4** | You have reached a conclusion and want to unmute. How do you do this using your hands? |
| **Q5** | The call has been going on for quite a while and you have a short break. In the meantime you do not want the client to see what you are doing and therefore want to turn the camera off. How do you turn it off using your hands? |
| **Q6** | The break is over and you have to turn on the camera again. How do you do this ' using your hands? |
| **Q7** | Before you end the meeting you want to ask a question without interrupting the current discussion. How would you do this using your hands? |
| **Q8** | The meeting is finally over. How would you end the call using your hands? |

By doing this we could elicit qualitative gestures based on eight different tasks. The think-aloud method helped us gain insight into problems they had making gestures based upon a task or why they chose a given gesture. Furthermore, it also helped to learn if there were any biases that the individual might have had. The selection of tasks was guided by multiple criteria. Firstly, tasks which are often used during virtual meeting scenarios were prioritized. Secondly, tasks which are already mapped to keys in Zoom, which is available in Appendix A.9, were prioritized. Lastly, if a task involves multiple actions, it should be avoided. Therefore the task *Share screen* was not used due to having to use the `alt + s` mapping and then afterwards the specific screen the user wanted to share had to be selected, which in turn would result in a convoluted Wizard of Oz experience. After they were asked to fill out the rest of the form regarding their experiences with virtual meeting platforms and what they thought about using hand gestures in virtual meetings.

### 4.2.4 Data analysis

To have an overview of the results, we conducted data analysis of the resulting data of the gesture elicitation. Before starting the actual data analysis we fixed incorrect birthdays, which were identified due to participants setting their birth year to 2024. By contacting the few individuals who made the mistake we could correct it and proceed with the data analysis.

For the Data analysis, we used Microsoft Excel and a custom Python Jupyter notebook defined in Appendix A.1.

We ensured to eliminate any duplicate gestures that were previously defined. Excel allowed us to gather preliminary data as well as define the gestures done during the experiment. Each task could then have its sub-sheet in a single Excel file to look at the gestures used for said task. The sheets also contained the tasks' respective agreement rates, AR. The AR shows the probability of two individuals from a group picking the same gesture for a task. The equation we used is the one used by Brito *et al.* [22] displayed in Equation 4.1

$$AR(r, G_i) = \frac{a_i}{n_i} = \frac{\sum_{p=1}^{|G_i|} \sum_{q=p+1}^{|G_i|} \delta_{p,q}(r)}{\frac{1}{2}|G_i|(|G_i| - 1)} \tag{4.1}$$

Table 4.2: The table shows what each symbol of the agreement rate formula denotes

| **Symbol** | **Denotes** |
|---|---|
| $r$ | Referent/Task |
| $G\_i$ | The group |
| $\|G\_i\|$ | Cardinality of the group |
| $a\_i$ | Number of pairs in agreement |
| $n\_i$ | Total number of pairs |
| $\delta_{p,q}$ | Kronecker's notation (1 if p and q are in agreement else 0) |

Table 4.2 denotes the meaning of each symbol. This calculates the AR for a referent or task $r$. It shows the number of pairs in agreement divided by the total number of pairs and therefore the percentage of pairs in agreement. For example, if everyone picks the same gesture for a task, the AR is 1, since all pairs agree. Whereas if everyone picks a different gesture, it is 0.

The first part of the Python Jupyter Notebook allowed us to see if there were any obvious correlations in our dataset. This part of the notebook is displayed in the exploratory part of the Appendix A.1.2.

Furthermore, we created correlation matrices, as shown in Appendix A.1.3, and conducted dimensionality reduction on our data by Principal Component Analysis, PCA, shown in Appendix A.1.5. The correlation matrix used Cramers V statistic for categorical-categorical association described by IBM [23]. Cramers V had to be used since the data was mostly category-based and not value-based. For the PCA it was needed to tokenize the data. The matrices and PCAs enabled us to see more in-depth if there might be any correlations or obvious variations in all the observations.

## 4.3 A/B Testing

To further clarify our results from the elicitation study, we conducted an A/B test as described by Tomitsch *et al.* [16]. The test consists of having two versions of the same product. In this case, there were two different gestures for each task in our program. Then the two versions are evaluated and compared. In our case, we evaluated the number of people choosing one gesture over the other.

### 4.3.1 Participants

Participants N = 104, were all university students. They came from different fields of study but mostly IT as shown in Appendix A.29. 55 were male, 47 were female and 2 were of other genders. The age distribution in the study was as follows. The females' ages were $24.72 \pm 2.32$, while the men's ages were $24.93 \pm 2.94$, as defined in Appendix A.31.

We limited the participants to only students to match our prior participant group. The sample size was initially set to 102 to match the participants of the gesture elicitation study and to reach a statistical power of 0.8, as described in Subsection 4.2.1. Two individuals answered the questionnaire before closing the form resulting in 104 participants. Diversity was strived for by dispersing the Microsoft form in various places.

### 4.3.2 Apparatus

To efficiently reach 102 participants who wanted to partake in our survey, we used various social media platforms to disseminate the form. These consisted of various Facebook groups, as well as LinkedIn and Instagram posts. This approach allowed us to acquire the number of responses needed within slightly over one week, requiring less effort compared to the gesture elicitation.

When completing the survey, users were presented with a Microsoft form displayed in Appendix A.10. This would gather both preliminary data and the gesture they found most suitable for each task.

### 4.3.3 Procedure

To gather more quantitative data we conducted an A/B test. First, there was preliminary data the participants had to fill out, to get demographic and other cultural and environmental factors that might affect how they would answer our experiment. Afterwards, participants ranked the two most commonly used gestures, derived from the gesture elicitation study. When two gestures were equally common for a task the gesture which had it's reversible gesture linked to the opposite task was selected. Reversible means a gesture which can be used in opposite directions to control the inverse task, e.g. volume up and volume down. Figure 4.2 displays the proposed gestures for each task.

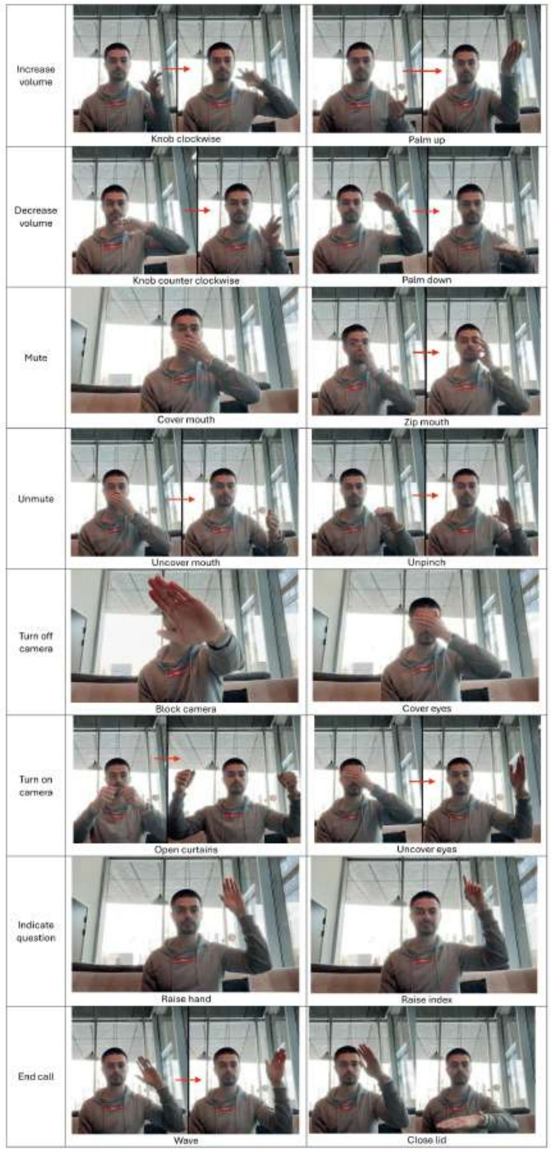



Figure 4.2: The figure shows the top two gestures for each task extracted from the results of the gesture elicitation. (Scaled-up version is available in Appendix A.16)

### 4.3.4 Data analysis

To have an overview of the results, we conducted data analysis of the resulting data of the A/B test. In the data set, we had two data points that had input an incorrect birthday, as they had the birth year set to 2024. Unfortunately, we were incapable of fixing these data points as the Microsoft form was anonymous. Instead, we circumvented this by not taking these data points into account when analysing data in terms of age. The rest of the data points were otherwise of good quality.

For the Data Analysis, we used Microsoft Excel and a slightly modified Python Jupyter notebook, available in Appendix A.2, compared to the one we used in the gesture elicitation, available in Appendix A.1.

The notebook would still enable the creation of exploratory data analysis, correlation matrices, and PCAs - which is mostly similar to what was done in Subsection 4.2.4.

## 4.4 Usability test

To test our MVP based on data from the previous two experiments, we conducted a usability test. We conducted it as described by Tomitsch *et al.* [16] with a few deviations, as clarified in Subsection 4.4.3.

### 4.4.1 Participants

Participants N = 5, were all university students. 4 came from an IT background while 1 came from Engineering. 4 were male and 1 was female. The age distribution in the study was as follows. The females' age was $23 \pm 0$, while the men's age was $22.5 \pm 1$, as calculated in Appendix A.32.

We limited the participants to only students to match our prior participant group.

The number of participants was limited to five because our application is an MVP, and as Tomitsch *et al.* [16] described it as the minimum amount you can create a usability test with and that five participants are optimal to give most findings for a usability test and that increasing the number of participants does not make a big difference in the number of findings, as described by Strba [24].

### 4.4.2 Apparatus

The participants were placed in a room with a computer connected to a virtual meeting with our MVP running. The MVP is described in Chapter 5. We used smartphones for minimal interference with the laptops to record audio during the usability test.

### 4.4.3 Procedure

When we had placed the participant in front of the computer with a Zoom meeting and our MVP running, we started recording the meeting. We then followed the interview guide as described in Appendix A.26, where we first introduced the scenario, and the gestures available for use. This way our usability test procedure differs from the one described by Tomitsch *et al.* [16] because they specifically say that the user should not be introduced to what to do in specific scenarios at all. We decided to introduce the gesture set since the use of gestures as an interface is not widespread in the world. We did not, however, tell the participants what task the gestures were used for beforehand. After this we instructed the user to complete the different tasks in our MVP, like muting, unmuting, increasing the volume etc. When we were through all the tasks, we asked the user the post-test questions as seen in the interview guide. These include aspects that the user liked and disliked about the experience, and what they would improve about the MVP. This concludes the usability test.

### 4.4.4 Results analysis

To get the relevant results of the tests we went through the audio recordings of the tests, and transcribed them. This meant that we could refer to the points that the participants mentioned for the different questions in the text.

# Chapter 5

# Minimal Viable Product

We designed the MVP to extend a virtual meeting client to focus on implementing hand-gesture recognition, and not a virtual meeting application. We considered three popular meeting clients as candidates, namely, Google Meet, Microsoft Teams and Zoom. We chose these three because they were popular in our gesture elicitation study, as defined in Appendix A.24, and because they match the professional work setting we explore in the study.

We compared how many keyboard shortcuts were available in the three applications since this would make it easy to toggle functionalities in our MVP. Said comparison is available in the table in Appendix A.9. We chose Zoom because it was the most popular in our elicitation study as well as the A/B test, we display these results in Appendix A.24 and Appendix A.25, and because it provided the most keyboard shortcuts for relevant tasks.

## 5.1 Structure of the program

Our MVP is built on the Mediapipe Framework and OpenCV and uses a virtual camera to provide the camera output to Zoom. The complete source code of our MVP is available in Appendix A.23

The way it works is, that the MVP uses the MediaPipe Gesture Recognizer by Google [18], which uses our trained model and tries to recognize the gestures on each frame from the video stream. We describe our model in Subsection 5.1.2. In our case, a live feed from the webcam was provided by OpenCV. We then took the live feed and created a virtual camera using the `pyvirtualcam` library. Then Zoom can connect to the virtual camera. This also makes it possible to turn on and off the camera in Zoom with gestures, due to Zoom no longer using the physical camera, and our program still using it to recognize gestures. The Gesture Recognizer analyses a frame and passes it on to a callback function `handle_gesture()` available in Figure 5.1.

```python
1  # For CONFIDENCE_FRAMES the gestures has to be recognized CONFIDENCE_CUT_OFF times
2  CONFIDENCE_FRAMES = 15
3  CONFIDENCE_CUT_OFF = 3
4
5  # specifies a array that has space for gestures
6  confidence_array = [""] * CONFIDENCE_FRAMES
7  iteration = 0
8
9  def handle_gesture(result: GestureRecognizerResult, output_image: mp.Image, timestamp_ms: int):
10     global iteration
11     global confidence_array
12     global CONFIDENCE_FRAMES
13     if (iteration >= CONFIDENCE_FRAMES):
14         most_common = get_most_common_element(confidence_array)
15         #print(f"\r{confidence_array}")
16         match_gesture(most_common)
17         confidence_array = [""] * CONFIDENCE_FRAMES
18         iteration = 0
19
20     # Iterate through the nested lists returned by result.gestures
21     for gesture_list in result.gestures:
22         for category in gesture_list:
23             if category.score > 0.6:
24                 confidence_array[iteration] = category.category_name
25             else:
26                 confidence_array[iteration] = "None"
27
28     iteration += 1
```

Figure 5.1: The figure displays the code block showing the `handle_gesture()` callback function

The function adds the category of the hand gesture to an array named `confidence_array` up until the array reaches a size of 15. The category can be an empty string, `none`, `palm_up`, `palm_down`, `zip_mouth`, `block_camera`, and `wave`. An empty string means that it did not recognize anything. *none* means that it analysed a hand gesture, but the gesture was not a part of the defined gesture set. The other ones represent the different gestures we defined for the model, described in Subsection 5.1.1. When the array reaches the size of 15, it counts up the occurrences of the defined gestures with the `get_most_common_element()` function used on line 14 of Figure 5.1. It then takes the maximum of all of these counts, and that is the winner gesture. If the winner gesture has occurrences greater than or equal to our cut-off of 3, that means that it is enough to be converted to an action in Zoom, and thus it gets passed on to the `match_gesture(gesture)` function available in Figure 5.2.

The function therefore matches the received gesture. If the gesture is *palm_ up*, or *palm_ down*, it uses the `change_volume()` function used in lines 6 and 8 of Figure 5.2. This function uses the `pyautogui` library to click the volume up or volume down button on the computer respectively. If the gesture is one of the other gestures, it calls the `match_lock_gesture()` function

```python
def match_gesture(gesture):
    print(gesture)
    # Check if cooldown since last gesture execution has passed
    match gesture:
        case "palm_up":
            change_volume('volumeup', 2)
        case "palm_down":
            change_volume('volumedown', 2)
        case "None":
            return
        case _:
            match_lock_gesture(gesture)
            return
```

Figure 5.2: The figure depicts the code block showing the `handle_gesture()` callback function

available in Figure 5.3, which sets an `RLock` to avoid race conditions, and a cooldown timer of 3 seconds, and uses the `pyautogui` library to call the relevant keyboard shortcut to complete the action relevant to the gesture.

### 5.1.1 Gesture set for the program

The gesture set we decided to use for the MVP is depicted in Figure 5.4. This set is primarily extracted from the results of our gesture elicitation and A/B tests. But for simplicity of implementation, we only used static variations of the gestures and changed some of the gestures to ensure less overlap with similar gestures. Since the keyboard shortcut for muting and unmuting the microphone is mapped as a toggle, we chose the same gesture for muting and unmuting. This means that we implemented the *zip mouth* gesture, shown in Figure 5.4, for both muting and unmuting the microphone. We chose this gesture since it is reversible with the *unpinch* gesture, and *unpinch* won by more in the *unmute* task than *cover mouth* won by in the *mute* task. This is the same for turning on and off the camera, where we implemented the very popular gesture from our A/B test of blocking the camera. For raise hand, we implemented *Raise index* even though it was the least popular in the A/B test. We chose it since it would overlap less with the wave and block camera gestures than if we chose the regular *raise hand* gesture depicted in Figure 4.2. For increasing and decreasing the volume, and ending the call we chose the most popular gesture from the A/B test as concluded in Section 6.2.

```python
def match_lock_gesture(gesture):
    current_time = time.time()
    global last_execution_time
    # Set a resource lock on last_execution time, to avoid race conditions.
    with lock:
        if current_time - last_execution_time >= cooldown:
            match gesture:
                case "zip_mouth":
                    pyautogui.hotkey('alt', 'a')
                case "wave":
                    pyautogui.hotkey('alt', 'q')
                    pyautogui.press('enter')
                    exit(0)
                case "index_up":
                    pyautogui.hotkey('alt', 'y')
                case "block_camera":
                    pyautogui.hotkey('alt', 'v')
                case _:
                    print(gesture)
                    return
            # Set new cooldown time
            last_execution_time = current_time
```

Figure 5.3: The figure depicts the code block showing the `match_lock_gesture()` callback function

### 5.1.2 Machine learning model

We trained the model using the Mediapipe Model Maker by Google [25] as described in the Mediapipe documentation by Google [26]. The Model Maker uses a technique called transfer learning, described by Gillis [27], which retrains an existing model with new data. This means that we only had to provide the model maker with a minimum of 100 images, explained in a video by Google [19], per gesture. This means that training the model takes very little time, but the model is still quite accurate. We gave the model images extracted from our gesture elicitation study videos, corresponding to the gesture set, that we specified in Figure 5.4 and described in Subsection 5.1.1. The videos had faces obfuscated before they were extracted into images, and fed to the model. In total, we fed the model 11033 images, for recognising 7 different categories. For the increase, decrease volume and zip mouth tasks, we fed the model images of different phases in the completion of the gestures, due to them being dynamic and the model only recognizing static gestures. The *block camera* and *raise index* gestures are static, so we fed the model all available images from the videos. For the *wave* gesture we only fed images where the user had rotated the hand to the left or right. This way the model does not overlap much with the *block camera* gesture, and the two are more distinct from each other. Furthermore, we gave the model a set of *none* gestures that

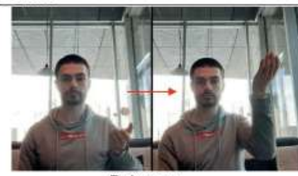
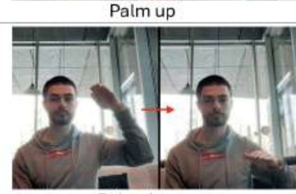
| Task | Gesture |
|---|---|
| Increase volume |  Palm up |
| Decrease volume |  Palm down |
| Mute/Unmute |  Zip mouth |
| Turn camera off/on |  Block camera |
| Raise hand |  Raise index |
| End meeting call |  Wave |

Figure 5.4: The figure shows the defined gesture set for the MVP.

do not represent any gestures that are a part of our gesture set.

## 5.2 Future extensions

As the application is an MVP, we did not get to implement everything we would have wanted. The following describes what features we would have liked to implement, and how it would be implemented.

### 5.2.1 Dynamic gesture set training

As the finalized gesture set from our elicitation and A/B tests resulted in a gesture set that primarily includes dynamic gestures, we would have liked to implement a gesture set that used said dynamic gestures for the program to recognize. To do this we would have to make a set of time series for each gesture, that would be used with a long short-term memory, LSTM, neural network to train a model. The time series would be composed of the number of frames that show the user performing the relevant gesture. 21 landmarks depicted in Figure 5.5 that cover the hand and the $x$ and $y$ values of all of these 21 points. The $x$ and $y$ values represent where the points are relative to the perspective of the camera.

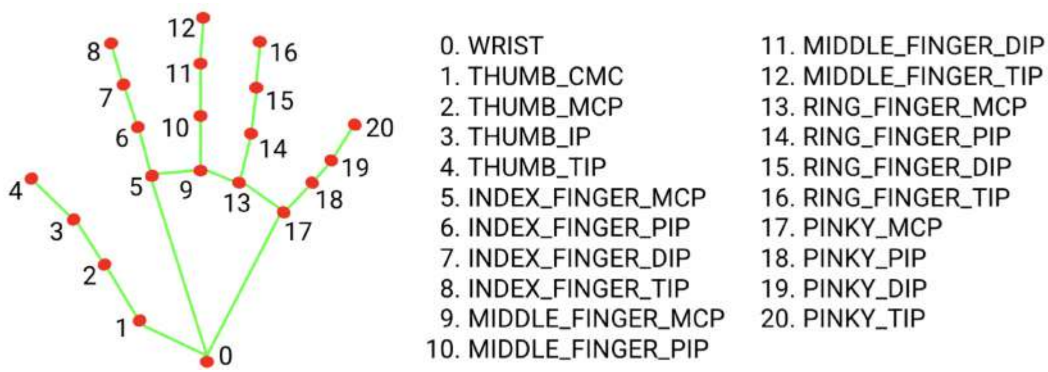



Figure 5.5: The figure shows the 21 hand landmarks used to train models. Source: Mediapipie documentation

The LSTM would then be used to make a time series classification, that can recognize any certain gesture from hand landmarks, such that if a user does a dynamic gesture, the model would recognize this time series of hand landmarks and classify them as the gesture the model was trained to classify it as. This would replace our current model, which processes each frame's landmarks, and compares them to a single gesture the model was trained to recognize. Meaning, that if we used this LSTM model we would compare a time series of landmarks, instead of a single snapshot in time of landmarks. This would enable us to handle dynamic gestures instead of only static ones as our current MVP does.

For example, if we want to train a waving gesture, we would have to have a video clip of a hand waving from side to side and extract a time series of hand landmarks from this video. The landmarks of the video are composed of the fingers rotating from side to side, while the landmark *0 WRIST* available in Figure 5.5 would have roughly the same coordinates throughout the time series, due to being close to the rotational center. Our program would then use the model to compare any similar time series it sees the user do to this, and if it is similar with a certain confidence rate, classify the time series as a wave.

### 5.2.2 Visual feedback and delay of task activation

For a good user experience, we would have liked to implement visual feedback for performing gestures. That is, for example, if a user raises the hand to trigger the raise hand action, a text saying *raise hand*, and a loading bar appears on the screen. The user then needs to keep performing the gesture until the loading bar is full, and the program then raises the hand in Zoom. This feature would be similar to the one that is already implemented in Zoom when interacting with predefined hand gestures. This way we would provide visual feedback to the user performing a gesture, which can improve user experience in hand-gesture-based interactions.

# Chapter 6

# Exploratory Data Analysis

## 6.1 Gesture elicitation

The gesture elicitation resulted in a large gesture library mapped to the eight tasks. We display a snippet in Appendix A.8.

A total of 141 gestures were defined. This is without taking into account if they have been used multiple times for different tasks. We recorded just below 14 hours of video, shown in Appendix A.35, which takes up 43.52 gigabytes of data. The statistical power achieved was 0.7545. This was calculated with ai-therapy.com. It was then answered over two times, both set with the following; *Sample groups* to *Independant groups*, *Number of tails* to *One*, *Effect size* to 0.5, *Significance level* to 0.05. The first time with *Group 1 size* to 64 and *Group 2 size* to 36, which resulted in a statistical power of 0.770 to get the statistical power given the two groups males and females. The second time with *Group 1 size* to 72 and *Group 2 size* to 30, which resulted in a statistical power of 0.739 to get the statistical power given the two groups of Danish native speakers and others. The numbers are shown in Appendix A.33. The mean can then be found to display the average statistical power of the dataset based on the two major groups.

Without sorting any of the data based on the demographics or other factors the agreement rate for each of the tasks looked as displayed in Table 6.1. Which generally shows a low AR.

Based on the agreement rate and the number of unique gestures the consensus in the participants group was generally low. This does not apply to the *Indicate question* task, which was expected as the *Raise hand* and *Raise index*, as we depicted in Appendix A.16, is commonly used in everyday life by a multitude of cultures.

Table 6.1: The table shows AR and UG based on all the data points. AR denotes the agreement rate for each tasks and the method as well as the formula is defined in Equation 4.1. UG denotes the unique gestures and describes the amount of gestures proposed for a task

| **Task** | **AR** | **UG*** |
|---|---|---|
| Increase volume | 0.1033 | 21 |
| Decrease volume | 0.1309 | 18 |
| Mute | 0.0910 | 28 |
| Unmute | 0.0458 | 47 |
| Turn off camera | 0.0727 | 31 |
| Turn on camera | 0.0340 | 45 |
| Indicate question | 0.4754 | 4 |
| End call | 0.1523 | 30 |

Table 6.2: The table shows AR and UG which are filtered based on different demographic values

(a) Scandinavian countries (defined as Denmark, Norway, and Sweden)

| **Task** | **AR** | **UG*** |
|---|---|---|
| Increase volume | 0.1062 | 17 |
| Decrease volume | 0.1311 | 14 |
| Mute | 0.0961 | 24 |
| Unmute | 0.0303 | 38 |
| Turn off camera | 0.0849 | 28 |
| Turn on camera | 0.0373 | 38 |
| Indicate question | 0.4810 | 3 |
| End call | 0.1577 | 26 |

(b) Denmark

| **Task** | **AR** | **UG*** |
|---|---|---|
| Increase volume | 0.1036 | 16 |
| Decrease volume | 0.1277 | 13 |
| Mute | 0.0971 | 23 |
| Unmute | 0.0288 | 37 |
| Turn off camera | 0.0921 | 27 |
| Turn on camera | 0.0341 | 38 |
| Indicate question | 0.4821 | 3 |
| End call | 0.1611 | 24 |

(c) Non danes

| **Task** | **AR** | **UG*** |
|---|---|---|
| Increase volume | 0.1061 | 14 |
| Decrease volume | 0.1326 | 14 |
| Mute | 0.0882 | 15 |
| Unmute | 0.1226 | 16 |
| Turn off camera | 0.0685 | 14 |
| Turn on camera | 0.0414 | 20 |
| Indicate question | 0.4657 | 3 |
| End call | 0.1383 | 13 |

(d) Females

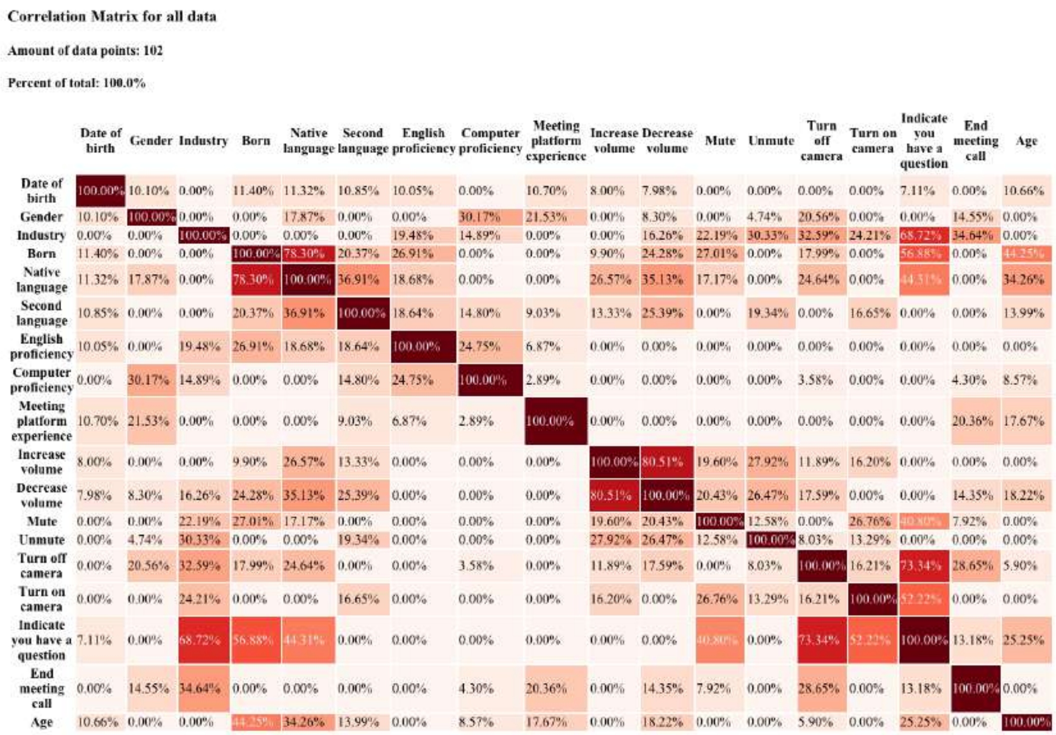
| **Task** | **AR** | **UG*** |
|---|---|---|
| Increase volume | 0.1061 | 14 |
| Decrease volume | 0.1326 | 14 |
| Mute | 0.0882 | 15 |
| Unmute | 0.1226 | 16 |
| Turn off camera | 0.0685 | 14 |
| Turn on camera | 0.0414 | 20 |
| Indicate question | 0.4657 | 3 |
| End call | 0.1383 | 13 |

(e) Males

| **Task** | **AR** | **UG*** |
|---|---|---|
| Increase volume | 0.1176 | 18 |
| Decrease volume | 0.1423 | 14 |
| Mute | 0.0853 | 21 |
| Unmute | 0.0392 | 34 |
| Turn off camera | 0.0819 | 24 |
| Turn on camera | 0.0360 | 34 |
| Indicate question | 0.4658 | 3 |
| End call | 0.1464 | 20 |

When the data points are filtered, displayed in Table 6.2, the ARs do not change significantly compared to the non-filtered, displayed in Table 6.1. *Unmute* is the exception to this that goes from 0.0910, non-filtered, to 0.1226 both when filtering on non-Danes and females. This observation about the AR can be attributed to random variation or noise, due to not being a consistent change across more tasks. These observations suggest that demographics do not play a significant role in the consensus of gestures chosen. ARs do not show if there might be other tendencies in specific demographics or prior chosen gestures.

We then created PCAs and correlation matrices based on the data to see if there might be other nonobvious correlations. The main ones to look at are the two in Figure 6.1 and Figure 6.2. More were created, shown in Appendix A.11 and Appendix A.13.



**Correlation Matrix for all data**

**Amount of data points: 102**

**Percent of total: 100.0%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 10.10% | 0.00% | 11.40% | 11.32% | 10.85% | 10.05% | 0.00% | 10.70% | 8.00% | 7.98% | 0.00% | 0.00% | 0.00% | 0.00% | 7.11% | 0.00% | 10.66% |
| **Gender** | 10.10% | 100.00% | 0.00% | 0.00% | 17.87% | 0.00% | 0.00% | 30.17% | 21.53% | 0.00% | 8.30% | 0.00% | 4.74% | 20.56% | 0.00% | 0.00% | 14.55% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 19.48% | 14.89% | 0.00% | 0.00% | 16.26% | 22.19% | 30.33% | 32.59% | 24.21% | 68.72% | 34.64% | 0.00% |
| **Born** | 11.40% | 0.00% | 0.00% | 100.00% | 78.30% | 20.37% | 26.91% | 0.00% | 0.00% | 9.90% | 24.28% | 27.01% | 0.00% | 17.99% | 0.00% | 56.88% | 0.00% | 44.25% |
| **Native language** | 11.32% | 17.87% | 0.00% | 78.30% | 100.00% | 36.91% | 18.68% | 0.00% | 0.00% | 26.57% | 35.13% | 17.17% | 0.00% | 24.64% | 0.00% | 44.31% | 0.00% | 34.26% |
| **Second language** | 10.85% | 0.00% | 0.00% | 20.37% | 36.91% | 100.00% | 18.64% | 14.80% | 9.03% | 13.33% | 25.39% | 0.00% | 19.34% | 0.00% | 16.65% | 0.00% | 0.00% | 13.99% |
| **English proficiency** | 10.05% | 0.00% | 19.48% | 26.91% | 18.68% | 18.64% | 100.00% | 24.75% | 6.87% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Computer proficiency** | 0.00% | 30.17% | 14.89% | 0.00% | 0.00% | 14.80% | 24.75% | 100.00% | 2.89% | 0.00% | 0.00% | 0.00% | 0.00% | 3.58% | 0.00% | 0.00% | 4.30% | 8.57% |
| **Meeting platform experience** | 10.70% | 21.53% | 0.00% | 0.00% | 0.00% | 9.03% | 6.87% | 2.89% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 20.36% | 17.67% |
| **Increase volume** | 8.00% | 0.00% | 0.00% | 9.90% | 26.57% | 13.33% | 0.00% | 0.00% | 0.00% | 100.00% | 80.51% | 19.60% | 27.92% | 11.89% | 16.20% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 7.98% | 8.30% | 16.26% | 24.28% | 35.13% | 25.39% | 0.00% | 0.00% | 0.00% | 80.51% | 100.00% | 20.43% | 26.47% | 17.59% | 0.00% | 0.00% | 14.35% | 18.22% |
| **Mute** | 0.00% | 0.00% | 22.19% | 27.01% | 17.17% | 0.00% | 0.00% | 0.00% | 0.00% | 19.60% | 20.43% | 100.00% | 12.58% | 0.00% | 26.76% | 40.80% | 7.92% | 0.00% |
| **Unmute** | 0.00% | 4.74% | 30.33% | 0.00% | 0.00% | 19.34% | 0.00% | 0.00% | 0.00% | 27.92% | 26.47% | 12.58% | 100.00% | 8.03% | 13.29% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 20.56% | 32.59% | 17.99% | 24.64% | 0.00% | 0.00% | 3.58% | 0.00% | 11.89% | 17.59% | 0.00% | 8.03% | 100.00% | 16.21% | 73.34% | 28.65% | 5.90% |
| **Turn on camera** | 0.00% | 0.00% | 24.21% | 0.00% | 0.00% | 16.65% | 0.00% | 0.00% | 0.00% | 16.20% | 0.00% | 26.76% | 13.29% | 16.21% | 100.00% | 52.22% | 0.00% | 0.00% |
| **Indicate you have a question** | 7.11% | 0.00% | 68.72% | 56.88% | 44.31% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 40.80% | 0.00% | 73.34% | 52.22% | 100.00% | 13.18% | 25.25% |
| **End meeting call** | 0.00% | 14.55% | 34.64% | 0.00% | 0.00% | 0.00% | 0.00% | 4.30% | 20.36% | 0.00% | 14.35% | 7.92% | 0.00% | 28.65% | 0.00% | 13.18% | 100.00% | 0.00% |
| **Age** | 10.66% | 0.00% | 0.00% | 44.25% | 34.26% | 13.99% | 0.00% | 8.57% | 17.67% | 0.00% | 18.22% | 0.00% | 0.00% | 5.90% | 0.00% | 25.25% | 0.00% | 100.00% |

Figure 6.1: The figure shows a correlation matrix for the entire unfiltered dataset

Examining the matrix correlations available in Figure 6.1 between two demographic variables, and between the task *Indicate you have a question* and other variables do not provide valuable insight. This is partly due to the demographic variables being out of scope, and *Indicate you have a question* only having four unique gestures in total, where two of the gestures were only chosen by 2 individuals. Thus any observed correlations may be more indicative of chance than anything else.

The slight correlations that are worth noting are the ones between *Increase volume* and *Decrease volume* with 80.51%, as well as between *Turn off camera* and *End meeting call* with 28.65%. The reasonings for these correlations are different. For the task *Turn off camera* and *End meeting call* it is mainly due to most choosing the *wave* gesture for the last mentioned task. Which easily results in shown correlations that might mostly be based on chance. For changing the volume the correlation is due to the tendency of the individuals

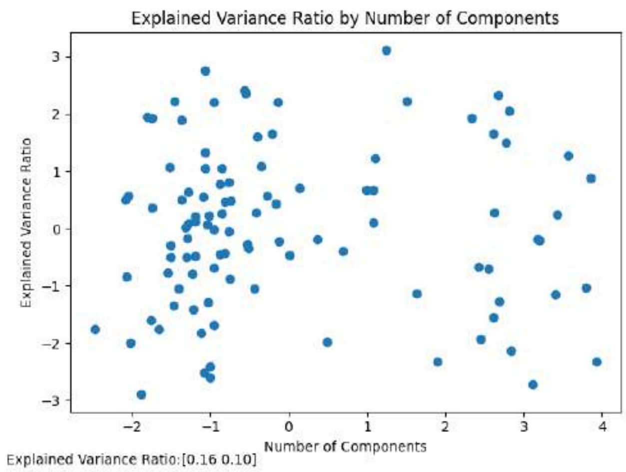



Figure 6.2: The figure shows the PCA for our dataset plotted

choosing reversible gestures. This includes gestures such as the *knob clockwise* and *counter-clockwise* and *palm up* and *down*, as shown in Appendix A.16. In Table 6.3 the percentage of individuals using reversible gestures can be seen for the tasks that had a natural inverse. There was a lower correlation for the tasks that interacted with the microphone and camera, though still significant.

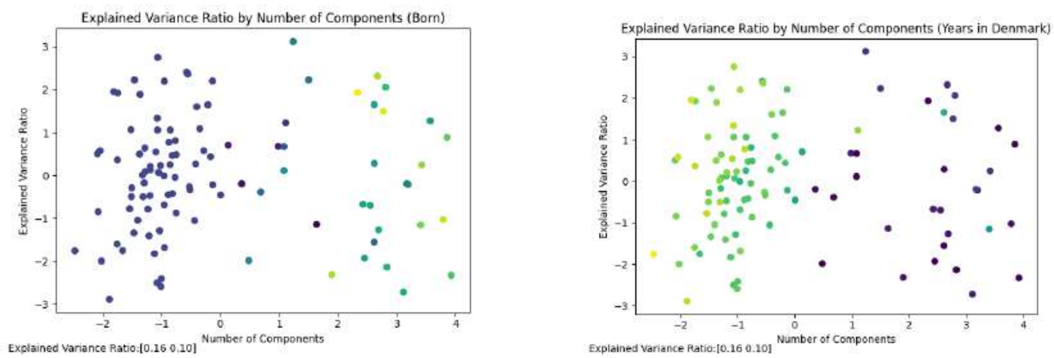
Table 6.3: The table shows the approximate amount of reversible gestures used in the gesture elicitation study

| **Task 1** | **Task 2** | **% used reversible gestures** |
|---|---|---|
| Increase volume | Decrease volume | ~75.5 (shown in Appendix A.17) |
| Mute | Unmute | ~36.3 (shown in Appendix A.17) |
| Turn camera off | Turn camera on | ~35.3 (shown in Appendix A.17) |

Toggleable tasks, *mute* to *unmute* and *turn off camera* to *turn on camera*, also had a noticeable amount using the same gesture for toggling it on and off. Individuals used toggleable gestures 10.68% of the time for *mute/unmute* and 11.76% of the time for *turn off camera/turn on camera*, as displayed in Appendix A.18. This further suggests a wish for a streamlined gesture-to-task vocabulary.

The distribution between static and dynamic gestures was also stark. Dynamic gestures were chosen in favour of static gestures by most individuals. Out of the 141 defined gestures 85.11% were dynamic whilst 14.89% were static, which is defined in Section 3.2. This was also prevalent when comparing between tasks, shown in Appendix A.22. The only task with a majority of static gestures is *Indicate you have a question* with 99% of individuals choosing static over dynamic. The other tasks which came close to being more equally distributed were *Mute* and *Turn off camera*. They respectively had 38.38% and 45.54% for using static gestures This implies that individuals tend toward using dynamic gestures, except when a gesture is defined in societal use cases such as the *Indicate you have a question*.

We applied dimensionality reduction using PCA on the dataset, going from 17 dimensions to 2. The PCA is available in Figure 6.2, where two clusters are displayed, with a more dense cluster on the left-hand side. Figure 6.3a and Figure 6.3b colour the nodes based on where participants were born and how many years they have been in Denmark respectively, it is clear that the two clusters are based on whether the participants are from Denmark or not.




(a) The figure is coloured by where participants were born.

(b) The figure is coloured by how many years participants have been in Denmark

Figure 6.3: The figures show PCAs from Gesture elicitation data

The few outliers mainly consist of the following. One, was where individuals were either born outside of Denmark and have not spent a significant amount of time in the specific country or Denmark. Two was where they were born in Denmark and have not spent considerable time there.

In most PCA for each of the filters, there were no obvious correlations as shown in Appendix A.13. When compared to the action of turning off the camera there was a slight tendency for the Danes to choose the gesture *Block camera*. The gesture is shown in Appendix A.16, and the tendency is displayed in Figure 6.4 by the purple circles. This could indicate slight cultural influence in the gesture chosen by individuals.

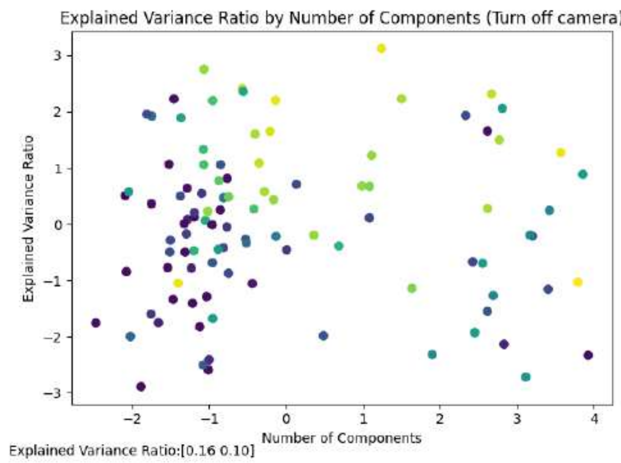



Figure 6.4: The figure shows the PCA from the gesture elicitation study coloured by which gesture was chosen for turning off the camera

## Discussion

During the experiment phase and in the feedback section of the form some individuals brought up worries about the social acceptance and false triggers of hand-gesture-based interaction.

Participant 1 mentioned:

> *I was thinking a lot about whether it's "rude" to visibly mute or turn off the camera of someone.*
> *- Quote from the feedback part of the gesture elicitation form*

Participant 2 mentioned:

> *I would be afraid of doing "fatal" hand gestures if they were enabled in a meeting, such as ending the call or kicking someone out of the meeting. I think it is fair to keep these only accessible by clicking a button on a laptop. In general, [in] most meetings I would have my own laptop nearby and if so, would prefer to use this for navigation. However, in settings with a meeting microphone and multiple people present together, it might be useful to be able to manage the call easily.*
> *- Quote from the feedback part of the gesture elicitation form*

A few also hesitated when about to perform gestures such as the *zip mouth*, as we displayed in Appendix A.16, and asked to which extent they should consider the social situation of the Wizard of Oz.

In general, the feedback towards the Wizard of Oz was positive. They also expressed curiosity about the future implementation and the potential opportunities it could offer such as for people with disabilities.

## 6.2 A/B test

The statistical power achieved was 0.7135. This was calculated with ai-therapy.com. It was then answered over two times, both set with the following; *Sample groups* to *Independant groups*, *Number of tails* to *One*, *Effect size* to 0.5, *Significance level* to 0.05. The first time with *Group 1 size* to 55 and *Group 2 size* to 47, which resulted in a statistical power of 0.804 to get the statistical power given the two groups males and females. The second time with *Group 1 size* to 85 and *Group 2 size* to 19, which resulted in a statistical power of 0.623 to get the statistical power given the two groups of Danish native speakers and others. The numbers are shown in Appendix A.33. The mean can then be found to display the average statistical power of the dataset based on the two major groups.

The results of the A/B test in terms of which gestures were chosen are depicted in Figure 6.5. In these results, it is shown that most categories had a winner or at least a slight winner. The exception was the gestures for the *Camera on* task, where answers were split equally. Of notable results, we can see that the *Camera off* task had the most dominant winner at 78% with the gesture *Block camera*. Other dominant winners were *Raise hand* in the *Indicate question* task, and *Wave* in the *End call* task.

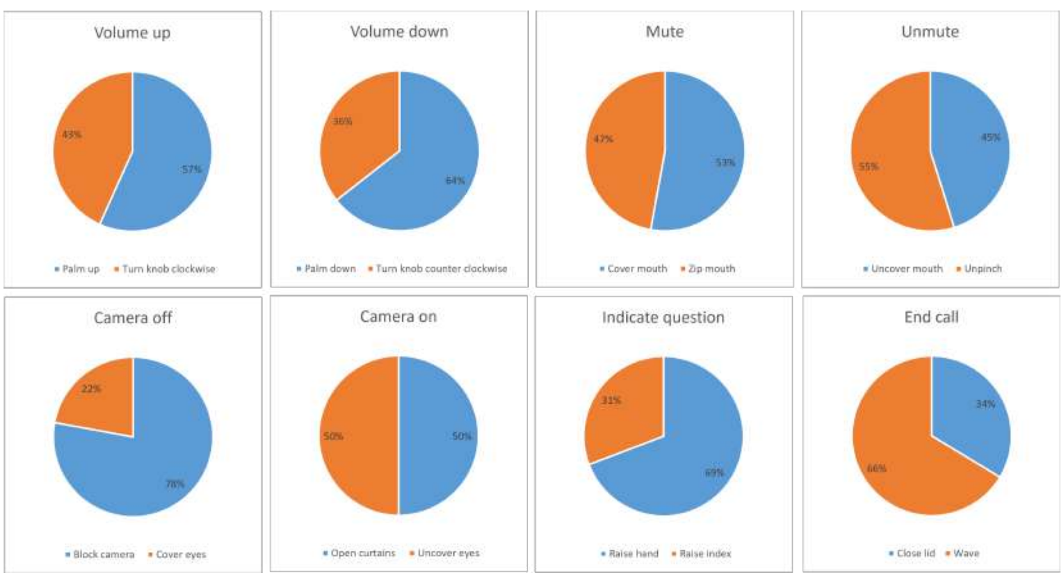



Figure 6.5: The figure shows the results from the A/B test in terms of the percentage of people picking a gesture, the gestures are depicted in Figure 4.2

We conducted exploratory data analysis on this data, which included a comparable analysis to the gesture elicitation analysis. The correlation matrix for the data is available in

Figure 6.6. Notable of the results was, that *Increase volume* and *Decrease volume* were 79% correlated. *Mute* and *Unmute* were 32% correlated, and *Turn off camera* and *Turn on camera* were 31% correlated. These correlations support the mentioned tendencies, from the gesture elicitation study towards the use of reversible and toggleable gestures.

**Correlation Matrix for all data**

**Amount of data points: 104**

**Percent of total: 100.0%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 9.87% | 0.00% | 10.29% | 14.99% | 5.82% | 14.07% | 7.68% | 3.78% | 0.00% | 0.00% | 1.11% | 0.27% | 0.00% | 1.37% | 0.00% | 14.00% | 14.74% |
| **Gender** | 9.87% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 20.38% | 30.09% | 11.74% | 0.00% | 18.54% | 20.79% | 0.00% | 0.00% | 0.00% | 9.08% | 19.46% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 12.19% | 20.47% | 31.38% | 54.35% | 21.68% | 0.00% | 0.00% | 22.48% | 0.00% | 13.13% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Born** | 10.29% | 0.00% | 0.00% | 100.00% | 79.85% | 24.85% | 0.00% | 0.00% | 21.54% | 6.12% | 0.00% | 0.00% | 4.24% | 0.00% | 9.90% | 0.00% | 0.00% | 43.35% |
| **Native language** | 14.99% | 0.00% | 12.19% | 79.85% | 100.00% | 12.68% | 27.92% | 0.00% | 29.24% | 11.35% | 0.00% | 0.00% | 0.00% | 0.00% | 16.71% | 16.48% | 7.28% | 37.58% |
| **Second language** | 5.82% | 0.00% | 20.47% | 24.85% | 12.68% | 100.00% | 0.00% | 24.08% | 0.00% | 22.52% | 23.71% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 10.33% | 0.00% |
| **English proficiency** | 14.07% | 20.38% | 31.38% | 0.00% | 27.92% | 0.00% | 100.00% | 11.24% | 9.85% | 6.33% | 15.61% | 0.00% | 7.94% | 0.00% | 0.00% | 10.45% | 9.30% | 0.00% |
| **Computer proficiency** | 7.68% | 30.09% | 54.35% | 0.00% | 0.00% | 24.08% | 11.24% | 100.00% | 8.78% | 3.10% | 0.00% | 12.47% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 6.13% |
| **Meeting platform experience** | 3.78% | 11.74% | 21.68% | 21.54% | 29.24% | 0.00% | 9.85% | 8.78% | 100.00% | 4.60% | 0.00% | 6.31% | 8.73% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 6.12% | 11.35% | 22.52% | 6.33% | 3.10% | 4.60% | 100.00% | 78.78% | 0.00% | 0.00% | 5.89% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 18.54% | 0.00% | 0.00% | 0.00% | 23.71% | 15.61% | 0.00% | 0.00% | 78.78% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 13.80% | 0.00% | 0.00% |
| **Mute** | 1.11% | 20.79% | 22.48% | 0.00% | 0.00% | 0.00% | 0.00% | 12.47% | 6.31% | 0.00% | 0.00% | 100.00% | 32.13% | 12.01% | 0.00% | 2.29% | 7.21% | 0.00% |
| **Unmute** | 0.27% | 0.00% | 0.00% | 4.24% | 0.00% | 0.00% | 7.94% | 0.00% | 8.73% | 0.00% | 0.00% | 32.13% | 100.00% | 10.63% | 16.70% | 0.00% | 4.84% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 13.13% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 5.89% | 0.00% | 12.01% | 10.63% | 100.00% | 31.05% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 1.37% | 0.00% | 0.00% | 9.90% | 16.71% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 16.70% | 31.05% | 100.00% | 3.40% | 7.25% | 0.00% |
| **Indicate you have a question** | 0.00% | 9.08% | 0.00% | 0.00% | 16.48% | 0.00% | 10.45% | 0.00% | 0.00% | 0.00% | 13.80% | 2.29% | 0.00% | 0.00% | 3.40% | 100.00% | 0.00% | 0.00% |
| **End meeting call** | 14.00% | 19.46% | 0.00% | 0.00% | 7.28% | 10.33% | 9.30% | 0.00% | 0.00% | 0.00% | 0.00% | 7.21% | 4.84% | 0.00% | 7.25% | 0.00% | 100.00% | 9.52% |
| **Age** | 14.74% | 0.00% | 0.00% | 43.35% | 37.58% | 0.00% | 0.00% | 6.13% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 9.52% | 100.00% |

Figure 6.6: The figure shows the correlation matrix for all the unfiltered data from the A/B test

When we looked further into the data, we found that 99% of everyone that picked the *Palm up* gesture for the task *Increase volume* picked the *Palm down* gesture for the task *Decrease volume*, shown in Appendix A.19. Of the ones who chose the *Turn knob clockwise* gesture 81% chose the *Turn knob counter clockwise* gesture and 19% of them chose *Palm down*, as was calculated in Appendix A.20. This supports the previously mentioned tendency towards using reversible gestures, which was discussed in the gesture elicitation Section 6.1 as well as the prior paragraph.

We also applied dimensionality reduction using PCA on the A/B dataset, going from 19 dimensions to 2. The PCAs, displayed in Appendix A.14, did not yield any results of great interest, but there is one big cluster of people from Denmark, and outliers who are not from Denmark.

## 6.3 Usability test

The results of the usability test interviews have been transcribed in Appendix A.27. It confirmed a few problems we had suspected that individuals might point out during the test, mainly it did not recognize the gestures reliably and that there was a wish for visual feedback. The findings will be highlighted in more depth below about what was pointed out as well as what general solutions could consist of.

### 6.3.1 Improvements

All participants mentioned that the program should recognise the gestures more reliably as written in Appendix A.27. They thought that the program did not necessarily do what they wanted right away when performing a gesture and, in some instances, performed the wrong action. This would involve general improvement of the machine learning model but could also be improved with further testing of the tweakable confidence variables in the MVP. Integration of dynamic gestures could potentially also enhance confidence.

### 6.3.2 Visual feedback

Individuals 1 and 2 call for visual feedback when performing gestures, as transcribed in Appendix A.27.1 and Appendix A.27.2. They want more visual feedback, to ensure that they are performing the right gesture, or that the MVP is registering anything at all. Individuals 4 and 5 did not mention visual feedback but instead mentioned: "I did not know how to unmute..."Appendix A.27.4, and "... when I found out exactly how I should do it [the gesture], it worked..."Appendix A.27.5. This shows that they were sometimes slightly unsure how to perform the gesture and which to use for the task wanted. This could therefore partly be helped with visual feedback.

### 6.3.3 Additional results

Additionally, participants 2, 3, and 4 pointed out the gestures to be well chosen and natural to use for the tasks provided, even though they were only shown the gesture without its task mapping. Individual four said: "Of the gesture set you showed me it was obvious what each gesture should be used for."Appendix A.27.4.

# Chapter 7

# Discussion

This section will investigate and discuss the results of our experiments. We will furthermore evaluate if the results addressed the problem statement.

After we had completed the gesture elicitation and analysis of the data it was apparent that there is no natural gesture-to-task vocabulary, due to the amount of different gestures that were proposed. This could be concluded due to the low agreement rate and the number of unique gestures, as shown in Table 6.1, and was also backed up based on our correlation matrices and PCAs, depicted in Figure 6.1 and Figure 6.2 which does not show any obvious correlation between demographics and gestures chosen. Though a few tendencies were identified such as in the PCA Figure 6.4, which displayed a tendency for Danes to choose the *Block camera* gesture for turning off the camera. There could therefore exist some tendencies but among university students, mainly based at the IT University of Copenhagen, it seems to be unlikely. Wu *et al.* [4] also identified tendencies between Chinese and American individuals. So the results might have differed slightly if the research had been done on a more global scale.

The results of the A/B test concluded that the individual tasks did not have a gesture that won by more than 80%, as shown in Figure 6.5. The closest was the task *camera off* with 78%. Thus a hand-gesture-based language seems to function mostly like spoken or sign language in the way that it has to be learned. Languages of course have some fundamentals to make it easier such as onomatopoeic words. Therefore the right question is not if there is a natural language but what would make a good framework for building a gesture-to-task-vocabulary. We present our findings for such a framework in the following sections.

## 7.1 Reversible gestures

The first aspect of this framework would be that for tasks that work in both directions, such as increasing and decreasing the volume, it was discovered that a large segment of individuals preferred reversible gestures. We define reversible gestures as gestures that have a natural inverse. The inverse for the gesture *palm up* is *palm down*. Both gestures are displayed in Figure 5.4. This is apparent, described in the results of the A/B test Section 6.2 as participants have a strong tendency to pick the inverse gesture, to the one they picked in the inverse task. As seen in the results of the A/B test, 99% of participants that picked the *palm up* gesture for the task *increase volume* also picked *palm down* for *decrease volume*. In general, tasks with an inverse such as increasing and decreasing volume, muting and unmuting, and turning on and off the camera, exhibited a correlation as displayed in the correlation matrix depicted in Figure 6.6. In the gesture elicitation results it is also apparent from the correlation matrix that participants picked reversible gestures, especially in the tasks *increase volume* and *decrease volume*, where the two are correlated by 80%. As such it is apparent from our experiments that participants want to use reversible gestures for a potential gesture-to-task vocabulary.

## 7.2 Experiental bias

The second aspect would be to mimic the task's analogue interface, as the phenomena of skeuomorphism, which is when an interface mimics its real-world counterpart. This was observed as when individuals tended towards using a few popular responses there were often analogue interfaces or signs that already existed. More generally, an experiential bias can help define gestures, which was also discussed by Ali *et al.* [5]. The *index finger up* and *palm up* gestures are two of these since it is used regularly to indicate to peers that you have a question. This is also apparent as these two gestures were used heavily in our gesture elicitation study for the task *Indicate that you have a question*. In general, the *Indicate that you have a question* task had the lowest number of unique gestures proposed in the gesture elicitation by far as shown in Table 6.1. This shows how generally accepted the *palm up* and *index finger up* gestures are for the task of indicating that you have a question. A few other examples include the *knob clockwise* and *knob counterclockwise* and *wave*. These are respectively found on volume dials on various sound interfaces and when saying goodbye.

## 7.3 Dynamic gestures

The third aspect would be to use dynamic gestures for tasks that do not already have static gestures defined for their societal use case. This observation in the gesture elicitation study showed that when individuals had to come up with a gesture themselves, they tended to use dynamic gestures. This is apparent from our results, as 85% of proposed gestures were dynamic. Furthermore, this is also visible when comparing the number of participants picking dynamic and static gestures for each task. As shown in Appendix A.22, all tasks except *Indicate that you have a question* have a majority of participants picking a dynamic gesture. However, it may be possible to use static gestures for specific tasks, as a few of the tasks were close to an equal distribution. For the task *Indicate that you have a question*, 99% of participants picked a static gesture. This shows that if they are properly defined in society, static gestures are possible to use, although when picking, participants generally chose dynamic gestures.

## 7.4 Haptic/visual feedback

The fourth is less related to the vocabulary and more giving feedback to the client when performing gestures. As seen in the results of the usability test in Section 6.3, users call for more feedback when performing gestures, to see if the MVP has recognised what the user has performed. One way could be to provide visual feedback as described in Subsection 5.2.2. Another way could be to provide feedback when performing gestures with haptics. Haptics could be taken advantage of, as discussed by Vaquero-Melquor *et al.* [28]. This would in turn improve the user experience of the product as the user would receive instant feedback on what the system is recognizing without forcing the user to look at the display or generally cluttering up the display with information.

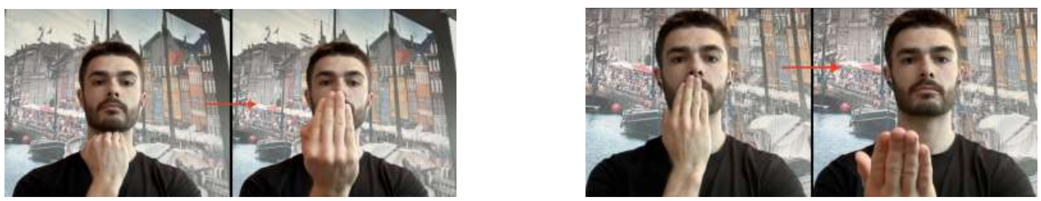
## 7.5 Machine learning model

We would also like to highlight the importance of a well-trained machine-learning model. Users highlighted in the usability test that the MVP was too unreliable in recognizing the gestures. It is therefore very important to have a well-trained machine learning model, and for the use case of dynamic gestures, it should also be trained with a Long short-term memory, LSTM, neural network. This would enable the gestures to be recognized dynamically and the training data would be videos or time series of images instead of static images of gestures. We described LSTM in further detail in Subsection 5.2.1. Therefore, to have a well-working program with hand gesture interaction, the machine learning model should be trained properly and implemented well.

## 7.6 Additional findings

Additionally, the development of a vocabulary would have to address various challenges. One of these challenges is social acceptance. This is because some gestures chosen for the tasks could be perceived as rude in the conversation, as described in Section 6.1. One solution could be to strive for socially acceptable gestures. This would create new problems, such as difficulties with proposing gestures that still live up to the second aspect of our proposed framework as described in Section 7.2. An example of this is the task *mute* in our gesture elicitation experiment, where several of the most popular gestures could also be used to gesticulate to another person that they should be silent, examples can be seen in the gestures linked to mute in Appendix A.16.

Another problem would be to take different communities and cultures into account. E.g. the gesture for *me ne fott proprio*, shown in Figure 7.1a, from Italian looks similar to a gesture proposed often for *unmute*, as shown in Figure 7.1b.



(a) The figure shows the gesture *me ne fott proprio* which is considered a rude hand gesture in Italian

(b) The figure shows a common gesture used for unmuting during the gesture elicitation study

Figure 7.1: The figures display two hand gestures which look similar but have two completely different meanings

Therefore creating a globally acceptable vocabulary would take a lot of work and could make it less intuitive. Another solution could be to be optimistic and hope that gestures, based on the framework's second aspect described in Section 7.2, would become socially acceptable as adoption grows. Another major challenge, also pointed out during the interviews, would be to ensure no false positives. This could partly be addressed programmatically by performing a gesture for a specified amount of time before being recognised or by deviating from the second aspect of the framework to have gestures that are different from most social use cases such that a user does not perform them by accident.

As pointed out by Uva *et al.* [14] it is a good consideration to take the fatigue of performing a gesture into consideration:

> *In particular, long sessions of midair interactions may lead to upper limb fatigue, a condition known as "the gorilla arm effect"*
> *- Uva et al. [14]*

It was also pointed out by one of the authors that when testing the MVP, as described in Chapter 5, the gesture of turning the camera on and off got increasingly more uncomfortable as time passed.

## 7.7 Suggestions for future work

Future work could go in many directions as the research paper presents the large data set "GestureVerse: Exploring the Multiverse of Interactive Gestures" together with analyses and the MVP.

**The first direction** could be to gather data from another university based in another country. This would enable them to compare findings in a more varied cultural setting, like Wu *et al.* [4] did in their study. This would result in proving or disproving the findings that showed minor demographic influence, mostly within a single university, in the gestures chosen within the domain of hybrid and virtual meetings.

**The second direction** could be to implement the dynamic gestures from Figure 8.1 into a program. This would be a major improvement from our MVP which was only able to recognize static gestures. It would improve the usability of the program.

**The third direction** could be to propose a gesture-to-task vocabulary based upon the framework discussed in Chapter 7. This could provide a comfortable gesture set which could also allow for easily distinguishable gestures for a hand landmarked setup to use.

These directions would provide a lot of value for hand-gesture-based interaction in hybrid meetings potentially paving the path for fully integrating gesture interactions into virtual meeting platforms.

# Chapter 8

# Conclusion

Throughout the work, we have discussed the lack of natural gesture commands with 80% consensus on each task even when filtering based on demographics. This was clear after doing data analysis of the gesture elicitation study. The only natural gestures which existed were for indicating you have a question, due to already having a societal use-case, while the other tasks resulted in various suggested gestures. This means that gesticulation works more like a normal language, which if left untaught individuals will mimic what they have previously been exposed to. Instead of a natural set, other tendencies for suggested gestures were identified.

Both in the data analysis of the gesture elicitation study and the A/B test it was identified that there were tendencies towards the use of specific types of gestures, discussed in Chapter 7. The findings were that individuals preferred to use; reversible gestures, gestures using experiential bias, and dynamic gestures.

Reversible gestures are gestures with a natural opposite like with *Palm up* and *Palm down*. This observation was especially prevalent when individuals were asked to adjust the volume, described in Section 6.1.

Experiential bias also had a significant effect as individuals often either mimicked an analogue interface or used commonly recognized gestures, such as *Index finger up* for indicating they had a question.

The use of dynamic gestures was also predominantly chosen during gesture elicitation. Indicating a preference for movement though static gestures may suit specific tasks with clear societal norms, as for the task *Indicate you have a question*.

Furthermore, as pointed out in Section 7.6, both based on the usability tests, our own experience, and other papers that the MVP should be more precise in detection and if gestures are to be performed over a longer period their fatigue should be taken into account.

Given the sample size of participants, 102 for the gesture elicitation and 104 for the A/B test, we achieved a statistical power of 0.7545, calculated in Section 6.1, and 0.7135, calculated in Section 6.2 we can assume that our results are representative of the general university student. Furthermore, most cited papers use a sample size well below 55 as used in Pulina *et al.* [12]. The exception is the paper from Ali *et al.* [5] with a sample size of 167. Our dataset described in Section 6.1, has been evaluated and analysed through multiple methodologies. This resulted in a robust gesture set. Additionally, we propose a framework for creating a gesture set which strives to be easily learnable, and could therefore provide a valuable stepping stone for future work.

Our research study shows the unlikelihood of finding a natural gesture-to-task vocabulary among university students. Which was shown in the lack of consensus. Therefore we want to propose the following for future work. Future work should propose a vocabulary based on our framework. Memorability and guessability tests of the vocabulary should then be performed to evaluate the performance and potential aspects that could be improved, as described by Andolina *et al.* [11]. After it will be essential to make multiple virtual meetings clients use this language to make people familiar with the vocabulary and thus make it natural. Our study presents the dataset "GestureVerse: Exploring the Multiverse of Interactive Gestures", including just under 14 hours, or 43.52 gigabytes of face-obfuscated videos of 102 participants from our gesture elicitation, and demographical data of both 102 participants from the elicitation and 104 from the A/B test.

## Finalized gesture set

The finalized gesture set is depicted in Figure 8.1. This was based on the two most popular gestures for each task from the gesture elicitation study and then an A/B test.

The MVP was trained on a slightly modified gesture set, as discussed in Chapter 5. After modifications, most participants still found it natural, as described in Section 6.3.

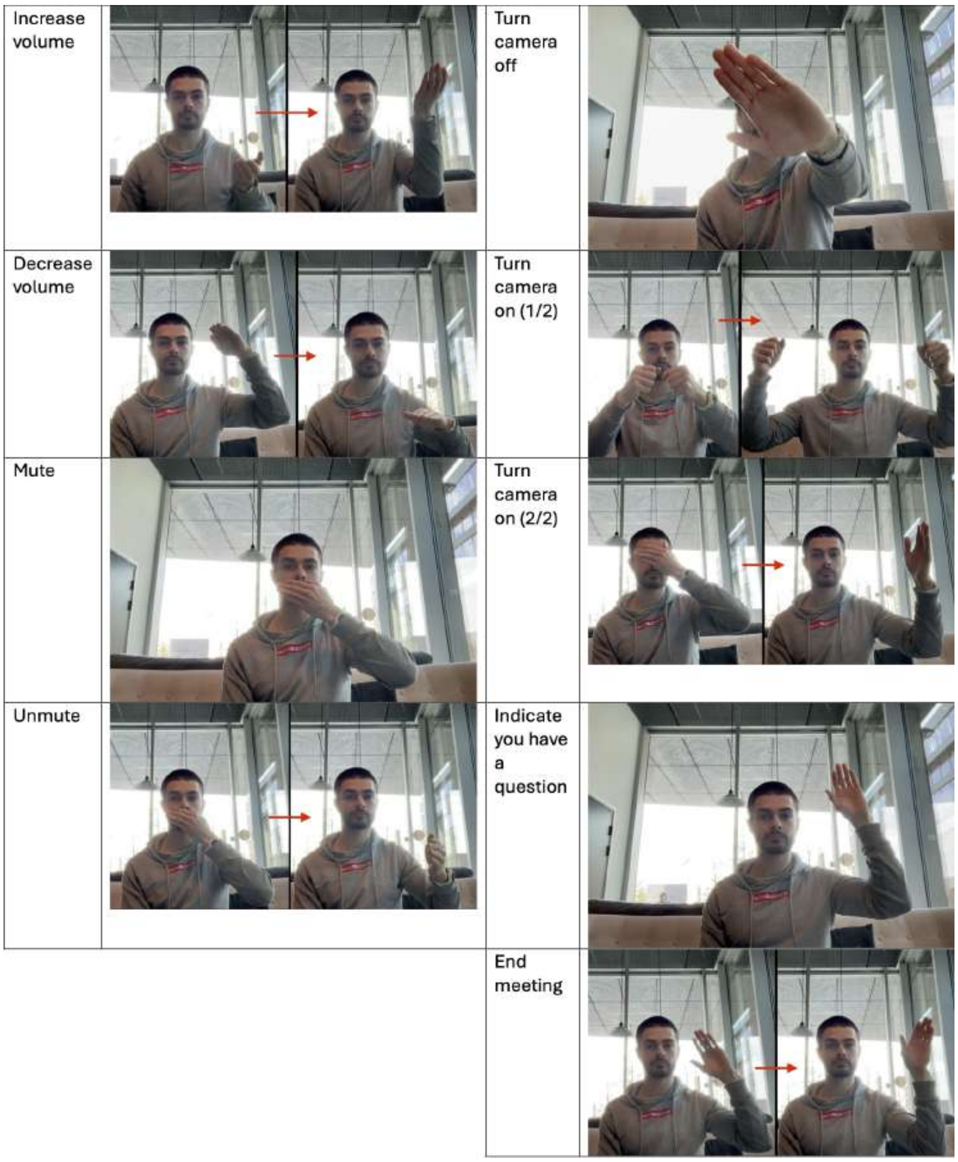



Figure 8.1: The figure shows the most popular gesture set from the data set "GestureVerse: Exploring the Multiverse of Interactive Gestures" (equal distribution of individuals that voted for two different gestures for turning on the camera)

# Bibliography


[1] Zhifan He et al. "A User-Defined Gesture Set for Natural Interaction in a Smart Kitchen Environment". In: *2020 13th International Symposium on Computational Intelligence and Design (ISCID)*. 2020, pp. 122–125. DOI: 10.1109/ISCID51228.2020.00034.

[2] Andreas Löcken et al. "User-centred process for the definition of free-hand gestures applied to controlling music playback". In: *Multimedia Syst.* 18 (Feb. 2012), pp. 15–31. DOI: 10.1007/s00530-011-0240-2.

[3] Anna Pereira et al. "A User-Developed 3-D Hand Gesture Set for Human-Computer Interaction". In: *Human factors* 57 (June 2015), pp. 607–21. DOI: 10.1177/0018720814559307.

[4] Huiyue Wu et al. "Influence of cultural factors on freehand gesture design". In: *International Journal of Human-Computer Studies* 143 (2020), p. 102502. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2020.102502. URL: https://www.sciencedirect.com/science/article/pii/S107158192030104X.

[5] Abdullah Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. ""I Am Iron Man": Priming Improves the Learnability and Memorability of User-Elicited Gestures". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445758. URL: https://doi-org.ep.ituproxy.kb.dk/10.1145/3411764.3445758.

[6] Huiyue Wu, Jianmin Wang, and Xiaolong Zhang. "User-centered gesture development in TV viewing environment". In: *Multimedia Tools and Applications* 75 (Nov. 2014), pp. 1–28. DOI: 10.1007/s11042-014-2323-5.

[7] Radu-Daniel Vatavu and Jean Vanderdonckt. "What Gestures Do Users with Visual Impairments Prefer to Interact with Smart Devices? And How Much We Know About It". In: *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*. DIS' 20 Companion. Eindhoven, Netherlands: Association for Computing Machinery, 2020, pp. 85–90. ISBN: 9781450379878. DOI: 10.1145/3393914.3395896. URL: https://doi-org.ep.ituproxy.kb.dk/10.1145/3393914.3395896.

[8] Santiago Villarreal-Narvaez et al. "Brave New GES World: A Systematic Literature Review of Gestures and Referents in Gesture Elicitation Studies". In: *ACM Comput. Surv.* 56.5 (2024). ISSN: 0360-0300. DOI: 10.1145/3636458. URL: https://doi-org.ep.ituproxy.kb.dk/10.1145/3636458.

[9] Jung In Koh et al. "Developing a Hand Gesture Recognition System for Mapping Symbolic Hand Gestures to Analogous Emojis in Computer-Mediated Communication". In: *ACM Trans. Interact. Intell. Syst.* 9.1 (Mar. 2019). ISSN: 2160-6455. DOI: 10.1145/3297277. URL: https://doi.org/10.1145/3297277.

[10] Vik Parthiban et al. "Gestural-Vocal Coordinated Interaction on Large Displays". In: *Companion of the 2022 ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS '22 Companion. Sophia Antipolis, France: Association for Computing Machinery, 2022, pp. 26–32. ISBN: 9781450390316. DOI: 10.1145/3531706.3536457. URL: https://doi-org.ep.ituproxy.kb.dk/10.1145/3531706.3536457.

[11] Salvatore Andolina et al. "Experimenting with Large Displays and Gestural Interaction in the Smart Factory". In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. 2019, pp. 2864–2869. DOI: 10.1109/SMC.2019.8913900.

[12] Francesca Pulina and Fabio Paternò. "Supporting cross-device interactions with gestures between personal and public devices". In: *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. MUM '19. Pisa, Italy: Association for Computing Machinery, 2019. ISBN: 9781450376242. DOI: 10.1145/3365610.3368419. URL: https://doi-org.ep.ituproxy.kb.dk/10.1145/3365610.3368419.

[13] William Delamare et al. "On Gesture Combination: An Exploration of a Solution to Augment Gesture Interaction". In: *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. ISS '19. Daejeon, Republic of Korea: Association for Computing Machinery, 2019, pp. 135–146. ISBN: 9781450368919. DOI: 10.1145/3343055.3359706. URL: https://doi-org.ep.ituproxy.kb.dk/10.1145/3343055.3359706.

[14] Antonio Emmanuele Uva et al. "A User-Centered Framework for Designing Midair Gesture Interfaces". In: *IEEE Transactions on Human-Machine Systems* 49.5 (2019), pp. 421–429. DOI: 10.1109/THMS.2019.2919719.

[15] Daria Krasovskaya. *Complete Guide to Human-Computer Interaction (HCI)*. en-US. July 2023. URL: https://blog.uxtweak.com/what-is-human-computer-interaction/ (visited on 05/13/2024).

[16] M. Tomitsch et al. *Design. Think. Make. Break. Repeat. A Handbook of Methods.* Revised. Amsterdam: BIS Publishers, 2020.

[17] Amit Konar and Sriparna Saha. *Gesture Recognition: Principles, Techniques and Applications*. 1st ed. 2018. Studies in Computational Intelligence 724. Cham: Springer International Publishing : Imprint: Springer, 2018. ISBN: 9783319622125.

[18] Google. *Gesture recognition guide for Python*. Accessed on: 09-05-2024. 2023. URL: https://developers.google.com/mediapipe/solutions/vision/gesture_recognizer/python.

[19] Google for Developers. *How to train a new model for gesture recognition - ML on Android with MediaPipe*. Apr. 2023. URL: https://www.youtube.com/watch?v=vODSFXEP-XY.

[20] Jakob Nielsen. *Thinking Aloud: The #1 Usability Tool*. en. URL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/.

[21] Sara Ramaswamy and Maria Rosala. *The Wizard of Oz Method in UX*. en. URL: https://www.nngroup.com/articles/wizard-of-oz/.

[22] Icaro de Vasconcelos Brito et al. "Analysis of Cross-Cultural Effect on Gesture-Based Human-Robot Interaction". In: *International Journal of Mechanical Engineering and Robotics Research* (Jan. 2019), pp. 852–859. DOI: 10.18178/ijmerr.8.6.852-859.

[23] IBM. *Cramér's V*. Accessed on: 05-05-2024. 2024. URL: https://www.ibm.com/docs/en/cognos-analytics/12.0.0?topic=terms-cramrs-v.

[24] Marek Strba. *Usability Testing Studies: How Many Participants?* en-US. Jan. 2024. URL: https://blog.uxtweak.com/usability-studies-how-many-participants-are-enough/ (visited on 05/13/2024).

[25] Google. *MediaPipe Model Maker*. Accessed on: 2024-05-02. URL: https://developers.google.com/mediapipe/solutions/model_maker.

[26] Google. *MediaPipe Gesture Recognizer Customization*. Accessed on: 2024-05-02. URL: https://developers.google.com/mediapipe/solutions/customization/gesture_recognizer.

[27] Alexander S. Gillis. *What is Transfer Learning? | Definition from TechTarget*. en. URL: https://www.techtarget.com/searchcio/definition/transfer-learning (visited on 05/02/2024).

[28] Diego Vaquero-Melchor and Ana M. Bernardos. "Enhancing Interaction with Augmented Reality through Mid-Air Haptic Feedback: Architecture Design and User Feedback". eng. In: *Applied sciences* 9.23 (2019), pp. 5123–. ISSN: 2076-3417.

# Appendix A

# Appendix

## A.1 Correlation analysis program (gesture elicitation)

### A.1.1 Imports

```python
# Import the data analysis libraries.
import pandas as pd
from IPython.display import display as ipy_display, Markdown, HTML
from ydata_profiling import ProfileReport
from sklearn.preprocessing import LabelEncoder

# Import for correlation matrix
from pandas import DataFrame
import scipy.stats as ss
import numpy as np

import imgkit

# Import for PCA stuff
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from matplotlib.backends.backend_pdf import PdfPages
```

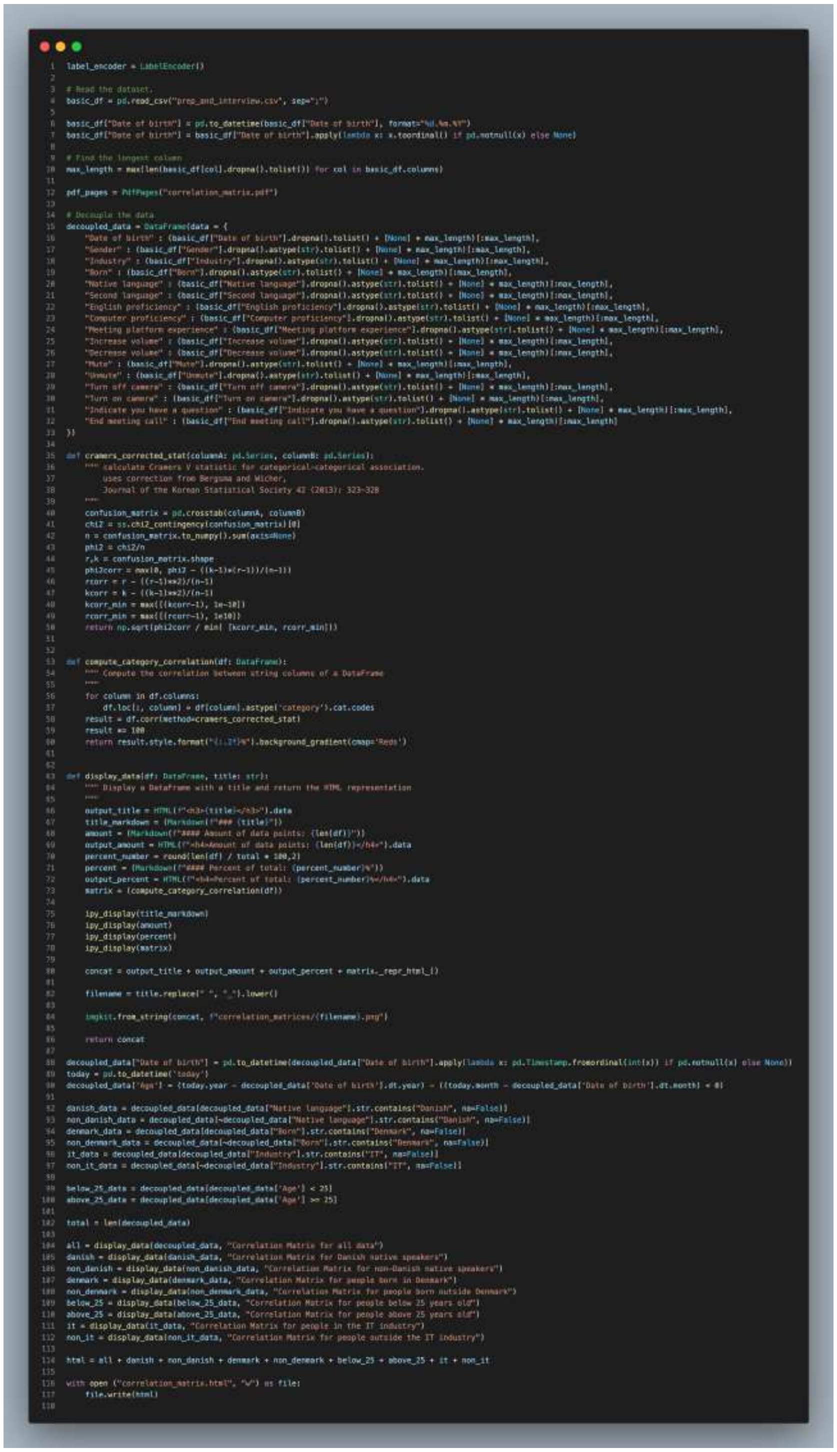
The figure shows the imports needed for the following code snippets (gesture elicitation)

### A.1.2 Creation of analysis

```
label_encoder = LabelEncoder()

# Read the dataset.
basic_df = pd.read_csv("prep_and_interview.csv", sep=";")
ipy_display(basic_df.info())
report = ProfileReport(basic_df, explorative=True)
report.to_notebook_iframe()
report.to_file("report.html")
```

The figure shows the code for creating the exploratory analysis (gesture elicitation)

### A.1.3 Correlation Matrices



```python
label_encoder = LabelEncoder()

# Read the dataset.
basic_df = pd.read_csv("prep_and_interview.csv", sep=";")

basic_df["Date of birth"] = pd.to_datetime(basic_df["Date of birth"], format="%d.%m.%Y")
basic_df["Date of birth"] = basic_df["Date of birth"].apply(lambda x: x.toordinal() if pd.notnull(x) else None)

# Find the longest column
max_length = max([len(basic_df[col].dropna().tolist()) for col in basic_df.columns])

pdf_pages = PdfPages("correlation_matrix.pdf")

# Decouple the data
decoupled_data = DataFrame(data = {
    "Date of birth" : (basic_df["Date of birth"].dropna().tolist() + [None] * max_length)[:max_length],
    "Gender" : (basic_df["Gender"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Industry" : (basic_df["Industry"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Born" : (basic_df["Born"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Native language" : (basic_df["Native language"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Second language" : (basic_df["Second language"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "English proficiency" : (basic_df["English proficiency"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Computer proficiency" : (basic_df["Computer proficiency"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Meeting platform experience" : (basic_df["Meeting platform experience"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Increase volume" : (basic_df["Increase volume"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Decrease volume" : (basic_df["Decrease volume"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Mute" : (basic_df["Mute"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Unmute" : (basic_df["Unmute"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Turn off camera" : (basic_df["Turn off camera"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Turn on camera" : (basic_df["Turn on camera"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Indicate you have a question" : (basic_df["Indicate you have a question"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "End meeting call" : (basic_df["End meeting call"].dropna().astype(str).tolist() + [None] * max_length)[:max_length]
})

def cramers_corrected_stat(columnA: pd.Series, columnB: pd.Series):
    """ calculate Cramers V statistic for categorical-categorical association.
        uses correction from Bergsma and Wicher,
        Journal of the Korean Statistical Society 42 (2013): 323-328
    """
    confusion_matrix = pd.crosstab(columnA, columnB)
    chi2 = ss.chi2_contingency(confusion_matrix)[0]
    n = confusion_matrix.to_numpy().sum(axis=None)
    phi2 = chi2/n
    r,k = confusion_matrix.shape
    phi2corr = max(0, phi2 - ((k-1)*(r-1))/(n-1))
    rcorr = r - ((r-1)**2)/(n-1)
    kcorr = k - ((k-1)**2)/(n-1)
    kcorr_min = max([(kcorr-1), 1e-10])
    rcorr_min = max([(rcorr-1), 1e10])
    return np.sqrt(phi2corr / min( [kcorr_min, rcorr_min]))


def compute_category_correlation(df: DataFrame):
    """ Compute the correlation between string columns of a DataFrame
    """
    for column in df.columns:
        df.loc[:, column] = df[column].astype('category').cat.codes
    result = df.corr(method=cramers_corrected_stat)
    result *= 100
    return result.style.format("{:.2f}%").background_gradient(cmap='Reds')


def display_data(df: DataFrame, title: str):
    """ Display a DataFrame with a title and return the HTML representation
    """
    output_title = HTML(f"<h3>{title}</h3>").data
    title_markdown = (Markdown(f"### {title}"))
    amount = (Markdown(f"#### Amount of data points: {len(df)}"))
    output_amount = HTML(f"<h4>Amount of data points: {len(df)}</h4>").data
    percent_number = round(len(df) / total * 100,2)
    percent = (Markdown(f"#### Percent of total: {percent_number}%"))
    output_percent = HTML(f"<h4>Percent of total: {percent_number}%</h4>").data
    matrix = (compute_category_correlation(df))

    ipy_display(title_markdown)
    ipy_display(amount)
    ipy_display(percent)
    ipy_display(matrix)

    concat = output_title + output_amount + output_percent + matrix._repr_html_()

    filename = title.replace(" ", "_").lower()

    imgkit.from_string(concat, f"correlation_matrices/{filename}.png")

    return concat

decoupled_data["Date of birth"] = pd.to_datetime(decoupled_data["Date of birth"].apply(lambda x: pd.Timestamp.fromordinal(int(x)) if pd.notnull(x) else None))
today = pd.to_datetime('today')
decoupled_data['Age'] = (today.year - decoupled_data['Date of birth'].dt.year) - ((today.month - decoupled_data['Date of birth'].dt.month) < 0)

danish_data = decoupled_data[decoupled_data["Native language"].str.contains("Danish", na=False)]
non_danish_data = decoupled_data[~decoupled_data["Native language"].str.contains("Danish", na=False)]
denmark_data = decoupled_data[decoupled_data["Born"].str.contains("Denmark", na=False)]
non_denmark_data = decoupled_data[~decoupled_data["Born"].str.contains("Denmark", na=False)]
it_data = decoupled_data[decoupled_data["Industry"].str.contains("IT", na=False)]
non_it_data = decoupled_data[~decoupled_data["Industry"].str.contains("IT", na=False)]

below_25_data = decoupled_data[decoupled_data['Age'] < 25]
above_25_data = decoupled_data[decoupled_data['Age'] >= 25]

total = len(decoupled_data)

all = display_data(decoupled_data, "Correlation Matrix for all data")
danish = display_data(danish_data, "Correlation Matrix for Danish native speakers")
non_danish = display_data(non_danish_data, "Correlation Matrix for non-Danish native speakers")
denmark = display_data(denmark_data, "Correlation Matrix for people born in Denmark")
non_denmark = display_data(non_denmark_data, "Correlation Matrix for people born outside Denmark")
below_25 = display_data(below_25_data, "Correlation Matrix for people below 25 years old")
above_25 = display_data(above_25_data, "Correlation Matrix for people above 25 years old")
it = display_data(it_data, "Correlation Matrix for people in the IT industry")
non_it = display_data(non_it_data, "Correlation Matrix for people outside the IT industry")

html = all + danish + non_danish + denmark + non_denmark + below_25 + above_25 + it + non_it

with open ("correlation_matrix.html", "w") as file:
    file.write(html)
```

The figure shows the code for creating correlation matrices (gesture elicitation)

### A.1.4 Tokenisation

```python
df = pd.read_csv('prep_and_interview.csv', sep=';')

df['Year of birth'] = pd.to_datetime(df['Date of birth'], format='%d.%m.%Y').dt.year
df = df.drop(columns=['Date of birth'])

# Assuming your DataFrame is named 'df'
# Assuming 'string_columns' is a list containing the names of string columns
string_columns = ['Gender', 'Industry', 'Born', 'Native language', 'Second language',
                  'English proficiency', 'Computer proficiency', 'Meeting platform experience',
                  'Increase volume', 'Decrease volume','Mute','Unmute','Turn off camera',
                  'Turn on camera','Indicate you have a question','End meeting call']

# Create a copy of the DataFrame to preserve the original data
df_encoded = df.copy()

# Iterate through each string column and apply LabelEncoder
for column in string_columns:
    encoder = LabelEncoder()
    df_encoded[column] = encoder.fit_transform(df[column])

# Step 4: Save the tokenized DataFrame to a new CSV file
df_encoded.to_csv('prep_and_interview_tokens.csv', index=False)
```

The figure shows code for tokenizing the CSV file with strings to numbers (gesture elicitation)

### A.1.5 PCA

```python
%matplotlib inline
pdf = PdfPages('PCA.pdf')
df = pd.read_csv("prep_and_interview_tokens.csv")

scaler = StandardScaler()
scaled = scaler.fit_transform(df), None
scaled_year = scaler.fit_transform(df), "Year of birth"
scaled_year_in_denmark = scaler.fit_transform(df), "Years in Denmark"
scaled_born = scaler.fit_transform(df), "Born"
scaled_native_language = scaler.fit_transform(df), "Native language"
scaled_second_language = scaler.fit_transform(df), "Second language"
scaled_inc_volume = scaler.fit_transform(df), "Increase volume"
scaled_dec_volume = scaler.fit_transform(df), "Decrease volume"
scaled_mute = scaler.fit_transform(df), "Mute"
scaled_unmute = scaler.fit_transform(df), "Unmute"
scaled_camera_off = scaler.fit_transform(df), "Turn off camera"
scaled_camera_on = scaler.fit_transform(df), "Turn on camera"
scaled_question = scaler.fit_transform(df), "Indicate you have a question"
scaled_end_call = scaler.fit_transform(df), "End meeting call"
scaled_industry = scaler.fit_transform(df), "Industry"

components = 2

def explained_variance_ratio(pca):
    return (f"Explained Variance Ratio:{pca.explained_variance_ratio_}")

def apply_pca(scaled_data):
    pca = PCA(n_components=components)
    pca.fit(scaled_data)
    variance = explained_variance_ratio(pca)
    transformed_data = pca.transform(scaled_data)
    return transformed_data, variance


def plot_pca(plot, array=None):
    scaled_data = plot[0]
    name = plot[1]
    transformed_data, variance = apply_pca(scaled_data)
    float_formatter = "{:.2f}".format
    np.set_printoptions(formatter={'float_kind':float_formatter})
    print(transformed_data)
    if name is not None:
        # If array is not None, use it, otherwise use the column from the dataframe
        color_values = array if array is not None else df[name]

        plt.scatter(transformed_data[:, 0], transformed_data[:, 1], c=color_values, cmap='viridis')
        plt.title(f'Explained Variance Ratio by Number of Components ({name})')
    else:
        plt.scatter(transformed_data[:, 0], transformed_data[:, 1], marker='o')
        plt.title('Explained Variance Ratio by Number of Components')
    plt.figtext(0,0.01,variance)
    plt.subplots_adjust(bottom=0.25)
    plt.xlabel("Number of Components")
    plt.ylabel("Explained Variance Ratio")
    plt.tight_layout()
    pdf.savefig()
    plt.show()

plot_pca(scaled)
plot_pca(scaled_year)
plot_pca(scaled_year_in_denmark)
plot_pca(scaled_born)
plot_pca(scaled_native_language)
plot_pca(scaled_second_language)
plot_pca(scaled_inc_volume)
plot_pca(scaled_dec_volume)
plot_pca(scaled_mute)
plot_pca(scaled_unmute)
plot_pca(scaled_camera_off)
plot_pca(scaled_camera_on)
plot_pca(scaled_question)
plot_pca(scaled_end_call)
plot_pca(scaled_industry)

pdf.close()
```

The figure shows code for creating PCAs (gesture elicitation)

## A.2 Correlation analysis program (A/B test)

```python
# Import the data analysis libraries.
import pandas as pd
from IPython.display import display as ipy_display, Markdown, HTML
from ydata_profiling import ProfileReport
from sklearn.preprocessing import LabelEncoder

# Import for correlation matrix
from pandas import DataFrame
import scipy.stats as ss
import numpy as np

# Import for PCA stuff
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from matplotlib.backends.backend_pdf import PdfPages
```

The figure shows the imports needed for the following code snippets (a/b test)

```python
label_encoder = LabelEncoder()

# Read the dataset.
basic_df = pd.read_csv("ab_test.csv", sep=";")
ipy_display(basic_df.info())
report = ProfileReport(basic_df, explorative=True)
report.to_notebook_iframe()
report.to_file("report.html")
```

The figure shows the code for creating the exploratory analysis (a/b test)

```python
label_encoder = LabelEncoder()

# Read the dataset.
basic_df = pd.read_csv("ab_test.csv", sep=";")

basic_df["Date of birth"] = pd.to_datetime(basic_df["Date of birth"], format="%d.%m.%Y")
basic_df["Date of birth"] = basic_df["Date of birth"].apply(lambda x: x.toordinal() if pd.notnull(x) else None)

# Find the longest column
max_length = max(len(basic_df[col].dropna().tolist()) for col in basic_df.columns)

pdf_pages = PdfPages("correlation_matrix.pdf")

# Decouple the data
decoupled_data = DataFrame(data = {
    "Date of birth" : (basic_df["Date of birth"].dropna().tolist() + [None] * max_length)[:max_length],
    "Gender" : (basic_df["Gender"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Industry" : (basic_df["Industry"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Born" : (basic_df["Born"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Native language" : (basic_df["Native language"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Second language" : (basic_df["Second language"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "English proficiency" : (basic_df["English proficiency"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Computer proficiency" : (basic_df["Computer proficiency"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Meeting platform experience" : (basic_df["Meeting platform experience"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Increase volume" : (basic_df["Increase volume"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Decrease volume" : (basic_df["Decrease volume"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Mute" : (basic_df["Mute"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Unmute" : (basic_df["Unmute"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Turn off camera" : (basic_df["Turn off camera"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Turn on camera" : (basic_df["Turn on camera"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "Indicate you have a question" : (basic_df["Indicate you have a question"].dropna().astype(str).tolist() + [None] * max_length)[:max_length],
    "End meeting call" : (basic_df["End meeting call"].dropna().astype(str).tolist() + [None] * max_length)[:max_length]
})

def cramers_corrected_stat(columnA: pd.Series, columnB: pd.Series):
    """ calculate Cramers V statistic for categorical-categorical association.
        uses correction from Bergsma and Wicher,
        Journal of the Korean Statistical Society 42 (2013): 323-328
    """
    confusion_matrix = pd.crosstab(columnA, columnB)
    chi2 = ss.chi2_contingency(confusion_matrix)[0]
    n = confusion_matrix.to_numpy().sum(axis=None)
    phi2 = chi2/n
    r,k = confusion_matrix.shape
    phi2corr = max(0, phi2 - ((k-1)*(r-1))/(n-1))
    rcorr = r - ((r-1)**2)/(n-1)
    kcorr = k - ((k-1)**2)/(n-1)
    kcorr_min = max([(kcorr-1), 1e-10])
    rcorr_min = max([(rcorr-1), 1e10])
    return np.sqrt(phi2corr / min( [kcorr_min, rcorr_min]))


def compute_category_correlation(df: DataFrame):
    """ Compute the correlation between string columns of a DataFrame
    """
    for column in df.columns:
        df.loc[:, column] = df[column].astype('category').cat.codes
    result = df.corr(method=cramers_corrected_stat)
    result *= 100
    return result.style.format("{:.2f}%").background_gradient(cmap='Reds')


def display_data(df: DataFrame, title: str):
    """ Display a DataFrame with a title and return the HTML representation
    """
    output_title = HTML(f"<h3>{title}</h3>").data
    title_markdown = (Markdown(f"### {title}"))
    amount = (Markdown(f"#### Amount of data points: {len(df)}"))
    output_amount = HTML(f"<h4>Amount of data points: {len(df)}</h4>").data
    percent_number = round(len(df) / total * 100,2)
    percent = (Markdown(f"#### Percent of total: {percent_number}%"))
    output_percent = HTML(f"<h4>Percent of total: {percent_number}%</h4>").data
    matrix = (compute_category_correlation(df))

    ipy_display(title_markdown)
    ipy_display(amount)
    ipy_display(percent)
    ipy_display(matrix)

    concat = output_title + output_amount + output_percent + matrix._repr_html_()

    filename = title.replace(" ", "_").lower()

    # imgkit.from_string(concat, f"correlation_matrices/{filename}.png")

    return concat

decoupled_data["Date of birth"] = pd.to_datetime(decoupled_data["Date of birth"].apply(lambda x: pd.Timestamp.fromordinal(int(x)) if pd.notnull(x) else None))
today = pd.to_datetime('today')
decoupled_data['Age'] = (today.year - decoupled_data['Date of birth'].dt.year) - ((today.month - decoupled_data['Date of birth'].dt.month) < 0)

danish_data = decoupled_data[decoupled_data["Native language"].str.contains("Danish", na=False)]
non_danish_data = decoupled_data[~decoupled_data["Native language"].str.contains("Danish", na=False)]
denmark_data = decoupled_data[decoupled_data["Born"].str.contains("Denmark", na=False)]
non_denmark_data = decoupled_data[~decoupled_data["Born"].str.contains("Denmark", na=False)]
it_data = decoupled_data[decoupled_data["Industry"].str.contains("IT", na=False)]
non_it_data = decoupled_data[~decoupled_data["Industry"].str.contains("IT", na=False)]

below_25_data = decoupled_data[decoupled_data['Age'] < 25]
above_25_data = decoupled_data[decoupled_data['Age'] >= 25]

total = len(decoupled_data)

all = display_data(decoupled_data, "Correlation Matrix for all data")
danish = display_data(danish_data, "Correlation Matrix for Danish native speakers")
non_danish = display_data(non_danish_data, "Correlation Matrix for non-Danish native speakers")
denmark = display_data(denmark_data, "Correlation Matrix for people born in Denmark")
non_denmark = display_data(non_denmark_data, "Correlation Matrix for people born outside Denmark")
below_25 = display_data(below_25_data, "Correlation Matrix for people below 25 years old")
above_25 = display_data(above_25_data, "Correlation Matrix for people above 25 years old")
it = display_data(it_data, "Correlation Matrix for people in the IT industry")
non_it = display_data(non_it_data, "Correlation Matrix for people outside the IT industry")

html = all + danish + non_danish + denmark + non_denmark + below_25 + above_25 + it + non_it

with open ("correlation_matrix.html", "w") as file:
    file.write(html)
```

The figure shows code for creating correlation matrices (a/b test)

```python
df = pd.read_csv('ab_test.csv', sep=';')

df['Year of birth'] = pd.to_datetime(df['Date of birth'], format='%d.%m.%Y').dt.year
df = df.drop(columns=['Date of birth'])

# Assuming your DataFrame is named 'df'
# Assuming 'string_columns' is a list containing the names of string columns
string_columns = ['Gender', 'Industry', 'Born', 'Native language', 'Second language',
                  'English proficiency', 'Computer proficiency', 'Meeting platform experience',
                  'Increase volume', 'Decrease volume','Mute','Unmute','Turn off camera',
                  'Turn on camera','Indicate you have a question','End meeting call']

# Create a copy of the DataFrame to preserve the original data
df_encoded = df.copy()

# Iterate through each string column and apply LabelEncoder
for column in string_columns:
    encoder = LabelEncoder()
    df_encoded[column] = encoder.fit_transform(df[column])

# Step 4: Save the tokenized DataFrame to a new CSV file
df_encoded.to_csv('ab_test_tokens.csv', index=False)
```

The figure shows code for tokenizing the CSV file with strings to numbers (a/b test)

```python
%matplotlib inline
pdf = PdfPages('PCA.pdf')
df = pd.read_csv("ab_test_tokens.csv")

scaler = StandardScaler()
scaled = scaler.fit_transform(df), None
scaled_year = scaler.fit_transform(df), "Year of birth"
scaled_year_in_denmark = scaler.fit_transform(df), "Years in Denmark"
scaled_born = scaler.fit_transform(df), "Born"
scaled_native_language = scaler.fit_transform(df), "Native language"
scaled_second_language = scaler.fit_transform(df), "Second language"
scaled_inc_volume = scaler.fit_transform(df), "Increase volume"
scaled_dec_volume = scaler.fit_transform(df), "Decrease volume"
scaled_mute = scaler.fit_transform(df), "Mute"
scaled_unmute = scaler.fit_transform(df), "Unmute"
scaled_camera_off = scaler.fit_transform(df), "Turn off camera"
scaled_camera_on = scaler.fit_transform(df), "Turn on camera"
scaled_question = scaler.fit_transform(df), "Indicate you have a question"
scaled_end_call = scaler.fit_transform(df), "End meeting call"
scaled_industry = scaler.fit_transform(df), "Industry"

components = 2

def explained_variance_ratio(pca):
    return (f"Explained Variance Ratio:{pca.explained_variance_ratio_}")

def apply_pca(scaled_data):
    pca = PCA(n_components=components)
    pca.fit(scaled_data)
    variance = explained_variance_ratio(pca)
    transformed_data = pca.transform(scaled_data)
    return transformed_data, variance


def plot_pca(plot, array=None):
    scaled_data = plot[0]
    name = plot[1]
    transformed_data, variance = apply_pca(scaled_data)
    float_formatter = "{:.2f}".format
    np.set_printoptions(formatter={'float_kind':float_formatter})
    print(transformed_data)
    if name is not None:
        # If array is not None, use it, otherwise use the column from the dataframe
        color_values = array if array is not None else df[name]

        plt.scatter(transformed_data[:, 0], transformed_data[:, 1], c=color_values, cmap='viridis')
        plt.title(f'Explained Variance Ratio by Number of Components ({name})')
    else:
        plt.scatter(transformed_data[:, 0], transformed_data[:, 1], marker='o')
        plt.title('Explained Variance Ratio by Number of Components')
    plt.figtext(0,0.01,variance)
    plt.subplots_adjust(bottom=0.25)
    plt.xlabel("Number of Components")
    plt.ylabel("Explained Variance Ratio")
    plt.tight_layout()
    pdf.savefig()
    plt.show()

plot_pca(scaled)
plot_pca(scaled_year)
plot_pca(scaled_year_in_denmark)
plot_pca(scaled_born)
plot_pca(scaled_native_language)
plot_pca(scaled_second_language)
plot_pca(scaled_inc_volume)
plot_pca(scaled_dec_volume)
plot_pca(scaled_mute)
plot_pca(scaled_unmute)
plot_pca(scaled_camera_off)
plot_pca(scaled_camera_on)
plot_pca(scaled_question)
plot_pca(scaled_end_call)
plot_pca(scaled_industry)

pdf.close()
```

The figure shows the code for creating PCAs for the A/B test

## A.3 Research protocol

The following pdf displays the research protocol followed for the literature review

**Research Protocol**

**Literature Review on Hand Gestures Used in Video-Based Applications**

### 1 Introduction

This document provides an overview of the steps to develop the literature review and presents some criteria for searching and cataloging results.

This research is part of an extensive study with the primary objective of proposing a universal hand gesture vocabulary for collaborative products and environments.

The current research will explore the recent literature to collect all hand gestures used in interaction, control, and manipulation in video-based applications in domains defined in the Scope Section.

### 2 Literature Research Objectives

- Collect and extract data from publications from as many countries as possible that present any hand gestures vocabulary used in video-based applications or studies in selected domains.
- Explore the set of hand gestures found applying some classifications and exploring cultural observations. Investigate hand gesture convergences and variations.
- Propose a hand gesture vocabulary set, based on the founds, to be applied in a user experience to study the feasibility of a universal vocabulary.

### 3 Object, Scope, and Domain

The following literature review has studied object-collecting hand gesture vocabulary, which means <u>any description of a hand gesture pose or temporal movement (including image, text or numeric description, graphic signal, etc.) used to perform an interaction with the related system and its description</u>.

The scope of the study is focused on the literature on video-based applications, indicating that <u>the core functionality revolves around the video processing of hand images</u> involving video cameras. Other sensors, such as rings, armbands, etc., also based on other technologies such as sound waves, infrared, etc, are excluded.

The main domain for the master project is focused on collaboration systems and environments for hybrid and virtual meetings. As this domain lacks publications, opening the search to <u>related domains that are paired in terms of equipment, systems, and interactions</u> is necessary. The related lists are a guide for included and excluded vide-based domains, but not limited to:

Included Domains

Meeting platforms (for virtual and hybrid meetings).

Collaborative products (systems).

Cameras in general (with HG).

Online education platforms.

Virtual, Augmented, and Mixed reality systems.

| | |
|---|---|
| Excluded Domains | Medical (surgery and related). |
| Manipulative interfaces in open scenarios (e.g., museums, shopping displays, etc.) | Automotive interfaces (navigation or media). |
| Industrial (project and manipulation of 3D objects). | Signal language. |

## 4 Literature Search Strategy

### 4.1 Terminology Definition

The current protocol starts with a pre-list of terms, keywords, and synonyms related to this study to compose the search script. The terms of any part of the search can be updated as other related terms are found during the search.

The object: *Hand gesture vocabulary*

a) Synonyms and variations for <u>hand gestures</u>:

*Hand gesture, hand gestures, gesture, gestural, mid-air gesture, in-air gesture, manual signal, hand-sign, manual gesture, gesture-based, hand motion, mid-air hand motion,*

b) Synonyms and variations for <u>vocabulary</u>:

*Vocabulary, set, lexicon, lexical, commands, control, pose, movement, list, repertoire,*

The scope: *Video-based application.*

c) Synonyms and variations for <u>video-based</u>:

*Video-based, visual-based, camera-based, video-driven, video-centric, video-enabled*

d) Synonyms and variations for <u>applications</u>:

*Application, software, system, program, solution, tool, platform, interaction, control, interface.*

The domain: Meeting platforms (virtual and hybrid meetings) and collaborative products (systems). Cameras in general (with HG). Online education platforms. Virtual, Augmented, and Mixed reality systems. Smart televisions.

e) Synonyms and variations for <u>virtual and hybrid, augmented etc</u>:

*Virtual, hybrid, augmented, mixed, online, web, remote, unified, synchronous, digital, web-based, online-based, immersive, cyberspace, distributed, Internet-mediated.*

f) Synonyms and variations for <u>meeting</u>:

*Meeting, conferencing, collaboration, collaborative, conference, conferencing, telepresence.*

g) Synonyms and variations for <u>platforms</u>:

*Platform, system, application, prototype, solution, tool, environment, room, framework, interface, portal, infrastructure, channel, product.*

h) Synonyms and variations for cameras:

*Camera, image devices, digital images, optical sensor, imaging system, capture device, imaging apparatus, visual recording equipment.*

i) Synonyms and variations for online (use as complement to item e):

*Online, e-, distance, remote, digital, web-based, internet-based, cyber.*

j) Synonyms and variations for education:
*Education, educational, learning,*

k) Synonyms and variations for virtual reality and others:

Reality, environment, VR, Virtual Reality, AR, Augmented Reality, MR, Mixed Reality, VR environment, AR environment, Cyber.

## 4.2 Inclusion and Exclusion Criteria

This search's inclusion and exclusion criteria are related to the quality of the information needed, cover date, and language.

Inclusion criteria:

a) Papers that are over the scope and domain defined in item 4.
b) Publication in English language.
c) Publication from 2019 to present.
d) Present one or more hand gestures with any description in a described context that can identified as a gesture vocabulary for the related artifact.

Exclusion criteria:

a) Papers that are out of the scope and domain defined in item 4.
b) The publication cites hand gestures as an interaction but does not describe them.
c) The hand gestures are far from the scope of this study.
d) The hand gestures presented are not enough described to be collected.

## 4.3 Databases

The search can explore publications that cover any kind of results, such as other research, study cases, frameworks, methodologies, techniques, prototypes, applications, systems, etc., for the last five years (2019 to 2014). The search string is used in the advanced search tool of the databases.

The online databases used are the ones provided by ITU that can be found at https://kub.kb.dk/itb/databaser, which include:

- ACM Digital Library
- IEEE
- Lecture Notes in Computer Science
- Springer
- Taylor and Francis Online Journals
- SAGE Journals Online
- Infomedia
- Google Scholar

The first step will be based on IEEE and ACM databases. If necessary to expand the results, the search will be repeated on other list databases until the study is complete.

### 4.4 Search String

- IEEE: (https://ieeexplore.ieee.org/search/advanced/command)

```
(("hand gesture*" OR "gestural" OR "mid-air" OR
"gesture-based") AND ("vocabular*" OR "lexic*" OR "gesture
set*" OR "gesture command*" OR "command set"))
```

- ACM:

```
AllField:("hand gesture*" OR "gestural" OR "mid-air" OR
"gesture-based") AND AllField:("vocabular*" OR "lexic*" OR
"gesture set*" OR "gesture command*" OR "command set")
```

## 5 Search Execution and Data Collection

- Conduct a systematic search of literature databases and sources using the defined search strategy.
- Divide the screen search process into blocks, for example, by date, domain, database, etc. If necessary, the search string can be adapted.
- Save all the possible papers to share using the following file nomenclature: FirstAutorSurnameYYYY.pdf (e.g. Barchovisk2022.pdf).
- The information extracted from the publications will be inserted in an online FORM that will be shared. Read all the information needed in the form before start reading the publications.
    - The data to be collected is closely related to the found hand gestures and some extra information regarding the domain, interaction, equipment, application, and country, among others, that will vary from case to case according to the publication type.

## 6 Synthesis of Knowledge and Data Exploration

To synthesize the findings from the literature review in the form of knowledge, it is necessary to organize the relevant information and apply curiosity over the research content, trying different analyses. Some questions to answer can be:

- What are the most common hand gestures in general?
- What are the most common gestures that can be seen in different domains? Are they used for the same kind of interaction?
- What are the most common interactions between the domains? Can a correlation be found between the command interaction and the hand gesture in the funds?
- How did the author report the use of the presented gestures? What measurement or method is used to evaluate the gestures or gesture interface?

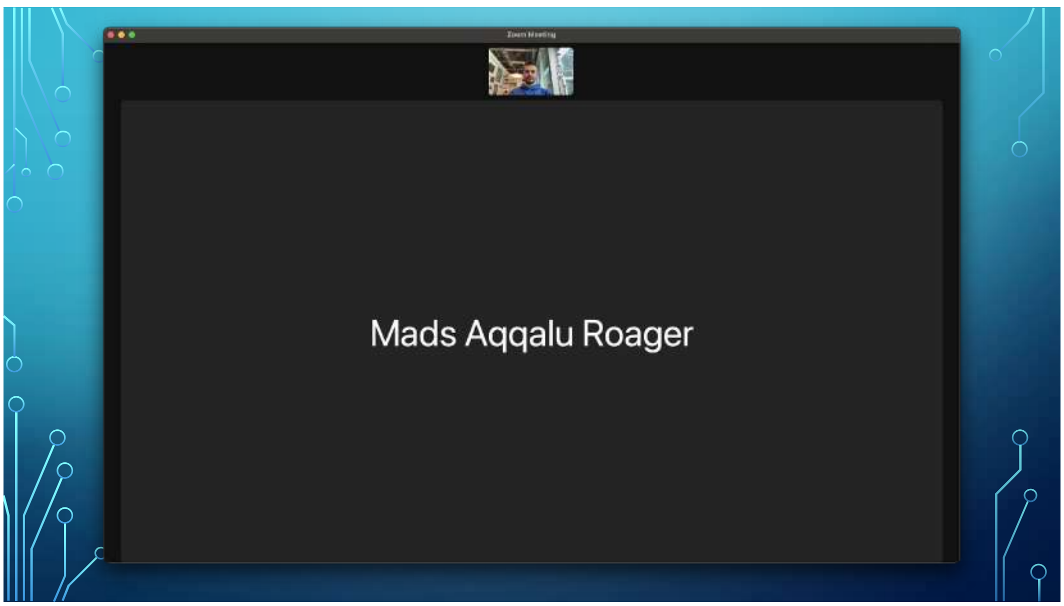
- Is it possible to identify some tendency of gestures in a group of publications with users from the same country or region?
- What are the most common hand gestures applications? Final products? Prototypes? Systems? Think here about how hand gestures have been explored in the fields.
- Think about clustering the data to explore common commands or gestures with different meanings to analyze.

It is possible to transform the collected information into citation sentences of knowledge. Relate the findings by answering the previous questions and adding more extra information to enrich the knowledge. After it is possible to discuss each case. Follow this sample with an imaginary citation example:

- Question: What are the most common hand gestures in general?
- Citation: *According to our research, the most common hand gesture found among 42 publications in different domains is related to the "rise hand gesture" that was largely used in meeting platforms (Silva2022, Souza2019, Campben2002, Cillasd2023) followed by online education platforms (Author5YEAR, Author6YEAR, Author7YEAR) and Virtual reality (Author9YEAR, Author9YEAR). However, the "rise hand gesture" is applied for different commands across those platforms. The most common interaction is related to asking for talk, and the most isolated case is related to turning off the camera. More examples can be seen in the table X.*

## A.4 Wizard of Oz presentation

The following pdf shows the Wizard of Oz simulation presented during the gesture elicitation




Mads Aqqalu Roager


Mads Aqqalu Roager


Mads Aqqalu Roager


Mads Aqqalu Roager


Mads Aqqalu Roager


Mads Aqqalu Roager


Mads Aqqalu Roager

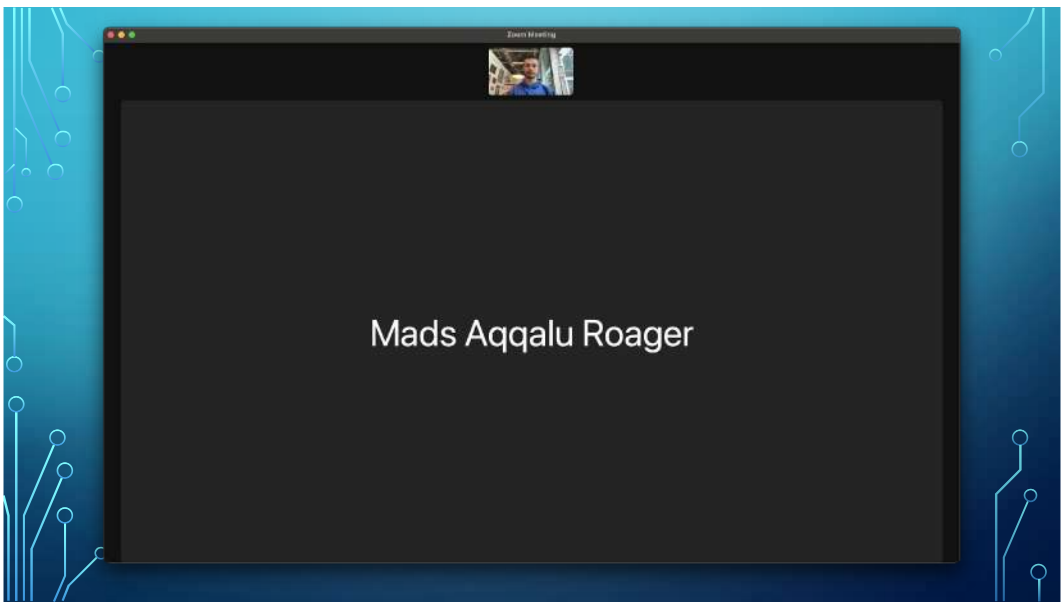

Mads Aqqalu Roager

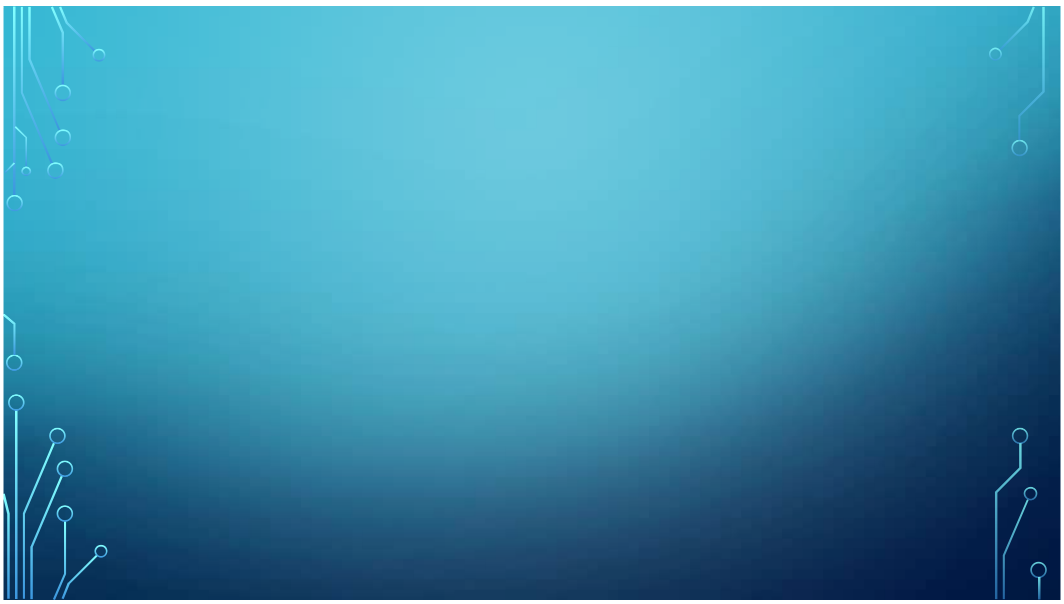

Zoom Meeting 40-minutes
View
L
Mads Aqqalu Roager
Mute
Start Video
Security
Participants
Chat
Share Screen
Summary
Record
Reactions
Apps
Whiteboards
Notes
End


Mads Aqqalu Roager


Mads Aqqalu Roager


Mads Aqqalu Roager


Mads Aqqalu Roager



## A.5 ChatGPT: vocabulary to csv

**You**

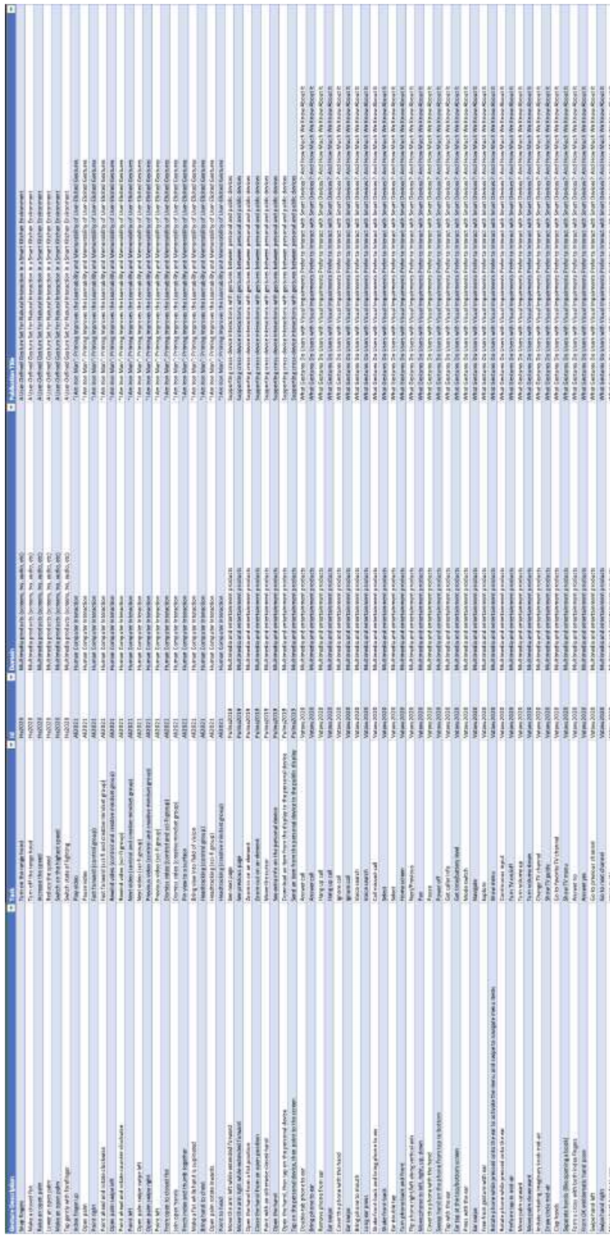
User-centered gesture development in TV viewing environment
Wu2016
Multimedia and enterteinment products (cameras, screens, TVs, audio, etc), Human Computer Interaction (foccus on method, project, etc. )
1 - Point at TV with one hand - Turn on TV (1)
2 - Two open palms facing the TV spreading outwards, away from the body - Turn on TV (2)
3 - Point at TV with one hand - Turn on TV (1)
4 - Two open palms facing the TV moving inwards towards the body - Turn off TV (2)
5 - One opena palm facing TV - Menu (1)
6 - Finger snap - Menu (2)
7 - Okay sign - Confirm (1)
8 - Push open palm towards TV - Confirm (2)
9 - Swipe right - Next channel (1)
10 - Swipe left - Next channel (2)
11 - Swipe left - Previous channel (1)
12 - Swipe right - Previous channel (2)
13 - Open palms put together pointing up - Mute (1)
14 - Forming a T with two hands - Mute (2)
15 - Raise open palm facing TV - Volume up
16 - Lower open palm facing TV - Volume down

**ChatGPT**

| Gesture Description | Task Description | ID | Domain | Publication Title |
|---|---|---|---|---|
| Point at TV with one hand | Turn on TV (1) | Wu2016 | Multimedia and Entertainment, Human Computer Interaction | User-centered gesture development in TV viewing environment |
| Two open palms facing the TV spreading outwards, away from the body | Turn on TV (2) | Wu2016 | Multimedia and Entertainment, Human Computer Interaction | User-centered gesture development in TV viewing environment |
| Two open palms facing the TV moving inwards towards the body | Turn off TV (2) | Wu2016 | Multimedia and Entertainment, Human Computer Interaction | User-centered gesture development in TV viewing environment |
| One open palm facing TV | Menu (1) | Wu2016 | Multimedia and Entertainment, Human Computer Interaction | User-centered gesture development in TV viewing environment |
| Finger snap | Menu (2) | Wu2016 | Multimedia and Entertainment, Human Computer Interaction | User-centered gesture development in TV viewing environment |
| Okay sign | Confirm (1) | Wu2016 | Multimedia and Entertainment, Human Computer Interaction | User-centered gesture development in TV viewing environment |
| Push open palm towards TV | Confirm (2) | Wu2016 | Multimedia and Entertainment, Human Computer Interaction | User-centered gesture development in TV viewing |

The figure shows a part of the ChatGPT conversation to get the vocabulary extracted from each of the papers into a CSV format, to combine it into a single file afterwards

## A.6 Vocabulary from papers



The figure shows a snippet of the vocabulary extracted with ChatGPT, as mentioned in Appendix A.5, from all of the papers described in the related work section

## A.7 Gesure Elicitation form

The following pdf shows the Microsoft form used for the gesture elicitation study

# Gesture elicitation study

3. apr. 2024

This study aims to understand the user's perceptions regarding hand gestures used for the interaction process in the collaborative domain. The general demographic information collected in this research is confidential and will be used for internal academic research only.

* Påkrævet

### Pre experiment

# GDPR

**Hand Gesture Recognition Project: Authorization and Consent Form**

**Introduction:** This document outlines the terms and conditions of the authorization and consent for using your image and personal data in ITU and GN's Hand Gesture Recognition Project. Please read this document carefully before providing your consent.

**1. Project Overview:** The Hand Gesture Recognition Project aims to study cross-cultural hand gesture interactions in virtual and hybrid meeting contexts.

**2. Data Collection and Usage:**

- **Types of Data:** We will collect images and videos that include your hand gestures while participating in the activities. Additionally, we will collect personal information and opinions. Note that the face will be blurred out.
- **Purpose of Data Collection:**
  - Research and Development: The data collected will be used for research and development purposes related to hand gesture recognition technology and;
  - Demographic Analysis: The additional personal information will be analyzed to understand the relationships between demographic factors and hand gesture interactions.

**3. GDPR Compliance:** This data collection and processing comply with the General Data Protection Regulation rules. Your data will be treated with the utmost confidentiality, and personally identifiable information will be anonymized.

**4. Confidentiality:** I acknowledge that ITU and GN will take reasonable steps to maintain the confidentiality and security of the collected data.

**5. Consent:** I, the undersigned, hereby grant my voluntary and informed consent for the collection, processing, and use of my image data and the additional personal information specified above in the Hand Gesture Recognition Project conducted by Elizabete Munzlinger, Mads Aqqalu, and Lucas Frey Torres Hanson, associated with ITU and GN.

**6. Withdrawal of Consent:** I understand I have the right to withdraw my consent at any time by notifying Elizabete Munzlinger at munzlinger@itu.dk , Mads Aqqalu Roager at mroa@itu.dk , or Lucas Frey Torres Hanson at luha@itu.dk in writing. Withdrawal of consent will not affect the lawfulness of processing based on consent before its withdrawal.

1

Do you consent to the above? *

- ◯ Yes
- ◯ No

## Personal Information

All of your information is highly confidential and for internal use only. Please contact us if you have any concerns.

2

Full name
*

3

Date of birth *

4

What gender do you identify as? *

- Female
- Male
- Non-binary
- Other/Prefer not to say

5

Which will best describe your status? *

- ◯ Student
- ◯ Employed
- ◯ Unemployed
- ◯ Retired
- ◯ Andet

6

What industry do you work?

- [ ] Arts
- [ ] Business
- [ ] Commerce
- [ ] Education
- [ ] Energy
- [ ] Engineering
- [ ] Financial
- [ ] Government
- [ ] Hospitality/Healthcare
- [ ] IT (Information Technology)
- [ ] Manufacturing
- [ ] Media/Entertainment
- [ ] Non-profit/NGO
- [ ] Public service
- [ ] Retail
- [ ] Transportation
- [ ] Andet

7

Email *

## Origin/Residence Information

8

In which country where you born? *

9

How many years did you stay in said country? *

10

How many years have you been in Denmark? *

## Spoken Language

11

What is your native language? *

12

What is your second language (if any)?

13

How proficient are you at English? *

- ◯ Beginner (understand common phrases, ask simple questions, and engage in basic interactions)
- ◯ Intermediate (speak and understand reasonably well, use basic tenses)
- ◯ Advanced (fluently in almost any situation, and produce clear, detailed texts on challenging subject)
- ◯ Native speaker

## User Proficiency and Experience

14

How proficient are you at using a computer? *

- ◯ Sporadic user
- ◯ Regular user
- ◯ Advanced user
- ◯ Programmer/Developer

15

Which of the following virtual meeting platforms have you used before
*

- [ ] Bitrix24
- [ ] Cisco Webex
- [ ] Zoom
- [ ] Skype
- [ ] Livestorm
- [ ] Microsoft Teams
- [ ] Zoho meeting
- [ ] GoToMeeting
- [ ] Lucid Meetings
- [ ] HubSpot Meeting
- [ ] Adobe Connect
- [ ] WebinarJam
- [ ] Google Meet
- [ ] Join.me
- [ ] TeamViewer Meeting
- [ ] Top of Form
- [ ] Discord
- [ ] Andet

# Experiment

In this section we will have you complete a few tasks related to virtual meetings.

# Experiment Impressions

16

On a scale of 1 to 5 (1 being Bad, 5 being Good), how would you rate your experience with using meeting platforms? *

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|

17

Have you ever used a hand gestures interface in a technological context (e.g., for gaming or other applications)? *

- ◯ Yes
- ◯ No

18

If yes, please describe your experience. How effective and intuitive was the hand gestures interface, and for what purpose did you use it?

19

How likely will you to use hand gesture commands in meetings (1 not likely, 5 very likely)? *

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|

20

How relevant do you think it is to be able to perform below actions with the help of gestures? *

| | Very irrelevant | Somewhat irrelevant | Neutral | Somewl |
|---|---|---|---|---|
| Changing the volume | ○ | ○ | ○ | |
| Mute and unmute your side of the call | ○ | ○ | ○ | |
| Turn on and off camera | ○ | ○ | ○ | |
| Raising the hand | ○ | ○ | ○ | |
| End call | | | | |

How inclined would you be to use hand commands in a professional setting if they were available as a tool during meetings (1 very unlikely, 5 very likely)? *

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|

22

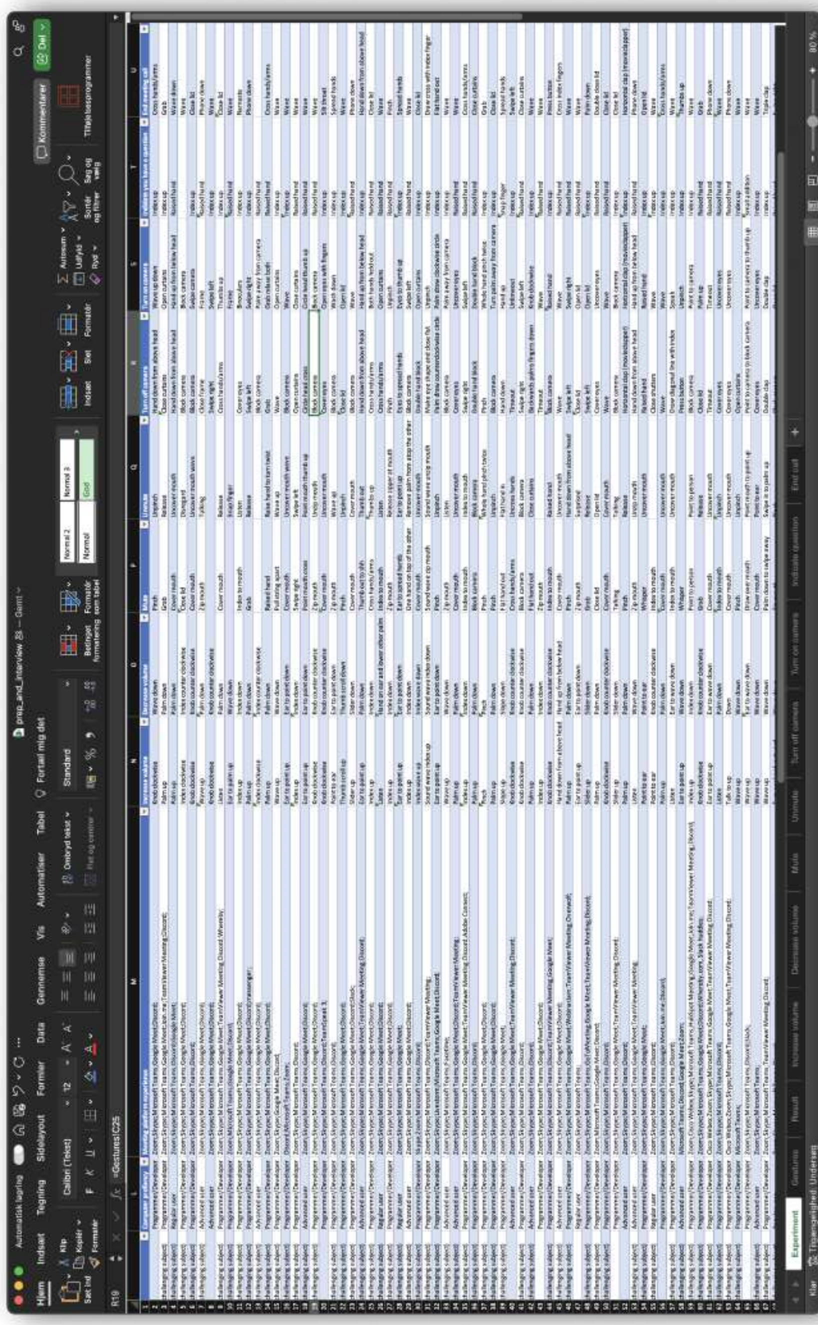
Are there specific situations or tasks in meetings where you believe hand commands could be particularly useful? Please provide examples together with an idea for a gesture.

# Concluding Feedback

23

Is there anything else you would like to share regarding your experience with technology, meetings, or potential improvements in collaboration tools?

## A.8 Data from elicitation study



The figure shows a snippet of the data gathered from the gesture elicitation study

## A.9 Keyboard shortcuts in virtual meeting clients

The table shows the available keyboard shortcuts available in virtual meeting clients

| | Mute/Unmute | End call | Toggle Camera on/off | Share screen | Start recording | Raise hand |
|---|---|---|---|---|---|---|
| **Zoom** | alt + a | alt + q | alt + v | alt + s | alt + r | alt + y |
| **Google Meet** | No binding | No binding | No binding | No binding | No binding | No binding |
| **Microsoft Teams** | ctrl + shift + m | ctrl + shift + h | ctrl + shift + o | ctrl + shift + e | No binding | ctrl + shift + k |

## A.10 A/B Test form

The following pdf shows the Microsoft form used for the A/B test

# A/B Test

* Påkrævet

### GDPR

**Hand Gesture Recognition Project: Authorization and Consent Form**

**Introduction:** This document outlines the terms and conditions of the authorization and consent for using your personal data in ITU and GN's Hand Gesture Recognition Project. Please read this document carefully before providing your consent.

**1. Project Overview:** The Hand Gesture Recognition Project aims to study cross-cultural hand gesture interactions in virtual and hybrid meeting contexts.

**2. Data Collection and Usage:**

- **Types of Data:** We will collect personal information and opinions.
- **Purpose of Data Collection:**
  - Research and Development: The data collected will be used for research and development purposes related to hand gesture recognition technology and;
  - Demographic Analysis: The additional personal information will be analyzed to understand the relationships between demographic factors and hand gesture interactions.

**3. GDPR Compliance:** This data collection and processing comply with the General Data Protection Regulation rules. Your data will be treated with the utmost confidentiality, and personally identifiable information will be anonymized.

**4. Confidentiality:** I acknowledge that ITU and GN will take reasonable steps to maintain the confidentiality and security of the collected data.

**5. Consent:** I, the undersigned, hereby grant my voluntary and informed consent for the collection, processing, and use of my image data and the additional personal information specified above in the Hand Gesture Recognition Project conducted by Elizabete Munzlinger, Mads Aqqalu, and Lucas Frey Torres Hanson, associated with ITU and GN.

**6. Withdrawal of Consent:** I understand I have the right to withdraw my consent at any time by notifying Elizabete Munzlinger at munzlinger@itu.dk , Mads Aqqalu Roager at mroa@itu.dk , or Lucas Frey Torres Hanson at luha@itu.dk in writing. Withdrawal of consent will not affect the lawfulness of processing based on consent before its withdrawal.

**7. Duration of Consent:** This consent is *valid* for the duration of my participation in the Hand Gesture Recognition Project and a reasonable period thereafter for project evaluation and analysis, and following the principle in the national "Danish Code of Conduct for Research Integrity, pkt. 2.1." where the research data must be stored for a minimum of 5 years after publication for responsibility management purposes

1. Do you consent to the above? *

○ No

○ Yes

Student Status

2. Are you a university student? *

◯ Yes

◯ No

## Personal information

All of your information is highly confidential and for internal use only. Please contact us if you have any concerns.

3. Date of birth *

4. What gender do you identify as? *

- ○ Female
- ○ Male
- ○ Non-binary
- ○ Other/Prefer not to say

5. What industry do you work/study in? *

- ☐ Arts
- ☐ Business
- ☐ Commerce
- ☐ Education
- ☐ Energy
- ☐ Engineering
- ☐ Financial
- ☐ Government
- ☐ Hospitality/Healthcare
- ☐ IT (Information Technology)
- ☐ Manufacturing
- ☐ Non-profit/NGO
- ☐ Public service
- ☐ Transportation
- ☐ Andet

6. In which country where you born? *

7. How many years did you stay in said country (please only write the number and not "years")? *

8. How many years have you been in Denmark (please only write the number and not "years")? *

9. What is your native language? *

10. What is your second language (if any)? *

11. How proficient are you at English *

- ◯ Beginner (understand common phrases, ask simple questions, and engage in basic interactions)
- ◯ Intermediate (speak and understand reasonably well, use basic tenses)
- ◯ Advanced (fluently in almost any situation, and produce clear, detailed texts on challenging subject)
- ◯ Native speaker

12. How proficient are you at using a computer? *

- ◯ Sporadic user
- ◯ Regular user
- ◯ Advanced user
- ◯ Programmer/Developer

13. Which of the following virtual meeting platforms have you used before? *

- [ ] Bitrix24
- [ ] Cisco Webex
- [ ] Zoom
- [ ] Skype
- [ ] Livestorm
- [ ] Microsoft Teams
- [ ] Zoho meeting
- [ ] GoToMeeting
- [ ] Lucid Meetings
- [ ] HubSpot Meeting
- [ ] Adobe Connect
- [ ] WebinarJam
- [ ] Google Meet
- [ ] Join.me
- [ ] TeamVeiwer Meeting
- [ ] Top of Form
- [ ] Discord
- [ ] Andet

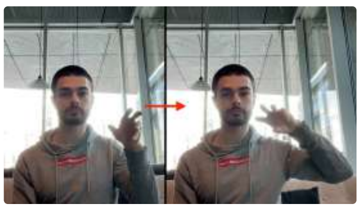
## Gesture Agreement

Thank you for helping us (Mads Roager and Lucas Hanson) with our bachelors, named Hand-Gesture-Based Interaction in Hybrid Meetings.
This survey will include two of the most commonly used gestures we have observed for specific tasks in our prior elicitation study.

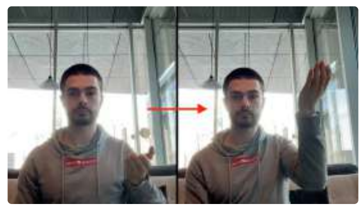
It is **important** that you understand the specific scenario which is:

Imagine you are sitting at a meeting at work. There are several of your coworkers sitting physically with you and a potential client is connected to the call. You are unable to reach the laptop connected to the meeting. But the laptop can recognize in air hand-gestures.

14. You have a hard time hearing what the client connected to the meeting says. You therefore want to increase the volume. Which of the following gestures would you expect to increase the volume? *



◯ Turn knob clockwise



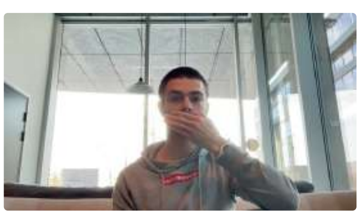
◯ Palm up

15. The volume is now a bit too loud. Which of the following gestures would you expect to decrease the volume? *

◯ Palm down

◯ Turn knob counter clockwise

16. After a bit of discussing back and forth with the client you want to discuss something with your coworkers without the client being able to hear what you are saying. Which of the following gestures would you expect to mute the microphone? *



◯ Cover mouth

◯ Zip mouth

17. You have reached a conclusion and want to unmute. Which of the following gestures would you expect to unmute the microphone? *

◯ Uncover mouth

◯ Unpinch

18. The call has been going on for quite a while and you have a short break. In the meantime you do not want the client to see what you are doing and therefore want to turn the camera off. Which of the following gestures would you expect to turn off the camera? *

◯ Cover eyes (can also be with two hands)

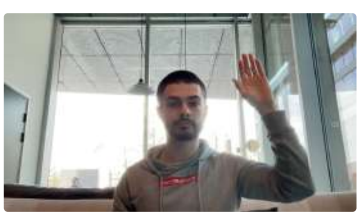
◯ Block camera

19. The break is over and you have to turn on the camera again. Which of the following gestures would you expect to turn on the camera? *

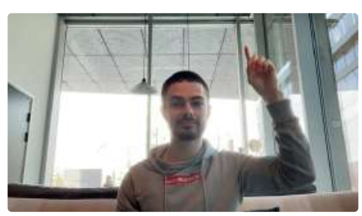
◯ Open curtains

◯ Uncover eyes (can also be done with two hands)

20. Before you end the meeting you just want to ask a question without interrupting the current discussion. Which of the following gestures would you expect to indicated you have question? *



◯ Raise hand



◯ Raise index

21. The meeting is finally over. Which of the following gestures would you expect to end the call?
*

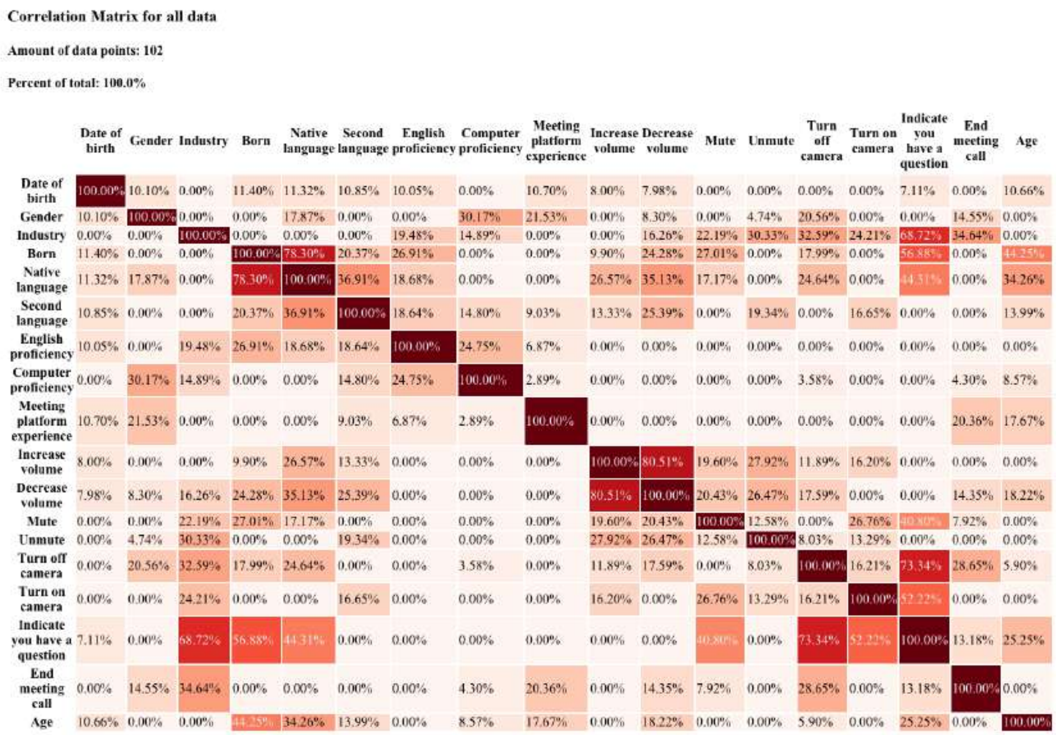
◯ Close lid

◯ Wave (can also be the open close hand wave)

Dette indhold er hverken oprettet eller påtegnet af Microsoft Corporation. De data, du indsender, sendes til ejeren af formularen.

Microsoft Forms

## A.11 Correlation Matrices (gesture elicitation)



**Correlation Matrix for all data**

**Amount of data points: 102**

**Percent of total: 100.0%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 10.10% | 0.00% | 11.40% | 11.32% | 10.85% | 10.05% | 0.00% | 10.70% | 8.00% | 7.98% | 0.00% | 0.00% | 0.00% | 0.00% | 7.11% | 0.00% | 10.66% |
| **Gender** | 10.10% | 100.00% | 0.00% | 0.00% | 17.87% | 0.00% | 0.00% | 30.17% | 21.53% | 0.00% | 8.30% | 0.00% | 4.74% | 20.56% | 0.00% | 0.00% | 14.55% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 19.48% | 14.89% | 0.00% | 0.00% | 16.26% | 22.19% | 30.33% | 32.59% | 24.21% | 68.72% | 34.64% | 0.00% |
| **Born** | 11.40% | 0.00% | 0.00% | 100.00% | 78.30% | 20.37% | 26.91% | 0.00% | 0.00% | 9.90% | 24.28% | 27.01% | 0.00% | 17.99% | 0.00% | 56.88% | 0.00% | 44.25% |
| **Native language** | 11.32% | 17.87% | 0.00% | 78.30% | 100.00% | 36.91% | 18.68% | 0.00% | 0.00% | 26.57% | 35.13% | 17.17% | 0.00% | 24.64% | 0.00% | 44.31% | 0.00% | 34.26% |
| **Second language** | 10.85% | 0.00% | 0.00% | 20.37% | 36.91% | 100.00% | 18.64% | 14.80% | 9.03% | 13.33% | 25.39% | 0.00% | 19.34% | 0.00% | 16.65% | 0.00% | 0.00% | 13.99% |
| **English proficiency** | 10.05% | 0.00% | 19.48% | 26.91% | 18.68% | 18.64% | 100.00% | 24.75% | 6.87% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Computer proficiency** | 0.00% | 30.17% | 14.89% | 0.00% | 0.00% | 14.80% | 24.75% | 100.00% | 2.89% | 0.00% | 0.00% | 0.00% | 0.00% | 3.58% | 0.00% | 0.00% | 4.30% | 8.57% |
| **Meeting platform experience** | 10.70% | 21.53% | 0.00% | 0.00% | 0.00% | 9.03% | 6.87% | 2.89% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 20.36% | 17.67% |
| **Increase volume** | 8.00% | 0.00% | 0.00% | 9.90% | 26.57% | 13.33% | 0.00% | 0.00% | 0.00% | 100.00% | 80.51% | 19.60% | 27.92% | 11.89% | 16.20% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 7.98% | 8.30% | 16.26% | 24.28% | 35.13% | 25.39% | 0.00% | 0.00% | 0.00% | 80.51% | 100.00% | 20.43% | 26.47% | 17.59% | 0.00% | 0.00% | 14.35% | 18.22% |
| **Mute** | 0.00% | 0.00% | 22.19% | 27.01% | 17.17% | 0.00% | 0.00% | 0.00% | 0.00% | 19.60% | 20.43% | 100.00% | 12.58% | 0.00% | 26.76% | 40.80% | 7.92% | 0.00% |
| **Unmute** | 0.00% | 4.74% | 30.33% | 0.00% | 0.00% | 19.34% | 0.00% | 0.00% | 0.00% | 27.92% | 26.47% | 12.58% | 100.00% | 8.03% | 13.29% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 20.56% | 32.59% | 17.99% | 24.64% | 0.00% | 0.00% | 3.58% | 0.00% | 11.89% | 17.59% | 0.00% | 8.03% | 100.00% | 16.21% | 73.34% | 28.65% | 5.90% |
| **Turn on camera** | 0.00% | 0.00% | 24.21% | 0.00% | 0.00% | 16.65% | 0.00% | 0.00% | 0.00% | 16.20% | 0.00% | 26.76% | 13.29% | 16.21% | 100.00% | 52.22% | 0.00% | 0.00% |
| **Indicate you have a question** | 7.11% | 0.00% | 68.72% | 56.88% | 44.31% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 40.80% | 0.00% | 73.34% | 52.22% | 100.00% | 13.18% | 25.25% |
| **End meeting call** | 0.00% | 14.55% | 34.64% | 0.00% | 0.00% | 0.00% | 0.00% | 4.30% | 20.36% | 0.00% | 14.35% | 7.92% | 0.00% | 28.65% | 0.00% | 13.18% | 100.00% | 0.00% |
| **Age** | 10.66% | 0.00% | 0.00% | 44.25% | 34.26% | 13.99% | 0.00% | 8.57% | 17.67% | 0.00% | 18.22% | 0.00% | 0.00% | 5.90% | 0.00% | 25.25% | 0.00% | 100.00% |

The figure shows the correlation matrix for all the data (gesture elicitation)

**Correlation Matrix for Danish native speakers**

**Amount of data points: 70**

**Percent of total: 68.63%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 12.22% | 0.00% | 12.40% | 12.22% | 13.13% | 12.22% | 0.00% | 14.34% | 9.45% | 8.56% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 13.02% |
| **Gender** | 12.22% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 39.49% | 16.61% | 14.01% | 21.04% | 0.00% | 0.00% | 21.64% | 0.00% | 0.00% | 20.70% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 31.55% | 7.36% | 0.00% | 0.00% | 18.91% | 17.84% | 43.11% | 20.15% | 0.00% | 59.67% | 23.86% | 0.00% |
| **Born** | 12.40% | 0.00% | 0.00% | 100.00% | 67.50% | 22.52% | 41.20% | 0.00% | 24.23% | 0.00% | 14.80% | 13.97% | 0.00% | 0.00% | 17.59% | 0.00% | 0.00% | 30.33% |
| **Native language** | 12.22% | 0.00% | 0.00% | 67.50% | 100.00% | 42.64% | 22.64% | 0.00% | 23.90% | 0.00% | 16.29% | 0.00% | 0.00% | 0.00% | 8.98% | 0.00% | 0.00% | 32.41% |
| **Second language** | 13.13% | 0.00% | 0.00% | 22.52% | 42.64% | 100.00% | 39.22% | 20.09% | 0.00% | 3.73% | 16.86% | 0.00% | 0.00% | 0.00% | 28.13% | 0.00% | 0.00% | 28.20% |
| **English proficiency** | 12.22% | 0.00% | 31.55% | 41.20% | 22.64% | 39.22% | 100.00% | 28.05% | 5.48% | 0.00% | 0.00% | 0.00% | 0.00% | 3.93% | 0.00% | 3.81% | 7.79% | 0.00% |
| **Computer proficiency** | 0.00% | 39.49% | 7.36% | 0.00% | 0.00% | 20.09% | 28.05% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 6.48% | 5.69% | 0.00% | 5.00% | 0.00% | 9.83% |
| **Meeting platform experience** | 14.34% | 16.61% | 0.00% | 24.23% | 23.90% | 0.00% | 5.48% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 2.84% | 27.05% |
| **Increase volume** | 9.45% | 14.01% | 0.00% | 0.00% | 0.00% | 3.73% | 0.00% | 0.00% | 0.00% | 100.00% | 81.10% | 23.42% | 33.82% | 18.28% | 13.53% | 0.00% | 17.67% | 0.00% |
| **Decrease volume** | 8.56% | 21.04% | 18.91% | 14.80% | 16.29% | 16.86% | 0.00% | 0.00% | 0.00% | 81.10% | 100.00% | 18.70% | 30.38% | 14.37% | 0.00% | 0.00% | 30.95% | 0.00% |
| **Mute** | 0.00% | 0.00% | 17.84% | 13.97% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 23.42% | 18.70% | 100.00% | 0.00% | 0.00% | 17.64% | 6.99% | 0.00% | 0.00% |
| **Unmute** | 0.00% | 0.00% | 43.11% | 0.00% | 0.00% | 0.00% | 0.00% | 6.48% | 0.00% | 33.82% | 30.38% | 0.00% | 100.00% | 29.40% | 32.16% | 49.97% | 9.32% | 0.00% |
| **Turn off camera** | 0.00% | 21.64% | 20.15% | 0.00% | 0.00% | 0.00% | 3.93% | 5.69% | 0.00% | 18.28% | 14.37% | 0.00% | 29.40% | 100.00% | 30.27% | 58.57% | 34.33% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 0.00% | 17.59% | 8.98% | 28.13% | 0.00% | 0.00% | 0.00% | 13.53% | 0.00% | 17.64% | 32.16% | 30.27% | 100.00% | 52.33% | 0.00% | 15.97% |
| **Indicate you have a question** | 0.00% | 0.00% | 59.67% | 0.00% | 0.00% | 0.00% | 3.81% | 5.00% | 0.00% | 0.00% | 0.00% | 6.99% | 49.97% | 58.57% | 52.33% | 100.00% | 0.00% | 12.47% |
| **End meeting call** | 0.00% | 20.70% | 23.86% | 0.00% | 0.00% | 0.00% | 7.79% | 0.00% | 2.84% | 17.67% | 30.95% | 0.00% | 9.32% | 34.33% | 0.00% | 0.00% | 100.00% | 0.00% |
| **Age** | 13.02% | 0.00% | 0.00% | 30.33% | 32.41% | 28.20% | 0.00% | 9.83% | 27.05% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 15.97% | 12.47% | 0.00% | 100.00% |

The figure shows the correlation matrix filtered on Danish native speakers (gesture elicitation)

**Correlation Matrix for non-Danish native speakers**

**Amount of data points: 32**

**Percent of total: 31.37%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 0.00% | 0.00% | 21.45% | 0.00% | 0.00% | 2.10% | 24.25% | 0.00% | 6.33% | 0.00% | 20.95% | 0.00% | 0.00% | 0.00% | 0.00% | 14.22% |
| **Industry** | 0.00% | 0.00% | 100.00% | 32.08% | 0.00% | 31.02% | 39.16% | 58.51% | 32.65% | 0.00% | 0.00% | 0.00% | 0.00% | 39.91% | 34.78% | 70.46% | 39.09% | 42.64% |
| **Born** | 0.00% | 0.00% | 32.08% | 100.00% | 55.27% | 0.00% | 24.43% | 0.00% | 31.64% | 0.00% | 0.00% | 20.34% | 13.02% | 0.00% | 36.56% | 41.77% | 0.00% | 24.68% |
| **Native language** | 0.00% | 21.45% | 0.00% | 55.27% | 100.00% | 0.00% | 43.86% | 0.00% | 32.66% | 40.60% | 41.46% | 13.34% | 0.00% | 15.02% | 13.75% | 0.00% | 0.00% | 0.00% |
| **Second language** | 0.00% | 0.00% | 31.02% | 0.00% | 0.00% | 100.00% | 0.00% | 37.85% | 32.38% | 31.69% | 30.05% | 0.00% | 24.49% | 23.61% | 0.00% | 0.00% | 29.10% | 0.00% |
| **English proficiency** | 0.00% | 0.00% | 39.16% | 24.43% | 43.86% | 0.00% | 100.00% | 0.00% | 26.20% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 17.93% | 0.00% | 0.00% | 0.00% |
| **Computer proficiency** | 0.00% | 2.10% | 58.51% | 0.00% | 0.00% | 37.85% | 0.00% | 100.00% | 20.45% | 0.00% | 0.00% | 0.00% | 0.00% | 18.30% | 0.00% | 0.00% | 29.87% | 0.00% |
| **Meeting platform experience** | 0.00% | 24.25% | 32.65% | 31.64% | 32.66% | 32.38% | 26.20% | 20.45% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 24.47% | 0.00% | 0.00% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 0.00% | 40.60% | 31.69% | 0.00% | 0.00% | 0.00% | 100.00% | 82.65% | 32.85% | 37.70% | 0.00% | 14.56% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 6.33% | 0.00% | 0.00% | 41.46% | 30.05% | 0.00% | 0.00% | 0.00% | 82.65% | 100.00% | 28.47% | 38.67% | 25.64% | 19.34% | 0.00% | 0.00% | 0.00% |
| **Mute** | 0.00% | 0.00% | 0.00% | 20.34% | 13.34% | 0.00% | 0.00% | 0.00% | 0.00% | 32.85% | 28.47% | 100.00% | 3.20% | 27.44% | 30.69% | 54.15% | 0.00% | 17.79% |
| **Unmute** | 0.00% | 20.95% | 0.00% | 13.02% | 0.00% | 24.49% | 0.00% | 0.00% | 0.00% | 37.70% | 38.67% | 3.20% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 39.91% | 0.00% | 15.02% | 23.61% | 0.00% | 18.30% | 0.00% | 0.00% | 25.64% | 27.44% | 0.00% | 100.00% | 25.91% | 48.44% | 0.00% | 32.02% |
| **Turn on camera** | 0.00% | 0.00% | 34.78% | 36.56% | 13.75% | 0.00% | 17.93% | 0.00% | 0.00% | 14.56% | 19.34% | 30.69% | 0.00% | 25.91% | 100.00% | 59.43% | 27.59% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 70.46% | 41.77% | 0.00% | 0.00% | 0.00% | 0.00% | 24.47% | 0.00% | 0.00% | 54.15% | 0.00% | 48.44% | 59.43% | 100.00% | 21.40% | 29.02% |
| **End meeting call** | 0.00% | 0.00% | 39.09% | 0.00% | 0.00% | 29.10% | 0.00% | 29.87% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 27.59% | 21.40% | 100.00% | 14.95% |
| **Age** | 0.00% | 14.22% | 42.64% | 24.68% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 17.79% | 0.00% | 32.02% | 0.00% | 29.02% | 14.95% | 100.00% |

The figure shows the correlation matrix filtered on non-danish native speakers (gesture elicitation)

**Correlation Matrix for people above 25 years old**

**Amount of data points: 33**

**Percent of total: 32.35%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 0.00% | 0.00% | 30.74% | 0.00% | 0.00% | 0.00% | 23.01% | 0.00% | 0.00% | 0.00% | 19.67% | 26.22% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 24.55% | 16.29% | 0.00% | 21.72% | 0.00% | 44.27% | 0.00% | 21.86% | 31.81% | 0.00% | 20.03% | 6.37% | 39.83% | 0.00% | 45.44% |
| **Born** | 0.00% | 0.00% | 24.55% | 100.00% | 73.36% | 39.14% | 44.60% | 0.00% | 0.00% | 5.06% | 24.49% | 38.43% | 0.00% | 34.00% | 0.00% | 54.95% | 0.00% | 48.55% |
| **Native language** | 0.00% | 30.74% | 16.29% | 73.36% | 100.00% | 67.74% | 0.00% | 0.00% | 30.00% | 0.00% | 0.00% | 6.01% | 33.85% | 40.46% | 0.00% | 21.78% | 0.00% | 34.57% |
| **Second language** | 0.00% | 0.00% | 0.00% | 39.14% | 67.74% | 100.00% | 0.00% | 0.00% | 25.63% | 0.00% | 10.76% | 0.00% | 38.84% | 30.26% | 17.58% | 0.00% | 0.00% | 19.35% |
| **English proficiency** | 0.00% | 0.00% | 21.72% | 44.60% | 0.00% | 0.00% | 100.00% | 53.10% | 13.48% | 0.00% | 0.00% | 12.07% | 0.00% | 0.00% | 0.00% | 0.00% | 5.86% | 0.00% |
| **Computer proficiency** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 53.10% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 14.28% | 23.19% | 9.59% | 5.35% | 6.34% |
| **Meeting platform experience** | 0.00% | 23.01% | 44.27% | 0.00% | 30.00% | 25.63% | 13.48% | 0.00% | 100.00% | 0.00% | 15.54% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 32.71% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 5.06% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 64.00% | 0.00% | 50.60% | 0.00% | 34.68% | 0.00% | 22.13% | 11.13% |
| **Decrease volume** | 0.00% | 0.00% | 21.86% | 24.49% | 0.00% | 10.76% | 0.00% | 0.00% | 15.54% | 64.00% | 100.00% | 0.00% | 13.72% | 0.00% | 23.40% | 0.00% | 21.99% | 33.38% |
| **Mute** | 0.00% | 0.00% | 31.81% | 38.43% | 6.01% | 0.00% | 12.07% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 6.09% | 13.71% | 37.40% | 0.00% | 20.96% | 0.00% |
| **Unmute** | 0.00% | 19.67% | 0.00% | 0.00% | 33.85% | 38.84% | 0.00% | 0.00% | 0.00% | 50.60% | 13.72% | 6.09% | 100.00% | 20.31% | 42.39% | 0.00% | 0.00% | 16.14% |
| **Turn off camera** | 0.00% | 26.22% | 20.03% | 34.00% | 40.46% | 30.26% | 0.00% | 14.28% | 0.00% | 0.00% | 0.00% | 13.71% | 20.31% | 100.00% | 0.00% | 55.11% | 0.00% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 6.37% | 0.00% | 0.00% | 17.58% | 0.00% | 23.19% | 0.00% | 34.68% | 23.40% | 37.40% | 42.39% | 0.00% | 100.00% | 0.00% | 29.50% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 39.83% | 54.95% | 21.78% | 0.00% | 0.00% | 9.59% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 55.11% | 0.00% | 100.00% | 27.52% | 35.14% |
| **End meeting call** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 5.86% | 5.35% | 0.00% | 22.13% | 21.99% | 20.96% | 0.00% | 0.00% | 29.50% | 27.52% | 100.00% | 0.00% |
| **Age** | 0.00% | 0.00% | 45.44% | 48.55% | 34.57% | 19.35% | 0.00% | 6.34% | 32.71% | 11.13% | 33.38% | 0.00% | 16.14% | 0.00% | 0.00% | 35.14% | 0.00% | 100.00% |

The figure shows the correlation matrix filtered on people above 25 years old (gesture elicitation)

**Correlation Matrix for people below 25 years old**

**Amount of data points: 69**

**Percent of total: 67.65%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 12.31% | 0.00% | 13.87% | 13.87% | 13.48% | 12.31% | 0.00% | 12.57% | 9.59% | 9.90% | 0.00% | 3.76% | 0.00% | 0.00% | 7.05% | 0.00% | 12.50% |
| **Gender** | 12.31% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 35.38% | 19.64% | 13.63% | 16.49% | 0.00% | 0.00% | 18.98% | 0.00% | 0.00% | 21.61% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 41.09% | 17.47% | 0.00% | 0.00% | 0.00% | 20.92% | 30.47% | 27.84% | 27.44% | 67.88% | 34.11% | 0.00% |
| **Born** | 13.87% | 0.00% | 0.00% | 100.00% | 90.29% | 29.70% | 11.33% | 0.00% | 15.92% | 18.31% | 29.25% | 26.26% | 0.00% | 2.00% | 0.00% | 43.07% | 0.00% | 21.68% |
| **Native language** | 13.87% | 0.00% | 0.00% | 90.29% | 100.00% | 19.31% | 0.00% | 0.00% | 0.00% | 24.88% | 35.31% | 30.98% | 0.00% | 19.97% | 9.49% | 41.84% | 0.00% | 16.04% |
| **Second language** | 13.48% | 0.00% | 0.00% | 29.70% | 19.31% | 100.00% | 13.93% | 31.81% | 17.04% | 23.10% | 28.72% | 0.00% | 17.69% | 0.00% | 11.99% | 0.00% | 0.00% | 10.23% |
| **English proficiency** | 12.31% | 0.00% | 41.09% | 11.33% | 0.00% | 13.93% | 100.00% | 9.32% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 4.56% | 9.06% | 0.00% |
| **Computer proficiency** | 0.00% | 35.38% | 17.47% | 0.00% | 0.00% | 31.81% | 9.32% | 100.00% | 3.56% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 15.41% |
| **Meeting platform experience** | 12.57% | 19.64% | 0.00% | 15.92% | 0.00% | 17.04% | 0.00% | 3.56% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 20.30% | 0.00% | 17.26% | 22.42% | 0.00% |
| **Increase volume** | 9.59% | 13.63% | 0.00% | 18.31% | 24.88% | 23.10% | 0.00% | 0.00% | 0.00% | 100.00% | 84.85% | 30.54% | 23.41% | 21.87% | 0.00% | 0.00% | 4.70% | 27.38% |
| **Decrease volume** | 9.90% | 16.49% | 0.00% | 29.25% | 35.31% | 28.72% | 0.00% | 0.00% | 0.00% | 84.85% | 100.00% | 33.56% | 27.44% | 27.47% | 0.00% | 0.00% | 19.86% | 18.68% |
| **Mute** | 0.00% | 0.00% | 20.92% | 26.26% | 30.98% | 0.00% | 0.00% | 0.00% | 0.00% | 30.54% | 33.56% | 100.00% | 8.07% | 0.00% | 23.28% | 53.88% | 0.00% | 19.77% |
| **Unmute** | 3.76% | 0.00% | 30.47% | 0.00% | 0.00% | 17.69% | 0.00% | 0.00% | 0.00% | 23.41% | 27.44% | 8.07% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 18.98% | 27.84% | 2.00% | 19.97% | 0.00% | 0.00% | 0.00% | 20.30% | 21.87% | 27.47% | 0.00% | 0.00% | 100.00% | 19.62% | 65.29% | 34.30% | 25.55% |
| **Turn on camera** | 0.00% | 0.00% | 27.44% | 0.00% | 9.49% | 11.99% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 23.28% | 0.00% | 19.62% | 100.00% | 58.15% | 0.00% | 12.92% |
| **Indicate you have a question** | 7.05% | 0.00% | 67.88% | 43.07% | 41.84% | 0.00% | 4.56% | 0.00% | 17.26% | 0.00% | 0.00% | 53.88% | 0.00% | 65.29% | 58.15% | 100.00% | 0.00% | 0.00% |
| **End meeting call** | 0.00% | 21.61% | 34.11% | 0.00% | 0.00% | 0.00% | 9.06% | 0.00% | 22.42% | 4.70% | 19.86% | 0.00% | 0.00% | 34.30% | 0.00% | 0.00% | 100.00% | 0.00% |
| **Age** | 12.50% | 0.00% | 0.00% | 21.68% | 16.04% | 10.23% | 0.00% | 15.41% | 0.00% | 27.38% | 18.68% | 19.77% | 0.00% | 25.55% | 12.92% | 0.00% | 0.00% | 100.00% |

The figure shows the correlation matrix filtered on people below 25 years old (gesture elicitation)

**Correlation Matrix for people born in Denmark**

**Amount of data points: 68**

**Percent of total: 66.67%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 12.50% | 0.00% | 0.00% | 12.50% | 13.25% | 12.40% | 0.00% | 14.21% | 9.78% | 8.90% | 0.00% | 3.63% | 0.00% | 0.00% | 0.00% | 0.00% | 13.13% |
| **Gender** | 12.50% | 100.00% | 0.00% | 0.00% | 55.41% | 0.00% | 0.00% | 36.10% | 24.53% | 15.54% | 28.25% | 0.00% | 0.00% | 28.70% | 0.00% | 0.00% | 15.64% | 0.96% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 45.66% | 11.41% | 0.00% | 0.00% | 16.82% | 27.18% | 42.08% | 0.00% | 14.47% | 59.47% | 27.54% | 0.00% |
| **Born** | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Native language** | 12.50% | 55.41% | 0.00% | 0.00% | 100.00% | 25.82% | 28.14% | 0.00% | 18.69% | 0.00% | 13.32% | 0.00% | 0.00% | 26.72% | 0.00% | 0.00% | 0.00% | 6.71% |
| **Second language** | 13.25% | 0.00% | 0.00% | 0.00% | 25.82% | 100.00% | 50.10% | 20.36% | 0.00% | 7.75% | 10.05% | 0.00% | 0.00% | 0.00% | 23.72% | 0.00% | 0.00% | 0.00% |
| **English proficiency** | 12.40% | 0.00% | 45.66% | 0.00% | 28.14% | 50.10% | 100.00% | 24.13% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 4.64% | 0.00% | 0.00% | 10.60% | 0.00% |
| **Computer proficiency** | 0.00% | 36.10% | 11.41% | 0.00% | 0.00% | 20.36% | 24.13% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 5.91% | 3.04% | 0.00% | 0.00% | 0.00% | 10.57% |
| **Meeting platform experience** | 14.21% | 24.53% | 0.00% | 0.00% | 18.69% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 11.12% | 18.30% |
| **Increase volume** | 9.78% | 15.54% | 0.00% | 0.00% | 0.00% | 7.75% | 0.00% | 0.00% | 0.00% | 100.00% | 81.02% | 22.18% | 32.40% | 17.32% | 14.24% | 0.00% | 19.88% | 0.00% |
| **Decrease volume** | 8.90% | 28.25% | 16.82% | 0.00% | 13.32% | 10.05% | 0.00% | 0.00% | 0.00% | 81.02% | 100.00% | 17.57% | 27.90% | 19.76% | 0.00% | 0.00% | 35.16% | 0.00% |
| **Mute** | 0.00% | 0.00% | 27.18% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 22.18% | 17.57% | 100.00% | 7.18% | 0.00% | 27.08% | 5.86% | 0.00% | 0.00% |
| **Unmute** | 3.63% | 0.00% | 42.08% | 0.00% | 0.00% | 0.00% | 0.00% | 5.91% | 0.00% | 32.40% | 27.90% | 7.18% | 100.00% | 25.37% | 24.17% | 46.05% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 28.70% | 0.00% | 0.00% | 26.72% | 0.00% | 4.64% | 3.04% | 0.00% | 17.32% | 19.76% | 0.00% | 25.37% | 100.00% | 32.78% | 56.01% | 36.54% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 14.47% | 0.00% | 0.00% | 23.72% | 0.00% | 0.00% | 0.00% | 14.24% | 0.00% | 27.08% | 24.17% | 32.78% | 100.00% | 51.48% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 59.47% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 5.86% | 46.05% | 56.01% | 51.48% | 100.00% | 0.00% | 13.25% |
| **End meeting call** | 0.00% | 15.64% | 27.54% | 0.00% | 0.00% | 0.00% | 10.60% | 0.00% | 11.12% | 19.88% | 35.16% | 0.00% | 0.00% | 36.54% | 0.00% | 0.00% | 100.00% | 0.00% |
| **Age** | 13.13% | 0.96% | 0.00% | 0.00% | 6.71% | 0.00% | 0.00% | 10.57% | 18.30% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 13.25% | 0.00% | 100.00% |

The figure shows the correlation matrix filtered on people born in Denmark (gesture elicitation)

**Correlation Matrix for people born outside Denmark**

**Amount of data points: 34**

**Percent of total: 33.33%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Date of birth | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Gender | 0.00% | 100.00% | 0.00% | 6.72% | 0.00% | 5.71% | 0.00% | 23.30% | 0.00% | 0.00% | 0.00% | 6.58% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 2.97% |
| Industry | 0.00% | 0.00% | 100.00% | 26.99% | 0.00% | 19.60% | 0.00% | 41.60% | 36.75% | 0.00% | 0.00% | 0.00% | 0.00% | 45.38% | 14.79% | 71.38% | 36.59% | 38.86% |
| Born | 0.00% | 6.72% | 26.99% | 100.00% | 53.50% | 0.00% | 28.72% | 11.47% | 27.76% | 0.00% | 13.59% | 0.00% | 0.00% | 0.00% | 26.81% | 42.02% | 0.00% | 14.21% |
| Native language | 0.00% | 0.00% | 0.00% | 53.50% | 100.00% | 25.24% | 22.02% | 0.00% | 26.12% | 37.32% | 36.12% | 20.16% | 23.70% | 24.43% | 13.06% | 16.07% | 0.00% | 0.00% |
| Second language | 0.00% | 5.71% | 19.60% | 0.00% | 25.24% | 100.00% | 0.00% | 24.78% | 31.30% | 33.28% | 34.11% | 0.00% | 37.90% | 26.36% | 0.00% | 0.00% | 12.72% | 2.97% |
| English proficiency | 0.00% | 0.00% | 0.00% | 28.72% | 22.02% | 0.00% | 100.00% | 14.80% | 16.50% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 18.50% | 0.00% | 10.44% | 0.00% |
| Computer proficiency | 0.00% | 23.30% | 41.60% | 11.47% | 0.00% | 24.78% | 14.80% | 100.00% | 8.85% | 0.00% | 0.00% | 0.00% | 0.00% | 14.10% | 0.00% | 0.00% | 28.59% | 0.00% |
| Meeting platform experience | 0.00% | 0.00% | 36.75% | 27.76% | 26.12% | 31.30% | 16.50% | 8.85% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 23.00% | 0.00% | 0.00% |
| Increase volume | 0.00% | 0.00% | 0.00% | 0.00% | 37.32% | 33.28% | 0.00% | 0.00% | 0.00% | 100.00% | 81.17% | 34.97% | 36.54% | 15.99% | 17.44% | 0.00% | 0.00% | 0.00% |
| Decrease volume | 0.00% | 0.00% | 0.00% | 13.59% | 36.12% | 34.11% | 0.00% | 0.00% | 0.00% | 81.17% | 100.00% | 34.63% | 35.22% | 23.97% | 25.66% | 0.00% | 0.00% | 8.33% |
| Mute | 0.00% | 6.58% | 0.00% | 0.00% | 20.16% | 0.00% | 0.00% | 0.00% | 0.00% | 34.97% | 34.63% | 100.00% | 8.85% | 21.39% | 36.47% | 34.59% | 0.00% | 0.00% |
| Unmute | 0.00% | 0.00% | 0.00% | 0.00% | 23.70% | 37.90% | 0.00% | 0.00% | 0.00% | 36.54% | 35.22% | 8.85% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Turn off camera | 0.00% | 0.00% | 45.38% | 0.00% | 24.43% | 26.36% | 0.00% | 14.10% | 0.00% | 15.99% | 23.97% | 21.39% | 0.00% | 100.00% | 0.00% | 57.52% | 0.00% | 27.25% |
| Turn on camera | 0.00% | 0.00% | 14.79% | 26.81% | 13.06% | 0.00% | 18.50% | 0.00% | 0.00% | 17.44% | 25.66% | 36.47% | 0.00% | 0.00% | 100.00% | 37.92% | 12.93% | 0.00% |
| Indicate you have a question | 0.00% | 0.00% | 71.38% | 42.02% | 16.07% | 0.00% | 0.00% | 0.00% | 23.00% | 0.00% | 0.00% | 34.59% | 0.00% | 57.52% | 37.92% | 100.00% | 19.51% | 25.25% |
| End meeting call | 0.00% | 0.00% | 36.59% | 0.00% | 0.00% | 12.72% | 10.44% | 28.59% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 12.93% | 19.51% | 100.00% | 10.94% |
| Age | 0.00% | 2.97% | 38.86% | 14.21% | 0.00% | 2.97% | 0.00% | 0.00% | 0.00% | 0.00% | 8.33% | 0.00% | 0.00% | 27.25% | 0.00% | 25.25% | 10.94% | 100.00% |

The figure shows the correlation matrix filtered on people born outside of Denmark (gesture elicitation)

**Correlation Matrix for people in the IT industry**

**Amount of data points: 81**

**Percent of total: 79.41%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 0.00% | 0.00% | 18.35% | 0.00% | 0.00% | 34.09% | 23.12% | 0.00% | 3.31% | 0.00% | 5.51% | 21.79% | 0.00% | 0.00% | 16.65% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 15.57% | 23.92% | 0.00% | 17.08% | 29.19% | 24.25% | 22.27% | 21.09% | 20.44% | 76.79% | 25.62% | 0.00% |
| **Born** | 0.00% | 0.00% | 0.00% | 100.00% | 77.91% | 23.77% | 28.85% | 0.00% | 0.00% | 0.00% | 15.17% | 40.31% | 0.00% | 0.00% | 5.54% | 41.25% | 0.00% | 39.61% |
| **Native language** | 0.00% | 18.35% | 0.00% | 77.91% | 100.00% | 37.01% | 0.00% | 0.00% | 16.96% | 23.43% | 27.35% | 28.33% | 0.00% | 0.00% | 0.00% | 39.69% | 0.00% | 31.48% |
| **Second language** | 0.00% | 0.00% | 0.00% | 23.77% | 37.01% | 100.00% | 0.00% | 10.65% | 15.83% | 9.55% | 10.20% | 0.00% | 0.00% | 0.00% | 6.89% | 0.00% | 0.00% | 26.76% |
| **English proficiency** | 0.00% | 0.00% | 15.57% | 28.85% | 0.00% | 0.00% | 100.00% | 24.62% | 0.85% | 0.00% | 0.00% | 5.56% | 0.00% | 7.02% | 0.00% | 0.00% | 10.02% | 0.00% |
| **Computer proficiency** | 0.00% | 34.09% | 23.92% | 0.00% | 0.00% | 10.65% | 24.62% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 3.98% | 6.74% | 0.00% | 0.00% | 7.55% | 12.08% |
| **Meeting platform experience** | 0.00% | 23.12% | 0.00% | 0.00% | 16.96% | 15.83% | 0.85% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 13.88% | 27.11% | 23.56% |
| **Increase volume** | 0.00% | 0.00% | 17.08% | 0.00% | 23.43% | 9.55% | 0.00% | 0.00% | 0.00% | 100.00% | 80.26% | 24.57% | 26.55% | 11.07% | 21.90% | 0.00% | 0.00% | 7.04% |
| **Decrease volume** | 0.00% | 3.31% | 29.19% | 15.17% | 27.35% | 10.20% | 0.00% | 0.00% | 0.00% | 80.26% | 100.00% | 28.93% | 22.72% | 10.23% | 0.00% | 0.00% | 23.56% | 27.38% |
| **Mute** | 0.00% | 0.00% | 24.25% | 40.31% | 28.33% | 0.00% | 5.56% | 0.00% | 0.00% | 24.57% | 28.93% | 100.00% | 16.25% | 0.00% | 22.69% | 46.57% | 16.98% | 25.38% |
| **Unmute** | 0.00% | 5.51% | 22.27% | 0.00% | 0.00% | 0.00% | 0.00% | 3.98% | 0.00% | 26.55% | 22.72% | 16.25% | 100.00% | 0.00% | 0.00% | 20.17% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 21.79% | 21.09% | 0.00% | 0.00% | 0.00% | 7.02% | 6.74% | 0.00% | 11.07% | 10.23% | 0.00% | 0.00% | 100.00% | 14.08% | 67.07% | 25.22% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 20.44% | 5.54% | 0.00% | 6.89% | 0.00% | 0.00% | 0.00% | 21.90% | 0.00% | 22.69% | 0.00% | 14.08% | 100.00% | 61.03% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 76.79% | 41.25% | 39.69% | 0.00% | 0.00% | 0.00% | 13.88% | 0.00% | 0.00% | 46.57% | 20.17% | 67.07% | 61.03% | 100.00% | 0.00% | 0.00% |
| **End meeting call** | 0.00% | 16.65% | 25.62% | 0.00% | 0.00% | 0.00% | 10.02% | 7.55% | 27.11% | 0.00% | 23.56% | 16.98% | 0.00% | 25.22% | 0.00% | 0.00% | 100.00% | 0.00% |
| **Age** | 0.00% | 0.00% | 0.00% | 39.61% | 31.48% | 26.76% | 0.00% | 12.08% | 23.56% | 7.04% | 27.38% | 25.38% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% |

The figure shows the correlation matrix filtered on people in the IT industry (gesture elicitation)

**Correlation Matrix for people outside the IT industry**

**Amount of data points: 21**

**Percent of total: 20.59%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 6.74% | 11.73% | 16.37% | 0.00% | 0.00% | 0.00% | 18.87% | 12.48% | 20.95% | 18.18% | 0.00% | 0.00% | 9.44% | 0.00% | 15.08% | 9.53% |
| **Industry** | 0.00% | 6.74% | 100.00% | 0.00% | 11.72% | 0.00% | 38.79% | 0.00% | 0.00% | 0.00% | 0.00% | 30.89% | 46.72% | 46.89% | 14.76% | 0.00% | 46.72% | 0.00% |
| **Born** | 0.00% | 11.73% | 0.00% | 100.00% | 96.36% | 38.54% | 16.26% | 0.00% | 0.00% | 22.01% | 31.76% | 0.00% | 0.00% | 44.07% | 0.00% | 61.39% | 21.23% | 43.49% |
| **Native language** | 0.00% | 16.37% | 11.72% | 96.36% | 100.00% | 45.54% | 28.64% | 0.00% | 0.00% | 22.01% | 33.79% | 0.00% | 0.00% | 42.46% | 0.00% | 35.19% | 0.00% | 25.32% |
| **Second language** | 0.00% | 0.00% | 0.00% | 38.54% | 45.54% | 100.00% | 16.45% | 27.04% | 0.00% | 36.68% | 36.85% | 0.00% | 46.29% | 22.08% | 34.64% | 0.00% | 0.00% | 0.00% |
| **English proficiency** | 0.00% | 0.00% | 38.79% | 16.26% | 28.64% | 16.45% | 100.00% | 33.25% | 0.00% | 18.72% | 0.00% | 0.00% | 1.57% | 0.00% | 0.00% | 0.00% | 27.11% | 13.08% |
| **Computer proficiency** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 27.04% | 33.25% | 100.00% | 0.00% | 24.86% | 9.56% | 7.55% | 18.05% | 20.67% | 0.00% | 19.11% | 0.00% | 12.68% |
| **Meeting platform experience** | 0.00% | 18.87% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 18.91% | 22.14% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Increase volume** | 0.00% | 12.48% | 0.00% | 22.01% | 22.01% | 36.68% | 18.72% | 24.86% | 18.91% | 100.00% | 93.09% | 29.59% | 24.12% | 0.00% | 10.44% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 20.95% | 0.00% | 31.76% | 33.79% | 36.85% | 0.00% | 9.56% | 22.14% | 93.09% | 100.00% | 14.60% | 18.54% | 0.00% | 32.00% | 0.00% | 0.00% | 0.00% |
| **Mute** | 0.00% | 18.18% | 30.89% | 0.00% | 0.00% | 0.00% | 0.00% | 7.55% | 0.00% | 29.59% | 14.60% | 100.00% | 28.52% | 45.01% | 54.63% | 0.00% | 0.00% | 18.50% |
| **Unmute** | 0.00% | 0.00% | 46.72% | 0.00% | 0.00% | 46.29% | 1.57% | 18.05% | 0.00% | 24.12% | 18.54% | 28.52% | 100.00% | 44.10% | 65.68% | 0.00% | 22.54% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 46.89% | 44.07% | 42.46% | 22.08% | 0.00% | 20.67% | 0.00% | 0.00% | 0.00% | 45.01% | 44.10% | 100.00% | 0.00% | 44.44% | 38.21% | 20.51% |
| **Turn on camera** | 0.00% | 9.44% | 14.76% | 0.00% | 0.00% | 34.64% | 0.00% | 0.00% | 0.00% | 10.44% | 32.00% | 54.63% | 65.68% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 0.00% | 61.39% | 35.19% | 0.00% | 0.00% | 19.11% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 44.44% | 0.00% | 100.00% | 28.91% | 34.40% |
| **End meeting call** | 0.00% | 15.08% | 46.72% | 21.23% | 0.00% | 0.00% | 27.11% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 22.54% | 38.21% | 0.00% | 28.91% | 100.00% | 0.00% |
| **Age** | 0.00% | 9.53% | 0.00% | 43.49% | 25.32% | 0.00% | 13.08% | 12.68% | 0.00% | 0.00% | 0.00% | 18.50% | 0.00% | 20.51% | 0.00% | 34.40% | 0.00% | 100.00% |

The figure shows the correlation matrix filtered on people outside the IT industry (gesture elicitation)

## A.12 Correlation matrices (A/B test)

The following pdf shows the correlation matrices from the A/B test. They are filtered on different variables as defined at the top of each page

**Correlation Matrix for all data**

**Amount of data points: 104**

**Percent of total: 100.0%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 9.87% | 0.00% | 10.29% | 14.99% | 5.82% | 14.07% | 7.68% | 3.78% | 0.00% | 0.00% | 1.11% | 0.27% | 0.00% | 1.37% | 0.00% | 14.00% | 14.74% |
| **Gender** | 9.87% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 20.38% | 30.09% | 11.74% | 0.00% | 18.54% | 20.79% | 0.00% | 0.00% | 0.00% | 9.08% | 19.46% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 12.19% | 20.47% | 31.38% | 54.35% | 21.68% | 0.00% | 0.00% | 22.48% | 0.00% | 13.13% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Born** | 10.29% | 0.00% | 0.00% | 100.00% | 79.85% | 24.85% | 0.00% | 0.00% | 21.54% | 6.12% | 0.00% | 0.00% | 4.24% | 0.00% | 9.90% | 0.00% | 0.00% | 43.35% |
| **Native language** | 14.99% | 0.00% | 12.19% | 79.85% | 100.00% | 12.68% | 27.92% | 0.00% | 29.24% | 11.35% | 0.00% | 0.00% | 0.00% | 0.00% | 16.71% | 16.48% | 7.28% | 37.58% |
| **Second language** | 5.82% | 0.00% | 20.47% | 24.85% | 12.68% | 100.00% | 0.00% | 24.08% | 0.00% | 22.52% | 23.71% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 10.33% | 0.00% |
| **English proficiency** | 14.07% | 20.38% | 31.38% | 0.00% | 27.92% | 0.00% | 100.00% | 11.24% | 9.85% | 6.33% | 15.61% | 0.00% | 7.94% | 0.00% | 0.00% | 10.45% | 9.30% | 0.00% |
| **Computer proficiency** | 7.68% | 30.09% | 54.35% | 0.00% | 0.00% | 24.08% | 11.24% | 100.00% | 8.78% | 3.10% | 0.00% | 12.47% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 6.13% |
| **Meeting platform experience** | 3.78% | 11.74% | 21.68% | 21.54% | 29.24% | 0.00% | 9.85% | 8.78% | 100.00% | 4.60% | 0.00% | 6.31% | 8.73% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 6.12% | 11.35% | 22.52% | 6.33% | 3.10% | 4.60% | 100.00% | 78.78% | 0.00% | 0.00% | 5.89% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 18.54% | 0.00% | 0.00% | 0.00% | 23.71% | 15.61% | 0.00% | 0.00% | 78.78% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 13.80% | 0.00% | 0.00% |
| **Mute** | 1.11% | 20.79% | 22.48% | 0.00% | 0.00% | 0.00% | 0.00% | 12.47% | 6.31% | 0.00% | 0.00% | 100.00% | 32.13% | 12.01% | 0.00% | 2.29% | 7.21% | 0.00% |
| **Unmute** | 0.27% | 0.00% | 0.00% | 4.24% | 0.00% | 0.00% | 7.94% | 0.00% | 8.73% | 0.00% | 0.00% | 32.13% | 100.00% | 10.63% | 16.70% | 0.00% | 4.84% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 13.13% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 5.89% | 0.00% | 12.01% | 10.63% | 100.00% | 31.05% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 1.37% | 0.00% | 0.00% | 9.90% | 16.71% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 16.70% | 31.05% | 100.00% | 3.40% | 7.25% | 0.00% |
| **Indicate you have a question** | 0.00% | 9.08% | 0.00% | 0.00% | 16.48% | 0.00% | 10.45% | 0.00% | 0.00% | 0.00% | 13.80% | 2.29% | 0.00% | 0.00% | 3.40% | 100.00% | 0.00% | 0.00% |
| **End meeting call** | 14.00% | 19.46% | 0.00% | 0.00% | 7.28% | 10.33% | 9.30% | 0.00% | 0.00% | 0.00% | 0.00% | 7.21% | 4.84% | 0.00% | 7.25% | 0.00% | 100.00% | 9.52% |
| **Age** | 14.74% | 0.00% | 0.00% | 43.35% | 37.58% | 0.00% | 0.00% | 6.13% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 9.52% | 100.00% |

**Correlation Matrix for Danish native speakers**

**Amount of data points: 90**

**Percent of total: 86.54%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 10.65% | 0.00% | 0.00% | 15.25% | 6.22% | 15.16% | 8.22% | 1.01% | 0.00% | 0.00% | 1.55% | 0.00% | 0.00% | 1.44% | 0.00% | 15.08% | 15.81% |
| **Gender** | 10.65% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 18.83% | 31.04% | 13.49% | 0.00% | 20.94% | 17.90% | 0.00% | 0.00% | 0.00% | 14.14% | 18.47% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 37.91% | 26.08% | 34.10% | 54.49% | 20.93% | 2.15% | 0.00% | 15.98% | 0.00% | 14.15% | 8.57% | 0.00% | 0.00% | 0.00% |
| **Born** | 0.00% | 0.00% | 0.00% | 100.00% | 52.68% | 45.11% | 19.14% | 0.00% | 16.74% | 6.13% | 0.00% | 4.46% | 7.27% | 0.00% | 0.00% | 0.00% | 10.92% | 6.44% |
| **Native language** | 15.25% | 0.00% | 37.91% | 52.68% | 100.00% | 0.00% | 32.50% | 0.00% | 11.30% | 8.28% | 9.83% | 0.00% | 0.00% | 0.00% | 14.26% | 25.43% | 16.67% | 2.12% |
| **Second language** | 6.22% | 0.00% | 26.08% | 45.11% | 0.00% | 100.00% | 0.00% | 28.05% | 0.00% | 21.92% | 27.29% | 0.00% | 0.00% | 0.00% | 0.00% | 10.25% | 4.96% | 0.00% |
| **English proficiency** | 15.16% | 18.83% | 34.10% | 19.14% | 32.50% | 0.00% | 100.00% | 14.67% | 11.56% | 10.11% | 20.04% | 0.00% | 6.50% | 0.00% | 0.00% | 13.74% | 7.78% | 0.00% |
| **Computer proficiency** | 8.22% | 31.04% | 54.49% | 0.00% | 0.00% | 28.05% | 14.67% | 100.00% | 10.49% | 8.36% | 8.55% | 6.13% | 0.00% | 5.90% | 0.00% | 0.00% | 0.00% | 5.84% |
| **Meeting platform experience** | 1.01% | 13.49% | 20.93% | 16.74% | 11.30% | 0.00% | 11.56% | 10.49% | 100.00% | 3.86% | 0.00% | 14.15% | 3.72% | 0.00% | 0.00% | 0.00% | 0.00% | 22.65% |
| **Increase volume** | 0.00% | 0.00% | 2.15% | 6.13% | 8.28% | 21.92% | 10.11% | 8.36% | 3.86% | 100.00% | 77.40% | 0.00% | 0.00% | 6.83% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 20.94% | 0.00% | 0.00% | 9.83% | 27.29% | 20.04% | 8.55% | 0.00% | 77.40% | 100.00% | 6.51% | 0.00% | 0.00% | 2.18% | 15.87% | 0.00% | 0.00% |
| **Mute** | 1.55% | 17.90% | 15.98% | 4.46% | 0.00% | 0.00% | 0.00% | 6.13% | 14.15% | 0.00% | 6.51% | 100.00% | 29.90% | 16.81% | 0.00% | 8.18% | 10.58% | 0.00% |
| **Unmute** | 0.00% | 0.00% | 0.00% | 7.27% | 0.00% | 0.00% | 6.50% | 0.00% | 3.72% | 0.00% | 0.00% | 29.90% | 100.00% | 4.37% | 16.64% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 14.15% | 0.00% | 0.00% | 0.00% | 0.00% | 5.90% | 0.00% | 6.83% | 0.00% | 16.81% | 4.37% | 100.00% | 32.50% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 1.44% | 0.00% | 8.57% | 0.00% | 14.26% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 2.18% | 0.00% | 16.64% | 32.50% | 100.00% | 0.00% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 14.14% | 0.00% | 0.00% | 25.43% | 10.25% | 13.74% | 0.00% | 0.00% | 0.00% | 15.87% | 8.18% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 5.11% |
| **End meeting call** | 15.08% | 18.47% | 0.00% | 10.92% | 16.67% | 4.96% | 7.78% | 0.00% | 0.00% | 0.00% | 0.00% | 10.58% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 8.03% |
| **Age** | 15.81% | 0.00% | 0.00% | 6.44% | 2.12% | 0.00% | 0.00% | 5.84% | 22.65% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 5.11% | 8.03% | 100.00% |

**Correlation Matrix for non-Danish native speakers**

**Amount of data points: 14**

**Percent of total: 13.46%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 11.39% | 0.00% | 0.00% | 0.00% | 0.00% | 14.91% |
| **Industry** | 0.00% | 0.00% | 100.00% | 12.97% | 0.00% | 37.14% | 0.00% | 53.81% | 50.92% | 0.00% | 0.00% | 37.73% | 0.00% | 32.39% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Born** | 0.00% | 0.00% | 12.97% | 100.00% | 81.65% | 30.28% | 40.82% | 16.45% | 37.02% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 40.82% | 40.82% | 0.00% | 33.73% |
| **Native language** | 0.00% | 0.00% | 0.00% | 81.65% | 100.00% | 33.23% | 50.00% | 0.00% | 42.60% | 20.89% | 0.00% | 0.00% | 0.00% | 0.00% | 28.54% | 0.00% | 0.00% | 15.64% |
| **Second language** | 0.00% | 0.00% | 37.14% | 30.28% | 33.23% | 100.00% | 0.00% | 0.00% | 54.77% | 0.00% | 0.00% | 0.00% | 14.43% | 44.88% | 0.00% | 0.00% | 0.00% | 11.47% |
| **English proficiency** | 0.00% | 0.00% | 0.00% | 40.82% | 50.00% | 0.00% | 100.00% | 0.00% | 31.62% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Computer proficiency** | 0.00% | 0.00% | 53.81% | 16.45% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 39.23% | 0.00% | 0.00% | 0.00% | 0.00% | 33.79% | 11.79% | 0.00% | 28.28% |
| **Meeting platform experience** | 0.00% | 0.00% | 50.92% | 37.02% | 42.60% | 54.77% | 31.62% | 0.00% | 100.00% | 0.00% | 0.00% | 50.00% | 20.89% | 50.00% | 28.54% | 0.00% | 15.91% | 0.00% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 0.00% | 20.89% | 0.00% | 0.00% | 39.23% | 0.00% | 100.00% | 69.35% | 11.39% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 18.44% |
| **Decrease volume** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 69.35% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Mute** | 0.00% | 0.00% | 37.73% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 50.00% | 11.39% | 0.00% | 100.00% | 11.39% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Unmute** | 0.00% | 11.39% | 0.00% | 0.00% | 0.00% | 14.43% | 0.00% | 0.00% | 20.89% | 0.00% | 0.00% | 11.39% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 32.39% | 0.00% | 0.00% | 44.88% | 0.00% | 0.00% | 50.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 0.00% | 40.82% | 28.54% | 0.00% | 0.00% | 33.79% | 28.54% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 0.00% | 40.82% | 0.00% | 0.00% | 0.00% | 11.79% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 19.50% | 21.52% |
| **End meeting call** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 15.91% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 19.50% | 100.00% | 17.39% |
| **Age** | 0.00% | 14.91% | 0.00% | 33.73% | 15.64% | 11.47% | 0.00% | 28.28% | 0.00% | 18.44% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 21.52% | 17.39% | 100.00% |

**Correlation Matrix for people born in Denmark**

**Amount of data points: 86**

**Percent of total: 82.69%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 10.91% | 0.00% | 0.00% | 15.62% | 4.31% | 15.52% | 8.09% | 2.57% | 0.00% | 0.00% | 1.50% | 0.85% | 0.00% | 1.63% | 0.00% | 15.43% | 16.22% |
| **Gender** | 10.91% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 19.22% | 32.47% | 13.28% | 0.00% | 18.35% | 23.25% | 0.00% | 0.00% | 0.00% | 15.98% | 12.45% | 5.87% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 31.75% | 26.01% | 52.15% | 19.53% | 7.25% | 0.00% | 21.20% | 0.00% | 19.30% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Born** | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 9.36% | 15.02% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 3.27% | 0.00% | 0.00% | 0.00% | 0.00% | 4.35% |
| **Native language** | 15.62% | 0.00% | 0.00% | 0.00% | 100.00% | 36.23% | 21.84% | 0.00% | 11.21% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 15.46% | 25.74% | 10.84% | 0.00% |
| **Second language** | 4.31% | 0.00% | 31.75% | 9.36% | 36.23% | 100.00% | 0.00% | 28.10% | 0.00% | 23.51% | 25.60% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **English proficiency** | 15.52% | 19.22% | 26.01% | 15.02% | 21.84% | 0.00% | 100.00% | 12.55% | 9.14% | 3.74% | 15.83% | 0.00% | 14.87% | 0.00% | 0.00% | 15.83% | 0.00% | 0.00% |
| **Computer proficiency** | 8.09% | 32.47% | 52.15% | 0.00% | 0.00% | 28.10% | 12.55% | 100.00% | 11.42% | 8.22% | 0.00% | 12.83% | 3.04% | 0.00% | 0.00% | 0.00% | 4.18% | 3.24% |
| **Meeting platform experience** | 2.57% | 13.28% | 19.53% | 0.00% | 11.21% | 0.00% | 9.14% | 11.42% | 100.00% | 5.60% | 0.00% | 6.38% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.06% |
| **Increase volume** | 0.00% | 0.00% | 7.25% | 0.00% | 0.00% | 23.51% | 3.74% | 8.22% | 5.60% | 100.00% | 78.30% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 18.35% | 0.00% | 0.00% | 0.00% | 25.60% | 15.83% | 0.00% | 0.00% | 78.30% | 100.00% | 12.34% | 0.00% | 0.00% | 7.39% | 19.56% | 0.00% | 0.00% |
| **Mute** | 1.50% | 23.25% | 21.20% | 0.00% | 0.00% | 0.00% | 0.00% | 12.83% | 6.38% | 0.00% | 12.34% | 100.00% | 26.51% | 12.81% | 0.00% | 7.51% | 19.18% | 0.00% |
| **Unmute** | 0.85% | 0.00% | 0.00% | 3.27% | 0.00% | 0.00% | 14.87% | 3.04% | 0.00% | 0.00% | 0.00% | 26.51% | 100.00% | 0.00% | 17.86% | 0.00% | 12.35% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 19.30% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 12.81% | 0.00% | 100.00% | 35.66% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 1.63% | 0.00% | 0.00% | 0.00% | 15.46% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 7.39% | 0.00% | 17.86% | 35.66% | 100.00% | 3.84% | 4.13% | 0.00% |
| **Indicate you have a question** | 0.00% | 15.98% | 0.00% | 0.00% | 25.74% | 0.00% | 15.83% | 0.00% | 0.00% | 0.00% | 19.56% | 7.51% | 0.00% | 0.00% | 3.84% | 100.00% | 0.00% | 6.47% |
| **End meeting call** | 15.43% | 12.45% | 0.00% | 0.00% | 10.84% | 0.00% | 0.00% | 4.18% | 0.00% | 0.00% | 0.00% | 19.18% | 12.35% | 0.00% | 4.13% | 0.00% | 100.00% | 9.17% |
| **Age** | 16.22% | 5.87% | 0.00% | 4.35% | 0.00% | 0.00% | 0.00% | 3.24% | 25.06% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 6.47% | 9.17% | 100.00% |

**Correlation Matrix for people born outside Denmark**

**Amount of data points: 18**

**Percent of total: 17.31%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 11.01% | 0.00% | 0.00% | 1.63% | 11.41% | 13.60% | 0.00% | 0.00% | 0.00% | 0.00% | 13.73% | 0.00% | 0.00% | 0.00% | 13.73% | 11.28% |
| **Industry** | 0.00% | 11.01% | 100.00% | 0.00% | 46.66% | 36.63% | 43.76% | 61.41% | 43.67% | 0.00% | 0.00% | 55.23% | 0.00% | 0.00% | 0.00% | 0.00% | 16.15% | 15.04% |
| **Born** | 0.00% | 0.00% | 0.00% | 100.00% | 55.90% | 25.60% | 0.00% | 0.00% | 9.50% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 36.12% | 19.36% | 0.00% | 38.62% |
| **Native language** | 0.00% | 0.00% | 46.66% | 55.90% | 100.00% | 0.00% | 63.25% | 13.10% | 32.86% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 50.54% | 0.00% | 0.00% | 0.00% |
| **Second language** | 0.00% | 1.63% | 36.63% | 25.60% | 0.00% | 100.00% | 0.00% | 0.00% | 24.06% | 0.00% | 0.00% | 12.50% | 0.00% | 49.58% | 0.00% | 28.38% | 0.00% | 34.29% |
| **English proficiency** | 0.00% | 11.41% | 43.76% | 0.00% | 63.25% | 0.00% | 100.00% | 0.00% | 40.82% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 18.46% |
| **Computer proficiency** | 0.00% | 13.60% | 61.41% | 0.00% | 13.10% | 0.00% | 0.00% | 100.00% | 0.00% | 22.26% | 0.00% | 9.85% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 11.01% |
| **Meeting platform experience** | 0.00% | 0.00% | 43.67% | 9.50% | 32.86% | 24.06% | 40.82% | 0.00% | 100.00% | 0.00% | 0.00% | 36.50% | 27.10% | 55.90% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 22.26% | 0.00% | 100.00% | 66.31% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 66.31% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Mute** | 0.00% | 0.00% | 55.23% | 0.00% | 0.00% | 12.50% | 0.00% | 9.85% | 36.50% | 0.00% | 0.00% | 100.00% | 40.59% | 0.00% | 0.00% | 0.00% | 0.00% | 2.08% |
| **Unmute** | 0.00% | 13.73% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 27.10% | 0.00% | 0.00% | 40.59% | 100.00% | 26.01% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 49.58% | 0.00% | 0.00% | 55.90% | 0.00% | 0.00% | 0.00% | 26.01% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 0.00% | 36.12% | 50.54% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 0.00% | 19.36% | 0.00% | 28.38% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 26.01% | 21.94% |
| **End meeting call** | 0.00% | 13.73% | 16.15% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 26.01% | 100.00% | 18.93% |
| **Age** | 0.00% | 11.28% | 15.04% | 38.62% | 0.00% | 34.29% | 18.46% | 11.01% | 0.00% | 0.00% | 0.00% | 2.08% | 0.00% | 0.00% | 0.00% | 21.94% | 18.93% | 100.00% |

**Correlation Matrix for people below 25 years old**

**Amount of data points: 67**

**Percent of total: 64.42%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 2.00% | 0.00% | 8.27% | 18.73% | 9.06% | 17.68% | 9.88% | 0.00% | 0.00% | 0.00% | 1.70% | 2.00% | 0.00% | 2.13% | 0.00% | 17.54% | 17.96% |
| **Gender** | 2.00% | 100.00% | 1.13% | 0.00% | 0.00% | 0.00% | 15.86% | 35.47% | 2.48% | 0.00% | 0.00% | 0.00% | 4.97% | 0.00% | 0.00% | 0.00% | 17.46% | 7.90% |
| **Industry** | 0.00% | 1.13% | 100.00% | 0.00% | 0.00% | 9.40% | 16.24% | 53.49% | 0.00% | 33.64% | 0.00% | 18.96% | 0.00% | 0.00% | 21.50% | 0.00% | 0.00% | 0.00% |
| **Born** | 8.27% | 0.00% | 0.00% | 100.00% | 74.16% | 26.35% | 0.00% | 0.00% | 31.17% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 18.37% | 0.00% | 0.00% | 15.67% |
| **Native language** | 18.73% | 0.00% | 0.00% | 74.16% | 100.00% | 15.07% | 33.56% | 0.00% | 38.23% | 0.00% | 8.70% | 0.00% | 0.00% | 0.00% | 21.38% | 21.17% | 0.00% | 6.17% |
| **Second language** | 9.06% | 0.00% | 9.40% | 26.35% | 15.07% | 100.00% | 0.00% | 7.57% | 0.00% | 27.75% | 19.87% | 0.00% | 0.00% | 0.00% | 0.00% | 12.54% | 11.54% | 0.00% |
| **English proficiency** | 17.68% | 15.86% | 16.24% | 0.00% | 33.56% | 0.00% | 100.00% | 13.14% | 11.96% | 10.54% | 17.92% | 0.00% | 10.09% | 0.00% | 0.00% | 0.00% | 20.52% | 0.00% |
| **Computer proficiency** | 9.88% | 35.47% | 53.49% | 0.00% | 0.00% | 7.57% | 13.14% | 100.00% | 8.48% | 22.23% | 12.88% | 0.00% | 0.00% | 0.00% | 12.40% | 8.61% | 20.93% | 13.56% |
| **Meeting platform experience** | 0.00% | 2.48% | 0.00% | 31.17% | 38.23% | 0.00% | 11.96% | 8.48% | 100.00% | 7.95% | 0.00% | 22.98% | 0.00% | 0.00% | 0.00% | 11.17% | 0.00% | 36.10% |
| **Increase volume** | 0.00% | 0.00% | 33.64% | 0.00% | 0.00% | 27.75% | 10.54% | 22.23% | 7.95% | 100.00% | 80.86% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 0.00% | 0.00% | 0.00% | 8.70% | 19.87% | 17.92% | 12.88% | 0.00% | 80.86% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 17.52% | 0.00% | 0.00% |
| **Mute** | 1.70% | 0.00% | 18.96% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 22.98% | 0.00% | 0.00% | 100.00% | 15.64% | 19.76% | 0.00% | 0.00% | 3.52% | 0.00% |
| **Unmute** | 2.00% | 4.97% | 0.00% | 0.00% | 0.00% | 0.00% | 10.09% | 0.00% | 0.00% | 0.00% | 0.00% | 15.64% | 100.00% | 27.66% | 22.47% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 19.76% | 27.66% | 100.00% | 31.32% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 2.13% | 0.00% | 21.50% | 18.37% | 21.38% | 0.00% | 0.00% | 12.40% | 0.00% | 0.00% | 0.00% | 0.00% | 22.47% | 31.32% | 100.00% | 8.37% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 0.00% | 0.00% | 0.00% | 21.17% | 12.54% | 0.00% | 8.61% | 11.17% | 0.00% | 17.52% | 0.00% | 0.00% | 0.00% | 8.37% | 100.00% | 8.74% | 3.42% |
| **End meeting call** | 17.54% | 17.46% | 0.00% | 0.00% | 0.00% | 11.54% | 20.52% | 20.93% | 0.00% | 0.00% | 0.00% | 3.52% | 0.00% | 0.00% | 0.00% | 8.74% | 100.00% | 14.23% |
| **Age** | 17.96% | 7.90% | 0.00% | 15.67% | 6.17% | 0.00% | 0.00% | 13.56% | 36.10% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 3.42% | 14.23% | 100.00% |

**Correlation Matrix for people above 25 years old**

**Amount of data points: 37**

**Percent of total: 35.58%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 19.72% | 20.29% | 23.82% | 0.00% | 37.08% | 48.68% | 40.04% | 0.00% | 0.00% | 11.38% | 12.37% | 0.00% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 23.04% | 61.81% | 29.35% | 46.59% | 48.70% | 0.00% | 0.00% | 18.32% | 0.00% | 30.71% | 26.04% | 0.00% | 24.01% | 0.00% |
| **Born** | 0.00% | 0.00% | 0.00% | 100.00% | 78.57% | 23.90% | 0.00% | 0.00% | 0.00% | 24.70% | 0.00% | 0.00% | 2.66% | 0.00% | 0.00% | 0.00% | 13.45% | 56.94% |
| **Native language** | 0.00% | 0.00% | 23.04% | 78.57% | 100.00% | 0.00% | 7.29% | 0.00% | 0.00% | 14.62% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 17.38% | 52.29% |
| **Second language** | 0.00% | 0.00% | 61.81% | 23.90% | 0.00% | 100.00% | 0.00% | 29.15% | 0.00% | 0.00% | 11.39% | 0.00% | 0.00% | 0.00% | 9.26% | 0.00% | 19.94% | 0.00% |
| **English proficiency** | 0.00% | 19.72% | 29.35% | 0.00% | 7.29% | 0.00% | 100.00% | 0.00% | 19.32% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 18.90% | 14.01% | 0.00% | 0.00% |
| **Computer proficiency** | 0.00% | 20.29% | 46.59% | 0.00% | 0.00% | 29.15% | 0.00% | 100.00% | 18.56% | 0.00% | 0.00% | 13.58% | 29.16% | 0.00% | 0.00% | 20.19% | 0.00% | 6.55% |
| **Meeting platform experience** | 0.00% | 23.82% | 48.70% | 0.00% | 0.00% | 0.00% | 19.32% | 18.56% | 100.00% | 32.79% | 39.02% | 11.90% | 0.00% | 3.40% | 0.00% | 0.00% | 22.15% | 0.00% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 24.70% | 14.62% | 0.00% | 0.00% | 0.00% | 32.79% | 100.00% | 67.57% | 0.00% | 0.00% | 19.77% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 37.08% | 0.00% | 0.00% | 0.00% | 11.39% | 0.00% | 0.00% | 39.02% | 67.57% | 100.00% | 9.68% | 0.00% | 8.12% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Mute** | 0.00% | 48.68% | 18.32% | 0.00% | 0.00% | 0.00% | 0.00% | 13.58% | 11.90% | 0.00% | 9.68% | 100.00% | 50.85% | 0.00% | 0.00% | 0.00% | 0.00% | 6.23% |
| **Unmute** | 0.00% | 40.04% | 0.00% | 2.66% | 0.00% | 0.00% | 0.00% | 29.16% | 0.00% | 0.00% | 0.00% | 50.85% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Turn off camera** | 0.00% | 0.00% | 30.71% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 3.40% | 19.77% | 8.12% | 0.00% | 0.00% | 100.00% | 16.36% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 26.04% | 0.00% | 0.00% | 9.26% | 18.90% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 16.36% | 100.00% | 0.00% | 0.00% | 9.82% |
| **Indicate you have a question** | 0.00% | 11.38% | 0.00% | 0.00% | 0.00% | 0.00% | 14.01% | 20.19% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% |
| **End meeting call** | 0.00% | 12.37% | 24.01% | 13.45% | 17.38% | 19.94% | 0.00% | 0.00% | 22.15% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 16.03% |
| **Age** | 0.00% | 0.00% | 0.00% | 56.94% | 52.29% | 0.00% | 0.00% | 6.55% | 0.00% | 0.00% | 0.00% | 6.23% | 0.00% | 0.00% | 9.82% | 0.00% | 16.03% | 100.00% |

**Correlation Matrix for people in the IT industry**

**Amount of data points: 59**

**Percent of total: 56.73%**

| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 13.36% | 0.00% | 6.62% | 14.29% | 0.00% | 13.36% | 0.00% | 0.00% | 13.25% | 13.25% | 13.25% | 13.25% | 0.00% | 0.00% | 0.00% | 13.25% | 14.59% |
| **Gender** | 13.36% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 35.86% | 18.28% | 0.00% | 0.00% | 18.31% | 6.02% | 0.00% | 8.73% | 0.00% | 0.00% | 7.10% | 16.02% |
| **Industry** | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 46.41% | 8.74% | 39.88% | 0.00% | 10.92% | 17.27% | 18.83% | 18.04% | 0.00% | 0.00% | 0.00% | 19.78% | 0.00% |
| **Born** | 6.62% | 0.00% | 0.00% | 100.00% | 84.71% | 27.80% | 29.79% | 0.00% | 31.97% | 17.56% | 9.98% | 0.00% | 1.53% | 0.00% | 1.53% | 0.00% | 0.00% | 35.00% |
| **Native language** | 14.29% | 0.00% | 0.00% | 84.71% | 100.00% | 34.11% | 32.09% | 0.00% | 30.16% | 11.41% | 13.99% | 0.00% | 17.77% | 0.00% | 18.17% | 23.01% | 0.00% | 36.30% |
| **Second language** | 0.00% | 0.00% | 46.41% | 27.80% | 34.11% | 100.00% | 0.00% | 0.00% | 17.41% | 15.72% | 25.91% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **English proficiency** | 13.36% | 35.86% | 8.74% | 29.79% | 32.09% | 0.00% | 100.00% | 0.00% | 17.01% | 16.79% | 17.82% | 0.00% | 0.00% | 5.05% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Computer proficiency** | 0.00% | 18.28% | 39.88% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 7.65% | 0.00% | 7.65% | 0.00% | 0.00% | 13.91% |
| **Meeting platform experience** | 0.00% | 0.00% | 0.00% | 31.97% | 30.16% | 17.41% | 17.01% | 0.00% | 100.00% | 0.00% | 0.00% | 3.73% | 5.80% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Increase volume** | 13.25% | 0.00% | 10.92% | 17.56% | 11.41% | 15.72% | 16.79% | 0.00% | 0.00% | 100.00% | 85.15% | 0.00% | 0.00% | 5.38% | 0.00% | 10.63% | 0.00% | 6.46% |
| **Decrease volume** | 13.25% | 18.31% | 17.27% | 9.98% | 13.99% | 25.91% | 17.82% | 0.00% | 0.00% | 85.15% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 18.24% | 0.00% | 1.89% |
| **Mute** | 13.25% | 6.02% | 18.83% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 3.73% | 0.00% | 0.00% | 100.00% | 38.49% | 0.00% | 0.00% | 0.00% | 0.00% | 6.73% |
| **Unmute** | 13.25% | 0.00% | 18.04% | 1.53% | 17.77% | 0.00% | 0.00% | 7.65% | 5.80% | 0.00% | 0.00% | 38.49% | 100.00% | 25.90% | 33.35% | 0.00% | 13.62% | 0.00% |
| **Turn off camera** | 0.00% | 8.73% | 0.00% | 0.00% | 0.00% | 0.00% | 5.05% | 0.00% | 0.00% | 5.38% | 0.00% | 0.00% | 25.90% | 100.00% | 16.36% | 14.25% | 0.00% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 0.00% | 1.53% | 18.17% | 0.00% | 0.00% | 7.65% | 0.00% | 0.00% | 0.00% | 0.00% | 33.35% | 16.36% | 100.00% | 23.68% | 22.46% | 4.99% |
| **Indicate you have a question** | 0.00% | 0.00% | 0.00% | 0.00% | 23.01% | 0.00% | 0.00% | 0.00% | 0.00% | 10.63% | 18.24% | 0.00% | 0.00% | 14.25% | 23.68% | 100.00% | 3.62% | 1.11% |
| **End meeting call** | 13.25% | 7.10% | 19.78% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 13.62% | 0.00% | 22.46% | 3.62% | 100.00% | 13.36% |
| **Age** | 14.59% | 16.02% | 0.00% | 35.00% | 36.30% | 0.00% | 0.00% | 13.91% | 0.00% | 6.46% | 1.89% | 6.73% | 0.00% | 0.00% | 4.99% | 1.11% | 13.36% | 100.00% |

**Correlation Matrix for people outside the IT industry**

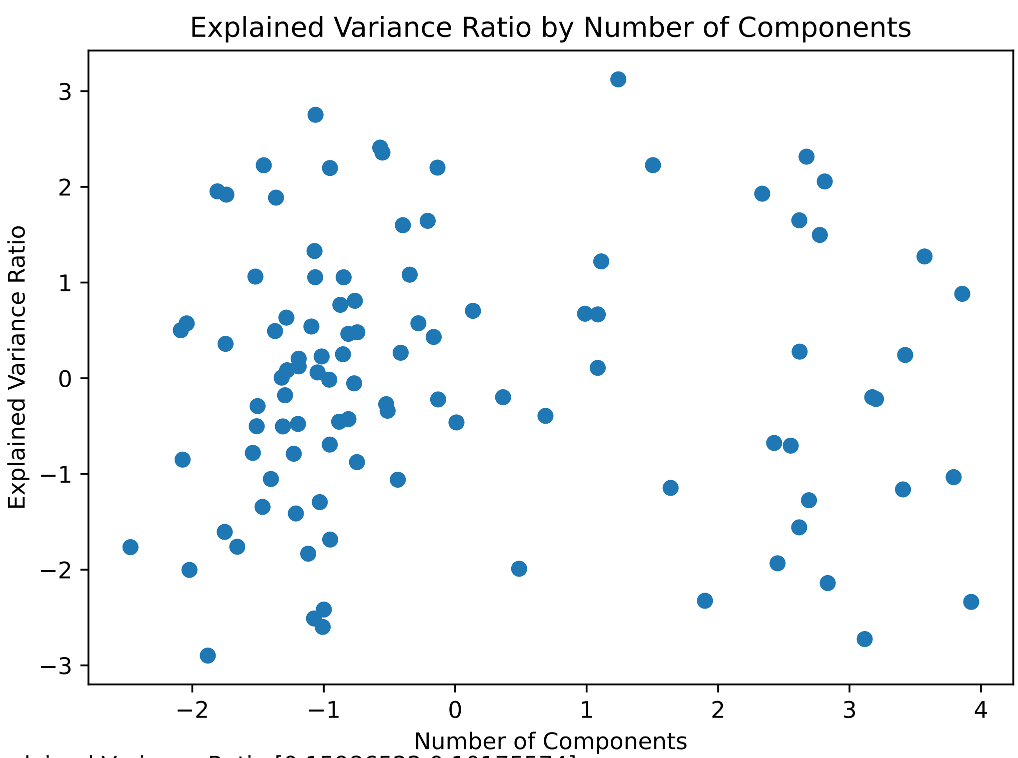
**Amount of data points: 45**

**Percent of total: 43.27%**

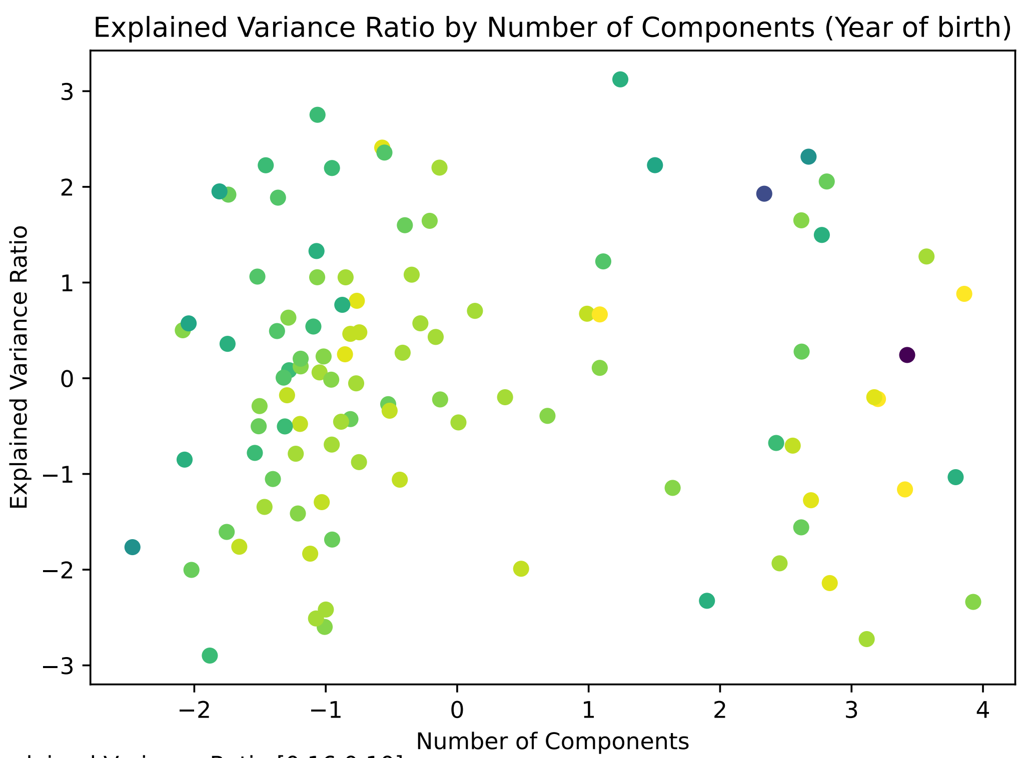
| | Date of birth | Gender | Industry | Born | Native language | Second language | English proficiency | Computer proficiency | Meeting platform experience | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indicate you have a question | End meeting call | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Date of birth** | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Gender** | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 41.70% | 23.53% | 0.00% | 23.27% | 16.78% | 0.00% | 22.19% | 0.00% | 7.15% | 30.36% | 9.11% |
| **Industry** | 0.00% | 0.00% | 100.00% | 18.86% | 37.23% | 0.00% | 25.18% | 51.10% | 12.69% | 0.00% | 0.00% | 16.03% | 0.00% | 39.05% | 0.00% | 10.57% | 0.00% | 25.03% |
| **Born** | 0.00% | 0.00% | 18.86% | 100.00% | 73.68% | 36.69% | 8.88% | 0.00% | 24.92% | 7.56% | 14.47% | 0.00% | 27.65% | 0.00% | 0.00% | 0.00% | 32.36% | 41.03% |
| **Native language** | 0.00% | 0.00% | 37.23% | 73.68% | 100.00% | 0.00% | 40.38% | 0.00% | 34.10% | 3.14% | 18.05% | 0.00% | 11.25% | 0.00% | 0.00% | 0.00% | 28.87% | 45.20% |
| **Second language** | 0.00% | 0.00% | 0.00% | 36.69% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 4.60% | 27.02% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 22.68% | 35.81% |
| **English proficiency** | 0.00% | 0.00% | 25.18% | 8.88% | 40.38% | 0.00% | 100.00% | 11.64% | 14.86% | 0.00% | 15.63% | 0.00% | 3.08% | 0.00% | 0.00% | 23.68% | 8.17% | 0.00% |
| **Computer proficiency** | 0.00% | 41.70% | 51.10% | 0.00% | 0.00% | 0.00% | 11.64% | 100.00% | 0.00% | 0.00% | 0.00% | 11.64% | 0.00% | 22.76% | 0.00% | 7.12% | 11.73% | 0.00% |
| **Meeting platform experience** | 0.00% | 23.53% | 12.69% | 24.92% | 34.10% | 0.00% | 14.86% | 0.00% | 100.00% | 41.58% | 32.35% | 0.00% | 0.00% | 0.00% | 0.00% | 16.65% | 0.00% | 35.82% |
| **Increase volume** | 0.00% | 0.00% | 0.00% | 7.56% | 3.14% | 4.60% | 0.00% | 0.00% | 41.58% | 100.00% | 67.08% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Decrease volume** | 0.00% | 23.27% | 0.00% | 14.47% | 18.05% | 27.02% | 15.63% | 0.00% | 32.35% | 67.08% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Mute** | 0.00% | 16.78% | 16.03% | 0.00% | 0.00% | 0.00% | 0.00% | 11.64% | 0.00% | 0.00% | 0.00% | 100.00% | 8.89% | 26.67% | 20.30% | 0.00% | 0.00% | 0.00% |
| **Unmute** | 0.00% | 0.00% | 0.00% | 27.65% | 11.25% | 0.00% | 3.08% | 0.00% | 0.00% | 0.00% | 0.00% | 8.89% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 5.63% |
| **Turn off camera** | 0.00% | 22.19% | 39.05% | 0.00% | 0.00% | 0.00% | 0.00% | 22.76% | 0.00% | 0.00% | 0.00% | 26.67% | 0.00% | 100.00% | 41.11% | 0.00% | 0.00% | 0.00% |
| **Turn on camera** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 20.30% | 0.00% | 41.11% | 100.00% | 0.00% | 0.00% | 0.00% |
| **Indicate you have a question** | 0.00% | 7.15% | 10.57% | 0.00% | 0.00% | 0.00% | 23.68% | 7.12% | 16.65% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 7.31% |
| **End meeting call** | 0.00% | 30.36% | 0.00% | 32.36% | 28.87% | 22.68% | 8.17% | 11.73% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 6.38% |
| **Age** | 0.00% | 9.11% | 25.03% | 41.03% | 45.20% | 35.81% | 0.00% | 0.00% | 35.82% | 0.00% | 0.00% | 0.00% | 5.63% | 0.00% | 0.00% | 7.31% | 6.38% | 100.00% |

## A.13 PCAs (Gesture Elicitation)

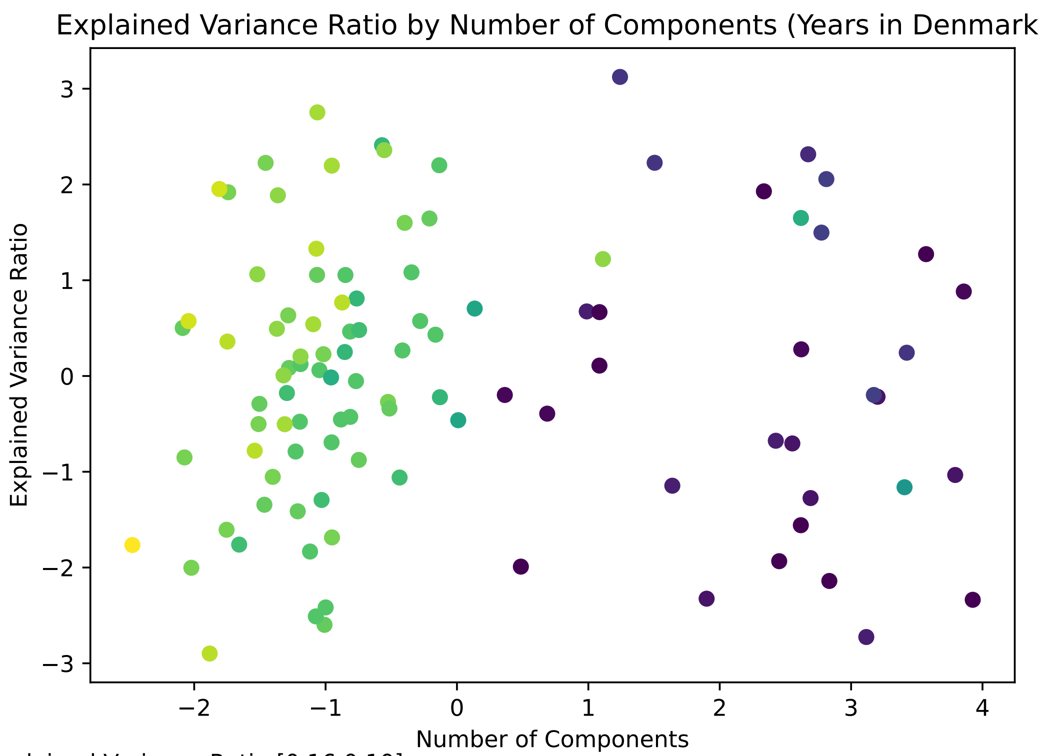
The following pdf shows the PCAs from the gesture elicitation. They are filtered on different variables as defined at the top of each page



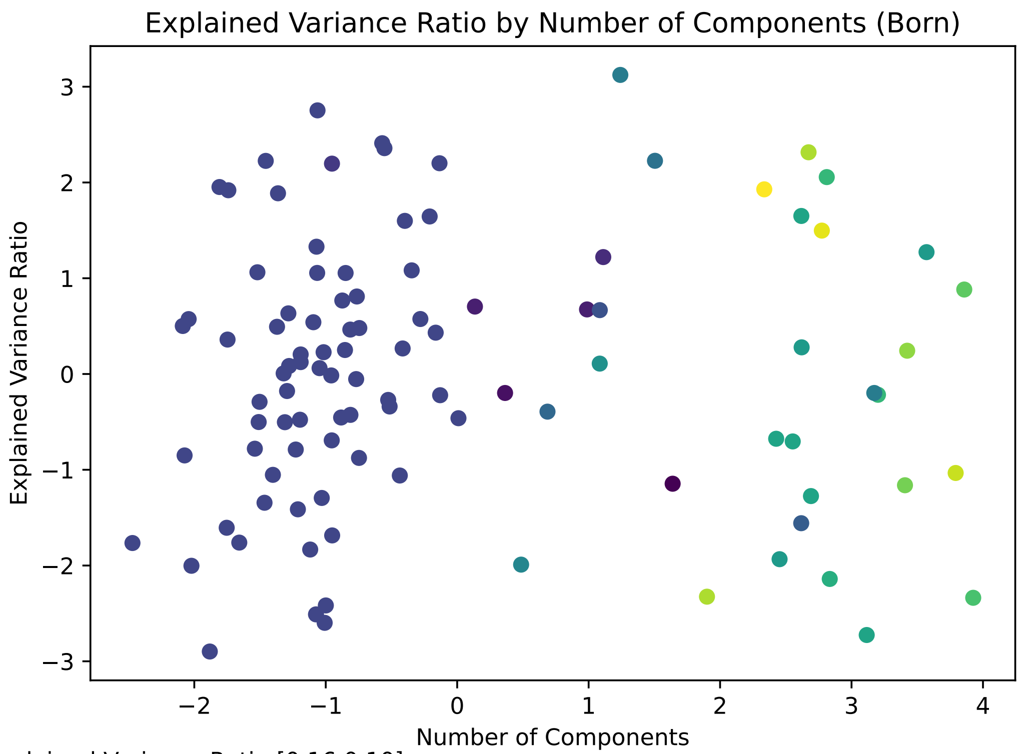

Explained Variance Ratio:[0.15986522 0.10175574]



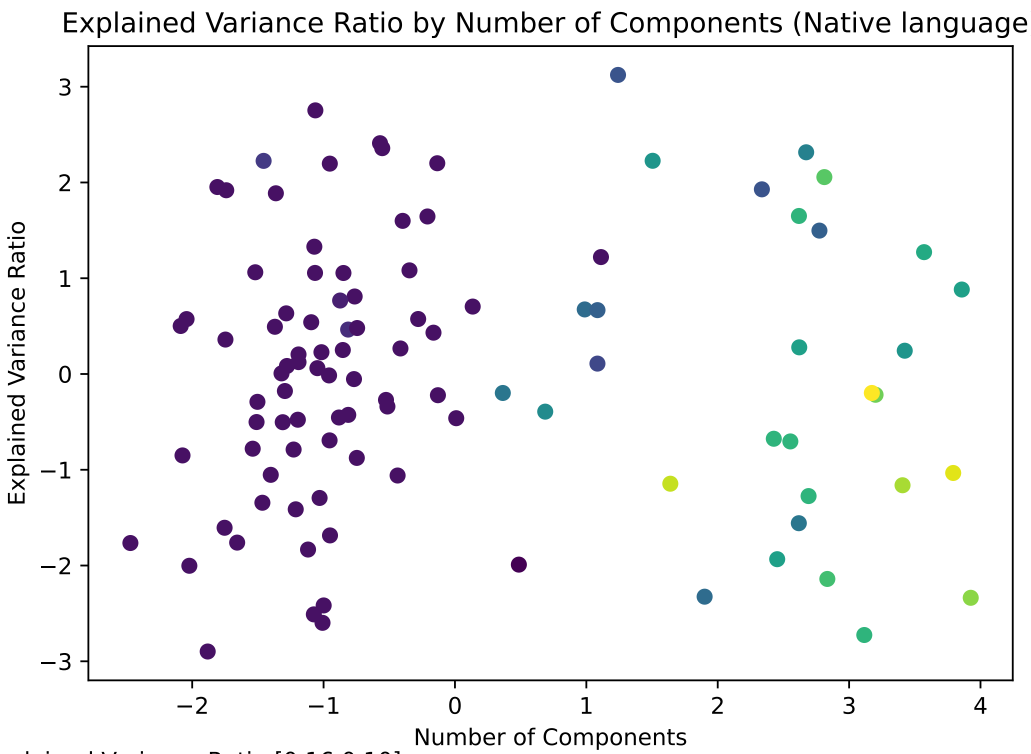

Explained Variance Ratio:[0.16 0.10]



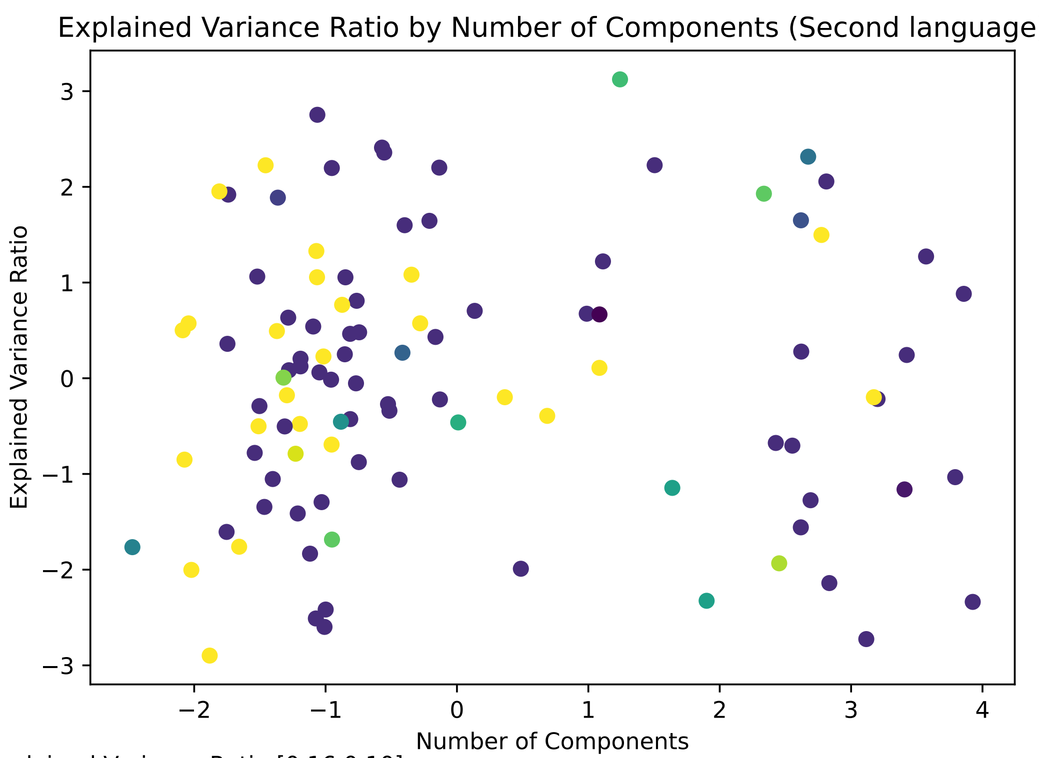

Explained Variance Ratio:[0.16 0.10]



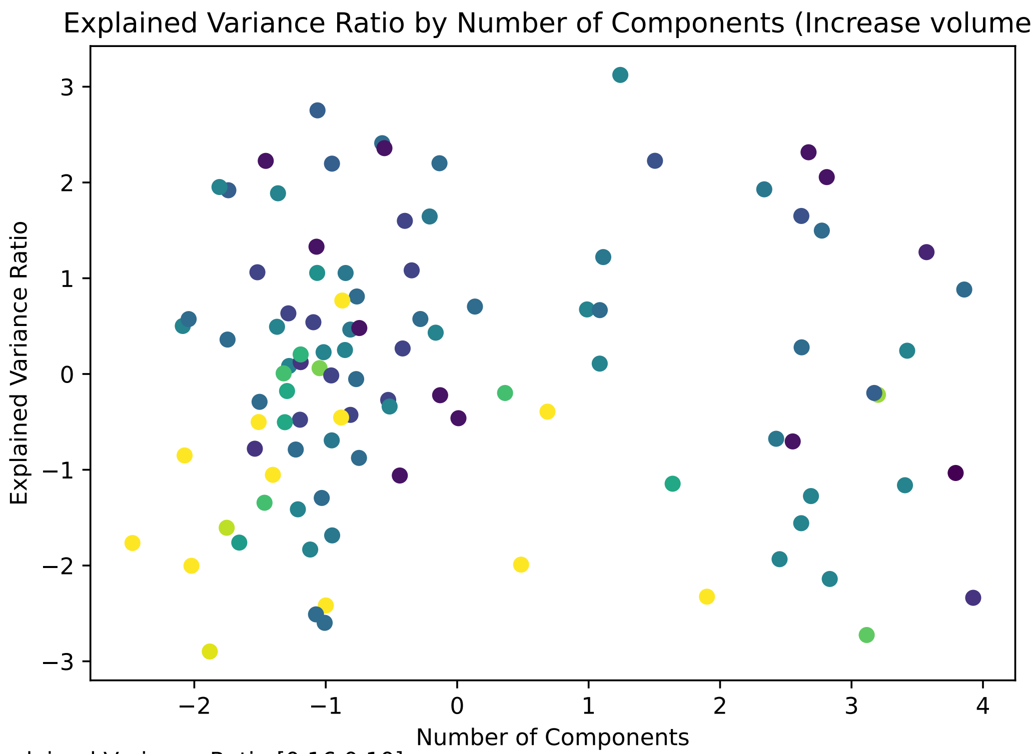

Explained Variance Ratio:[0.16 0.10]



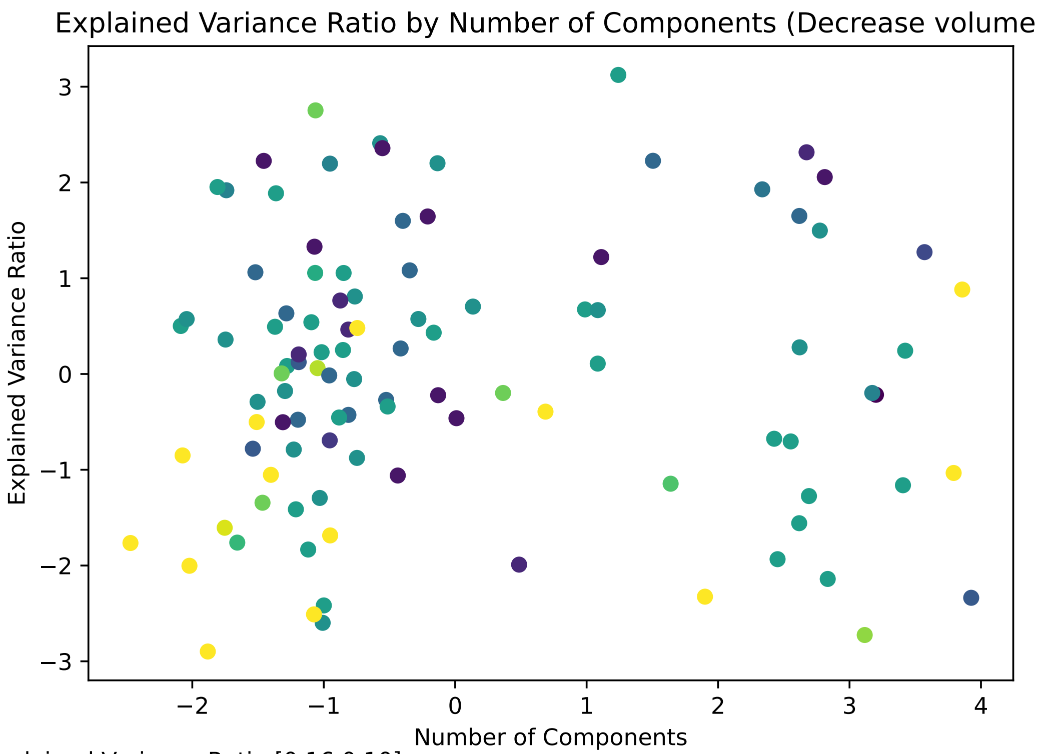

Explained Variance Ratio:[0.16 0.10]



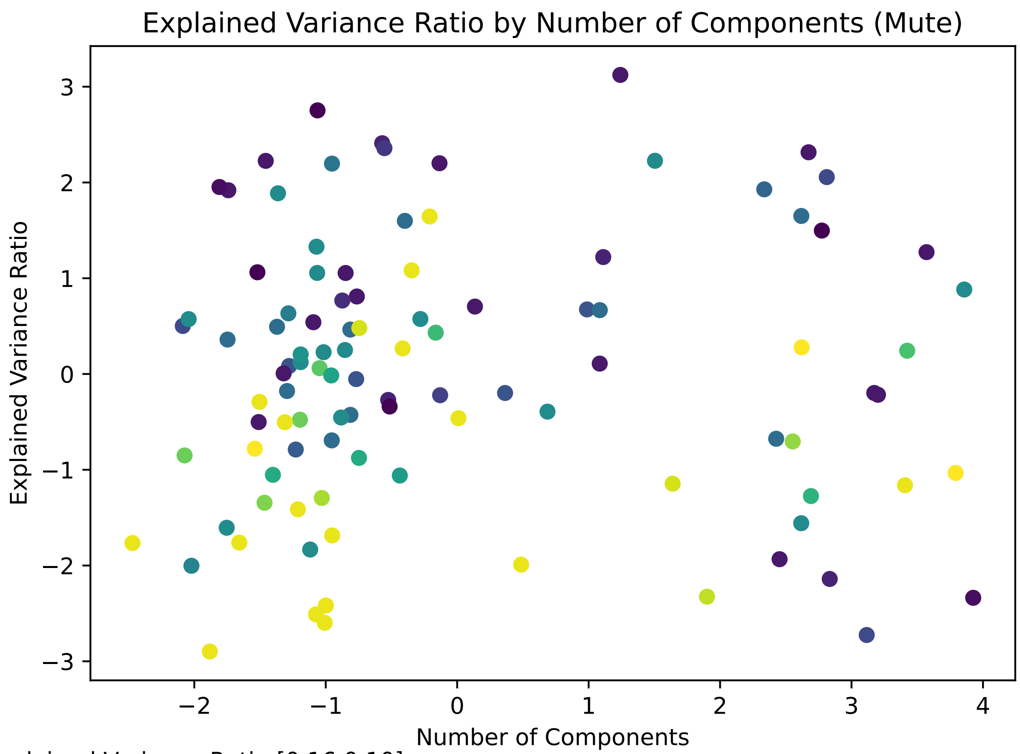

Explained Variance Ratio:[0.16 0.10]



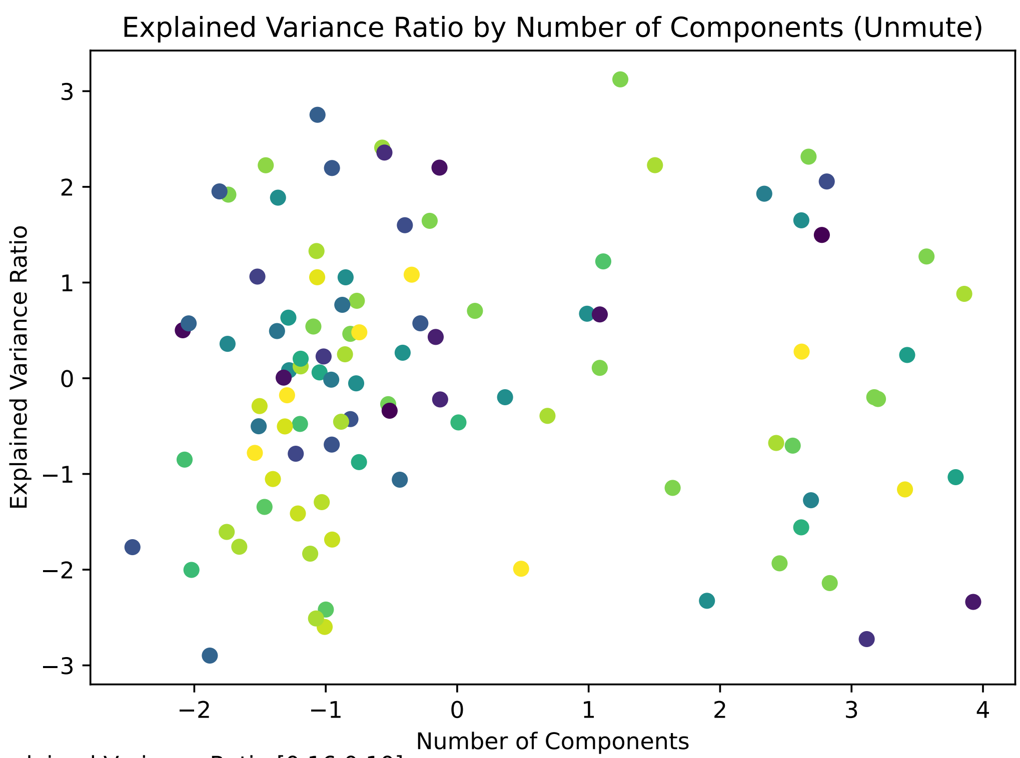

Explained Variance Ratio:[0.16 0.10]



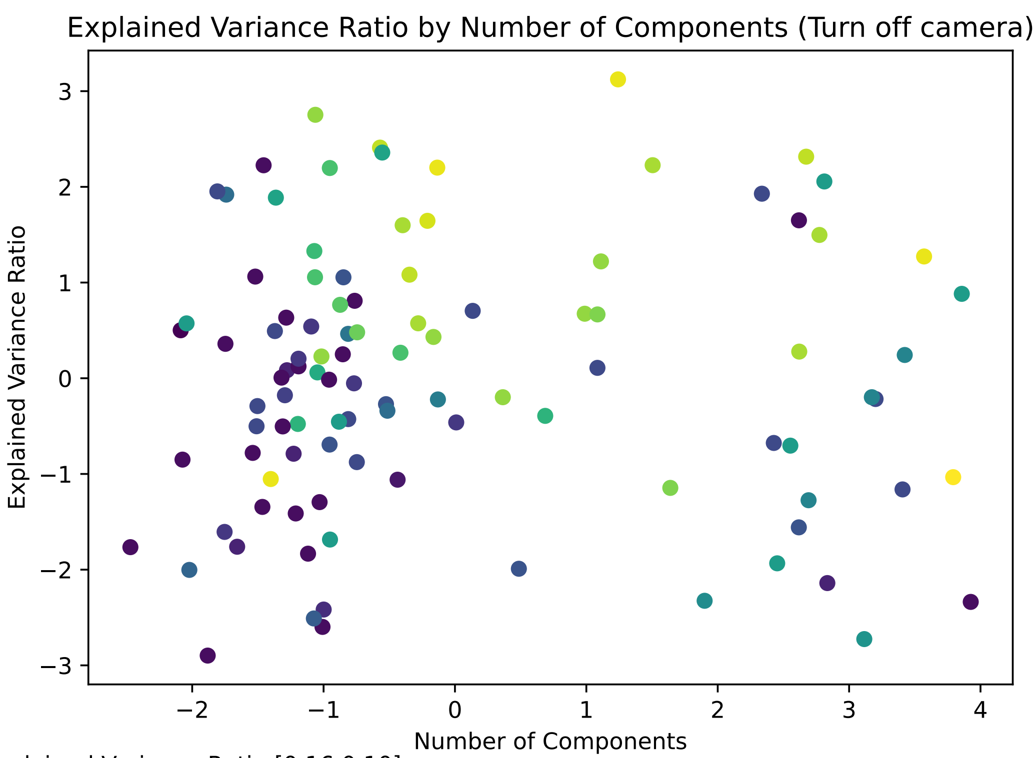

Explained Variance Ratio:[0.16 0.10]



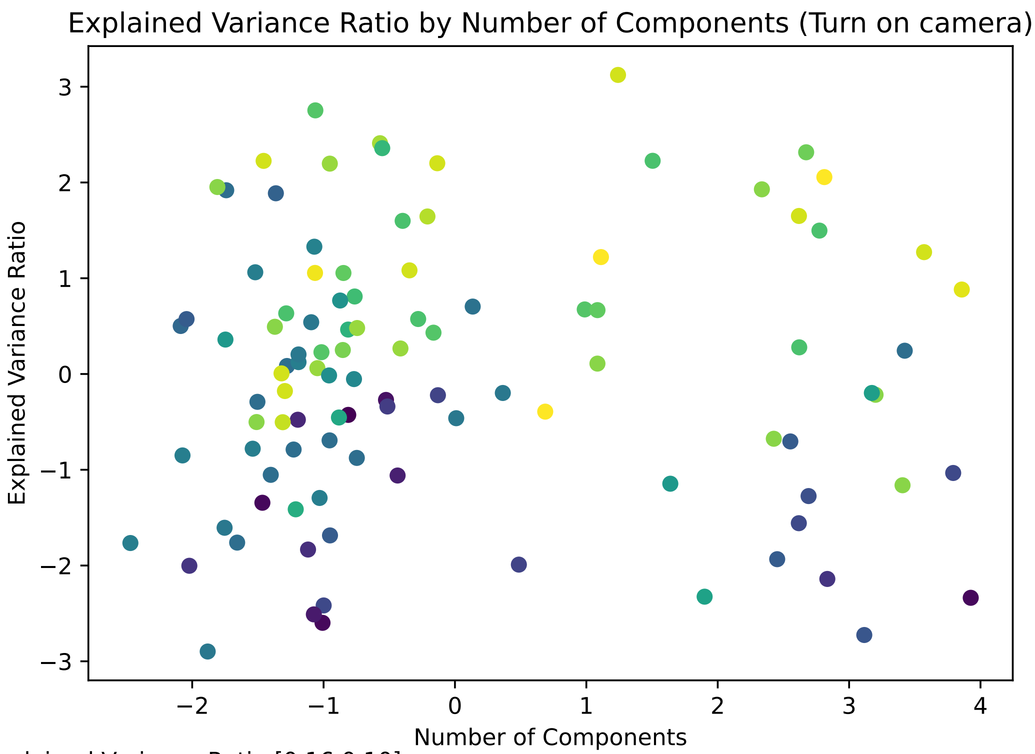

Explained Variance Ratio:[0.16 0.10]



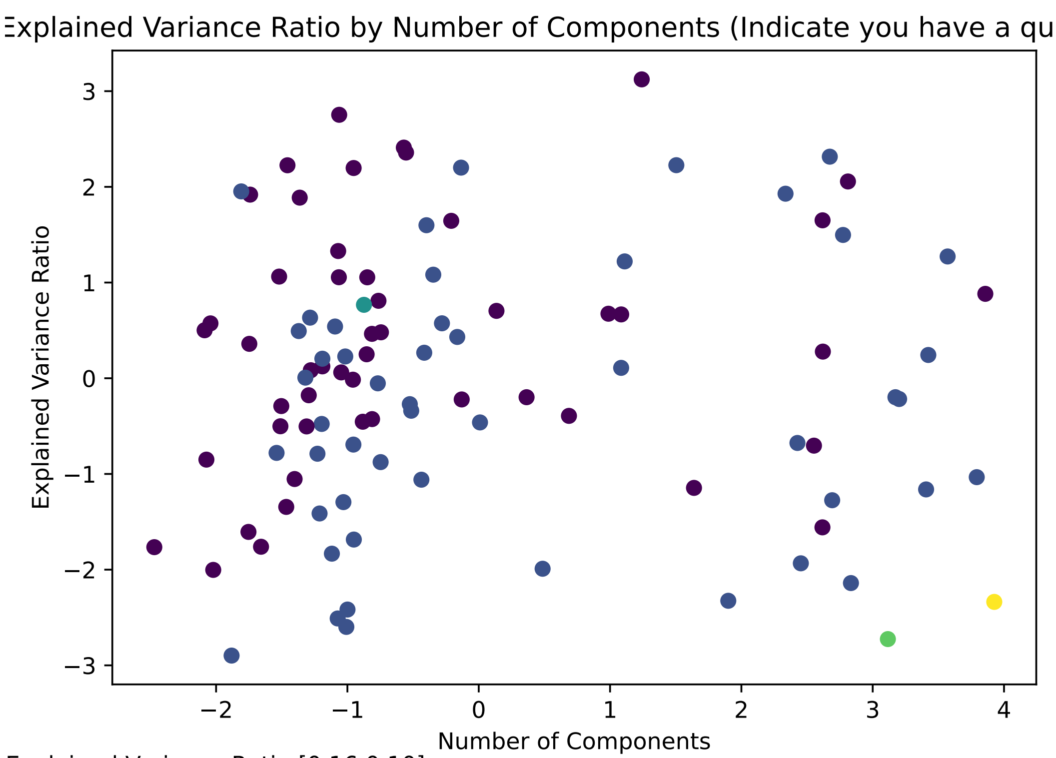

Explained Variance Ratio:[0.16 0.10]



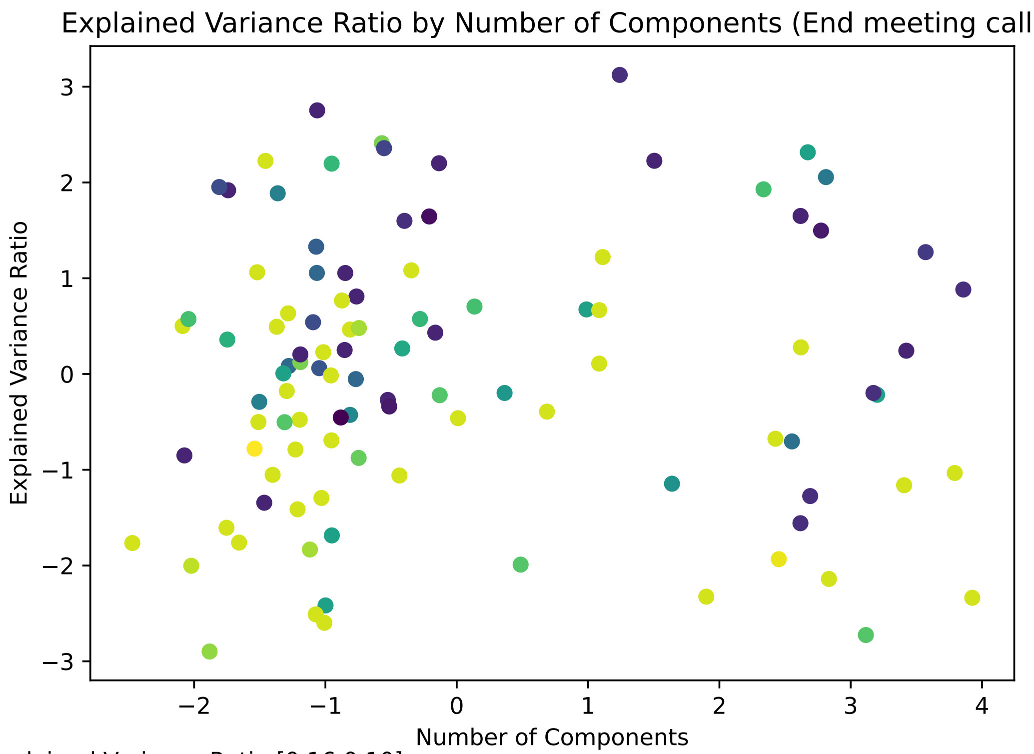

Explained Variance Ratio:[0.16 0.10]



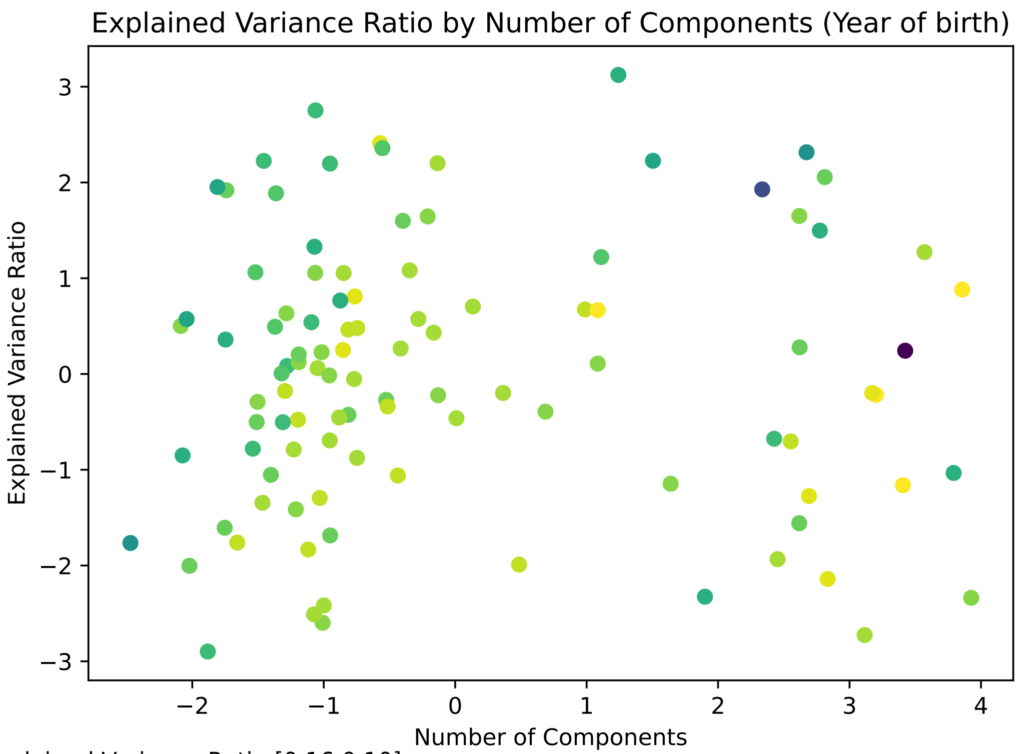

Explained Variance Ratio:[0.16 0.10]



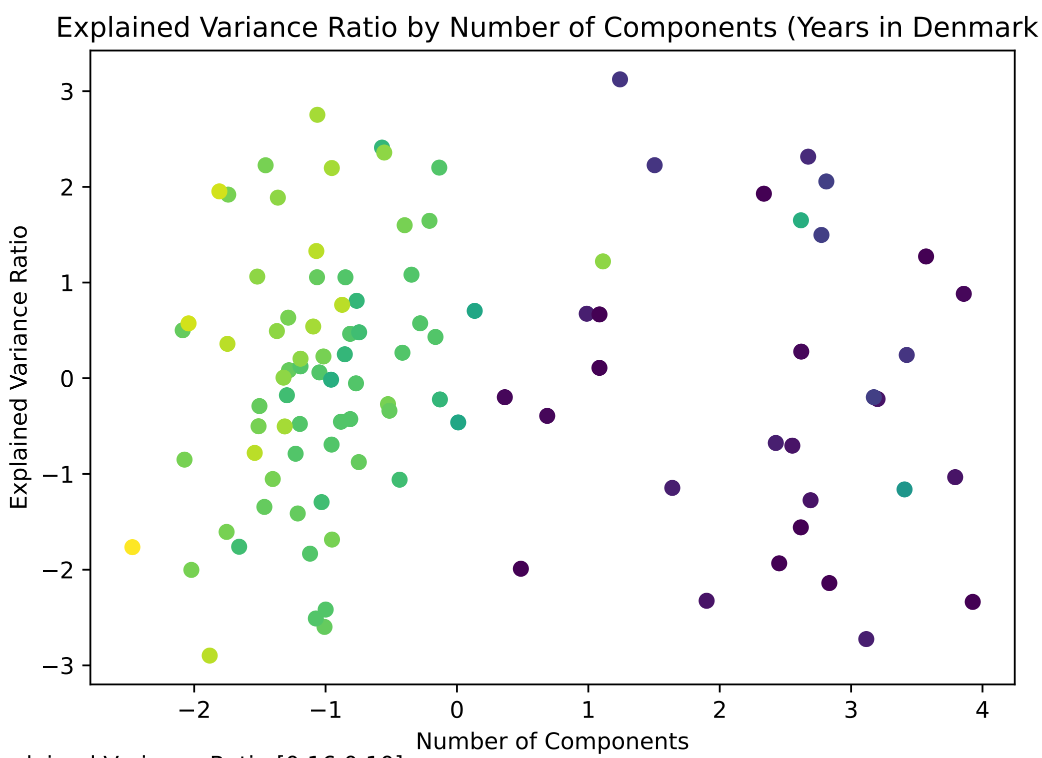

Explained Variance Ratio:[0.16 0.10]



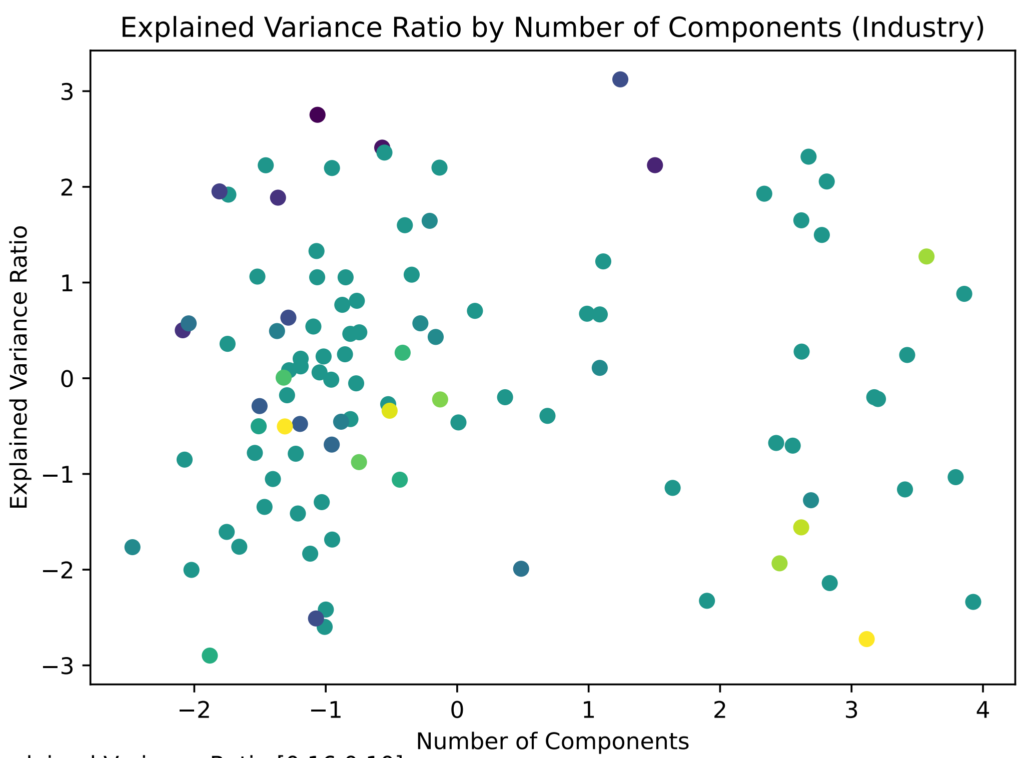

Explained Variance Ratio:[0.16 0.10]




Explained Variance Ratio:[0.16 0.10]



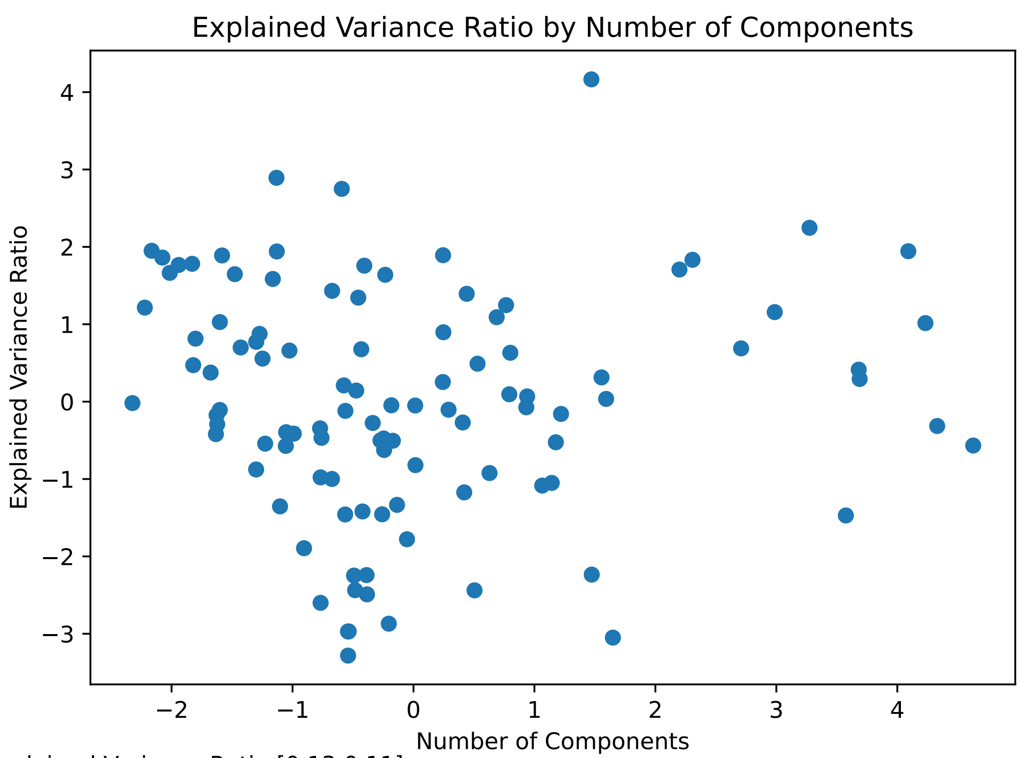

Explained Variance Ratio:[0.16 0.10]



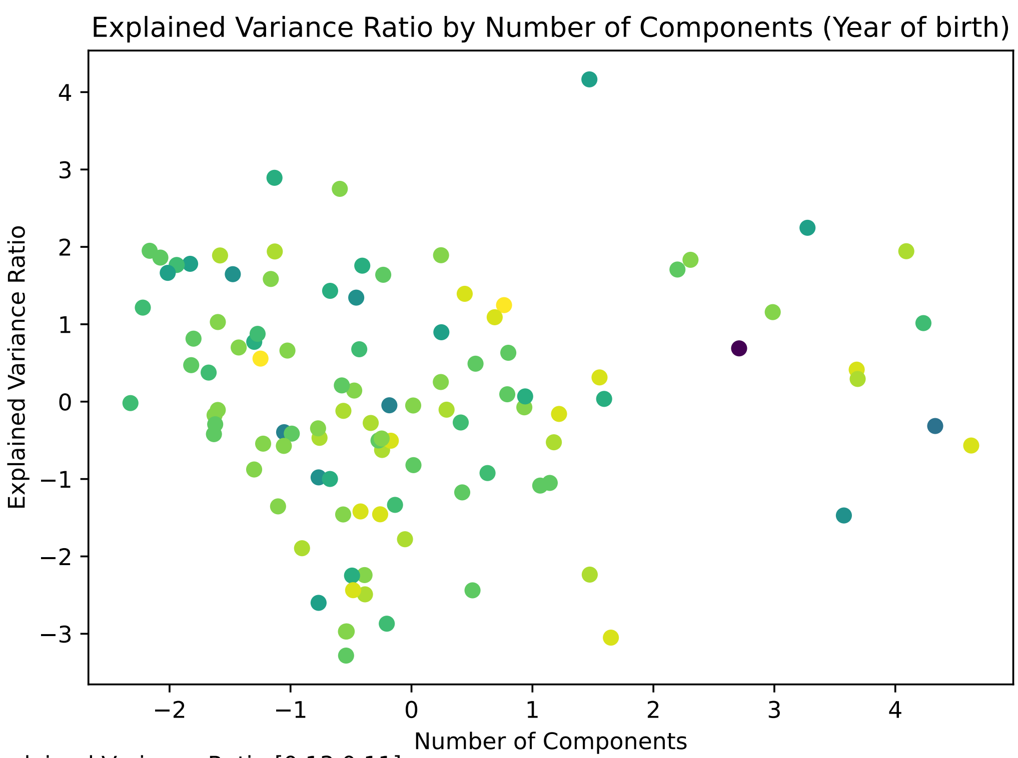

Explained Variance Ratio:[0.16 0.10]

## A.14 PCAs (A/B test)

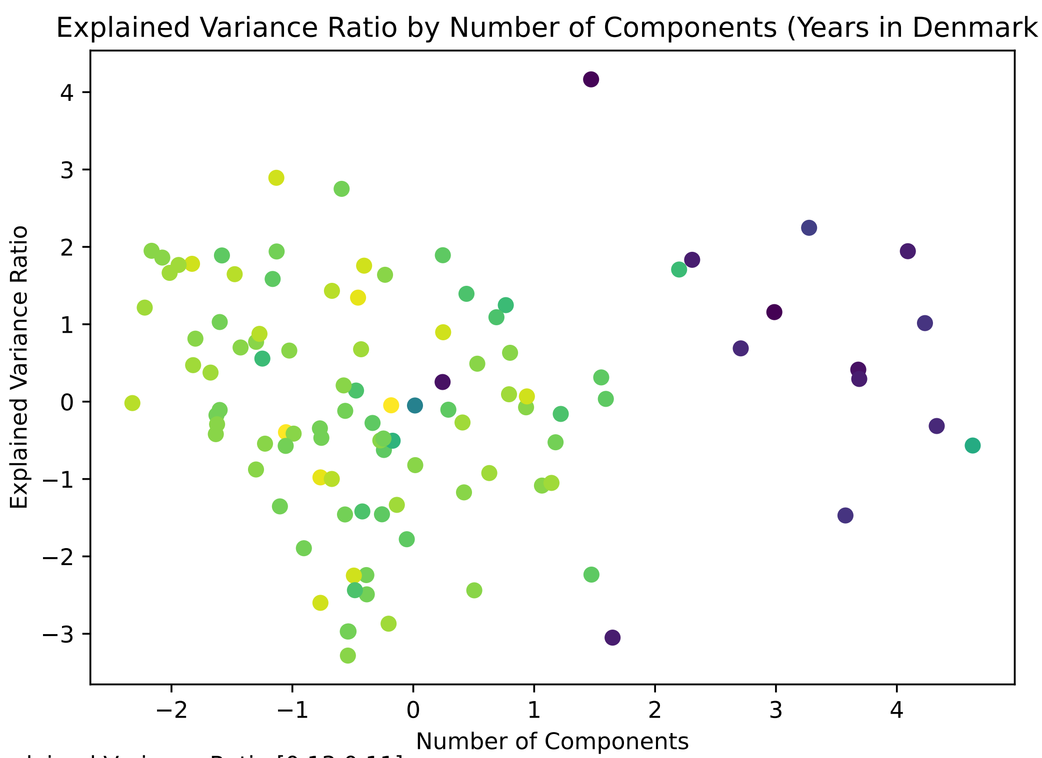
The following pdf shows the PCAs from the A/B test. They are filtered on different variables as defined at the top of each page



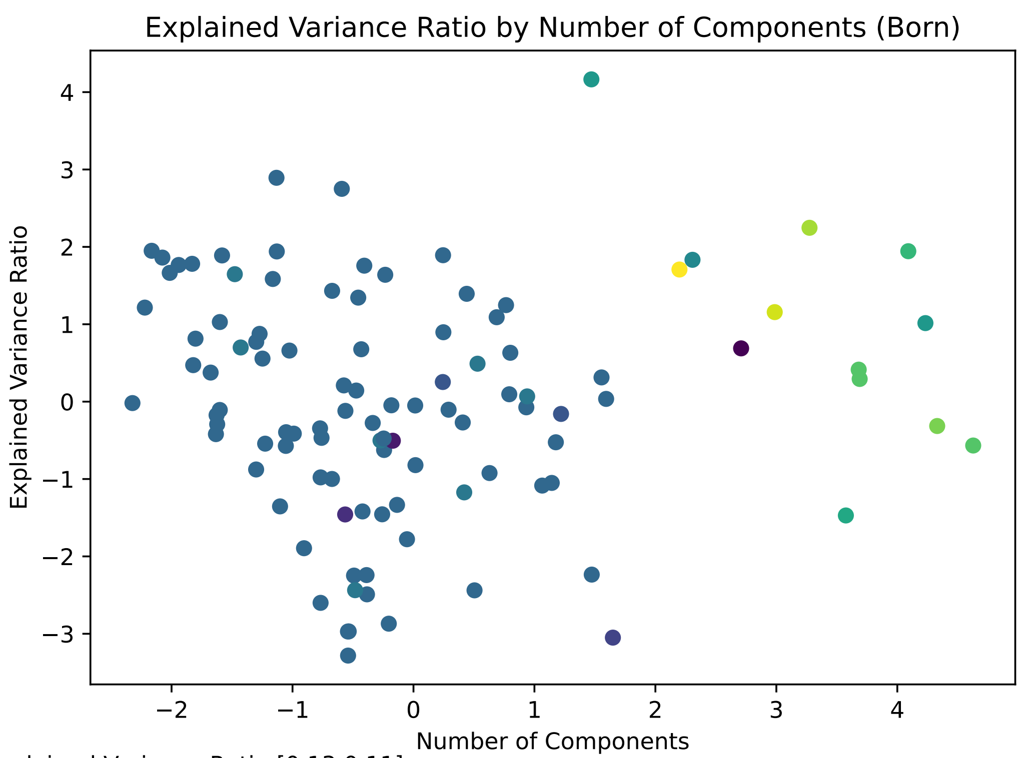

Explained Variance Ratio:[0.13 0.11]



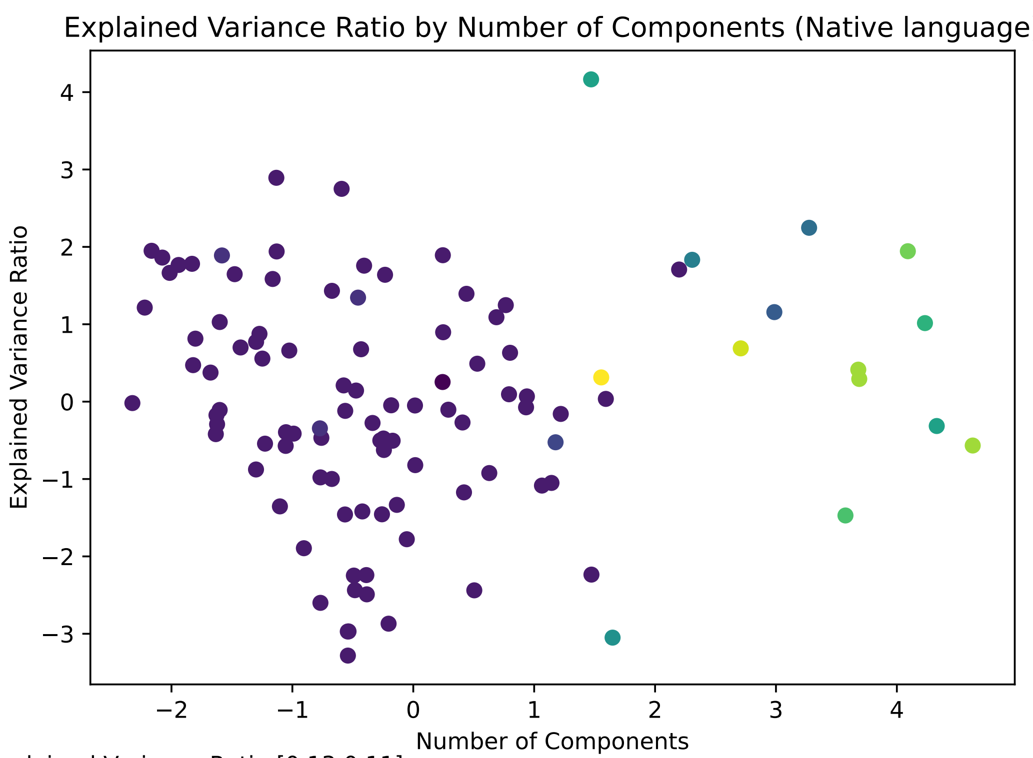

Explained Variance Ratio:[0.13 0.11]



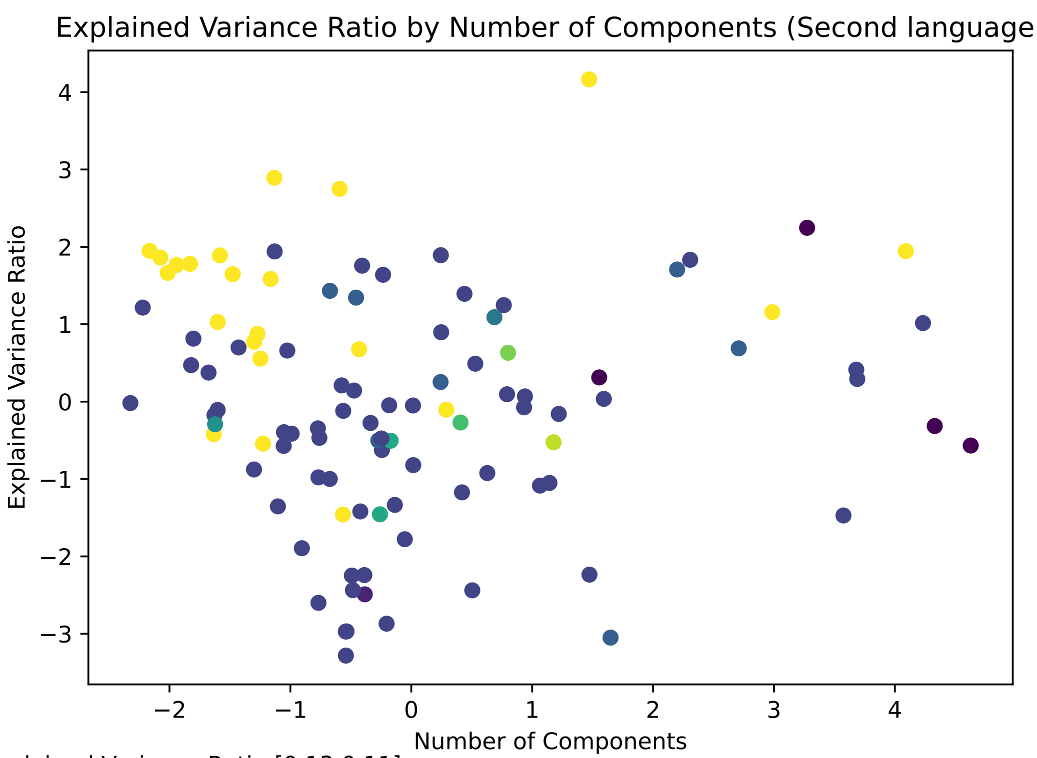

Explained Variance Ratio:[0.13 0.11]



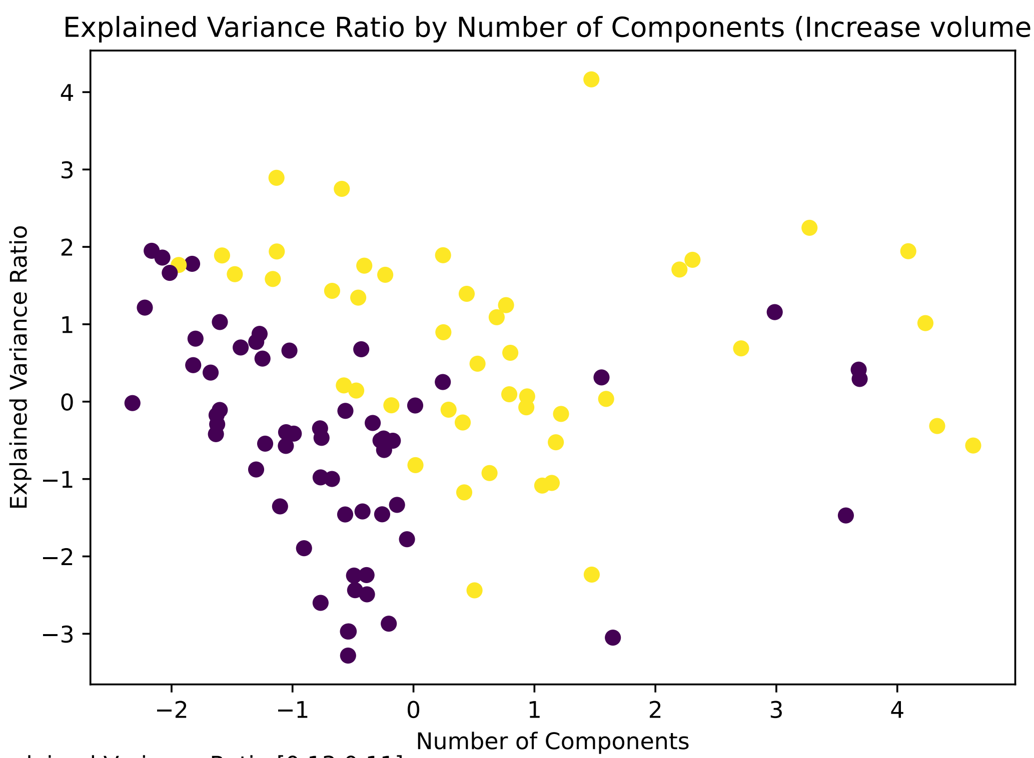

Explained Variance Ratio:[0.13 0.11]



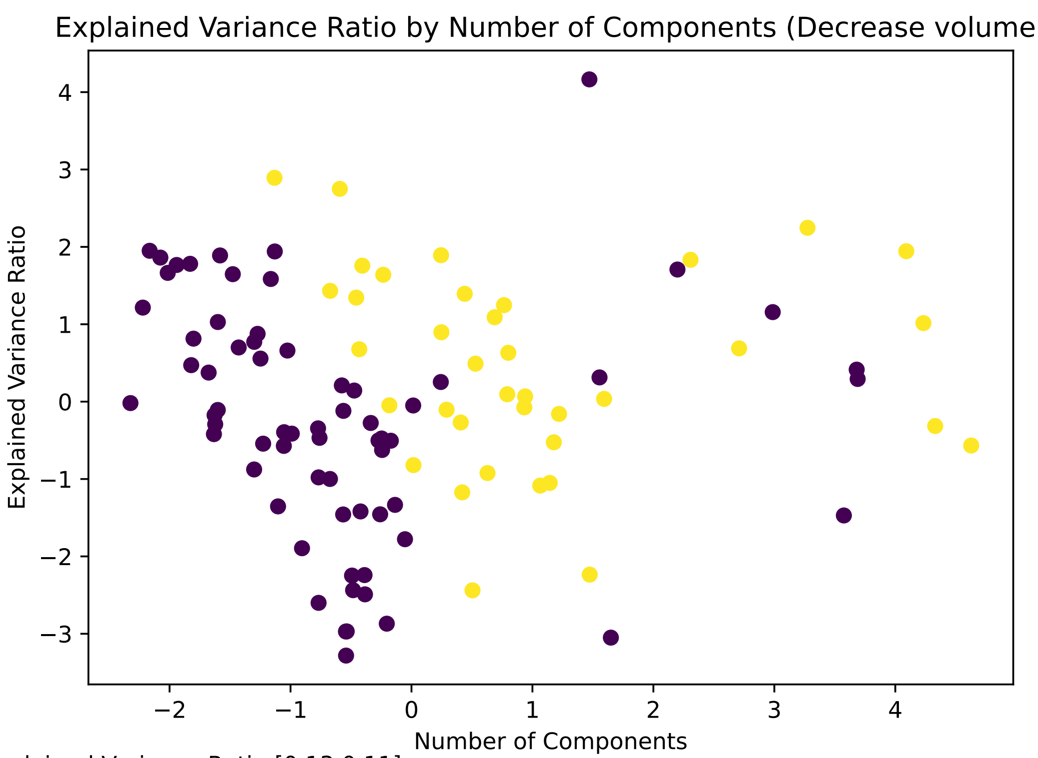

Explained Variance Ratio:[0.13 0.11]



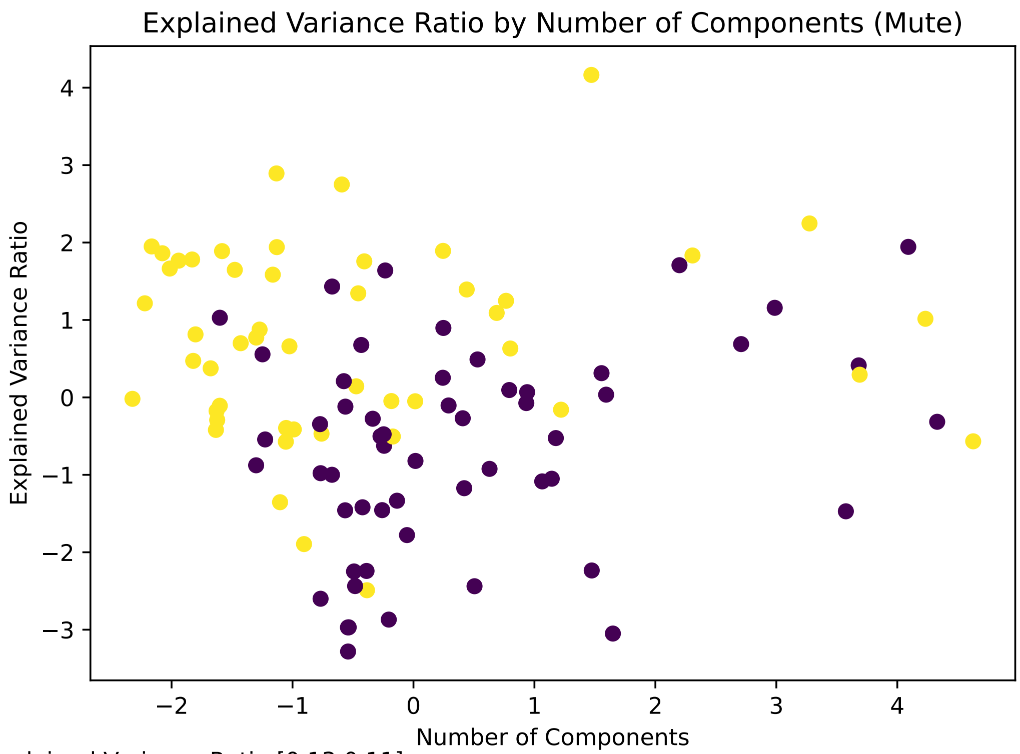

Explained Variance Ratio:[0.13 0.11]



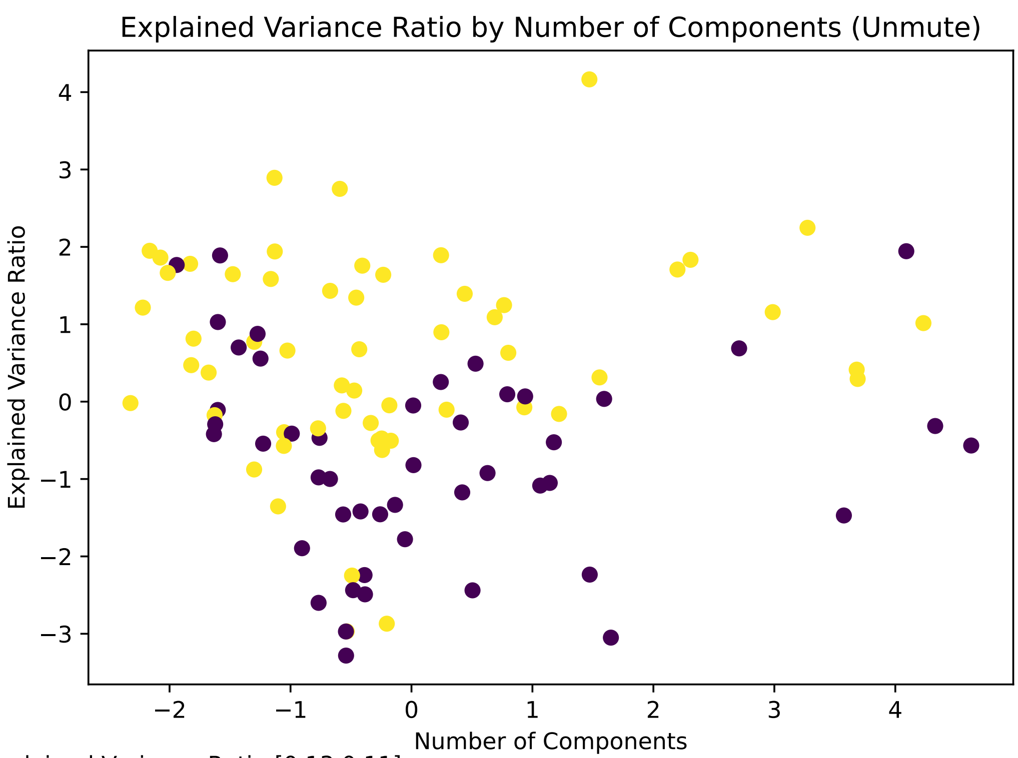

Explained Variance Ratio:[0.13 0.11]



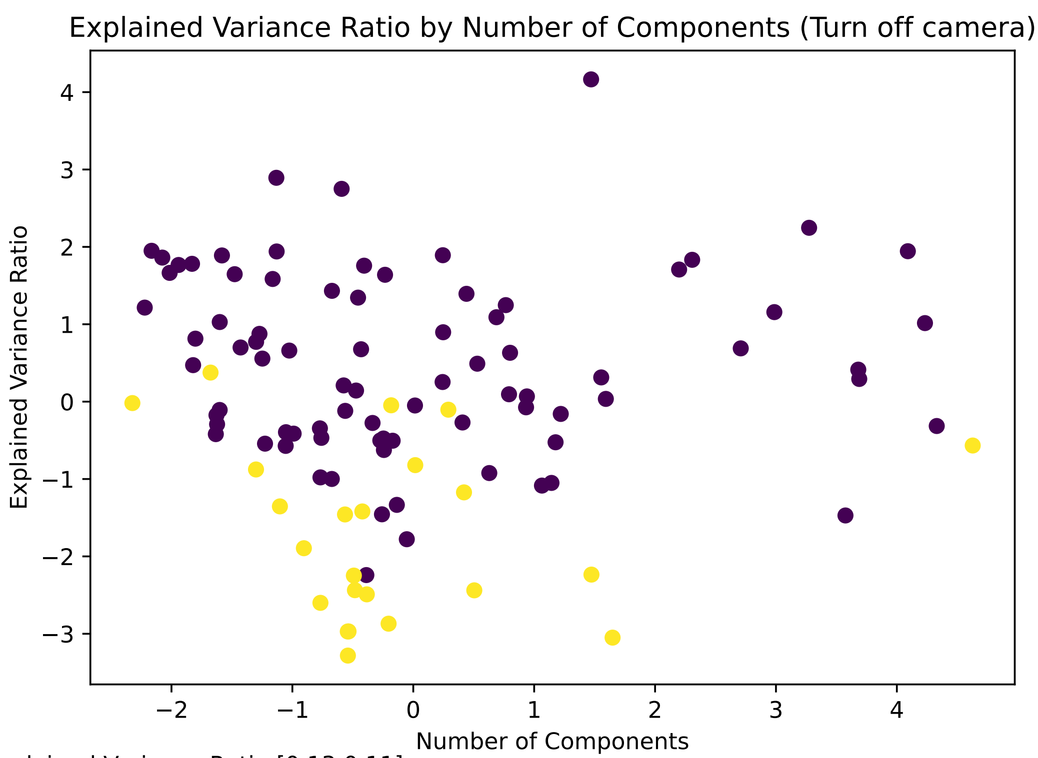

Explained Variance Ratio:[0.13 0.11]



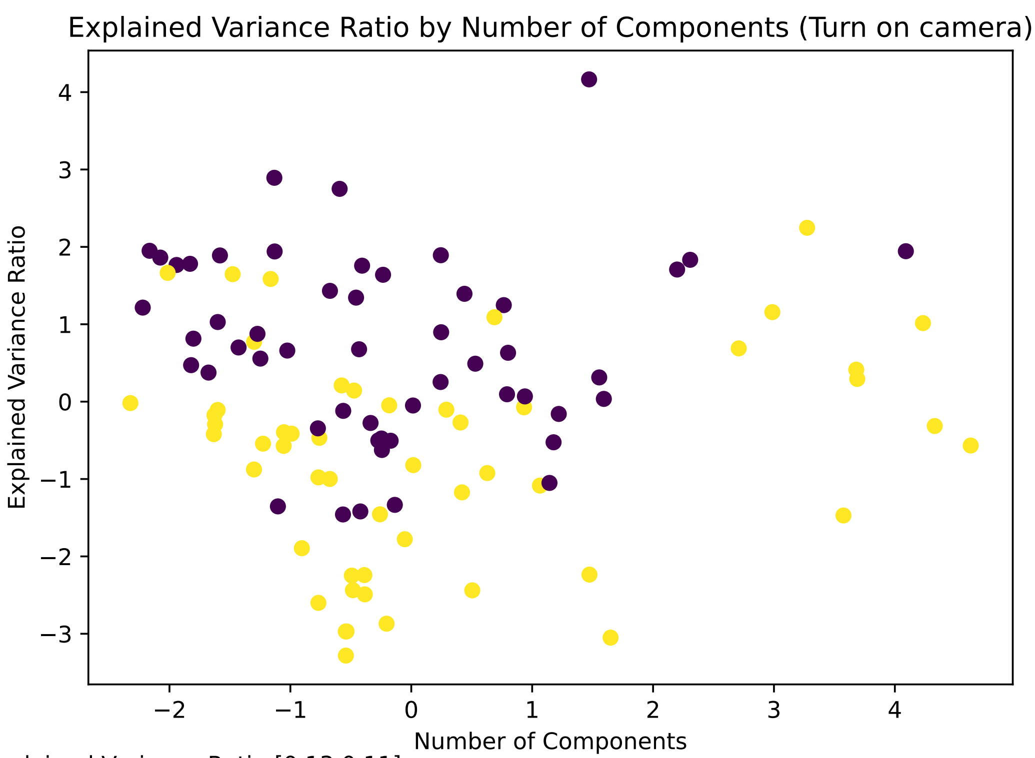

Explained Variance Ratio:[0.13 0.11]



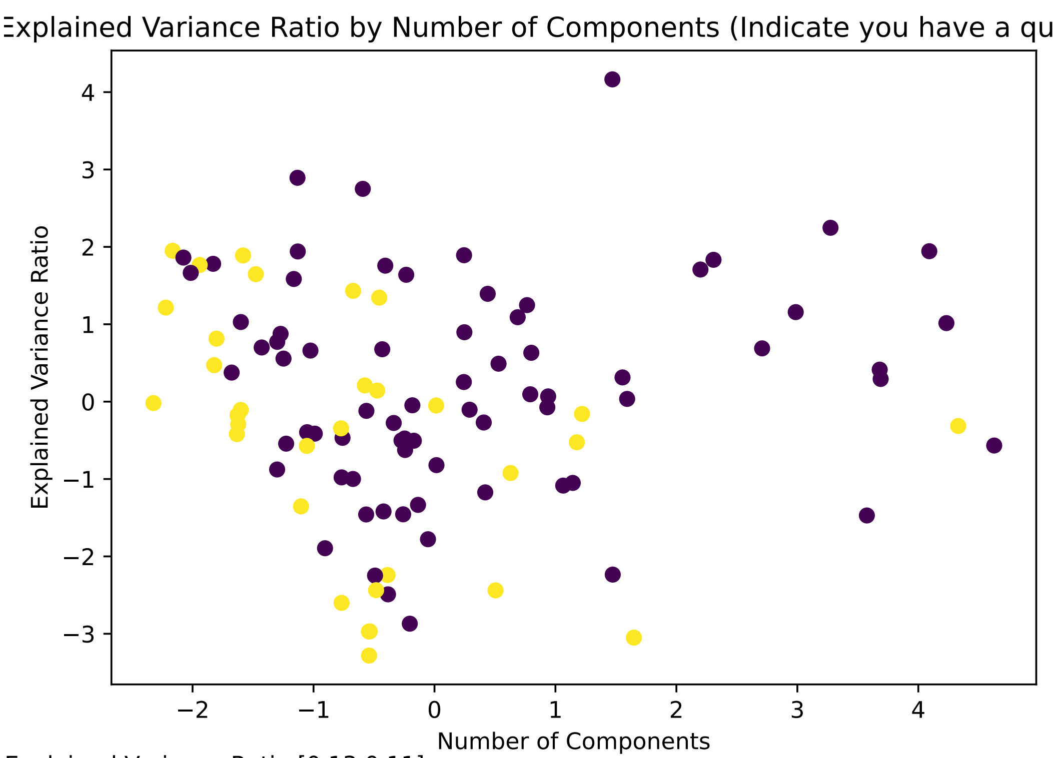

Explained Variance Ratio:[0.13 0.11]



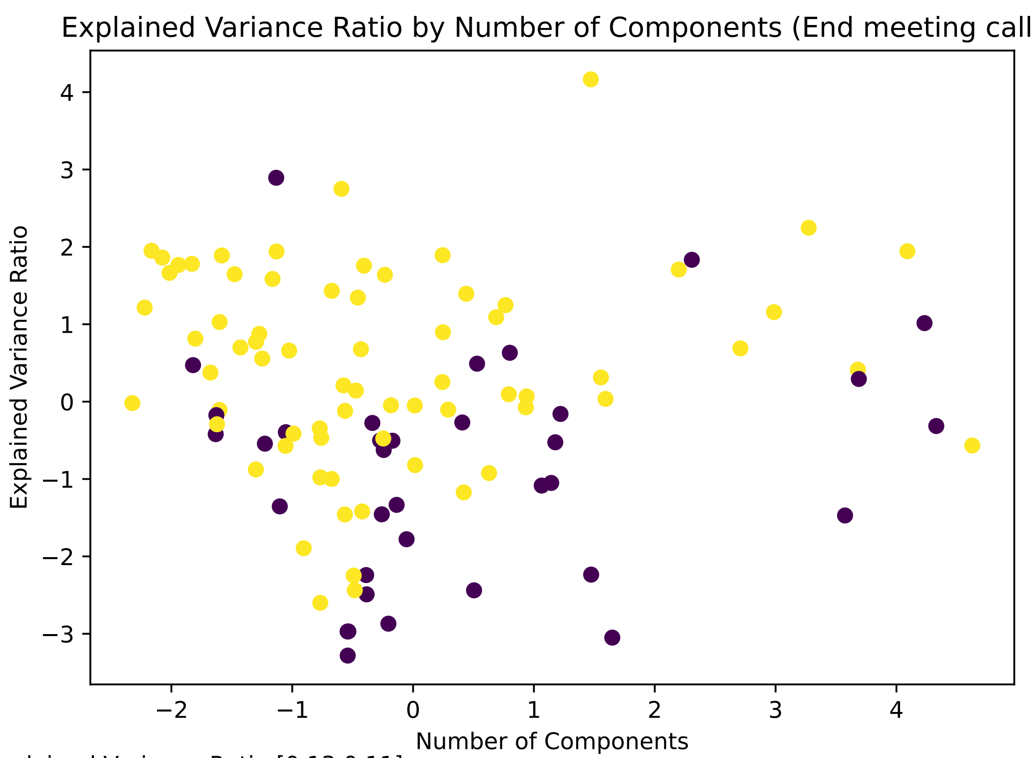

Explained Variance Ratio:[0.13 0.11]



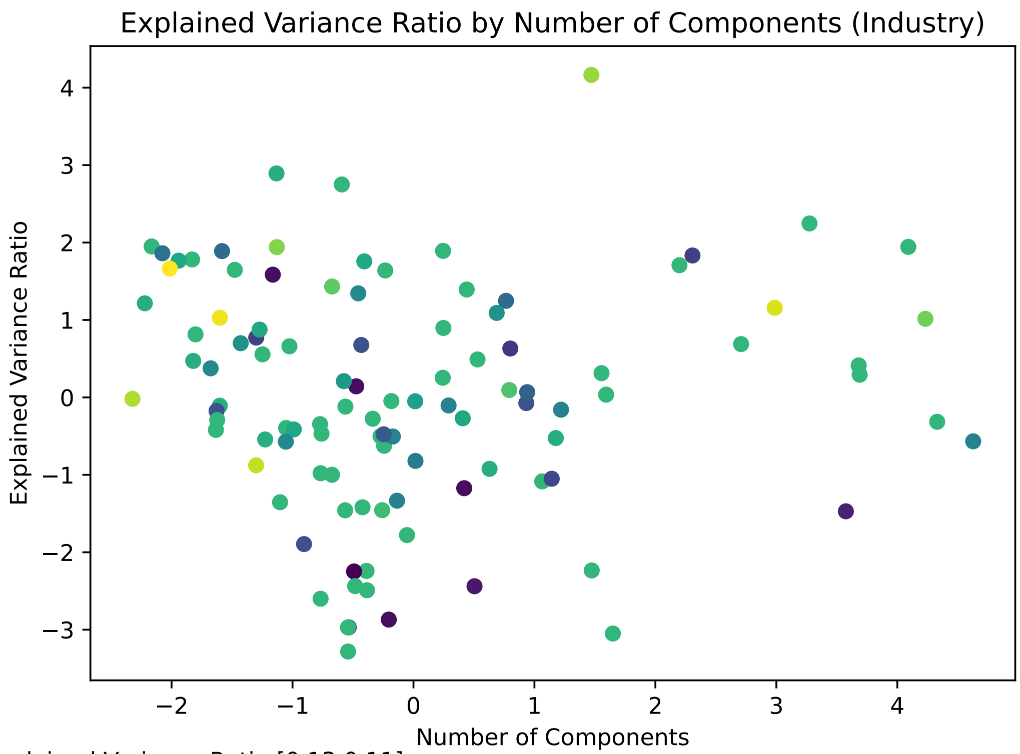

Explained Variance Ratio:[0.13 0.11]




Explained Variance Ratio:[0.13 0.11]

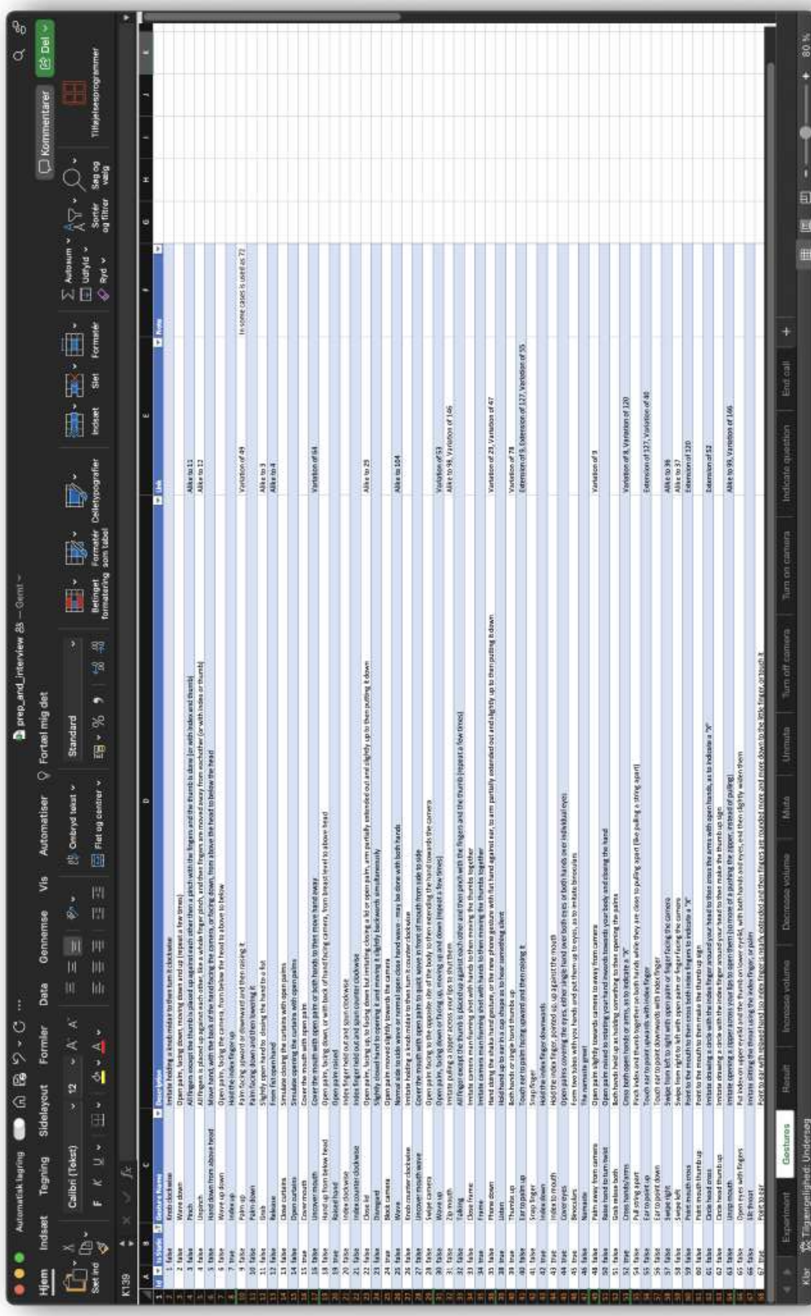



Explained Variance Ratio:[0.13 0.11]




Explained Variance Ratio:[0.13 0.11]

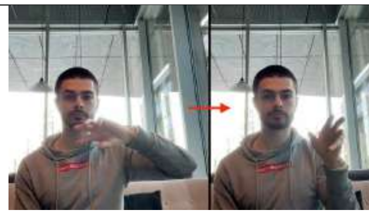
## A.15 Elicitation proposed gestures page

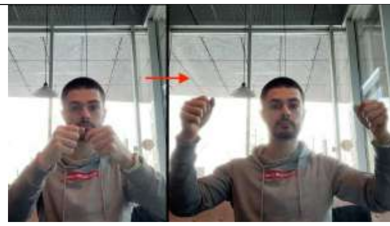
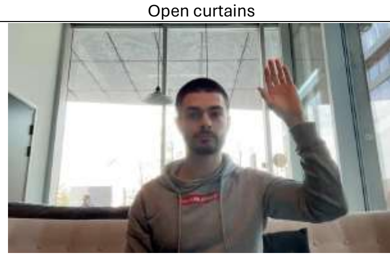
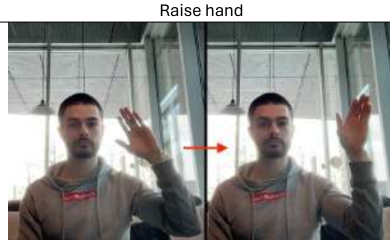
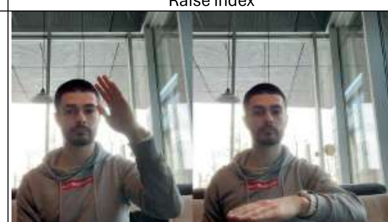


The figure shows a snippet of the elected gestures that were proposed during the gesture elicitation

## A.16 Proposed gestures for A/B test

The following pdf shows the two most popular gestures from the gesture elicitation study which were then used for the A/B test

| | | |
|---|---|---|
| Increase volume |  Knob clockwise |  Palm up |
| Decrease volume |  Knob counter clockwise |  Palm down |
| Mute |  Cover mouth |  Zip mouth |
| Unmute |  Uncover mouth |  Unpinch |
| Turn off camera |  Block camera |  Cover eyes |

| | | |
|---|---|---|
| Turn on camera |  Open curtains |  Uncover eyes |
| Indicate question |  Raise hand |  Raise index |
| End call |  Wave |  Close lid |

## A.17 Amount and percentage of reversible gestures used

| | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Palm up down | Knob | Wave | Point | Index | Slider | Thumb scroll | Slope | Hand from he | Pinch | Index clock | SUM | |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 83 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 84 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 85 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 86 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 87 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 88 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 89 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 90 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 91 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 92 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 93 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 94 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 95 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | |
| 96 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 97 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 99 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 100 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 101 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 102 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 103 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 104 | | | | | | | | | | | Sum of sum | 77 | 75,49% |

The figure shows the amount and percentage of reversible gestures used for changing the volume in the gesture elicitation study

| | AH | AI | AJ | AK | AL | AM | AN | AO | AP | AQ | AR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Pinch | Grab | Cover mouth | Swipe | Zip | Flat hand | Cross | Lid | Timeout | SUM | |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 4 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 83 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 84 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 86 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 88 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | |
| 89 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 90 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 91 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 92 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 94 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 95 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 96 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 97 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 99 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 101 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 102 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 103 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 104 | | | | | | | | | Sum of sum | 37 | 36,27% |

The figure shows the amount and percentage of reversible gestures used for muting and unmuting in the gesture elicitation study

| | AS | AT | AU | AV | AW | AX | AY | AZ | BA | BB | BC | BD | BE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Curtains | Hands | Swipe | Swipe | Curtains | Lid | Pinch | Draw clock | Cover | Timeout | Grab | SUM | |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 83 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | |
| 84 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 85 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 86 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 88 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 89 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 90 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 91 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 92 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 93 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 94 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 95 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 96 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 97 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 99 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | |
| 101 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | |
| 102 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | |
| 103 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 104 | | | | | | | | | | | Sum of sum | 36 | 35,29% |

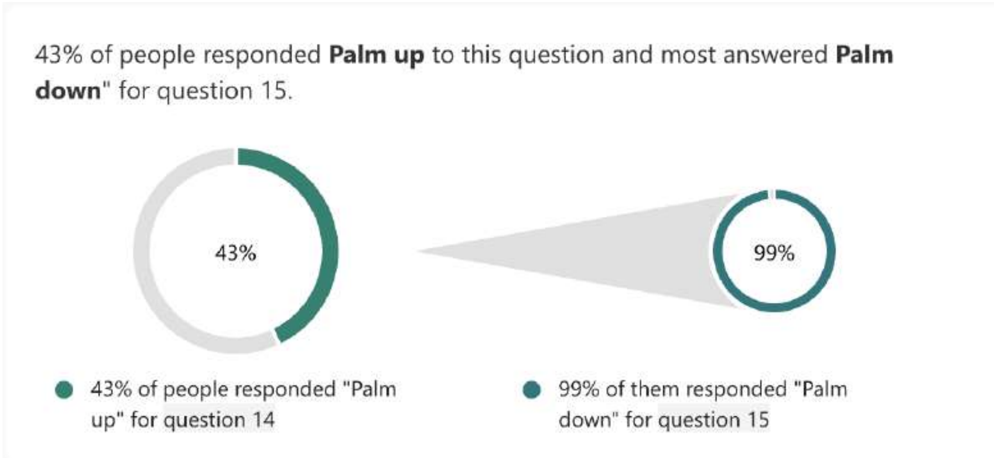
The figure shows the amount and percentage of reversible gestures used for toggling the camera in the gesture elicitation study

## A.18 Amount and percentage of toggleable gestures used

| | BI | BJ | BK | BL | BM | BN |
|---|---|---|---|---|---|---|
| 1 | | Toggle mute | | | Toggle camera | |
| 2 | | 0 | | | 0 | |
| 3 | | 0 | | | 0 | |
| 4 | | 0 | | | 0 | |
| 83 | | 0 | | | 0 | |
| 84 | | 0 | | | 0 | |
| 85 | | 1 | | | 0 | |
| 86 | | 0 | | | 0 | |
| 87 | | 0 | | | 0 | |
| 88 | | 0 | | | 0 | |
| 89 | | 0 | | | 0 | |
| 90 | | 0 | | | 1 | |
| 91 | | 0 | | | 0 | |
| 92 | | 0 | | | 0 | |
| 93 | | 0 | | | 0 | |
| 94 | | 0 | | | 0 | |
| 95 | | 0 | | | 0 | |
| 96 | | 0 | | | 0 | |
| 97 | | 0 | | | 0 | |
| 98 | | 0 | | | 0 | |
| 99 | | 0 | | | 0 | |
| 100 | | 0 | | | 0 | |
| 101 | | 0 | | | 0 | |
| 102 | | 0 | | | 0 | |
| 103 | | 1 | | | 0 | |
| 104 | Sum | 11 | 10,68% | Sum | 12 | 11,76% |

The figure shows the amount and percentage of toggleable gestures used for toggling the microphone and the camera on and off

## A.19 A/B test volume control task




The figure shows that most who chose *palm up* also chose *palm down* for changing the volume. Also, it was 57% that chose the gesture but is registered as less due to individuals being disallowed to answer the form due to their student status

## A.20 Volume reversibility in the A/B test

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Already known facts | | | |
| 2 | Amount who chose Palm down | 67 | Amount who chose Knob counter clockwise | 37 |
| 3 | Percentage who chose Palm down | 64% | Percentage who chose Knob counter clockwise | 36% |
| 4 | Amount who chose Palm up | 59 | Amount who chose Knob clockwise | 45 |
| 5 | Percentage who chose Palm up | 57% | Percentage who chose Knob clockwise | 43% |
| 6 | Percentage of the ones that chose Palm up and then Palm down | 99% | | |
| 7 | Percentage of the ones that chose Palm up and then Knob counter clockwise | 1% | | |
| 8 | | | | |
| 9 | Extrapolated | | | |
| 10 | Percentage of the ones that chose Knob clockwise and then Knob counter clockwise | 81% | Formula used: 1-B2 | |
| 11 | Amount of the ones that chose Knob clockwise and then Knob counter clockwise | 29,94 | Formula used: (D2*B10) | |
| 12 | Percentage of the ones that chose Knob clockwise and then Palm down | 19% | Formula used: (B2-(B4*B6))/D4 | |
| 13 | Amount of the ones that chose Knob clockwise and then Palm down | 12,79 | Formula used: (B2*B12) | |

The figure shows based on the data acquired from Appendix A.19 the percentage that chose Knob clockwise also chose the reversible

## A.21 Static/dynamic gesture distribution (gesture elicitation)

| | G | H | I |
|---|---|---|---|
| 164 | | Sum | Percentage |
| 165 | Dynamic gestures | 120 | 85,11% |
| 166 | Static gestures | 21 | 14,89% |

The figure shows the static vs dynamic gesture distribution in the gesture elicitation study

## A.22 Static/dynamic gesture distribution in tasks (gesture elicitation)

| | Increase volume | Decrease volume | Mute | Unmute | Turn off camera | Turn on camera | Indidate you have a question | End meeting call |
|---|---|---|---|---|---|---|---|---|
| Dynamic: | 81 | 89 | 61 | 81 | 55 | 84 | 1 | 90 |
| Percentage: | 79,41% | 87,25% | 61,62% | 84,38% | 54,46% | 84,85% | 0,99% | 89,11% |
| Static: | 21 | 13 | 38 | 15 | 46 | 15 | 100 | 11 |
| Percentage: | 20,59% | 12,75% | 38,38% | 15,63% | 45,54% | 15,15% | 99,01% | 10,89% |

The figure shows the distribution of static vs dynamic gestures for each task in the gesture elicitation study

## A.23 MVP Sourcecode

The listing shows the source code for the MVP as described in Chapter 5

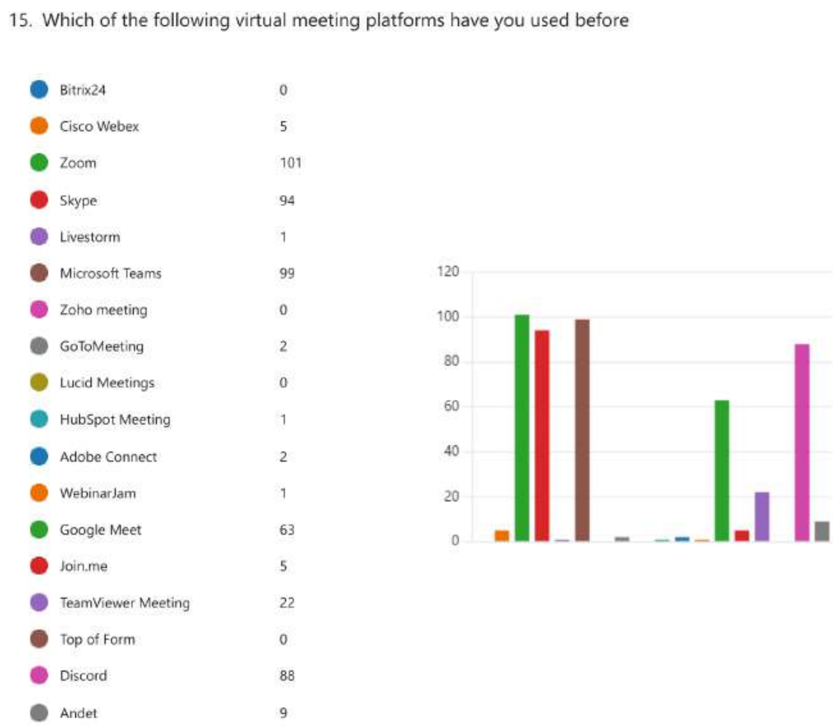
```
from gettext import npgettext
import os
import time
import cv2
import sys
import pyvirtualcam
import mediapipe as mp
import numpy as np
import pyautogui
import threading

# For CONFIDENCE_FRAMES the gestures has to be recognized CONFIDENCE_CUT_OFF
    times
CONFIDENCE_FRAMES = 15
CONFIDENCE_CUT_OFF = 3

# specifies a array that has space for gestures
confidence_array = [""] * CONFIDENCE_FRAMES
iteration = 0
dict = ["palm_up", "palm_down", "index_up",
        "zip_mouth", "block_camera", "wave"]

script_dir = os.path.dirname(__file__)
model_path = os.path.join(script_dir, "./model/gesture_recognizer.task")

debug = False

if len(sys.argv) > 1:
    arg1 = sys.argv[1]
    if arg1 == "debug":
        debug = True

BaseOptions = mp.tasks.BaseOptions
GestureRecognizer = mp.tasks.vision.GestureRecognizer
GestureRecognizerOptions = mp.tasks.vision.GestureRecognizerOptions
GestureRecognizerResult = mp.tasks.vision.GestureRecognizerResult
VisionRunningMode = mp.tasks.vision.RunningMode

cooldown = 3
last_execution_time = 0
lock = threading.RLock()

def change_volume(str, times):
    for _ in range(times):
        pyautogui.press(str)

def match_lock_gesture(gesture):
```

```
    current_time = time.time()
    global last_execution_time
    # Set a resource lock on last_execution time, to avoid race conditions.
    with lock:
        if current_time - last_execution_time >= cooldown:
            match gesture:
                case "zip_mouth":
                    pyautogui.hotkey('alt', 'a')
                case "wave":
                    pyautogui.hotkey('alt', 'q')
                    pyautogui.press('enter')
                    exit(0)
                case "index_up":
                    pyautogui.hotkey('alt', 'y')
                case "block_camera":
                    pyautogui.hotkey('alt', 'v')
                case _:
                    print(gesture)
                    return
            # Set new cooldown time
            last_execution_time = current_time


def match_gesture(gesture):
        print(gesture)
        # Check if cooldown since last gesture execution has passed
        match gesture:
            case "palm_up":
                change_volume('volumeup', 2)
            case "palm_down":
                change_volume('volumedown', 2)
            case "None":
                return
            case _:
                match_lock_gesture(gesture)
                return

def get_most_common_element(arr):
    global CONFIDENCE_CUT_OFF
    global dict
    uses = [0, 0, 0, 0, 0, 0]

    for i in arr:
        match i:
            case "palm_up":
                uses[0] += 1
            case "palm_down":
                uses[1] += 1
            case "index up":
                uses[2] += 1
```

```
            case "zip_mouth":
                uses[3] += 1
            case "block_camera":
                uses[4] += 1
            case "wave":
                uses[5] += 1
            case _:
                pass

    max_value = max(uses)
    max_index = uses.index(max_value)

    if max_value < CONFIDENCE_CUT_OFF:
        return "None"
    else:
        return dict[max_index]


def handle_gesture(result: GestureRecognizerResult, output_image: mp.Image,
    timestamp_ms: int):
    global iteration
    global confidence_array
    global CONFIDENCE_FRAMES
    if (iteration >= CONFIDENCE_FRAMES):
        most_common = get_most_common_element(confidence_array)
        #print(f"\r{confidence_array}")
        match_gesture(most_common)
        confidence_array = [""] * CONFIDENCE_FRAMES
        iteration = 0

    # Iterate through the nested lists returned by result.gestures
    for gesture_list in result.gestures:
        for category in gesture_list:
            if category.score > 0.6:
                confidence_array[iteration] = category.category_name
            else:
                confidence_array[iteration] = "None"

    iteration += 1


options = GestureRecognizerOptions(
    base_options=BaseOptions(model_asset_path=model_path),
    running_mode=VisionRunningMode.LIVE_STREAM,
    result_callback=handle_gesture)
with GestureRecognizer.create_from_options(options) as recognizer:
    with pyvirtualcam.Camera(width=1280, height=720, fps=30) as cam:
        # Create a video capture object
        cap = cv2.VideoCapture(0)
        cap.set(3, 1280)
```

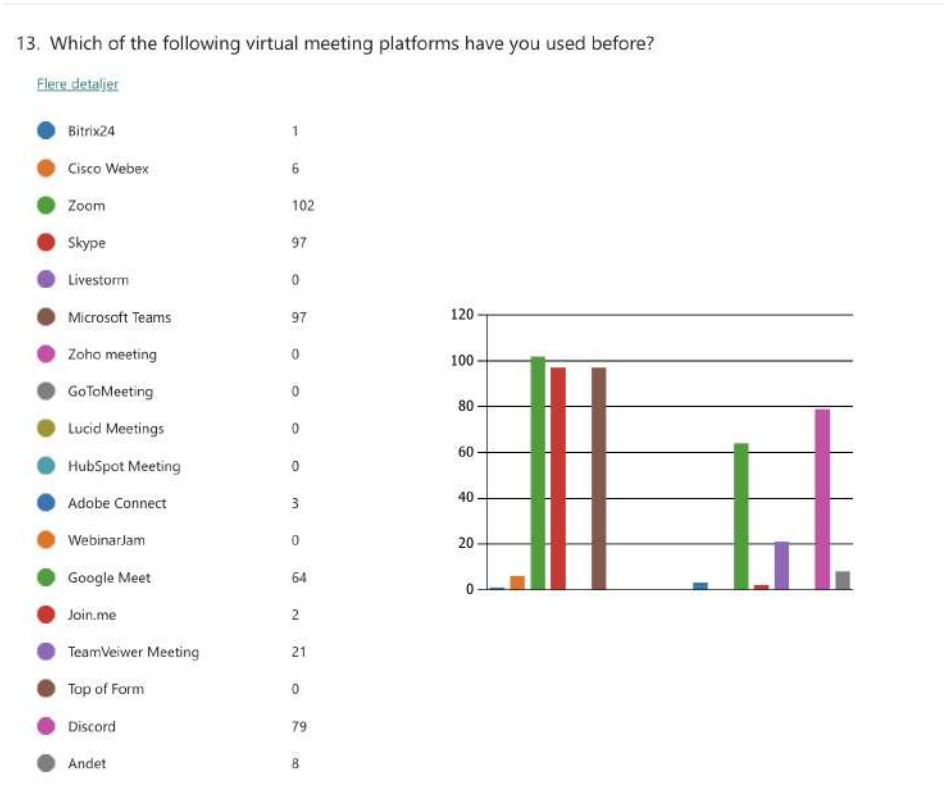
```python
    cap.set(4, 720)

    # Check if the webcam is opened correctly
    if not cap.isOpened():
        raise IOError("Cannot open webcam")

    frame_timestamp_ms = 0
    # Start an infinite loop to read frames from the webcam
    while True:
        # Read a frame from the webcam
        ret, frame = cap.read()

        # Convert the frame received from OpenCV to a MediaPipes Image object.
        mp_image = mp.Image(image_format=mp.ImageFormat.SRGB, data=frame)

        # The gesture recognizer must be created with the live stream mode.
        recognized = recognizer.recognize_async(
            mp_image, int(time.time() * 1000))

        # Convert from BGR to RGB for showing the webcam
        vframe = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)

        cam.send(vframe)
        cam.sleep_until_next_frame()

        # Display the frame in a window
        if debug:
            cv2.imshow('Input', frame)

        frame_timestamp_ms += 1

        # Check for the Esc key to exit the loop
        if cv2.waitKey(1) == 27:
            break

# Release the video capture object and close all windows
cap.release()
cam.close()
cv2.destroyAllWindows()
```

## A.24 Meeting platform experience (Gesture Elicitation)




The figure shows the distribution of which meeting platforms participants had experience with, in the gesture elicitation study

## A.25 Meeting platform experience (A/B test)




The figure shows the distribution of which meeting platforms participants had experience with from the A/B test

## A.26 Usability test interview guide

The following pdf shows the interview guide followed for the usability test

### Interview guide – Usability test

1. Explain the scenario.

"Imagine you are sitting at a hybrid meeting. All your coworkers are sitting around the table, and you are seated far away from the computer. Therefore, when you want to change certain settings or indicate certain things you do not want to interrupt the meeting. Luckily Zoom now recognises gestures."

2. Show the gestures that can be used (palm down, palm up, block camera, index up, wave)

3. Start the meeting and program.

4. Questions:

   1. You have a hard time hearing what the client connected to the meeting says. You therefore want to increase the volume. How would you do this using only your hands?

   2. The volume is now a bit too loud. How would you decrease the volume using your hands?

   3. After a bit of discussing back and forth with the client you want to discuss something with your coworkers without the client being able to hear what you are saying. How would you mute your microphone using your hands?

   4. You have reached a conclusion and want to unmute. How do you do this using your hands?

   5. The call has been going on for quite a while and you have a short break. In the meantime, you do not want the client to see what you are doing and therefore want to turn the camera off. How do you turn it off using your hands?

   6. The break is over, and you have to turn on the camera again. How do you do this using your hands?

   7. Before you end the meeting you just want to ask a question without interrupting the current discussion. How would you do this using your hands?

   8. The meeting is finally over. How would you end the call using your hands?

5. What issues did you experience?

6. What did you like?

7. How did you find the experience generally?

8. What improvements would you like to see?

## A.27 Usability test (transcriptions)

The following subsections show the transcriptions of participants' answers to the usability test questions. They are transcribed in Danish and translated into English with their respective meaning in mind

### A.27.1 Usability test - Individual 1

Feedback in the original language (Danish):

> *Jeg følte ikke den registrerede når jeg gjorde noget med det same. Så det krævede flere forsøg og jeg vidste ikke om jeg gjorde de rigtige gestures nødvendigvis eller ej. Det kunne måske være rart med noget visuelt hjælp at have din hånd skal være sådan her.*
>
> *Jeg ved ikke hvor langt væk fra kameraet for det virker. For det virker til at nogle gange skal man helt tæt og nogle gange er man måske for hurting. Så det kunne være rart med en indikator der siger noget med "Slow down" hvis man gør det for hurtigt. Og ja så det også svært når kameraet er slukket for så ved jeg ikke hvad jeg egentlig gør i dens vinkel.*
>
> *Jeg kan godt lide at lave gestures. Men nogle af dem føles også lidt unaturlige. Altså som om man skal matche en meget bestemt bevægelse. Ligesom den der farvel gesture. Men det er meget nice at der er en vifte af forskellige commands man kan lave med gestures.*
>
> *Den gjorde det som den skulle. Altså jeg behøvede ikke at røre ved computeren så det var meget rart. Jeg kan godt se hvordan det kunne være meget brugbart hvis man har travlt med andet også man lige skulle pause hurtigt eller jeg skal lige skrue op for at høre. Så det giver også en hurtigere mulighed end at hvis man skulle behøve at tage computeren frem og så pludselig skal skrue op og ned og sådan noget. Det er også svært hvis man har forskellige systemer tænker jeg. Hvis der er nogle der siger du skal lige trykke på denne knap for at mute. Og så får man ikke fat i hvilken knap og så er det lettere bare at gøre sådan her [udfører mute gesture] og muter den.*
>
> *Nogle visualiseringer som help. Hvis man havde en gesture for det. Så popper der nogle billeder frem eller animationer frem der viser du skal gøre dette for at mute. Og så måske også en feedback der siger du har gjort det her. Så du ved at du har gjort det rigtigt. Måske en tracker der siger din hånd er rigtig sådan her når du er så tæt på skærmen. Så den siger du gør det godt når du er lige så tæt på.*
>
> *Cool tool. Would use. Would recommend to a friend*

Translated with the meaning of the original transcription in mind:

> *I felt like the program did not register when I made a gesture instantly. So it took multiple tries and I did not necessarily know if I had performed the correct gesture. It could be nice with visual help that could show the hand pose.*

> *I do not know how far from the camera I had to be for the recognizer to work. Sometimes it seems like you have to be really close and other times you perform the gesture too fast. So it could be nice with an indicator that says "Slow down" if you perform it too quickly. It is also difficult when the camera is off since I do not know what I do from its perspective.*
>
> *I like performing gestures. But some of them feel slightly unnatural. It feels like you have to match a very specific movement. Like the goodbye gesture. But it is also nice with a wide range of different commands that can be made with gestures.*
>
> *The program did what it should. I did not need to touch the computer which was nice. I see how it could be useful if you are busy with something and then you quickly have to pause or turn down the volume to hear. So it makes for a quicker alternative than if you would have to take the computer up and suddenly had to turn the volume up or down. I believe it is also difficult when on different operating systems. If someone says that you have to press this button to mute and you do not get which button. Then it would be easier to do this [performs gesture] and it mutes.*
>
> *Some visualizations could help. If you had a gesture for it. Pictures or animations could pop out that show you what do do to mute. And then some feedback that says you have done this so you know you have done the correct gesture. Maybe a tracker that says that your hand is positioned correctly when this close to the screen. So it says you do it correctly when performing it at a specific distance.*
>
> *Cool tool. Would use. Would recommend to a friend.*

### A.27.2 Usability test - Individual 2

Feedback in the original language (Danish):

> *Jeg syntes det er mærkeligt at tænke på at den kan se mig når den er slukket. Det syntes jeg helt klart er lidt kontra med hvad jeg regner med. Når jeg slukker kameraet forventer jeg alt er slukket. Og så er der jo et eller andet med at så sidder man og klør sig i håret og laver et eller andet og sidder sådan her og så gør den et eller andet. Hov der havde jeg lige en hånd oppe eller et eller andet. Det jo ikke alt den skal fange. Så spørgsmålet er hvor sensitiv den skal være eller hvor tydelig arm bevægelserne skal være og sådan noget. Før den skal reagere på det. I hvert fald for sådan nogle features som er mere destruktive, sluk og tænd kameraet, mute unmute er indgribende hvis ikke man er sikker på er det brugeren gerne vil. Og generelt sådan nogle småting med at så gjorde den det ene og så gjorde noget andet når min hensigt var fx at slukke kameraet.*
>
> *Jeg syntes det er sejt men jeg er også lidt usikker på hvor meget jeg ville bruge det fordi der vil vel altid være en som sidder tæt på computeren som kan nå eller så går man lige op og trykker. Jeg tænker use-casen skal man lige overveje for man laver lidt et "fool out of yourself" når man skal lave gestures foran andre mennesker i samme lokale eller foran et kamera for at styre det her møde.*

> *Jeg tænker det her med at den registrere mine gestures korrekt. Så måske en eller anden form for "din gesture skal vare i så lang tid". Så et loading ikon, ligesom Fifa når du skal trykkes confirm til et eller andet så skal du holde den inde i et sekund. Mens du kan se en visuel indikator på at den loader. Så sådan en eller anden form for visuel indikation af at man er i gang med at gøre en gesture og fortsætte med den indtil loading indikatoren er færdig.*

Translated with the meaning of the original transcription in mind:

> *I think it is weird to think that it can see me when the camera is off. It is contrasted with what I believe should happen. When I turn off the camera I expect it all to be turned off. And then there is something about when you itch your hair or something else then it may act upon that. So if I just suddenly put my hand up or something. It is not everything that should be picked up. So the question is how sensitive it should be and how obvious your arm movements should be before the program reacts. At least on more destructive features such as turning off and turning on the camera as well as mute and unmute which are more invasive if you are not sure that is what the user wants. Generally also minor things such as the program did one thing when my intention was something else.*

> *I think it is cool but I am also a bit uncertain on how much I would use it since there will nearly always be one sitting close to the computer who can reach or else I can go up to it and click. I think the use case should be considered since you make a fool out of yourself when you perform gestures in front of other people in the same room or in front of a camera to control the meeting.*

> *I think the thing about recognizing my gestures correctly could be an improvement. So maybe something about how long a gesture should be performed could help. It could be a loading icon, like in FIFA, that when you click confirm then you have to hold down the button for one second, whilst you can see a visual indicator that it is loading. So a visual indication of you are performing a gesture and then continue until the loading icon is done.*

### A.27.3 Usability test - Individual 3

Feedback in the original language (Danish):

> *Følsomheden er ikke mega optimal, det føles som om jeg skal sidde meget præcist før at den reagerer, og det føles ikke super naturligt og så er jeg er bekymret for om skru ned gesturen kan socialt misforstås.*

> *Jeg synes at gestures er godt valgt, og de føles meget naturligt i forhold til hvad jeg forventede de ville gøre, og ideen med det er meget god. Når det virkede var det meget tilfredsstillende*

> *Følsomheden med gestures skal nok tunes, så vil jeg være tilfreds.*

Translated with the meaning of the original transcription in mind:

> *The sensitivity of the recognition is not optimal, it feels like I need to sit in a very precise way before it reacts, and it does not feel very natural. I'm also worried that the decrease volume gesture could be misunderstood.*
>
> *I think the gestures are chosen well, and they feel very natural to what I expected they would do. The idea of the program is very nice and when it worked it was very satisfying to use.*
>
> *The sensitivity should be tuned, then I would be content with using it.*

### A.27.4 Usability test - Individual 4

Feedback in the original language (Danish):

> *Jeg kom til at ryge ud af opkaldet da jeg prøvede at slukke for videoen. Den kunne ikke så godt lide at skrue op til at starte med. Jeg kunne ikke finde ud af at unmute den selv efter jeg havde mutet mig selv.*
>
> *Jeg syntes at det gav mening. Af dem du viste var det ret tydeligt at der skulle gøres med de forskellige.*
>
> *Når det virkede så var det meget lækkert.*
>
> *At den er lidt mere konsistent i hvad den gør.*

Translated with the meaning of the original transcription in mind:

> *I accidentally shut down the meeting call when trying to turn the video off. The program did not like to turn the volume down at the beginning. I did not know how to unmute even after I had mutet myself.*
>
> *I felt like it made sense. Of the gesture set you showed me it was obvious what each gesture should be used for.*
>
> *It was pretty cool when it worked.*
>
> *That it was more consistent in what it recognized.*

### A.27.5 Usability test - Individual 5

Feedback in the original language (Danish):

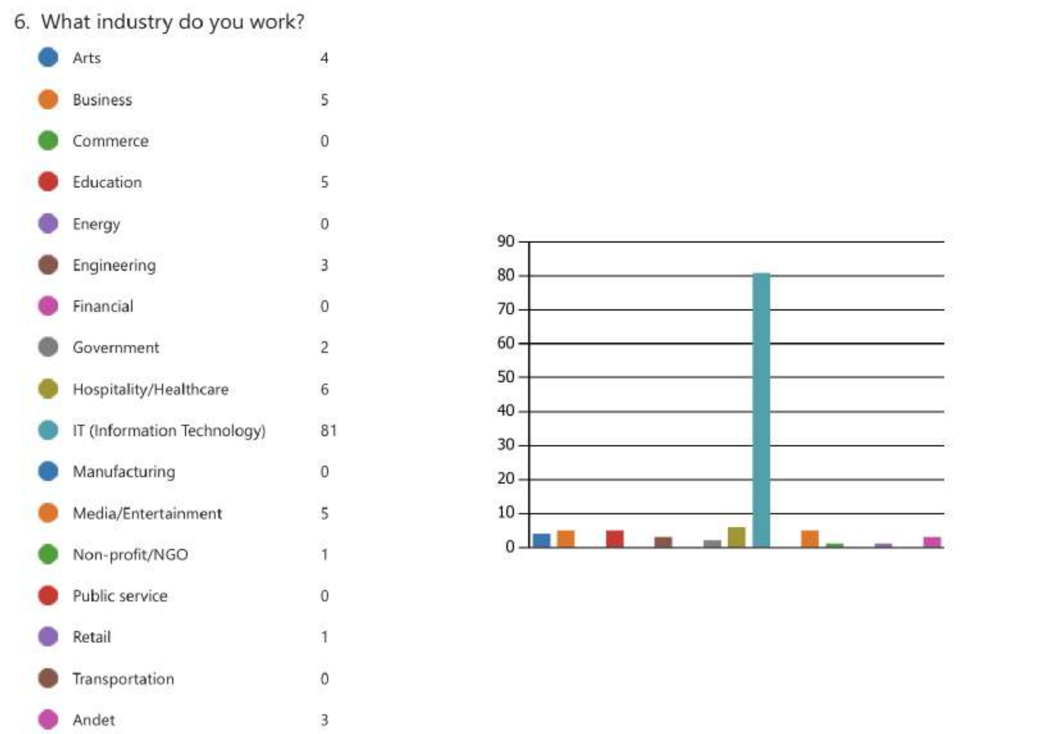
> *Jeg synes den havde svært ved at skelne volumen, om den skulle op eller ned. Og så var den lidt langsom til at registrere at slukke kameraet og lukke mødet, men når jeg lige vidste præcis hvordan jeg skulle gøre så virkede det.*

> *Altså jeg synes det er meget sjovt at man kan det, og jeg synes også den var overraskende god til at genkende. Jeg tror også at hvis man har hybrid møder så kunne det være brugbart.*
>
> *Den er meget sensitiv, når jeg skal slukke kameraet, så skal jeg holde fingrene på en helt bestemt måde, og hvis man skal bruge tid på at holde hånden på en helt bestemt måde så er det jo ligeså hurtigt bare at gå hen og trykke på knappen. Men det skal selvfølgelig heller ikke være sådan at den reagerer på noget som slet ikke er et forsøg på at udføre en gesture.*

Translated with the meaning of the original transcription in mind:

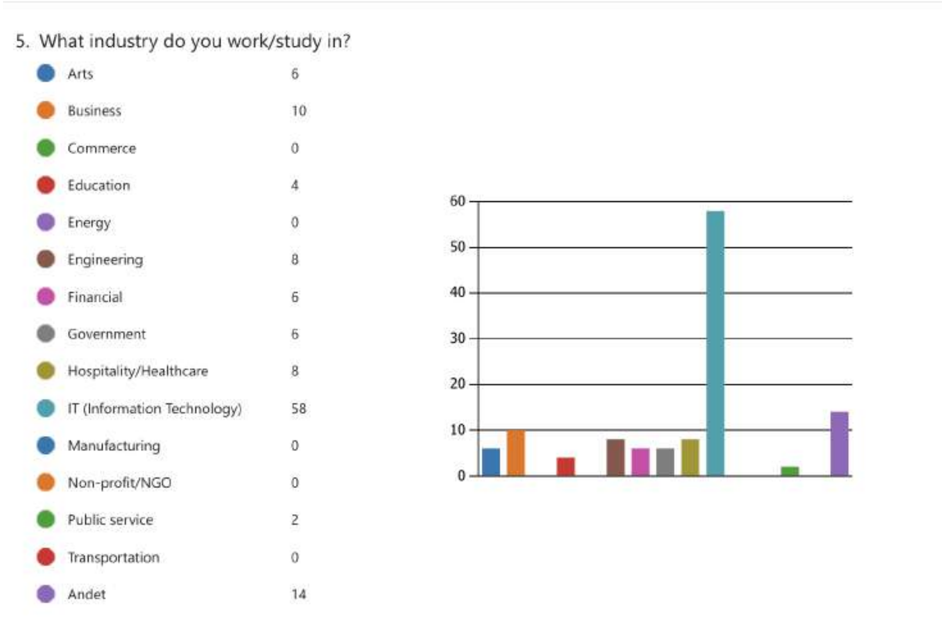
> *I think it had a hard time distinguishing between the increase and decrease volume gestures. And then it was a bit slow at registering that I wanted to turn off the camera and end the meeting. But when I found out exactly how I should do it, it worked out.*
>
> *I think it's very fun that you can do gesture interaction, and I think it was surprisingly good at recognizing. I also think that if you have hybrid meetings it could be useful.*
>
> *It is a bit sensitive. When I need to turn off the camera, I need to hold the fingers in a very specific way for it to work, and then it is just as fast to go and press the button myself. But of course, it should not recognize a gesture when I make a random movement with my hand.*

## A.28 Industry distribution (gesture elicitation)




The figure shows the industry distribution of the participants in the gesture elicitation study

## A.29 Industry distribution (A/B test)




The figure shows the industry distribution of the participants in the A/B test

## A.30 Age distribution (gesture elicitation)

| | General | Females | Males |
|---|---|---|---|
| Std deviation | 3,22112788 | 2,3204019 | 3,66643544 |
| Mean | 24,8184158 | 24 | 25 |

The figure shows the age distribution for all data, females, and males in the gesture elicitation study

## A.31 Age distribution (A/B test)

| | General | Females | Males |
|---|---|---|---|
| Std deviation | 2,454296857 | 1,821614572 | 2,940543612 |
| Mean | 24,86147059 | 24,72170213 | 24,93264151 |

The figure shows the age distribution for all data, females, and males in the A/B test

## A.32 Age distribution (usability test)

| Participant | Gender | Age | | General | Females | Males |
|---|---|---|---|---|---|---|
| 1 | Male | 22 | Std deviation | 0,89442719 | 0 | 1 |
| 2 | Male | 24 | Mean | 22,6 | 23 | 22,5 |
| 3 | Male | 22 | | | | |
| 4 | Male | 22 | | | | |
| 5 | Female | 23 | | | | |

The figure shows the age distribution for all data, females, and males from the usability test

## A.33 Gender/language distribution (gesture elicitation)

| | | | |
|---|---|---|---|
| Males | 64 | Native danish speakers | 72 |
| Females | 36 | Non native danish speakers | 30 |
| Sum | 100 | Sum | 102 |

The figure shows the gender and native language distribution in the gesture elicitation study. The gender does not sum to 102 since two were of other genders

## A.34 Gender/language distribution (A/B test)

| | | | |
|---|---|---|---|
| Male | 55 | Native danish speakers | 85 |
| Female | 47 | Non native danish speaker | 19 |
| Sum | 102 | Sum | 104 |

The figure shows the gender and native language distribution in the A/B test. The gender does not sum to 104 since two were of other genders

## A.35 Video time

```
~/Desktop via 🐍 v3.12.3 took 3s
● ❯ /opt/homebrew/bin/python3 /Users/lucasfreytorreshanson/Desktop/time.py
--------------------------------------
Total length of all videos:
In seconds: 50236.964669 seconds
In minutes: 837.2827444833333 minutes
In hours: 13.954712408055554 hours
--------------------------------------
```

The figure shows the output of a program that calculated the total amount of video time we gathered at the gesture elicitation study

## A.36 Bachelor Notion workspace

The following pdf shows our Notion workspace. Notion was used to organise all our work as well as keep an overview of the schedule. This ensured efficiency due to all information being gathered in a single place

# Bachelor - Hand-Gesture-Based Interaction in Hybrid Meetings

Drive↗ | Discord↗ | Course↗ | Bachelor↗ | Articles↗ | Fabricio↗ | Proposal↗ | GitHub↗

> **Semester 6 - 15ects, 20hours**

Censor: Troels Andreasen

**course calender**

| Aa subject | class | date | eventDone | lecturer | type |
|---|---|---|---|---|---|
| Conclusion and thesis | bachelor | @March 30, 2024 → May 15, 2024 | ☑ | Fabricio Batista Narcizo | study |
| Meeting 10 | bachelor | @May 12, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| Meeting 9 | bachelor | @April 25, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| Presentation 1 | bachelor | @April 24, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| Meeting 8 | bachelor | @April 15, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| MVP | bachelor | @April 8, 2024 → April 19, 2024 | ☑ | Fabricio Batista Narcizo | study |
| Meeting 7 | bachelor | @April 8, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| User tests | bachelor | @March 18, 2024 → April 9, 2024 | ☑ | Fabricio Batista Narcizo | study |
| Meeting 6 | bachelor | @March 18, 2024 1:30 PM | ☑ | Fabricio Batista Narcizo | meeting |
| Prep and experiment | bachelor | @February 20, 2024 → March 18, 2024 | ☑ | Fabricio Batista Narcizo | study |

| subject | class | date | eventDone | lecturer | type |
|---|---|---|---|---|---|
| Meeting 5 | bachelor | @March 4, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| Meeting 4 | bachelor | @February 26, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| Meeting 3 | bachelor | @February 19, 2024 | ☑ | Fabricio Batista Narcizo | meeting |
| Meeting 2 | bachelor | @February 14, 2024 1:30 PM | ☑ | Fabricio Batista Narcizo | meeting |
| Desk research | bachelor | @January 29, 2024 → February 19, 2024 | ☑ | Fabricio Batista Narcizo | study |
| Meeting 1 | bachelor | @February 7, 2024 1:30 PM → 2:00 PM | ☑ | Fabricio Batista Narcizo | meeting |

Link til eksempel på LSTM https://www.fabricionarcizo.com/supervision/luthje2023/Luthje2023.pdf

▼ Description from Fabricio about what the bachelor should include

1. Review the features/interactions available in business collaboration products (hardware), especially with hand gesture control and without it.
2. Review the features/commands available on video communications platforms (software)
   This means organizing the information as a table and classifying the items with all the features/interactions. Please include an image or description of the gesture and all the relevant information for the items that already have hand gestures .
   Do not forget to cite the references/source of information.
3. Prepare and apply a user interview, a method in Human-Computer Interaction (HCI) for understanding users' needs, preferences, and behaviors.
   This means preparing the following:
   A document with the Interview plan, including: a) Purpose and Goals b) Select/Describe Participants d) Prepare the interface to show c) Prepare Questions d) Plan of recordings (video and notations) e) Authorization of participation (according to GDPR)
   After: a) Processing records, b) Analyzing data collected, c) Looking for common patterns to implement at the prototype.
4. Implement a prototype using MediaPipe
5. Run a usability test
   This means writing a usability test plan to check if the results (users' choices/intuition) reflect an excellent usability and learning curve.
   A document with the usability test plan, including a) Purpose and Goals b) Select/Describe Participants d) Prepare the interface to show c) Prepare Scenarios of use and tasks d) Plan of recordings (video and notations) e) Authorization of participation (according to GDPR)
6. Draw Conclusions and write the thesis

▼ The Art of Data Science

The Art of Data Science
The book covers R software development for building data science tools. As the field of data science evolves, it has become clear that software development skills are essential for producing useful data science results and products. You will obtain rigorous training in the R language, including the skills for handling complex data, building R packages and
https://bookdown.org/rdpeng/artofdatascience/

▼ UCP

| Client | Quit call (or end call if host) | Mute | Camera toggle | Share screen | Record | Chat | Raise/l hand |
|---|---|---|---|---|---|---|---|
| Zoom | alt+q (enter) | alt+a | alt+v | alt+s (enter) | alt+r | alt+h | alt+y |
| Google Meet | No binding | No binding | No binding | No binding | No binding | No binding | No bind |
| Teams | ctrl+shift+h | ctrl+shift+m | ctrl+shift+o | ctrl+shift+e | No binding | No binding | ctrl+sh |

Fabricio mentioned that it would be interesting to give an option to which screen to share (e.g. either the left or the right option).

Give an option to react with an emoji.

**Already implemented hand gestures:**

| Client/Reaction | Raise hand | Like |
|---|---|---|
| Zoom | Open palm | Thumbs up |
| Google Meet | Open palm (only for enterprise etc.) | None |
| Teams | None | None |

**todo**

| Name | Assign | Status |
|---|---|---|
| Write related work section | Lucas Frey Torres Hanson Mads | Done |
| Organize interview results in Excel sheet | Mads Lucas Frey Torres Hanson | Done |
| Start general introduction | Lucas Frey Torres Hanson | Done |
| Do interviews | Mads Lucas Frey Torres Hanson | Done |
| Plan Interviews and how to reach participants | Lucas Frey Torres Hanson Mads | Done |
| Finish hypothesis | Mads | Done |
| Make form for interviews | Lucas Frey Torres Hanson | Done |
| Perform Pilot tests | Mads Lucas Frey Torres Hanson | Done |
| Write story for interview | Mads | Done |
| Finish research question(s) | Mads | Done |
| Analyze data from desk research form | Lucas Frey Torres Hanson | Done |
| Look through rest of articles from Elizabete | Lucas Frey Torres Hanson Mads | Done |
| Make gesture-emoji table for UCPs | Mads | Done |
| Review and edit GDPR document | Lucas Frey Torres Hanson | Done |

## A.37 GestureVerse: Exploring the Multiverse of Interactive Gestures

The figure shows an illustration created by Dall-E3 that should visualize the dataset name "Exploring the Multiverse of Interactive Gestures"