Augmenting video conferencing tools to support intercultural communication

Master thesis in Informatics

Arthur Toenz
Saint-Louis, France
Matriculation number 09-056-128

Institute for Computer Science
University of Zurich, Switzerland
Prof. Dr. Elaine M. Huang

Tutor: Helen Ai He

Thesis handed in: 17.12.2015

Abstract

Intercultural communication is challenging as every culture has its own set of implicit rules and norms related to communication. Intercultural training aims at addressing this problem by supporting people in learning how to better communicate with people from other cultures. One approach for intercultural training is the training of culture-general skills. This thesis is an initial exploration of how to augment videoconferencing tools to train culture-general skills for dyadic conversation. I used the Intel RealSense 3D camera to track the users' nonverbal cues and compute nonverbal behaviour from these, in order to compare both interlocutors' behaviours and present them live feedback. I designed interactive visualizations for a sample of three nonverbal cues (proxemics, smiling and expressiveness) and conducted a qualitative study to evaluate their comprehensibility and influence on real-time communication. From the results of the pilot study, I found that there is a trade-off between inciting users to adapt to each other and making them more aware of their own behaviour.

Acknowledgments

Julie for her patience and support
Helen for the tutoring, chats, ideas and support
Jagoda for the tips on Information Visualization
Patrick for reviewing and spellchecking

Contents

Index of figures

Index of tables

Chapter 1 – Introduction

Modern society and technical progress result in the need to cooperate worldwide. As a consequence, people from different cultural backgrounds come more and more into contact in the modern workplace. Different cultures mean different ways of communicating, and effective intercultural communication is becoming an important skill. However, cross-cultural communication is not innate: it must be learned and practised [1]. Intercultural training aims at helping learning effective cross-cultural communication. Intercultural training is defined as the development of intercultural competences, which is the ability to act and relate appropriately and effectively in various cultural contexts [2]. A traditional approach to intercultural training is culture-specific training, in which communication competences and appropriate behaviours are learned for each new culture one encounters. In culture-specific training, virtual agents are used to simulate intercultural communication and train such skills. But this approach does not scale anymore with the mass of different cultures one has to deal with every day. A way to bridge this gap is to train culture-general skills, which will be useful when communicating with any other culture. Culture-general skills are competences that are transferable across diverse cultural contexts [3]. He et al (2015) identified two culture-general skills that could be trained: awareness, the ability to perceive and interpret culture-specific feedback, and encoding adaptation, the ability to adapt to other people's behaviours by mimicking them.

Another major issue of culture-specific training is that it requires time and commitment on the part of the learners. In order to overcome this problem, learning would need to be intuitive and happen in-situ, i.e. when the learner is actually having conversation with people from different cultural background. Computer-Mediated Communication (CMC) media like video conferencing are a good example of where in-situ learning could be appropriate [4].

In this thesis, I do an initial exploration of how to train Culture-General Skills in Computer-Mediated communication, and especially in the context of video conferencing tools. In particular, I focus on the communication of two persons over a video chat. Specifically, I augment a video chat with live feedback regarding the nonverbal communication behaviour of the interlocutors. The goal is to help the users perceive nonverbal feedback by making them aware of differences in nonverbal behaviour between them. By providing them with feedback on what they do versus what their communication partner does, the users might then utilize this information to choose to adapt or not. The design of intuitive, non-prescriptive technologies is one of the main goals pursued when producing visual feedback on nonverbal behaviours [4]. In this thesis I address the overarching research question:

How can video conferencing tools be augmented to train culture-general skills?

In particular I explored the extension of video conferencing tools through visual feedback on the interlocutors' nonverbal communication behaviour. I went on to address the following research questions:

RQ1: How can we design intuitive visualizations of nonverbal behaviour during real-time video chat?

When videoconferencing, the primary task of the user is to be active in the conversation. Thus the goal of visualizations on nonverbal behaviours is to be intuitive and non-distracting,

to allow the user to pay attention to his primary task. The visualization method that used can have an influence on the effectiveness of communication [5]; therefore it is important to design visualizations carefully.
To address research question RQ1, in *Chapter 5 – Visualizing nonverbal behaviour*, I developed different visualization designs through an iterative process. Then I implemented the most promising ones and evaluated them in *Chapter 6 – Pilot study*. Finally I reflect on the results of the pilot study in *Chapter 7 – Findings / Analysis*.

RQ2: How can such visualizations of nonverbal behaviour help to train the culture-general skills of awareness and adaptation?
In particular, I investigate how the visualizations I designed can help users mimicking and become more self-aware about their nonverbal communication behaviours.
To address research question RQ2, I conducted a pilot study as presented in *Chapter 6 – Pilot study*. In *Chapter 7 – Findings / Analysis* I analyse the results of the pilot study to understand how the visualizations influenced the users' awareness and adaptation, using self-reported interview results.

Thesis overview

The remainder of this thesis describes the research outlined above.

Chapter 2 – Intercultural communication reviews related work in the fields of intercultural and nonverbal communication.

In ***Chapter 3 – Technical setup*** I introduce the nonverbal cues I chose to explore and propose a review of the state of the art technology applied the scenario of the detection and comparison of nonverbal behaviours in videoconferencing tools.

Chapter 4 – Implementation describes the system I developed for tracking and comparing nonverbal behaviours in videoconferencing tools.

Chapter 5 – Visualizing nonverbal behaviour details the design process through which I went to produce interactive visualizations.

In ***Chapter 6 – Pilot study*** I present the evaluation I conducted with my tutor to assess the comprehensiveness of the tool I designed and how it influences the users' communication.

Chapter 7 – Findings / Analysis proposes an analysis of the results from the pilot study presented in *Chapter 6 – Pilot study* with regard to my research questions.

In ***Chapter 8 – Future Work*** I reflect on the challenges I encountered throughout my thesis and propose leads for further research.

In ***Chapter 9 – Conclusion*** I summarize my work and how it informs future research on visualizing nonverbal communication feedback in videoconferencing tools.

Chapter 2 – Intercultural communication

Culture and communication

Culture can be defined as "an accumulated pattern of values, beliefs and behaviours shared by an identifiable group of people with a common history and a verbal and nonverbal symbol system" [6]. There are different levels on which culture can be regarded, such as national, regional or organizational [7].

Different cultures need to be able to communicate together; this is what *intercultural communication* is for [6]. Intercultural communication happens when two or more individuals from different cultures communicate together, and goes essentially as follows: person P1 encodes a message in one cultural context, and then person P2 decodes it in another cultural context [6], [4]. As the encoding and decoding are not the same across cultures, mismatches can occur, and frequently result in breakdowns in communication [6].

Nonverbal communication in particular differs across cultures [8]. Nonverbal communication has been defined as communication without words. It includes apparent behaviours with facial expressions, eyes, touching and tone of voice, as well as dress, posture and the use of space and time [9]. The cultural differences in nonverbal communication are often the cause for miscommunication [10]: as culture is unconscious to us, it is easy to apply and project our own cultural encoding-decoding schemes to other cultures. Encoding is the way one transmits a message to a conversation partner in the context of one's own culture, through verbal and nonverbal channels, whereas decoding is the way one understands and interprets a message within the context of one's own culture. As different cultures have different ways to encode and decode messages, it is interesting to look at ways to help better understanding one's interlocutor's nonverbal communication patterns.

Barriers to cross-cultural communication

He at al [4] identified two main barriers to intercultural communication. On one hand there is the **Awareness** problem: it is difficult to perceive and interpret nonverbal feedback when encoded in a foreign cultural background. On the other hand there is **Adaptation**: once you have perceived there is a mismatch in the encoding-decoding scheme of the conversation, you have to know how and when to adapt to the communication partner.

Awareness of culturally-relevant feedback and Encoding adaption are culture-general skills as they are helpful when communicating with any other culture. Culture-general skills are competences that are generalizable and usable when communicating with any other culture. This is in contrast to culture-specific skills, which describe the knowledge of appropriate behaviours and communicational patterns specific to a particular culture. *Table 1 – Design goals for training culture-general skills (He et al 2015)* presents design goals for training the abovementioned culture-general skills as elaborated in [4].

CULTURE-GENERAL SKILLS	DESIGN GOALS
(C1) Developing Awareness: Perceiving and interpreting culturally-relevant feedback	**(D1) Train perception:** - Highlight culturally-relevant feedback during in-situ intercultural communication - Help people interpret such feedback as positive or negative
(C2) Adapting Encoding: Mimicking culturally-relevant behaviors	**(D2) Train encoding ability**: - Guide people in what culturally-preferred behaviors to mimic, how to mimic and when to mimic - Visualize observable differences between the two interlocutors

Table 1 – Design goals for training culture-general skills (He et al 2015)

In this thesis, I implemented parts of design goals D1 and D2 from *Table 1*. To train perception (design goal D1), I wanted to create visualizations that help the users perceive the nonverbal part of the communication and become more aware of their own behaviour.
To train encoding ability (design goal D2), I wanted my visualizations to facilitate the act of comparing the user's behaviour to his/her communication partner. When presented with elements of comparison, the user can choose to adapt or not.

Chapter 3 – Technical setup

In this section, I first present method concept I explored. Then I depict the nonverbal cues I chose to explore. Finally, I propose a review of the state of the art technology applied to the scenario of visualizing nonverbal behaviours in the context of videoconferencing tools.

Method

My goal was to visualize differences in nonverbal behaviour in order to help the user noticing culture-general feedback and gain self-awareness about his own behaviour. The goal is that when the user notices a difference in nonverbal behaviour through the use of the tool, he/she will pay attention to the culture-general feedback from his interlocutor. He/she then may choose to adapt or not.

One of the questions He et al [4] presented was "Should technologies train awareness explicitly or implicitly?". In the visualization designs and by conducting a pilot study, I explored implications of both implicit versus explicit difference display.
He et al [4] also inquired "Should technologies offer one-sided or simultaneous training?". In this thesis, I focused on simultaneous training.

Concretely, I wanted to visualize the difference in nonverbal communication between the two interlocutors of a video conferencing tool. The tool should help users to become more aware of their behaviour and to perceive their interlocutor's nonverbal feedback, but without being prescribing about what would be an appropriate way to behave. The users should be able to choose if they want to use this information to interpret the nonverbal feedback and adapt their behaviour or not.

Selecting nonverbal cues to track

As the human face and body is hugely expressive, the design space for visualization of differences in nonverbal communication is very large. An important choice to address when designing visualizations for supporting communication is thus the selection of the communication cues one wants to track. As in the scenario of videoconferencing tools most of the communicated information is visual, I focused on visual nonverbal cues.

Cultural anthropology literature defines a plentiful of visual nonverbal behaviours which were proven to differ across cultures, and thus can be observed in order to interpret culturally-relevant feedback: the use of smiling, facial expressions, hand gestures, the use of space (proxemics), movement (kinesics), etc.

Exploring the complete design space is out of scope for this thesis; this is why I chose to focus on detecting and tracking three of them: **smiling**, **proxemics**, and an aggregate measure which I termed "**expressiveness"**.

Expressiveness I defined as being a measure of how expressive a user talks and behaves, and expressiveness is an aggregation of many nonverbal cues: it includes eyebrow movements, smiling, the use of hand gestures, emotional display, proxemics, and head movements.

Selecting a face-tracking technology

Defining requirements

In order to select the technology that would support the face-tracking and facial expression detection, different decision factors were identified:

- Price was an important dimension as most of the available products were aimed at market research and thus were costly.
- Accuracy of detection: I wanted the solution not to be too much subject to error propagation. Filtering of low-intensity nonverbal cues, extraction of nonverbal behaviour, comparison of behaviours; each of these steps would be affected by an error-prone detection, and the overall quality of the tool would suffer from it.
- What the tool should detect:
    - Minimum requirements:
        - Facial states and expressions, e.g. mouth open/closed, eyebrows raised/lowered, smiling
        - Positions of important facial landmarks as the eyes, mouth, nose, chin, cheeks, eyebrows.
        - Micro-expressions: the ability to detect very short expressions
        - 3D head pose: position and rotation of the face
    - Nice to have[1]:
        - The 7 basic emotions: happiness, sadness, surprise, fear, anger, disgust, and contempt [11]
        - Action Units as per the Facial Action Coding System (FACS) [12], [13]
        - Gaze direction and eye closure
        - Hand tracking

[1] At the beginning of my thesis, I was interested in training cultural display rules, but had to change and go with something simpler as the technology was not ready for this. I was interested in detecting FACS action unit and emotions as I wanted scientifically backed detection. But the goal of my thesis changed to be focusing on culture-general skills, these criteria was not required anymore.

- Robustness to face rotation is also important for nonverbal communication behaviour tracking as interlocutors might move during the conversation, and the use of space and position (proxemics) is part of the nonverbal communication.
- The tool needed to be able to do live analysis and keep up with a normal camera framerate. Many of the available face-tracking products in the market were eliminated because they only accepted low framerates.

Exploring options

Multiple options were found to meet the minimal important requirements, Microsoft Kinect[2], Intel RealSense[3] and VisageSDK FaceTrack[4] being the most promising ones for our scenario. All of the three are much more affordable than commercial solutions traditionally aimed at market research (e.g. Emotient, NViso, Noldus, Affdex): RealSense and Kinect because they only require to buy a compatible device (Intel RealSense F200 camera or any laptop featuring it for the RealSense, or the Kinect camera for Kinect) while their SDK's are free of charge, and VisageSDK as its license is available at a reasonable price.

Table 1 presents a comparison of the three tools along the abovementioned requirements:

Detection	Intel RealSense	Microsoft Kinect	VisageSDK FaceTrack
Facial expressions & states	Yes	~Yes	Yes
Facial landmarks	Yes	Yes	Yes
Micro-expressions	Yes	Yes	No
3D head pose	Yes	Yes	Yes
6/7 basic emotions	Yes	Yes	No
FACS Action Units	No	No	No
Gaze direction	Yes	Yes	Yes
Robustness to face rotation	Very good	Very good	Good
Live video analysis	30fps	30fps	<30fps

Table 2 – Comparison of three face-tracking tools with respect to their detection capabilities.

Intel RealSense SDK and Microsoft Kinect are very much similar when compared along the requirements of table 1.

I decided on the Intel RealSense SDK because on one side Kinect is more aimed at full-body tracking whereas RealSense is aimed at face- and hand-tracking, so more fitted to our scenario of nonverbal communication; and on the other side as VisageSDK is quite experimental and there is no big community of developers using it yet, when RealSense has a big community behind it where you can ask questions. Another important point is that RealSense technology, like the Kinect, is based on using a 3D-camera to get depth data, and thus has a great advantage over other software that would need to compute depth information from the images, like VisageSDK.

[2] https://dev.windows.com/en-us/kinect
[3] http://www.intel.eu/content/www/eu/en/architecture-and-technology/realsense-overview.html
[4] http://www.visagetechnologies.com/products/visagesdk/facetrack/

Table 2 presents a comparison of the three tools along additional decision criteria:

Decision criteria	Intel RealSense	Microsoft Kinect	VisageSDK
Target use	Face & hands	Full body	Face
Developer community	Yes	Yes	No
Bases on a 3D Camera	Yes	Yes	No

Table 3 – Comparison of three face-tracking tools with respect to additional decision criteria.

The Intel RealSense SDK and its brand camera allow the tracking of many visual elements of nonverbal communication:

- Head pose: position and rotation relative to the camera
- Mouth gestures, e.g. smiling, kissing, mouth openness
- Hand gestures, e.g. waving, A-OK
- Eyebrow movements: raising or lowering
- Gaze direction
- Emotional display[5]

I chose to focus on tracking proxemics, smiling and eyebrow movement. For proxemics, I selected the head as a reference point to track how the users move over the course of the conversation. The RealSense camera is equipped with a depth camera, so it is easy to accurately detect distance between the user and the camera. This is the dimension I used to evaluate proxemics. Concerning smiling and eyebrow movements, the RealSense SDK gives for each frame capture by its camera a numerical evaluation of the intensity of these facial movements. In the case of smiling, the intensity reflects how big a smile is; in the case of eyebrow movements it indicates how high the brows are raised or how low they are lowered.

Chapter 4 – Implementation

In this section I present the system I implemented for tracking nonverbal cues, computing and comparing nonverbal behaviours, and present the user with visual feedback. First, I consider the implementation of a videoconferencing tool. Then I go into more detail about the implementation of the detection and tracking system.

Implementation of a videoconferencing tool

As described in the previous section, I decided to use Intel RealSense. RealSense was much more polished in its C# implementation than in its Java implementation at the time of writing. Thus I had to select a C#-solution for video-conferencing. I chose and used the trial version of iConf.NET SDK by AVSPEED[6] because it is free, well documented and allows to easily build extendable videoconferencing tools.

Combining video conference and face-tracking

Once my video conferencing tool base was ready, I wanted to combine it with the RealSense detection. At this stage I had to make some important design decisions, like should I use the non-deterministic frame-rate of the RealSense camera as a time unit, and how I should organize the program structure so that each part is modular and can be extended and/or replaced.

[5] Only up to RSSDK R4

[6] http://www.avspeed.com/

In this section I first describe the concept of the system I implemented, and then I depict its architecture. After this, I discuss the communication structure I selected for exchanging data between peers. Then I express the challenges of detecting and comparing nonverbal behaviours. Finally I present the design process which I went through to produce the visualizations.

Concept

Basing on the three dimensions I discussed in the last paragraph of *Selecting a face-tracking* technology, I wanted to detect nonverbal behaviours for both users, compare the detected behaviours, and give live feedback through the use of visualizations. The overall concept of the system is the following: For each frame captured by the camera, the RealSense SDK produces values for the three dimensions I track (proximity, smiling, eyebrow movement). I then filter out the low values to remove noise and marginal data. Then, in order to detect nonverbal behaviours, I track the nonverbal cues associated with each nonverbal behaviour (see *Definitions*) over the course of the conversation. After this, I compare the behaviours of both users along three comparison dimensions: occurrence, length and intensity (see *Communication architecture*). When this is done I transmit the comparisons to the visualizations for display.

Definitions

In the rest of this document, I employ the term *nonverbal cue* as an observable signal of nonverbal communication. Examples are head pose and movement, smiling, touching and facial expressions. In this thesis I focus on visual nonverbal cues. Other categories of nonverbal cues like auditory or tactile nonverbal cues are out of the scope of this thesis, as for the former it would require a different type of technology than computer graphics tools, and the latter are non-existent in the scenario of videoconferencing tools.

Throughout the rest of this thesis, I use the term *nonverbal profile* as the set of nonverbal cues detected in a moment – in the case of videoconferencing, a frame – of conversation, associated with a point of time in the conversation. In each nonverbal profile, there will be at most one occurrence of each possible nonverbal cue. For instance: one's head can only have one pose at a time. A sequence of nonverbal profiles thus corresponds to the history of nonverbal cues for the corresponding sequence of frames. Such a sequence of nonverbal profiles I call a *nonverbal timeline*. The sequence of nonverbal profiles from the whole conversation I call the *global nonverbal timeline*.

I use the term *nonverbal behaviour* to designate a certain sequence of nonverbal profiles, containing certain nonverbal cues, over time. For example, a head nod is composed of a certain sequence of relative head positions and movements over a short period of time. A nonverbal behaviour is composed of a type (e.g. head nodding, smiling, eyebrow rising), a start and an end time (and thus a length), as well as an intensity (e.g. how big a smile is, how large a head nod is). A nonverbal behaviour does not objectively have a unique and constant intensity for all of its duration but one for each time point in its duration [13]. For the sake of simplicity, I chose to use the average of these intensities as the behaviour's overall intensity.
A collection of nonverbal behaviours having happened – i.e. having started and finished – between two points of time in the conversation I call a *behavioural timeline*. The collection

of all nonverbal behaviours having happened over the whole conversation I call the *global behavioural timeline*.

From nonverbal cues to visual feedback

Here I present the architecture of the tool and steps through which it goes, from detecting nonverbal cues to presenting the users with interactive visualizations. *Figure 1* – Class diagram depicts the organization of the different modules and what data they exchange. *Figure 2* – Sequence diagram describes the sequence of operations performed for each new frame. The following listing details each of the important steps.

1. **Detection**
   The detection of the raw data – the dimensions we track, in our case the proximity to the camera, the presence and intensity of a smile, and the raising or lowering of eyebrows – is done by the RealSense camera and the RealSense SDK.

2. **Filtering & Processing**
   The Central Monitor then charges the relevant specialized Monitors to filter and process the data produced by the RealSense camera, and to transform it into nonverbal cues which are then combined to produce the nonverbal profile for the current frame. All nonverbal profiles produced over time are registered in the global nonverbal timeline[7].

3. **Communication**
   Each peer transmits its current nonverbal profile to the other side of the conversation.
   Both nonverbal profiles, the locally produced one and the one received from the other user, are registered in the respective timelines.
   I detail the communication structure in section *Communication architecture*.

4. **Monitoring**
   Each side monitors both timelines and detects nonverbal behaviours in them, and adds them to the corresponding behavioural timeline. To detect nonverbal behaviour, I look for the presence and the intensities of the nonverbal cues constituting the behaviour in the monitored timeline.

5. **Comparison**
   Once all monitors are done monitoring the timeline, the Comparator computes the behavioural comparisons and transmits them to the Display Centre. I detail the comparison of nonverbal behaviours in section *Comparing nonverbal behaviours*.

6. **Display**
   The Display Centre then receives nonverbal comparisons and can present this information to the concerned visualizations. The visualizations are in turn responsible for the actual display of the comparison data. The visualizations also get direct information from the nonverbal profiles. I detail the different visualizations I produced in *Chapter 5 – Visualizing nonverbal behaviour*.

[7] In order to keep the space usage of the application trackable, old nonverbal profiles get dumped in a file after a while.

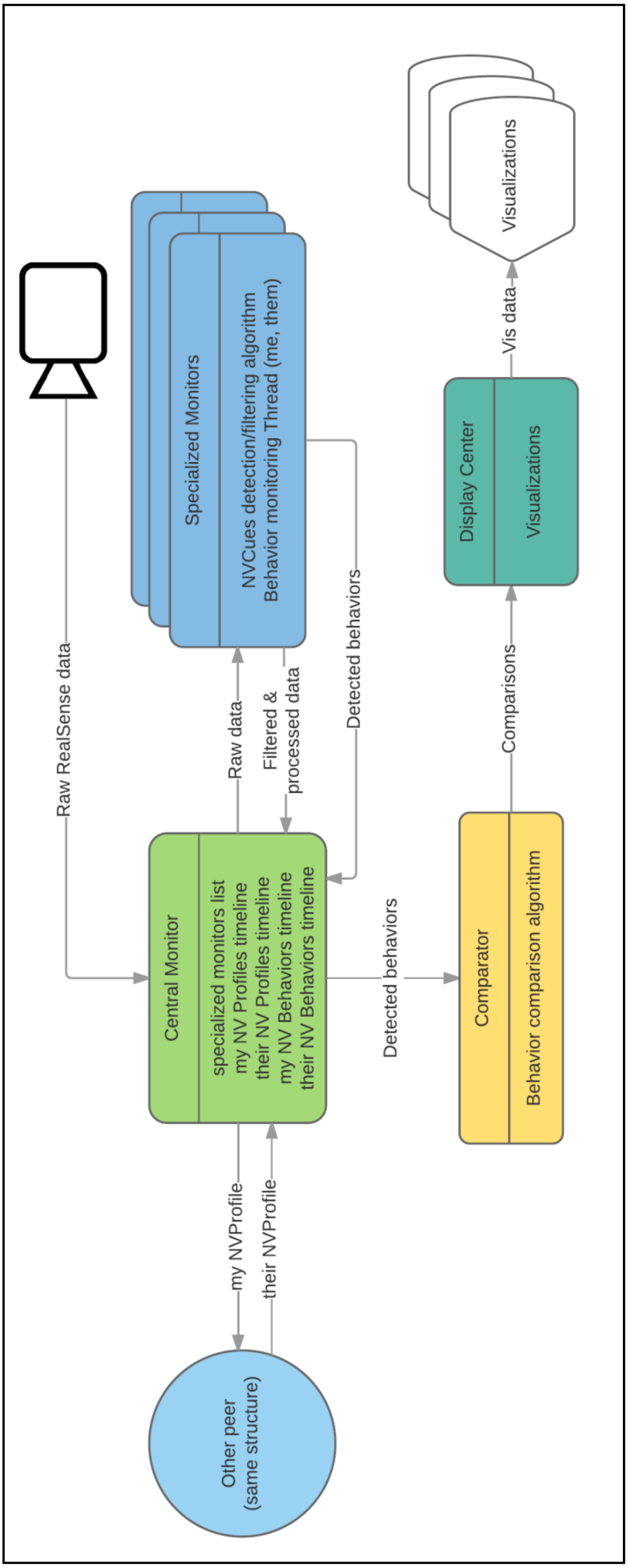


Figure 1 – Class diagram

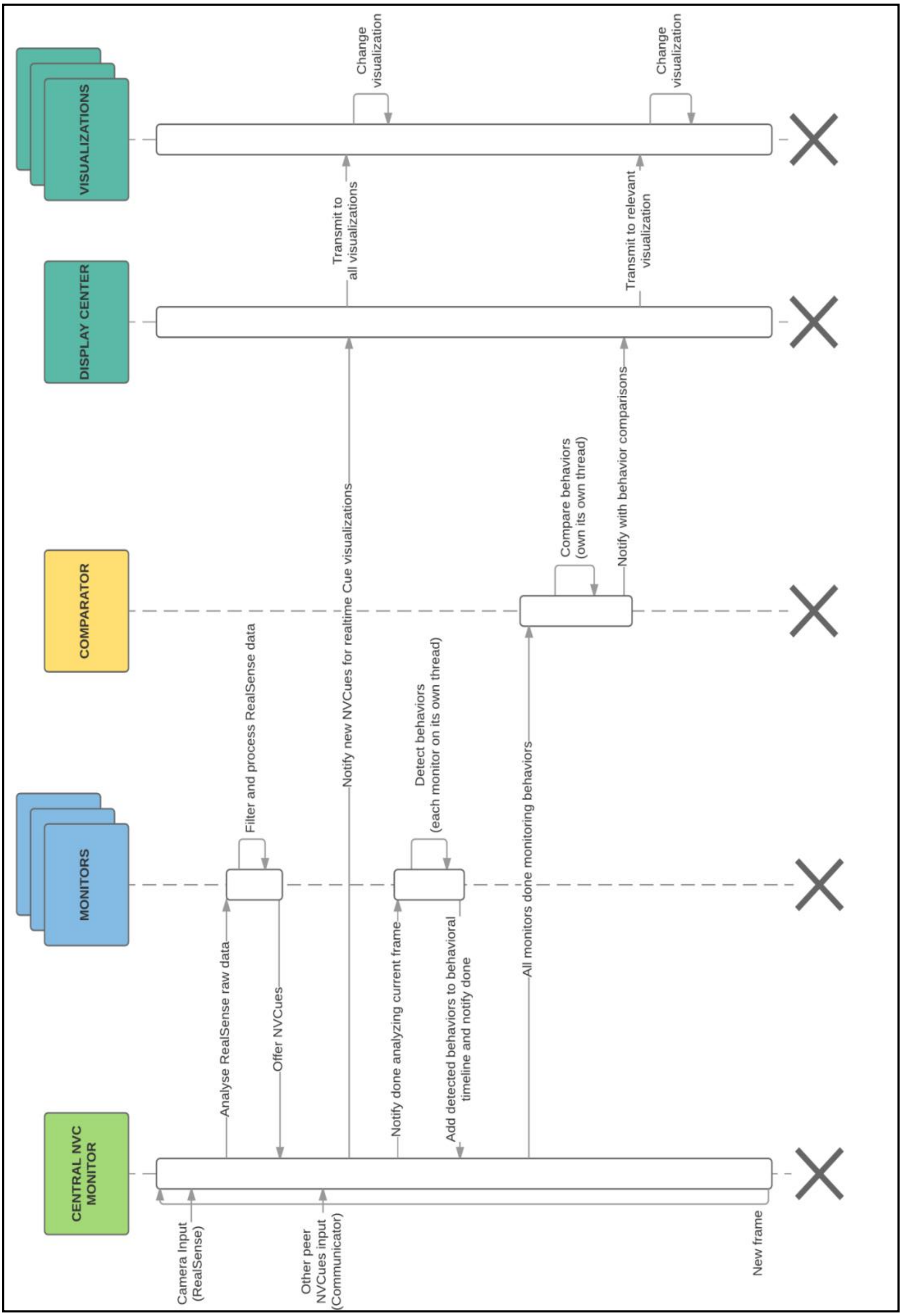


Figure 2 – Sequence diagram

Communication architecture

The question of how should the communication between both clients be done was central to the architecture design of the tool. Two alternatives were considered – Server-Client or Peer-to-Peer – both coming with a set of advantages and drawbacks. In this section I shortly present these and the rationale behind my decision of using a Peer-to-Peer structure.

Option 1: Server-Client

In our scenario, the Server-Client communication architecture would come with the following advantages:

- Centralization of computation: Only the server needs to do the heavy work.
- Synchronization: It is easy to make sure both clients are using the same time scale

But in the scenario of this study, the Server-Client architecture presents different drawbacks:

- Polling: Both peers would need to first send their nonverbal profiles, wait for the server to have both sides' data, do the computation, and then transmit results.
- Server as a third machine: The server would need to run on a separate machine, which would make the system more complex.

Option 2: Peer-to-Peer

The Peer-to-Peer communication architecture comes with the following advantages:

- No polling: Each peer sends his data to the other peer and can forget about it. As computation is done in-place, no need to wait for an answer from a central server.
- Less communication needed: As a result from the fact that no polling is needed, the amount of required communication is divided by two in a Peer-to-Peer system.
- Independency: As computation is done locally, each peer is independent in all considerations related to computing results, thus a more powerful peer would not get slowed down in its computations by a less powerful peer on the other side of the conversation.

The major drawbacks of the Peer-to-Peer system are:

- Synchronization: It is more difficult to make sure that the data received from the other side is referring to the exact same time in the conversation than the data produced locally.
- Computation redundancy: The computation needs to be done on both sides of the conversation.

Taking into consideration all of these factors, I chose to use a Peer-to-Peer architecture. First, independency is a very desirable property as the system relies on a different camera on each side, which work at different frame rate. This also includes the fact that the communication does not make an assumption on what kind of camera or data is transmitted and used, which would be more difficult to achieve in a Server-Client scenario. Second, computation redundancy is only problematic when the computation is heavyweight; which is not the case in this use case: the computation is merely composed of comparisons of intensities, lengths and occurrences of the nonverbal behaviours, in low quantity (in the order of 10-100 per frame). Third, synchronization is important but it can be assumed that the difference in frame rate between the cameras of both ends is marginal. Nevertheless it is important to make sure that both parties have a common "time zero" reference, so that it is possible to compare actions that happened on the same interval of time; for this purpose I implemented a simple three-way handshake connection establishment protocol where both peers agree on a time in near future (e.g. two seconds after connection establishment) to start the monitoring process.

Comparing nonverbal behaviours

Once nonverbal cues are detected and nonverbal behaviours extracted from them, the Comparator compares both users' behaviours and reports its findings to the Display Centre for visualization. The comparison is done over a certain time frame, and along three dimensions: **occurrence**, average **intensity**, and average **duration**. For instance, one can compare the intensities of the smiling behaviours of both users over the last twenty frames.

Categorizing difference: A key challenge to comparing nonverbal behaviours is to determine how to quantify difference along those three dimensions. The reason why this is important is that using continuous values for comparison is not intuitive.
For instance, if we want to compare the average duration of smiling behaviours between user A and user B over the last minute: if user A's smiles lasted on average 6 seconds whereas user B's smiles lasted 2 seconds on average, the difference would be of 4 seconds.
What can we understand about this comparison? Not much, apart from the fact that user A's smiles lasted on average 4 seconds later than user B's smiles. Raw data like numbers are is not meaningful for the user, thus there is a need for offering a level of interpretation when presenting the user with data on differences in nonverbal behaviours. This is why I decided to use four categories nonverbal behaviour: **equally**, **a bit more**, **more**, and **much more**. These categories allow building intuitive comparisons, for instance:

"On average, user A smiles much more than user B"

This is beneficial to visualization design as it allows communicating a clear message to the user. Specifically, one can design visualizations to display data in different modes according to how different users behave. For example, one could design a visualization to draw the users' attention more when one of them tends to smile much more or much more intensive than the other.
The main issue to this approach is that it requires defining "buckets" of values in order to allow such categorization. This can be articulated with the following:

How much is a bit more/more/much more?

In the scenario of the abovementioned example, this would mean: how much longer should A's smiles last longer than B's smiles to allow us to consider that A smiles a bit longer/longer/much longer than B? This partitioning needs to be done for all buckets, in the three dimensions, and for each nonverbal behaviour one wants to visualize. Another drawback of this approach is that the bucketing might differ across cultures, as for instance if people from culture C1 tend to stand very close to the camera when conversing whereas people from culture C2 tend to move a lot over the course of conversation; the definition of an objective and universal bucketing for comparison might be difficult.

The determination of these bucketing values would require a study in itself, which is out of the scope of this thesis. This why, in my visualizations, I used arbitrary values obtained by trial-and-error.

Chapter 5 – Visualizing nonverbal behaviour

In the previous sections I presented the technical setup and the implementation of the system I developed for detecting nonverbal cues, computing and comparing nonverbal behaviours, and presenting the users with visual feedback. In this section, I present the evolution of the visualization designs for the three nonverbal behaviours, from the state of rough ideas to concrete visualizations.

Design process

The design of the visualizations was an iterative process and went through the following steps.

Step 1: Brainstorming of design ideas

I brainstormed many visualizations ideas, based on the fact that the focus was on communicating the difference in behaviour. I tried to explore different paths for visualizing behaviour differences, among which textual, symbolic and graphical visualizations.

Step 2: Selection of the best designs and improvements

Then I discussed the design ideas with my tutor and we selected the most promising ones, based on the evaluation of possible advantages and drawbacks of each of them. Following this, I had another brainstorming phase where I looked for refinement possibilities for the selected designs to make them more concrete and intuitive. At that time I was already implementing mid-fidelity prototypes of the visualizations to be able to assess the feasibility of certain refinement ideas.

Step 3: Discussion with an Infovis specialist

After this, my tutor and I discussed the improved designs with an Information Visualization specialist[8]. New design ideas emerged from the discussion, and the already existing designs were refined further.

Step 4: Implementation

At the end, I implemented high-fidelity prototypes of the final visualization designs within my videoconferencing system.

In the rest of this section, I detail the iterative design process of visualizations for Proxemics, Smiling and Expressiveness behaviours, according to the process depicted above. I do not present all the visualization designs I brainstormed as there were many of them. See *Appendix A – Visualization designs sketches* for sketches of further visualization design ideas. I also depict the two versions of the user interface design in *Appendix B – User interface design*.

[8] J. Walny, doctoral candidate in computer science, University of Calgary (http://research.jagoda.ca/)

Visualization 1: Proxemics

Step 1: Brainstorming of design ideas

Proxemics is defined as “the interrelated observations and theories of man's use of space as a specialized elaboration of culture” [14]. In the scenario of videoconferencing tools, as the two communication parties are not actually sitting in the same room, the use of space would translate to the distance between the users and their respective cameras. Through designing visualizations for proxemics, I wanted to relate to the users’ behaviour with respect to their positioning in space with respect to the camera in front of them.
As it became clear after sketching some of the visualization design ideas, different information can be displayed regarding distance to the camera: current distance, history of past distances, average distance and deviation, etc.
Here I present different options I regarded for designing a visualization for proxemics, and reflect on their possible advantages and drawbacks.

Option 1

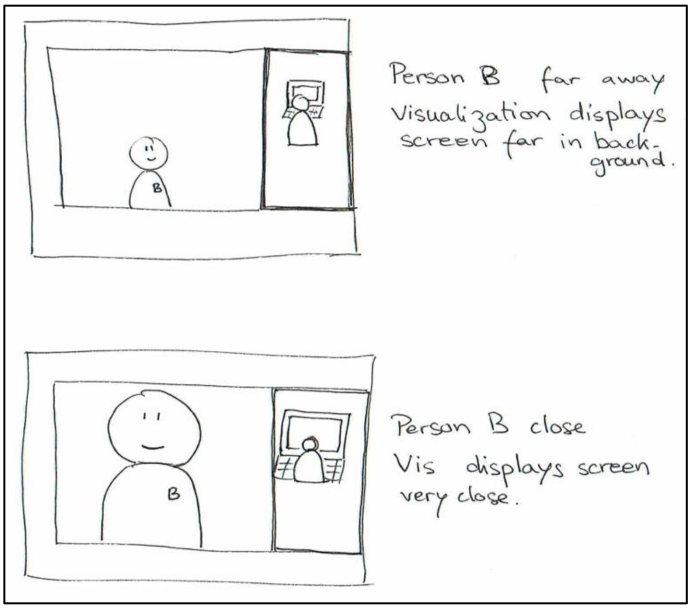


Figure 3 – Proxemics design 1

In this visualization design, the user gets a symbolic perspective view from the back of the other person, like if he was watching the scene from behind him.
To show the distance, the represented screen in made bigger or smaller according to the position of the person with respect to the camera.

Advantages: The visualization might make it clear that it depicts the distance to the camera.
Flaws: The visualization only shows one side of the conversation. It is not clear if seeing only how the other person is behaving can help the user getting more self-awareness.

Option 2

This visualization design is presenting the conversation as if it would happen in a face-to-face scenario, with the camera in the middle and seen from a side view. The small circle are markers to show how much one is close or far to the camera. The pawns would move from one marker to another based on the difference in distance to the users’ respective cameras.

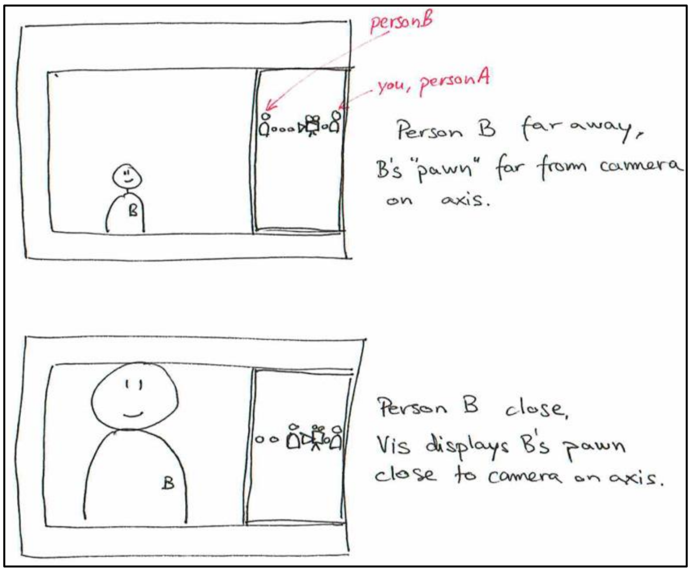


Figure 4 – Proxemics design 2

Advantages: The visualization follows the natural mapping of the real situation as it simulates a face-to-face conversation scenario.
Plus, the fact that the difference information is not communicated explicitly might have the effect that the user will compute it himself. And as the human brain excels at comparing things, this task does not represent much of an effort.
Flaws: With respect to the screen and in a right-hand coordinate system, the distance to the camera will be on the z-axis. This visualization makes it appear on the x-axis. Adaptation to this rotation might take some extra thought to the user.

Option 3

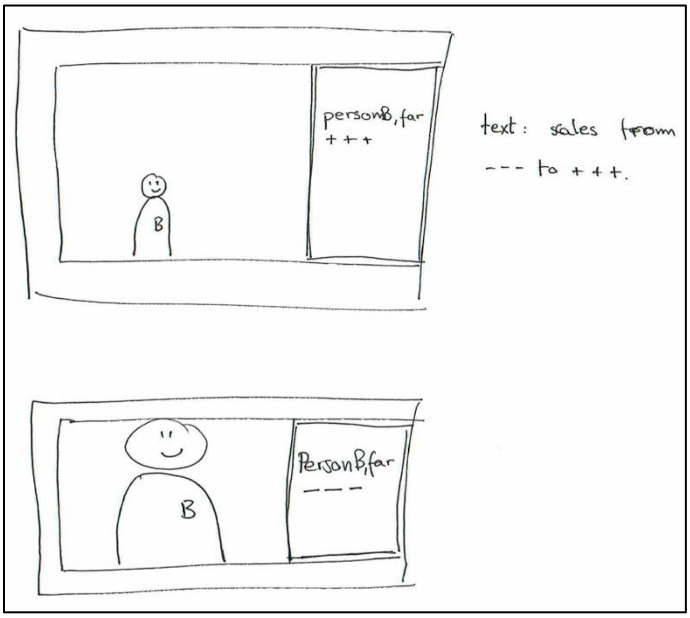


Figure 5 – Proxemics design 3

Using textual feedback was the basic idea of this visualization design.
When the other person tends to be further away from the camera, the visualization will display more "+" signs, and more "-" signs if the other person tends to be closer.

Advantages: A good label as e.g. "distance" would make it obvious what the visualization represents.
Flaws: The meaning of the "+" and "-" symbols might have an unwanted effect on the message conveyed by the visualization: "+" might get associated with "good" and "-" with "bad" and the visualization might then result in the users trying to get "+" signs.

Option 4

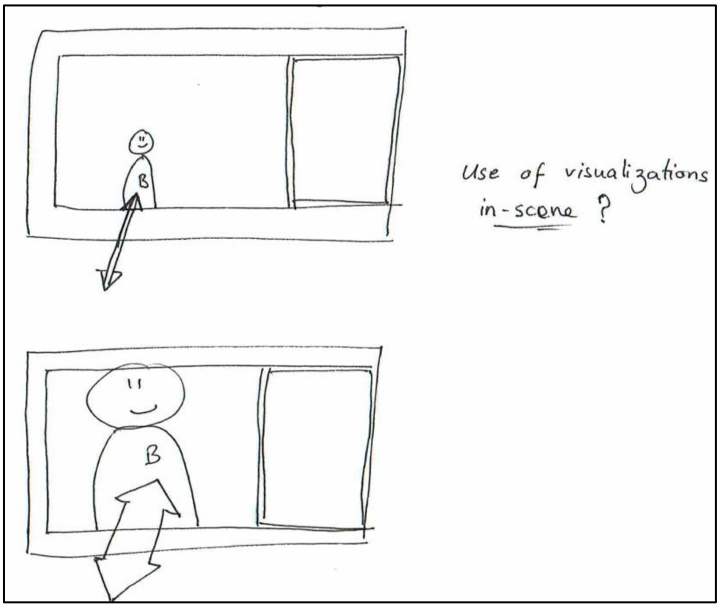


Figure 6 – Proxemics design 4

This visualization design uses graphical cues in-scene to represent distance. The 3D-arrow is simulating the direct representation between the actual persons if the conversation would be in face-to-face. The more distant the other user is, the thinner the arrow would become. Conversely, the closer the other user is, the thicker the arrow would get.

Advantages: By using direct representation in-scene, it is very clear what the visualization represents.

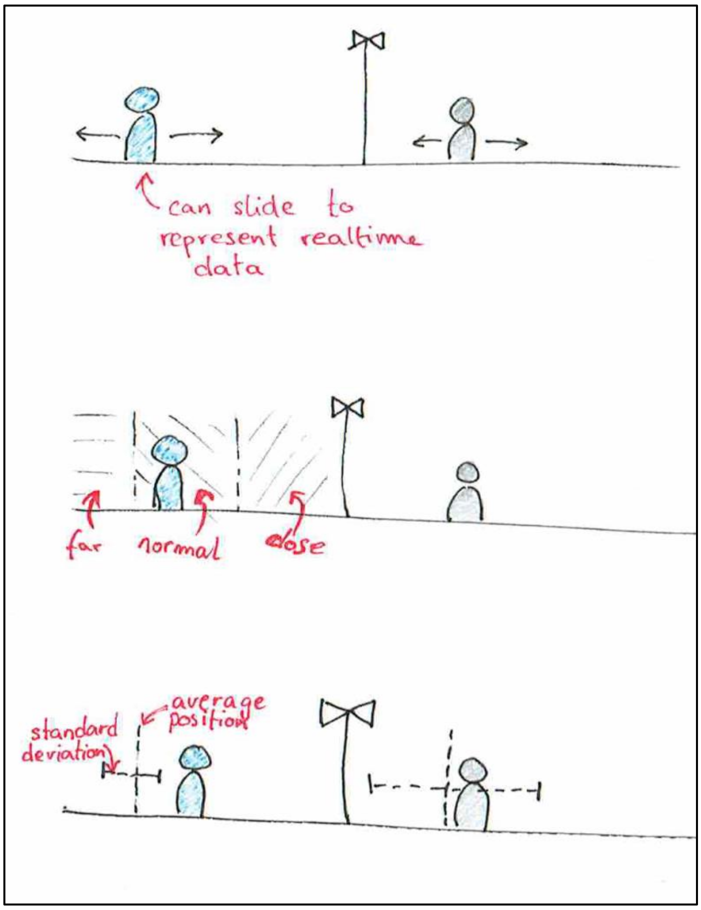
Flaws: It might be unclear to the user whether the visualization represents the difference in distance to the camera or the absolute distance between users.
Such a graphical visualization would require anchor points in the scene to draw the 3D arrow on the good spot. The videoconferencing technology I selected (see *Implementation of a videoconferencing tool*) did not allow video overlay, so I was not able to experiment with this idea.

Step 2: Selection of the best designs and improvements

Based on the analysis of possible advantages and flaws of the different design ideas, and on the discussion I had with my tutor, I selected the option 2 to be the most promising one and thus to be the one I would further investigate. Option 2 had the significant advantage to use natural mapping: it is intuitive as it represents a side-view of what would happen in a face-to-face conversation.

In the following I present the refinement ideas for the proxemics visualization designs.

The following improvements are focusing on the selected visualization design (*Option 2*).



Improvement 1: Sliding avatar

Improvement 2: Buckets

Improvement 3: History information

Figure 7 – Improvements for Proxemics design 2

Improvement 1: Sliding avatar

The initial design was reflecting only the difference in behaviour. The improvement idea was to use the same metaphor of two pawns representing the position to the camera, but to reflect the actual state of the conversation, i.e. the real-time state of distance to the camera for both users. This would be achieved by making the user avatar slide on the horizontal axis

to represent the actual distance to the camera. The visualization would then not represent any sort of difference in behaviour, but the user would be able to compute it himself.
The interest of such a modification is that it might become more obvious to the user that a change in his behaviour has an immediate impact on the visualization.

Improvement 2: Buckets

In the initial design, there were circle-shaped markers to help the user locate in space and give some basic indication of what can be a short, normal or long distance to the camera. The markers of the initial design were very discreet. The idea of this refinement was to make these buckets appear clearly, with different colours.
Each of the three zones would get illuminated whenever the user's avatar enters it.
This modification would emphasize the presence of these different virtual zones in the setting of a real conversation. A major drawback of this approach is that it would not take in consideration the fact that different cultures make different use of space (see [15]). For this reason I did not regard this improvement when defining the final visualization designs.

Improvement 3: History information

This improvement comes as a complement to improvement 1. Improvement 1's idea was to do a direct representation of what happens in the immediate conversation. A drawback of this approach is that it does not make use of data over time and thus does not offer information about general behaviour.
Improvement 3 thus aims at providing historical information, in the form of an average line (orange dashed) and a standard deviation line (green dashed). The average and standard deviation would reflect the recent past behaviour, but could also be used to present information spanning on the whole conversation.

Improvement 4: Opacity

Improvement 4 builds upon the implications of improvement 1.
Improvement 1 would affect the visualization so that it would not represent any comparison data explicitly. But the original aim of the visualizations is to help the user become more aware of differences in nonverbal communication between him and his communication partner. Thus an alteration on improvement 1 would be needed to encourage the user to pay attention to the difference in proximity to the camera.
Improvement 4 is about making the user more aware that there might be a behaviour difference: when the tool detects a high behavioural difference, the opacity of the visualization is maximal, but the more both users behave similarly, the more the visualization fades out to be less present in the user's field of view.

Step 3: Discussion with an Infovis specialist

Following the discussion with the Infovis specialist, different important factors were identified on how to design better visualizations. First, it was identified that it would be interesting to try visualizations with different levels of explicitness of the representation of differences in nonverbal behaviours: what are the implications of show differences implicitly or explicitly? Also, we found that, to emphasize difference when representing live-feedback on distance to the camera, it might be more intuitive to represent both users on the same side, to make the comparison be more visual.
Considering these factors, a new design option was established:

New Visualization: Option 5

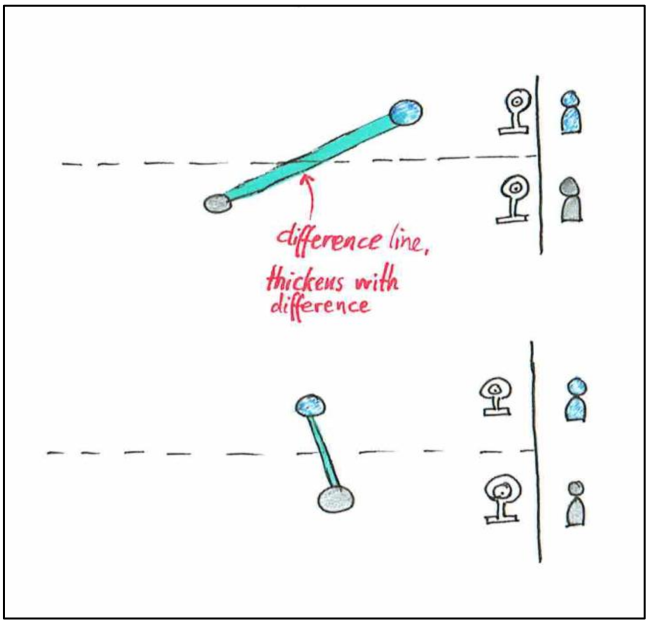
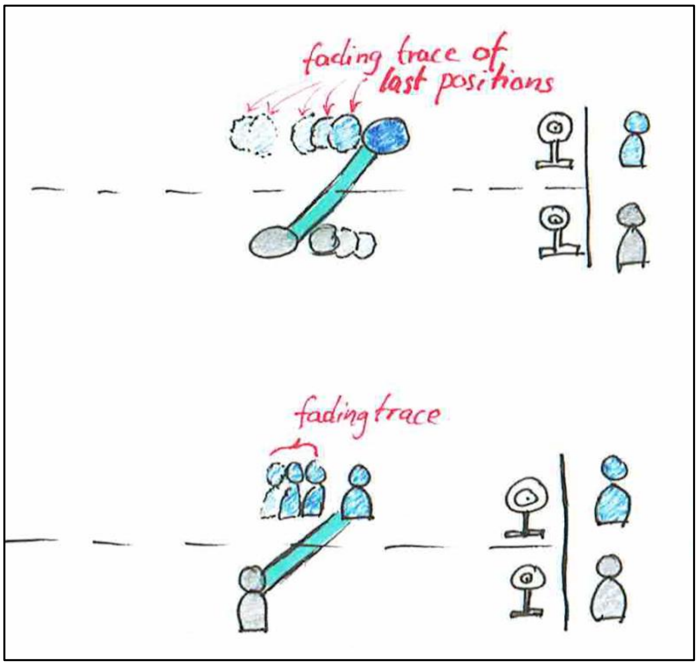


Figure 8 – Proxemics design 5

The most important information I wanted to make apparent with the visualizations was the fact that there are differences in nonverbal behaviour. For this reason, in this visualization design the focus is on presenting the difference, by the mean of a difference line linking both circles representing both users. The line would grow when the circles are far from one another, emphasizing difference; and shrink as they come closer to one another. This representation is allowed by the fact that both users' avatars are on the same side of the visualization, conversely as in visualization design 2 where a mirrored representation is used. A variation of improvement 3 was included in this visualization design in the form of a fading trace of previous position from the recent past.



Variation of improvement 3 : History information

Improvement 5 : Avatars instead of circles

Figure 9 – Improvements on proxemics design 5

Improvement 5: Avatars instead of circles

Option 5 used abstract circles to represent the users' position. I was interested in getting to know if this would be enough for users to interpret what data Proxemics 2 was showing; this is why I replicated Proxemics 2 into a new visualization Proxemics 3, with the only difference being the fact that Proxemics 3 would directly use the avatars in place of the circles.

Step 4: Implementation

Final visualization designs – Proxemics

Based on the improvements elicited in the design process, the following are the two final visualization design for proxemics. Note that Proxemics final 2 comes with a variation according to Improvement 5.
For the rest of the document, I will refer to Proxemics final 1 and 2 as simply Proxemics 1 and Proxemics 2 respectively. Proxemics 1 includes improvements 1, 3, and 4, and Proxemics 2 includes improvements 1, 3 and 5.

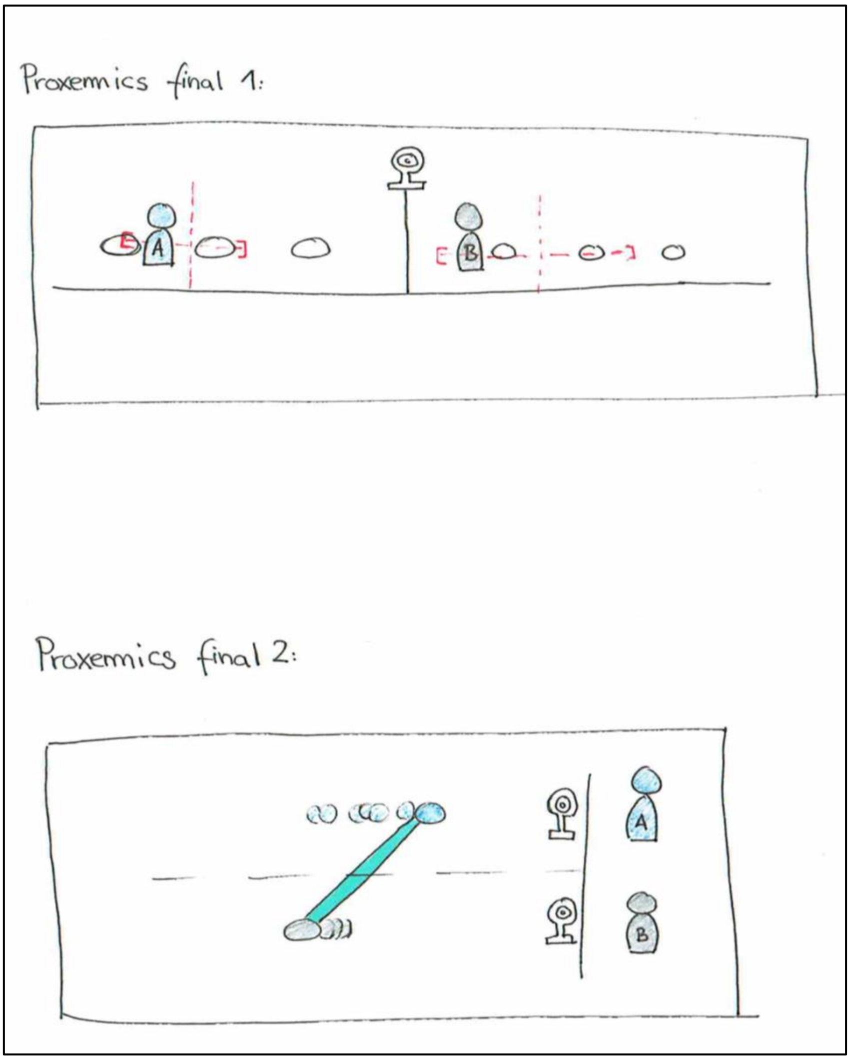


Figure 10 – Final visualization designs – Proxemics

Screenshots of the working visualizations

I implemented the final visualization designs into high-fidelity prototypes. The following are screenshots of the prototypes in working situation with two different settings: when both users have different behaviours, and when they have similar behaviours. Note that in the version the pilot study participants used, the difference line in Proxemics 2 was green.

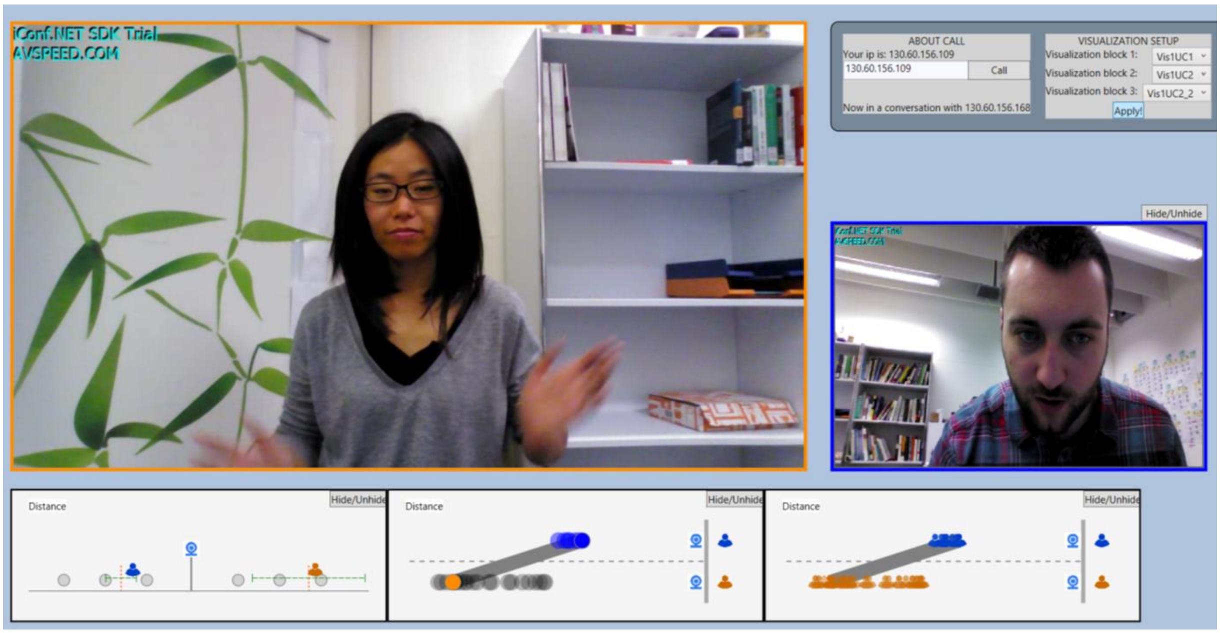


Figure 11 – Screenshot 1: Proxemics different behaviours

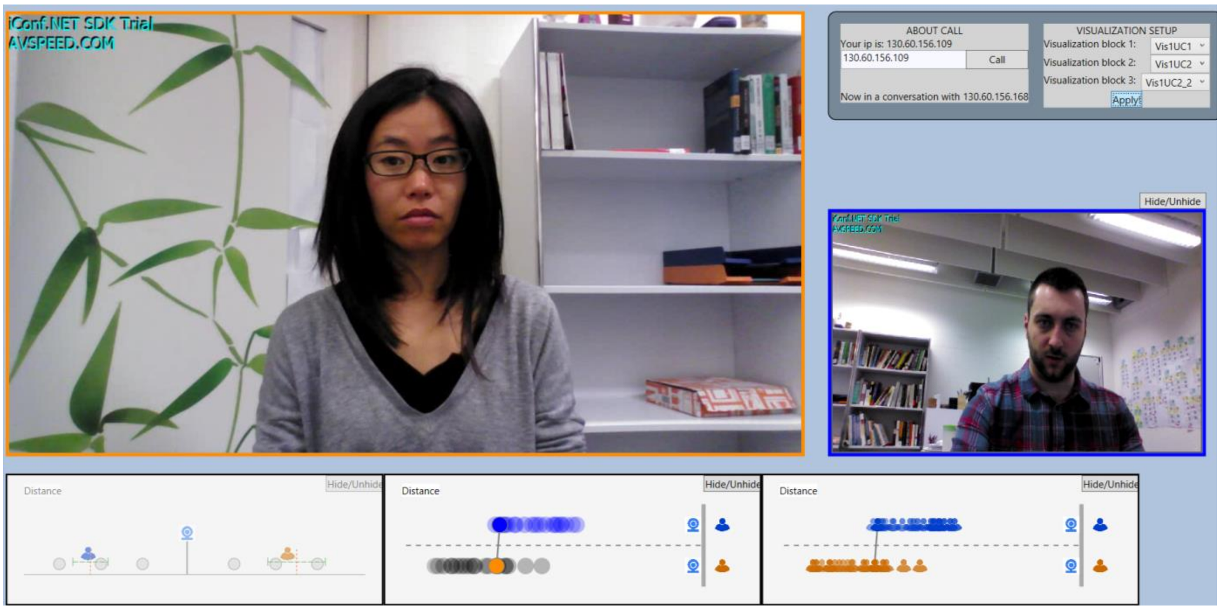


Figure 12 – Screenshot 2: Proxemics similar behaviours

Visualization 2: Smiling

Step 1: Brainstorming of design ideas

When brainstorming on how to display information on smiling, I identified three dimensions to explore: occurrence, duration, and intensity of smiles. I found it more intuitive to visualize information on occurrence and intensity of smiles, thus the following design options are mostly focusing on these two dimensions.

Option 1

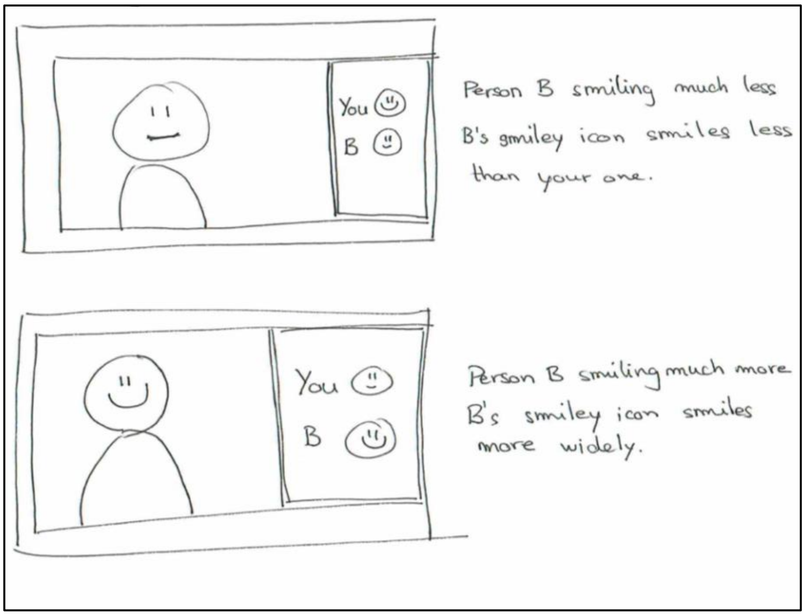


Figure 13 – Smiling design 1

In this visualization, smiling behaviour difference is associated with the size of a smiley face's smile. When one of the users is smiling more, his/her smiley icon would get a larger smile.

Advantages: It is easy to understand that the focus is on smiling.
Flaws: But it might be misleading as the visualization would be displaying information about occurrence that might be interpreted as information about intensity.

Option 2

This visualization is a simple and direct representation of what is happening in real time: how many smiles are done on each side. The more smiles one user does, the higher his bar is growing.

Advantages: Counting might be intuitive to the mind.
Flaws: It does not give any explicit information about difference in behaviour, and it only gives information about quantity/occurrence, and not about intensity of smiles.

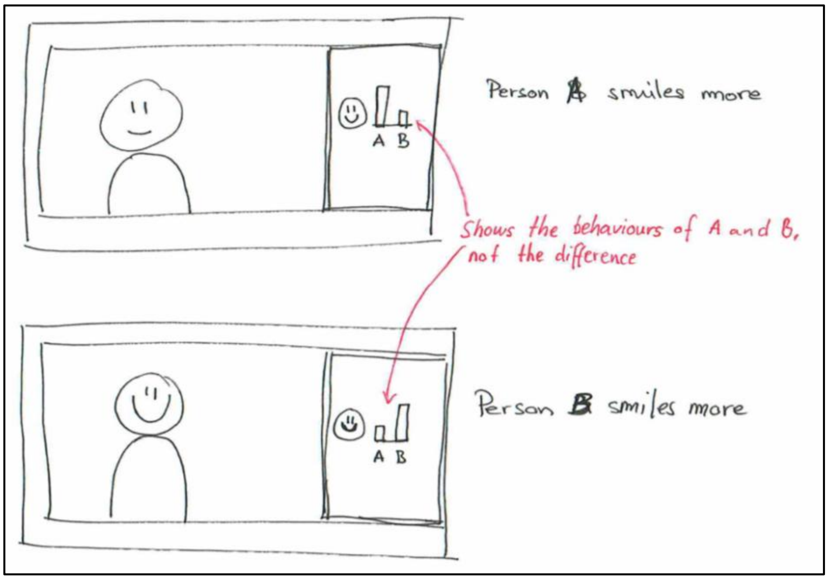


Figure 14 – Smiling design 2

Option 3

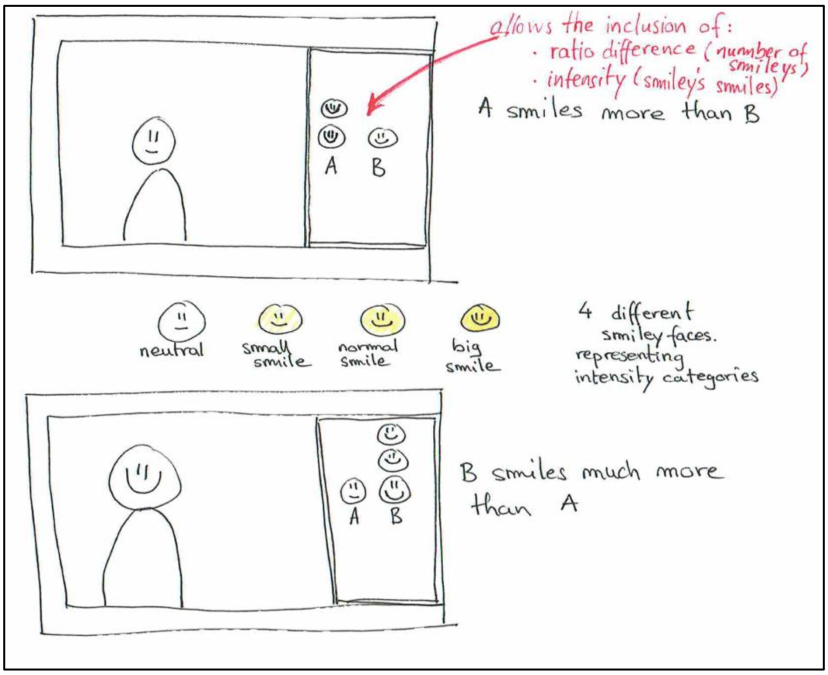


Figure 15 – Smiling design 3

I wanted to be able to show information about both occurrence and intensity differences; this is why I came up with a visualization that combines both. The more a user would smile more than his/her communication partner (in ratio) the more smiley faces will be one his/her side. With the same idea, 4 different smiley icons would depict the difference in intensity categories.

Advantages: This option combines both information.
Flaws: As it is a ratio comparison, there is a need for some side to stay at neutral (e.g. ratio is 1 to 4 or 1 to 3)

Step 2: Selection of the best designs and improvements

As for the proxemics visualizations, the analysis of possible advantages and flaws of the different design ideas and the discussion I had with my tutor helped me select smiling visualization design 3 to be the most promising one and thus to be the one I would further investigate.

Here I present refinement ideas for smiling design 3, which I chose to turn into new visualization designs.

Improvement 1: Separate occurrence and intensity comparison

Separating the visualization of occurrence and intensity comparisons into two subparts might be easier for the user to understand and result in the visualization being less confusing.

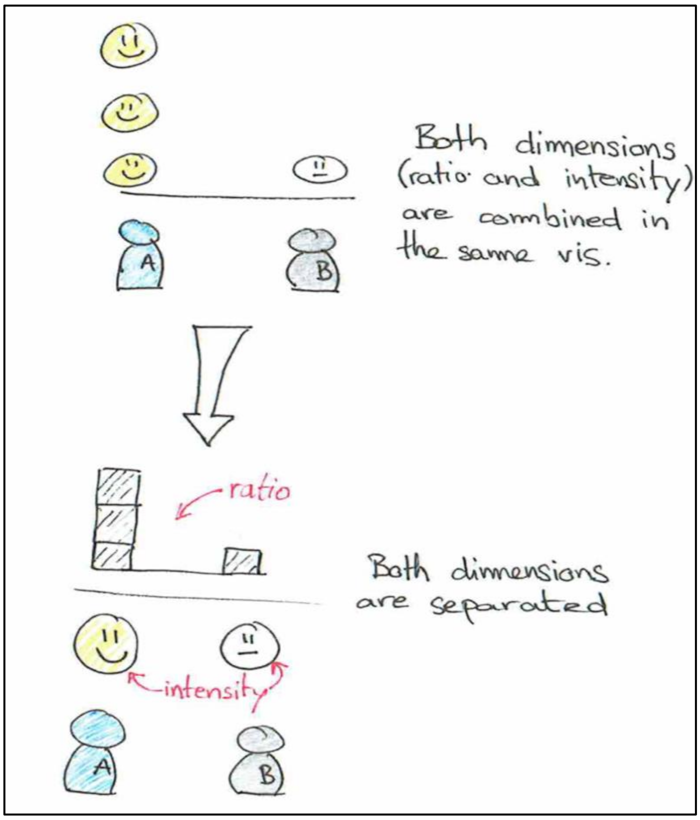


Figure 16 – Improvement 1 on Smiling 3: Smiling design 4

Improvement 2: Use a scale to express ratio of occurrence

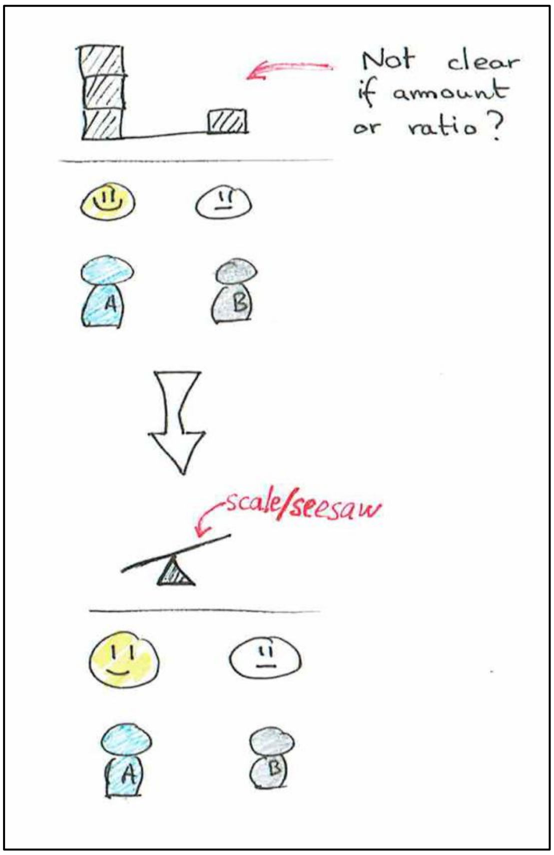


Figure 17 – Improvement 2 on Smiling 3: Smiling design 5

As in Option 4 of smiling visualization, bars could be interpreted as amount and not ratio; I wanted to make the notion of ratio appear more clearly in the visualization.
Thus I made a variant of the same idea with a scale or seesaw-like representation of ratio in smiling occurrences.

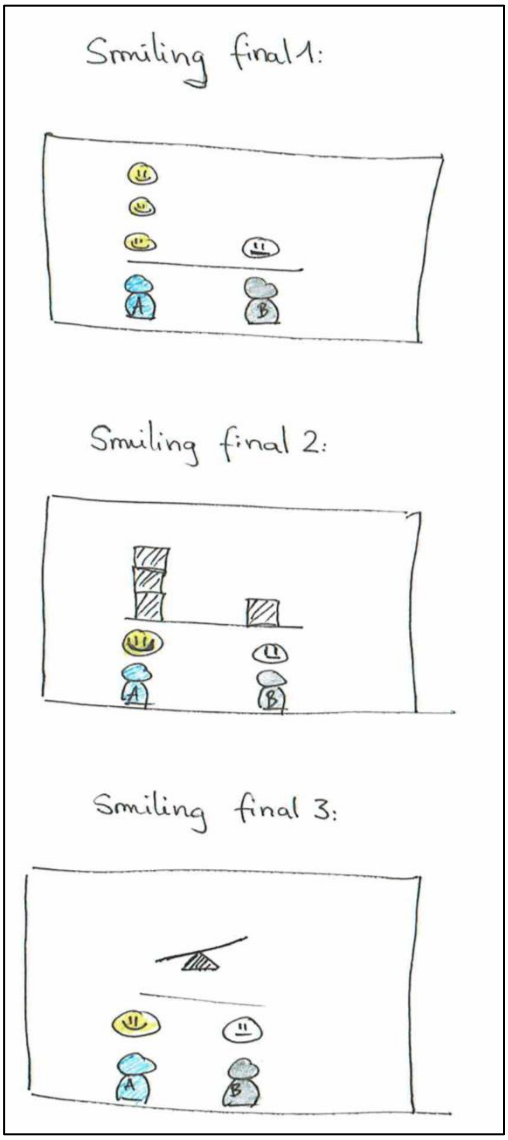


Figure 18 – Final visualization designs – Smiling

Step 3: Discussion with an Infovis specialist

Final visualization designs – Smiling

Based on the abovementioned improvements and the discussion with the Infovis specialsit, I decided to use the three variants of visualization design 3 as the final designs I would implement as high-fidelity prototypes.
For the rest of this thesis, I will refer to these three prototypes as Smiling 1, Smiling 2, and Smiling 3.

University of Zurich[UZH]

Step 4: Implementation

Screenshots of the working visualizations

I implemented the final visualization designs into high-fidelity prototypes. The following are screenshots of the prototypes in working situation with two different settings: when both users have different behaviours, and when they have similar behaviours.

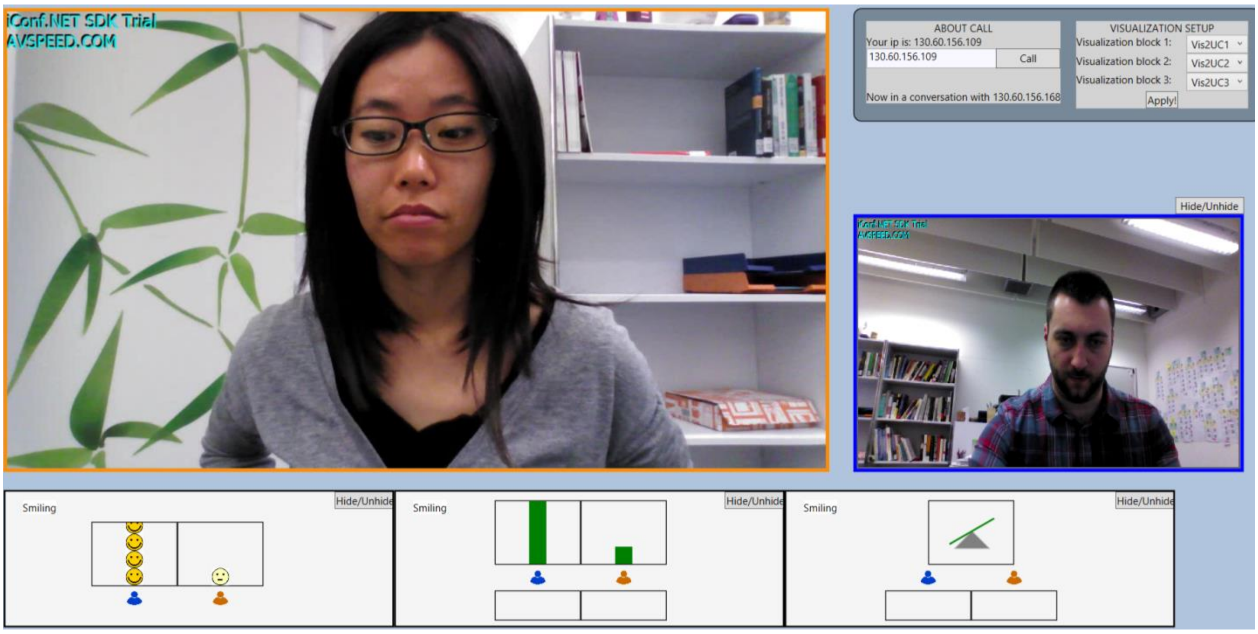


Figure 19 – Screenshot 3: Smiling different behaviours

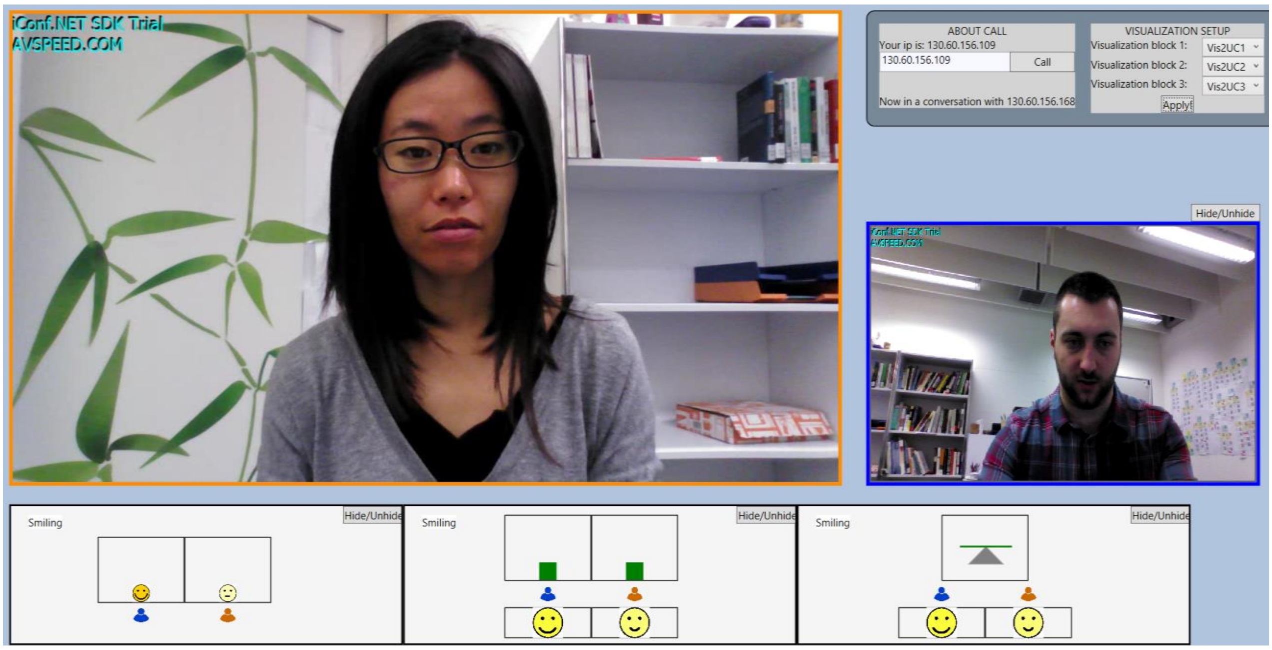


Figure 20 – Screenshot 4: Smiling similar behaviours

Visualization 3: Expressiveness

Step 1: Brainstorming of design ideas

The third nonverbal cue I was originally interested in tracking was eyebrow movement. But as I explain in this section, it turned out to be much more difficult to come up with interesting design ideas for eyebrow movement than for the two other behaviours. This is why I chose to change and explore an aggregate dimension including eyebrow movements instead of eyebrow movements alone. A reason for the need of this change is that eyebrow movements usually do not have a meaning per se, most of the time they are used to emphasize other facial expressions or are used in combination of other nonverbal cues to gain a meaning.

Option 1

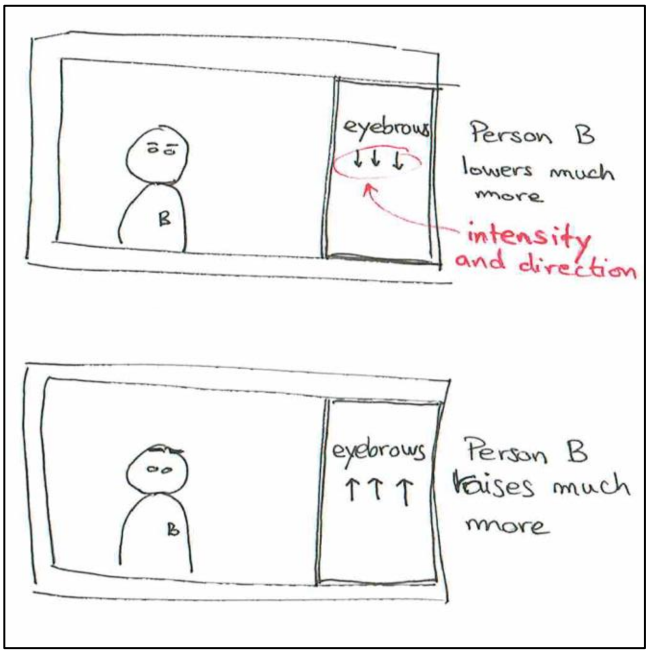


Figure 21 – Eyebrows design 1

In this visualization, arrows would indicate how much the other person acts differently than the user.

Advantages: The user does not need to compute the difference himself as is it explicitly given by then visualization.
Flaws: This visualization seems very prescriptive about what is the appropriate behaviour to have, i.e. mimicking.

Option 2

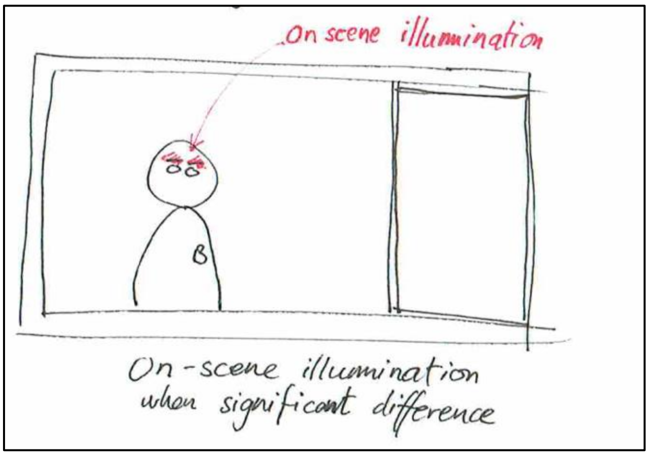


Figure 22 – Eyebrows design 2

This visualization would exploit on-scene illumination as in proxemics design 4, in order to draw attention to the eyebrow zone when significant difference is detected.

Advantages: On-scene, immediately clear what region of the face is concerned.
Flaws: Using a lot of on-scene visualizations might be distractive. Video overlay was not supported in the videoconferencing SDK I used.

Step 2: Selection of the best designs and improvements

Switch to Expressiveness

As the dimension of eyebrow movements did not elicit many interesting design ideas, I decided to try to track a different and more general nonverbal cue that would include eyebrow movements. I chose to try and go with Expressiveness, an experimental value I made up from different visual cues, including eyebrow movements, but also smiling behaviour, proximity to the camera, head movement, hand gestures and emotional display.
Based on this new value I developed the following visualization design ideas.

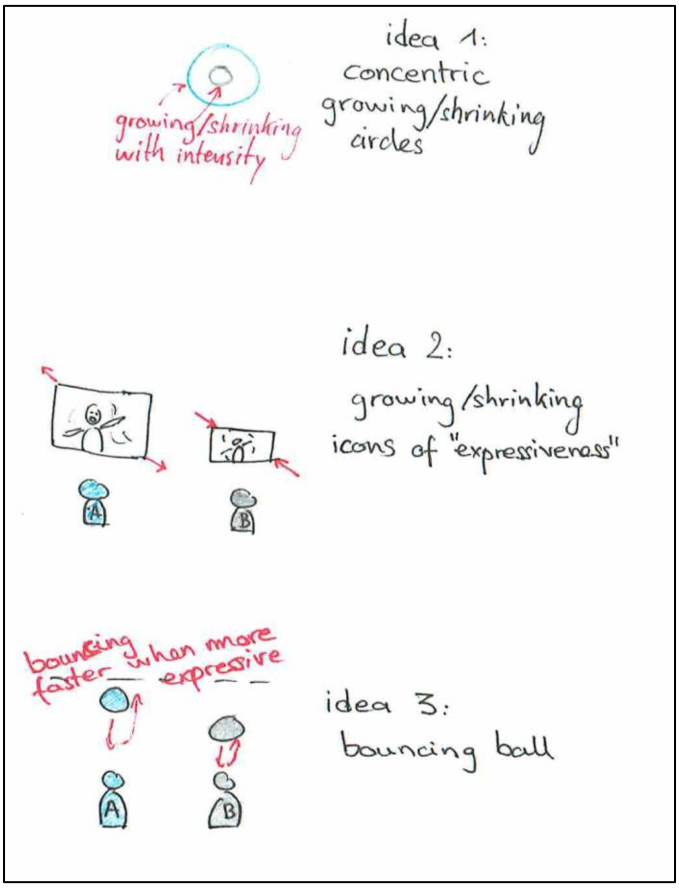


Figure 23 – New visualization design : Expressiveness

Design idea 1: Concentric circles that represent the intensity of Expressiveness of each communication participant with their radius. A big circle means the person is being very expressive whereas a small circle means the user is not very expressive.

Design idea 2: Same idea as for idea 1 but with growing/shrinking icons representing a very expressive person.

Design idea 3: Each user's expressiveness would be depicted as the speed of a bouncing ball: the more expressive, the fastest the ball would bounce. This idea I discarded as it might be annoying to the user.

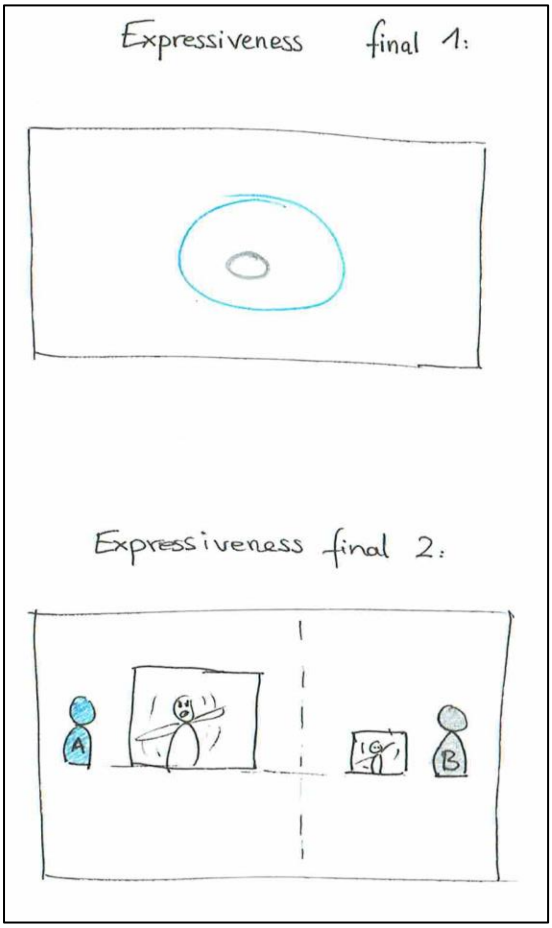


Figure 24 – Final visualization designs – Expressiveness

Step 3: Discussion with an Infovis specialist

The discussion with the Infovis specialist helped me selecting the two most interesting design ideas from the options 1 to 3. Option 3 was discarded as we believed it would have been very distractive.

Final visualization designs – Expressiveness

Following the decision of switching to Expressiveness, I chose to implement design idea 1 and 2 for Expressiveness.

Step 4: Implementation

Screenshots of the working visualizations

I implemented the final visualization designs into high-fidelity prototypes. The following are screenshots of the prototypes in working situation with two different settings: when both users have different behaviours, and when they have similar behaviours.

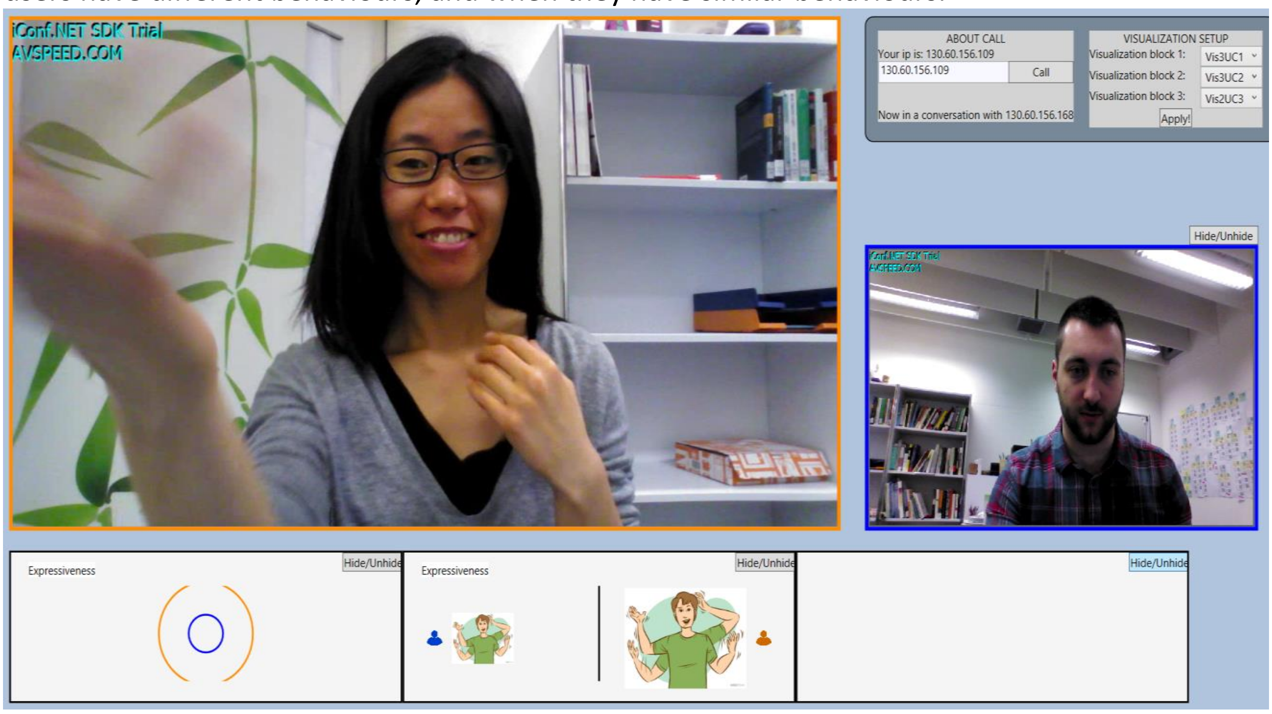


Figure 25 – Screenshot 5: Expressiveness different behaviours

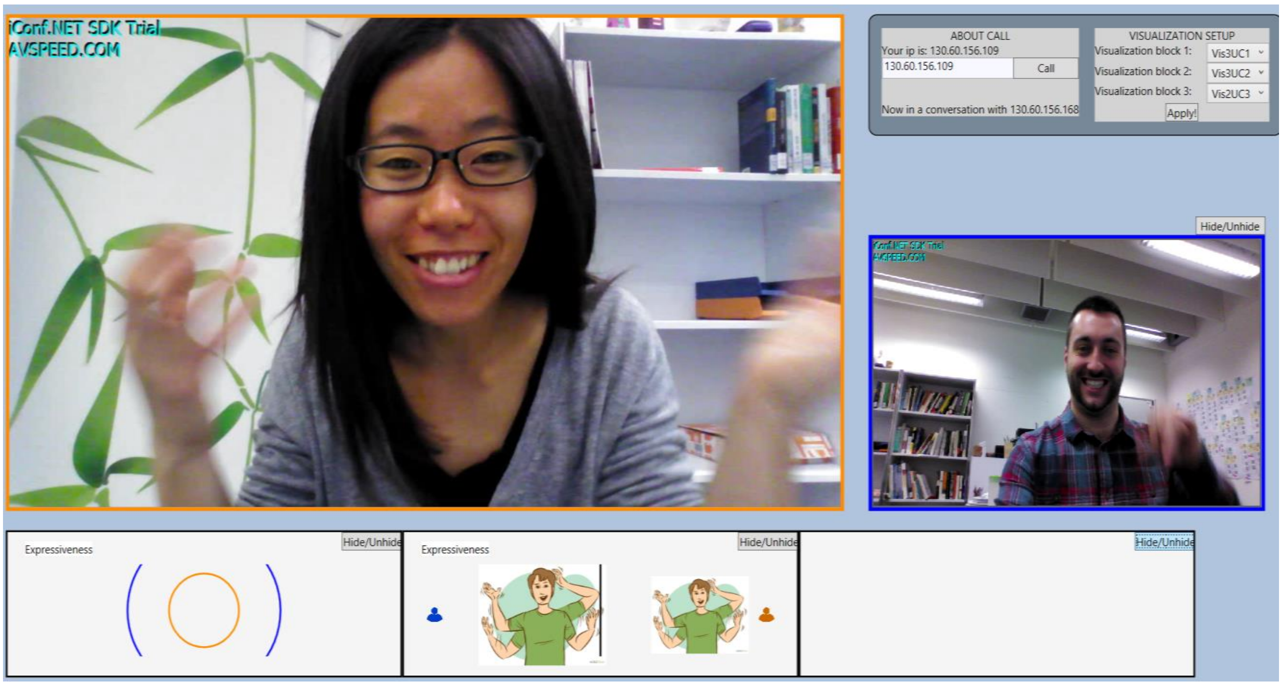


Figure 26 – Screenshot 6: Expressiveness similar behaviours

Chapter 6 – Pilot study

After implementing the visualization prototypes to a high-fidelity level, I conducted a pilot study together with my tutor.

Goal

The aim of this pilot study was to assess the comprehensibility and intuitiveness of the final set of visualization prototypes presented in *Chapter 5 – Visualizing nonverbal behaviour*, as well as getting low-level feedback about the visualizations themselves. We were also interested in getting high-level feedback about the study design as most of the participants had expertise in the Human-Computer Interaction field. The analysis of the results from the pilot study aimed at providing insights on the research questions identified in *Chapter 1 – Introduction*.

Participants

Category	Data
Age category	19-25: 0 26-35: 7 36-45: 1 46-55: 1
Country born	China: 1 Germany: 2 Japan: 1 Serbia: 1 Switzerland: 2 Norway: 1 Thailand: 1
Occupation	PhD Student: 4 Researcher: 1 Postdoc: 1 Administration: 1 Student: 1 Uni Assistant: 1
Lived in more than 2 countries	Yes: 6 No: 3
Use of face-to-face communication with people from other cultures	Daily: 7 Weekly: 2 Monthly: 0 Other: 0
Use of videoconferencing with people from other cultures	Daily: 1 Weekly: 2 Monthly: 3 Other: 3

Table 4 – Participant data

There were 9 persons participating in the pilot study. 5 of them were lab colleagues from the People and Computing (ZPAC) lab of the University of Zurich, whereas the 4 other participants were people working at the Institute for Computer Science (IFI) of the University of Zurich. The participants had different cultural backgrounds: there were 2 Germans, 2 Swiss and 1 of each of the following: Japanese, Chinese, Norwegian, Serbian, and Thai. 7 of the participants were between 26 and 35 years old, 1 was between 36 and 45 years old and 1 was between 46 and 55. Most of the participants were working in academics (7), one worked in administration (1), and one was a student (1). Most of the participants lived in more than 2 different countries (6), while few of them lived in 2 or less countries (3). Most of them mentioned they have daily face-to-face communication with people from different cultural background (7), while few of them had it weekly (2). The tendency of usage of video conferencing tools with people from different cultural background was more or less uniform: daily (1), weekly (2), monthly (3), and other (3).

Setup

The participants were asked to:

1. **Fill a pre-interview questionnaire** with demographics and on their use of video conferencing tools (see *Appendix D – Pilot study pre-interview questionnaire*).
2. **Watch videos of me and my tutor using the tool** following a set of scenarios to illustrate the visualizations.
   We then asked them to tell us what they understood from what was displayed, as well as general feedback. This step was useful as it informed us about the learn-effect induced by the visualizations, and allowed us allow the users to actually use the tool in the next step.
3. **Use the video conferencing** tool with me or my tutor to have an informal conversation.
   The main task was to have a conversation with me or my tutor, paying attention to the visualizations was more of a side task. This allowed us to gain knowledge on how distractive the visualizations are with respect to the main task of having a conversation over video chat.
4. **Answer questions about their experience** with the tool and about their general experience with video conferencing tools.

In order to avoid expectancy bias, my tutor was asking questions and I was taking notes.
With the participants' permission, the interviews were audio-recorded, and both I and my tutor took notes (Consent form: see Appendix B – User interface design).

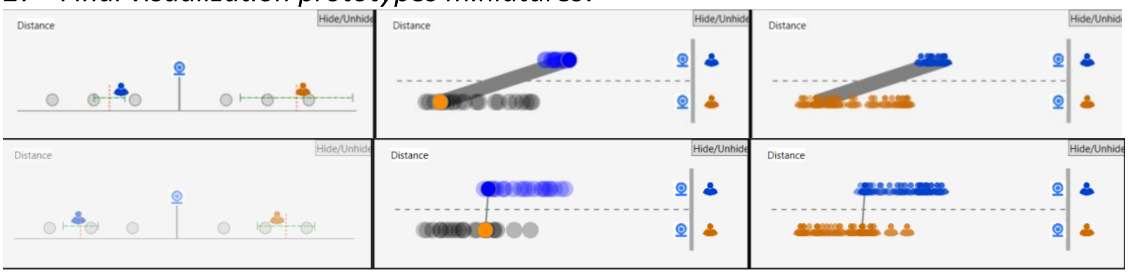
Study data analysis

In order to analyse the study data, I did an affinity analysis based on the notes taken over the course of the interviews. The next section details the results of the pilot study and the analysis I made of it.

Chapter 7 – Findings / Analysis

In this section I present the results from the pilot study and the insights I can conclude on visualization design for nonverbal communication in videoconferencing tools.
First, I resume what participants told us about their experience with videoconferencing tools with regard to nonverbal behaviour. Then, I propose a summary of general feedback on the visualizations and the nonverbal cues being shown in them. After this, I go in more details with low-level feedback we got on the visualizations. Finally, I present an analysis of the gathered study data and implications it can have on visualization design for nonverbal communication differences in the context of videoconferencing tools.

As a reference, the final visualization prototypes used in the pilot study are shown in *Figure 27 - Final visualization prototypes miniatures*.



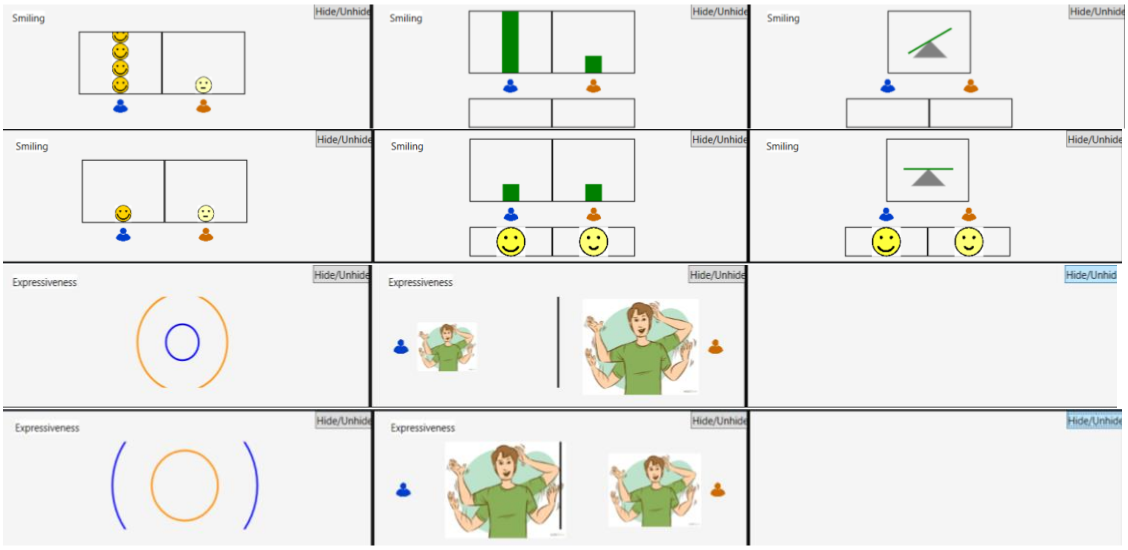


Figure 27 - Final visualization prototypes miniatures

Results from the pilot study

Past experience with videoconferencing tools with regard to nonverbal behaviour

The attention paid to nonverbal behaviour differs with the goal and setting of the conversation: According to our study participants, the attention they give to nonverbal communication differs with the goal and setting of the conversation, especially with its formality level. In particular, participants made a clear separation between social-oriented and work- or task-oriented conversations: in the latter they would tend to pay much more attention to nonverbal communication behaviours than in the former. Similarly, participants were mentioning the fact of knowing or not knowing the conversation partner as determinant. P3 in particular mentioned that when he does not know his communication partner, he wants to get to know his/her baseline behaviour, and thus pays attention to differences in nonverbal communication. Some participants said that they only pay attention to nonverbal communication consciously when the conversation takes place in a professional (thus task-oriented) setting.

The use of nonverbal communication differs with the language used: P6 also mentioned that his use of nonverbal communication differs according with the confidence he has with his verbal communication skills. When speaking in a language he is not native in, he would use nonverbal communication, especially gesturing, as a way to support conversation and compensate with a lack of words.

A lot of information is not available in videoconference: The nature of videoconferencing tools also can restrict parts of nonverbal communication, as P4 and P6 indicated: a lot of information gets lost, as for instance the camera does not show the whole person but mostly his/her face. This has the effect that part of the nonverbal communication is not visible when videoconferencing.

To interpret nonverbal communication you have to be able to see it: There exist different videoconferencing tools, and all of them have different structures, allowing different experiences. P3 mentioned that Google Hangouts was difficult to use with respect to paying attention to nonverbal communication, as it only displays the video of the person who is talking. It is thus more difficult to use it as one cannot see the other conversation partner(s), especially when you are talking as you cannot get visual feedback.

General feedback on the visualizations and the nonverbal cues being shown in them

Colour mapping consistency: All participants mentioned that the colour mapping was not consistent as in the final user interface the visualizations appeared under the video window, and thus the mapping was flipped, the right part of the visualizations representing the other person (on the left) and conversely. Nevertheless, all participants were able to figure out who was represented by each part of the visualizations, by paying attention to the colour mapping of the avatars to the video windows' frames, but mostly by observing the reaction between what happened in the actual video conference and what happened in the visualizations. Some participants proposed solutions to the mapping problem. One improvement would be to use the videoconference participants' nicknames to label the visualizations' icons representing them. Another possibility would be to make the avatar icons appear near to the corresponding video windows.

Visualizations can be distractive: Some users found the visualizations were distracting, as it was a bit out of the focus of the conversation view. Other participants found the visualizations were not too distractive because as P2 mentioned "the video window is prominent enough [for the visualizations not to be distractive]", but "it would be nice if the visualizations were in the main window". P3 and P8 remarked that it could be better to overlay visualizations on video but that it might be confusing as well.

Behaviour baseline shifting information: P2, P3 and P5 would be interested in getting information about their and their communication partner's baseline behaviour and see how they shift with respect to their usual behaviour during the conversation. Achieving this would require the establishment of nonverbal behaviour profile and would require taking into account data over the course of many conversations.

Visualization timeframe: Some users also questioned the timeframe of the visualizations. P8 explained that information might be more relevant over time as it is harder to keep track of what happens on the long term compared to seeing what is happening in the immediate present or recent past. P8 even said that he would trust his own perception better than the machine's on real time information, but not for keeping track of past actions. In the same idea, P5 put the finger on the fact that paying attention to real time feedback about the conversation can distract from the conversation itself.

Proxemics

Bad audio prompts to be closer: Many participants agreed on the fact that bad audio quality can be very influential on their distance to the camera, as being closer to the computer can help them hearing their conversation partner better or make them able to be heard better by the other party.

Proximity might have no intentional meaning: To many participants, it was not clear what distance to the camera can have for a meaning per se. P3 mentioned that body movements can be for personal relaxation, and do not necessarily have a meaning; it can be used to help thinking or just to keep oneself active.

Different environments mean different restrictions and distractions: 5 out of 9 participants mentioned that the environment – use of headphones, desks around, small place, luminosity – as well as technical issues – bad audio or video quality – can have an influence on the way they use space. P6 also pointed out that the fact that the participants of a videoconference have different environments – specifically different surroundings – as opposed to a face-to-face conversation, and thus are exposed to different sources of distraction, which can have an influence on the attention they give to the conversation in general and the nonverbal part of it in particular. P3 also indicated that the camera setup is important: often both communication participants would have their camera at a different angle, altering the face-to-face setup.

In videoconference with more than two people proxemics might not have the same meaning: In a videoconferencing scenario with more than two people, P4 stated that distance would not have the same meaning as in a one-to-one conversation, explaining that all persons would need to fit in the camera image.

Perspective problem: the camera is not an eye: Some participants noted the problem of perspective: what you see on the camera does not reflect what you would see in a face-to-face conversation, usually objects and persons seem further away in video conferencing that where they really are. One's distance to the camera does not affect how big one sees the other person, only the other person's distance to the camera has an effect on it. P9 added that the tool can help getting to know what the real state of things is. A good illustration of the perspective problem was proposed by P8: In a movie theatre, a close-up looks the same if you are sitting in front of the screen or if you sit in the rear of the room.

Smiling

Baseline behaviour shifting information: P2 commented that "it would be nice to encourage smiling when they [the conversation participants] critically tend not to smile".P2: Would be nice to encourage smiling when they critically tend not to smile; which goes in the same direction as comparing the users' baseline behaviour to the current one, as explained in *General feedback on the visualizations and the nonverbal cues being shown in them*.

User action needs to influence the visualizations in a timely fashion: Most participants were confused by the fact that the visualizations were not reflecting the immediate state of things, but the behaviour on the overall conversation. For instance they would have expected the visualizations to display nothing when no one is smiling. This is an indication that there is a need to make sure the users understand what timeframe each visualization is reflecting on.

Smiling has an inherent positive meaning, comparison could mean competition: Comparing to proxemics, many participants felt that smiling has an inherent positive meaning, that it is the "right thing to do" (P8), and thus that the displaying of smiley faces could be interpreted as a positive feedback, whereas no smiley face or a neutral one would be interpreted as a negative feedback, or at least as a worse state than having smileys. The comparison of positive feedback between both users was the reason why participants felt like there was a competition. Smiles were even referred as "points" by P8. This comes in contrast to the proxemics visualizations, which felt more like a "collaboration" by P8.

Expressiveness

A mix of multiple dimensions is difficult to grasp as a single one: Participants felt the concept of Expressiveness difficult to grasp, as it includes several separate dimensions. P3 said that capturing expressiveness in a single visualization might be difficult to understand because one does not know what actions would influence the visualization at any time, and thus it is more difficult to choose how to react to it in order to adapt to the conversation partner.

Maintaining a conversation and keeping an eye on a complex visualization is difficult: P2 also mentioned that turn taking in a conversation makes it difficult to look at the visualization while talking, especially if the visualization makes use of many visual cues.

Does expressiveness relate to the conversation or to the temperament: P4 pointed out that expressiveness might tell more about the temperament of a person than about the conversation in itself. Again comparing to baseline behaviour might be a good way to address this issue.

Low-level feedback on the visualizations

Proxemics 1

Visualization design Proxemics 1 was regarded as the most intuitive because it is a direct representation of what would happen in a face-to-face situation. P6 said for instance that it looks like "seeing the other person through a window". P2 mentioned that if the goal was to visualize difference it would be more intuitive to "align in the same coordinate", in the idea of Proxemics 2.

The standard deviation line was often interpreted as representing an error or uncertainty measure, but some participants understood the idea of an average line and its meaning on the history of the conversation.

All participants were not sure what the round markers were meaning.

Nearly all users didn't notice the difference in opacity. Some of those who noticed understood that it gives information about the fact that there is a behaviour difference. The rest of the participants noticing the difference in opacity interpreted it as giving a cue about how important that visualization is to pay attention to at different moments.

Proxemics 2

Proxemics 2 was regarded as less intuitive than Proxemics 1 in terms of knowing immediately what it represents, but better to emphasize that there is a difference in behaviour.
P1 asked "Why is the screen on the right?" because he would have expected it to be on the left as "usually magnitude goes to the right". This is in my opinion an important design principle to take into account for further visualization design.

All participants figured out that the green line was there to indicate difference in behaviour, but to most of them it was not clear what was to be done with it. One reason for this is that the colour of the line was confusing; which implies that the choice of colours is very important for every aspect of the visualization design. The second reason for the confusion was that the line would grow thicker as distance grew. Mostly it was interpreted as P2:"Green is good, more green is more good. […] I'd make the line become bigger". Thus, the combination of the colour green, which is culturally associated with positive feedback, and thickness of the line, might lead to the undesirable effect of inciting users to have different behaviours. P1 suggested as a solution to invert the thickness change in order to encourage mimicking.

All participants discerned that the fading trace had something to do with history and recent past actions, and most participants mentioned that the darker a trace is, the longer the user was there. This is not exactly the information I wanted to convey with this part of the visualization, but it was a nice side effect of the visualization. Overall, the participants had a very positive impression from this part of the visualization and said that it was very meaningful to know what happened in the recent past.

Most participants found improvement 5 on Proxemics 5, the use of avatars instead of circles in Proxemics 5, a better idea than using circles in order to immediately know who is who, although some participants found that this would be redundant as the avatar icon already is on the right side of the visualization for labelling.

Smiling 1

Most participants were confused by the fact that the stacking of smileys was not displaying the amount of smiles. Some participants did not understand what data was represented as they did not see a relation between what they did and the visualization, some even asked if the system was broken.

P5 was concerned by the fact that when someone smiles, he might be "Taking smiles away from the other person". Moreover, P1 said that "One plus one equals zero […] It is strange that my behaviour can delete the effect of their behaviour".
However, some participants liked this option better as it combines both information and is more neutral as it only shows what each side is doing without comparing them.

Smiling 2

P1 said that in this visualization "Bars do not represent quantity here because it went from one to three […] I'm not sure what they represent then […] But bars should represent quantity." Again participants were expecting a count of smiles instead of what the visualization shows.

But participants seemed less "offended" by the fact that "points" were taken away from one another than for Smiling 1. This has to do with the fact that the display of amount, or in our case the ratio of occurrence, was separated from the display of intensity of the smiling behaviour, which in the Smiling 1 was associated with points.

The lower part of the visualization was interpreted by some participants as representing the intensity or the kind of the smiles. The other participants regarded it as being an indication of which behaviour was displayed. This might have been caused by the fact that in the example video, the intensity part did not change a lot or at all.

Smiling 3

The scale was unanimously identified as describing the balance in smiling behaviour. All participants agreed on the fact that this visualization was the one encouraging mimicking the most, but also on the fact that they felt forced to adapt in order to maintain the balance. P1 said for instance "It looks like a battle of smiling […], a competition."

Expressiveness

P1 said that "[The first visualization] shows what happens relative to the other person, [the second visualization] shows more what happens in absolute." P5 mentioned that both visualizations do not really encourage balance as they only show what is happening without displaying any comparison information.

Analysis

Based on the results of the pilot study and the design experience, different inferences can be made on creating visualizations for supporting intercultural communication in the context of videoconferencing tools.

The analysis of the results presented in *Past experience with videoconferencing tools with regard to nonverbal behaviour*, especially the facts that the attention paid to nonverbal behaviour differs with the goal and setting of the conversation and that the use of nonverbal communication differs with the language used show that there is a need for adapting the display of nonverbal communication information feedback according to the context of the conversation. This can also mean that the development and use **potential** of tools to train culture-general skills **might be more significant in the context of professional communications**.

Also, an important point is that one of the main differences between a face-to-face conversation and a videoconference is that both parties are experiencing the communication exposed to different environments. This has on one hand the effect that both communication partners are not subject to the same physical constraints, such as a small room, bad audio or video quality, luminosity, the fact of using headphones, etc., which can reduce their ability to express their nonverbal communication feedback through proxemics. On the other hand, being exposed to different environments also means that each the interlocutors are subject to different sources of distractions. Basing on these observations, and in relation to **RQ1**, I consider that it is important to keep in mind that **the environment can have an impact on nonverbal communication**, especially in the scenario of videoconferencing tools where both interlocutors are exposed to different environments. This means that whatever data we extract with the help of face-tracking cameras, it might not bear the same meaning as it would do in a face-to-face setting. I believe that this point has to be kept in mind when designing any system that tracks nonverbal cues in a computer-mediated communication context.

Similarly, it is important to state that a videoconference setup is constrained by definition, suffering from the problem of perspective. It is indeed difficult to perceive all of the nonverbal communication of one’s conversation partner as the use of a web camera does not allow showing the full body, and thus part of the nonverbal communication might get lost. Perceiving the real distance to one’s interlocutor might also be impaired by the perspective effect caused by the fact that the screen cannot really reproduce the tridimensional feeling of a face-to-face conversation. The technology used in this thesis – Intel RealSense – comes in handy in this situation as it can produce depth information thanks to its depth camera. With respect to the abovementioned and regarding research question **RQ2**, I think that the HCI research community should look at **new technologies as ways to augment the videoconferencing experience** and help users overcome its restriction in order to **make it feel more like a face-to-face conversation**.

In order to address **RQ1,** two elements are to be taken in consideration. On the one hand it is important to determine what makes a visualization **intuitive**. According to the pilot study results, presenting the data according to the natural mapping, as in Proxemics 1 when it reproduced the situation of a face-to-face situation, as well as making sure that the user actions have a visible, quick and identifiable impact on the data shown are two determinant factors to the intuitiveness of a visualization. On the other hand, it is also crucial to know how to make visualizations **not be prescriptive** about what the appropriate behaviour is. For instance, many study participants felt forced to adapt when using Smiling 3 because they felt there was an incentive and thus a pressure to keep a balanced state. Consequently, I think that it is important to **assess the inherent meaning** of each **nonverbal dimension** one wants to visualize in order to be able to **produce intuitive and neutral visualizations**. This comes in opposition to **RQ2** when we want to train culture general skills and in particular encourage mimicking. Indeed, the visualizations which were described by the study participants to be encouraging mimicking the most were also the ones the participants regarded as being the most prescriptive. Based on these considerations I think it is important to keep in mind that there might be a **trade-off between inciting to mimic and showing information in a non-prescriptive way** when designing visualizations for supporting intercultural communication. In correlation to this hypothesis comes the implicitness or explicitness of the display of differences in nonverbal communication. In fact, the visualizations focusing on displaying differences explicitly (Proxemics 2, Smiling 3) were perceived as being more prescriptive – and thus better at inciting mimicking – than the visualizations focusing on displaying the state of things without interpreting, which were described as best for making users aware of their behaviour. With respect to the two culture-general skills described in *Chapter 2 – Intercultural* communication, and regarding these considerations, there might also exist a **trade-off between inciting to mimic and making users more aware of their own behaviour** in terms of the explicitness or implicitness of the display of differences in nonverbal communication.

Another key element is the **timeframe** which the visualizations refer to. As the results from the pilot study indicate, users might be confident about their ability to process the immediate nonverbal information, but they have less faith in their capacity to **keep track of nonverbal communication** on the longer run and understand patterns in it. This is, in my opinion and considering **RQ2**, one the most promising points where technology can support intercultural communication and communication in general. Multiple possibilities exist to add historical information in visualizations. For instance I explored the use of the average and deviation of position in Proxemics 1 and the use of a fading trace of last positions in Proxemics 2, but many other options could be investigated, as for instance using line graphs over the course of the conversation.

Finally, I think it is crucial to point out that **certain nonverbal behaviours might be associated with a semantic meaning** and should be treated with care when designing visualizations. We know from anthropological literature that smiling for instance has an inherent positive connotation. As opposed to proximity to the camera, which the pilot study participants did not perceive as having a real meaning per se. Nonverbal behaviours with a semantic meaning should be **visualized in a neutral way** in order to avoid transmitting the wrong message, as showed the fact that some users in the study felt that they were in a smiling competition. Likewise, **certain visual dimensions as for instance colour are critically decisive towards the resulting behaviour of users**. The fact that users were confused by the fact that the difference line was green in Proxemics 2 illustrates that point. These two points come in addition to address **RQ2**.

Chapter 8 – Future Work

Considering the insights presented in *Analysis*, I propose in this section leads for future visualization design of nonverbal communication in the context of videoconferencing tools.

First, it would be interesting to **explore more nonverbal cues** and to define **which are the most important** to visualize in order to support intercultural communication; this would help exploring and organizing the large design space. The determination of patterns in user behaviour and acceptance with regard to visualizations would allow establishing guidelines for future technology design.

In *Comparing nonverbal behaviours*, I identified one of the main challenges to visualize nonverbal behaviour comparison data: the **need of a categorization of difference**. I used the bucket words "equals", "a bit more", "more", and "much more" to allow this categorization. The main issue of this approach is that it requires defining values to delimit the buckets, and this for all nonverbal cues separately. It would be useful to do a study to **define these values** in an objective way and maybe to **consider the cultural differences** in this matter. Such a study would go more into the direction of social psychology and cultural anthropology.

Another attractive lead would be to **experiment with in-scene visualization**, i.e. the overlay of visualization on the video itself. Further design approaches would be available to in-scene visualization design as e.g. the highlighting of important areas in scene related to nonverbal communication. I proposed two possible design ideas in *Chapter 5 – Visualizing nonverbal behaviour*. Exploring this lead would also mean to inspect **potential distraction** effect of visualization overlay from the task of having a conversation. This is important for technology design as it can give insights about what are the limits between augmenting the communication experience and distracting from the main task of participating actively in a conversation.

During the design process of my visualizations and in the course of the pilot study I was often confronted with the topic of deciding on a timeframe on which to show information. It would be constructive to study the **implications of displaying information** about differences in nonverbal communication on the **immediate**, **semi-immediate** and **long-term** timescales.

Finally, based on the analysis of the results from the pilot study, I think it would be interesting to **evaluate the users' baseline behaviour**, i.e. how they tend to behave in general and over the course of multiple conversations, and produce visualizations on how they differ from these usual behaviours. This could have a positive impact on self-awareness and thus further help training culture-general skills. In the same line, a more general and broad concept would be to do the analysis of nonverbal behaviours for the different cultures and **establish cultural "profiles"** of nonverbal communication behaviours. Thanks to this it would be possible to **extract culture-specific nonverbal feedback** and produce visualizations that take into account cultural differences. Furthermore, establishing cultural profiles of nonverbal communication behaviours would possibly **allow informing future cultural anthropology literature** with **concrete and measurable data**. This was one of the initial project ideas me and my tutor were considering, but we recognized it would require much more time and resources than what I was able to use for my thesis.

Chapter 9 – Conclusion

In this thesis I did a first exploration of how to augment videoconferencing tools to train the culture-general skills of awareness and adaptation (see *Chapter 2 – Intercultural communication*). In order to do this, I wanted to produce visualizations of nonverbal communication behaviour differences in the context of dyadic videoconferencing. With this thesis I aimed at exploring the following two research questions:

How can we design intuitive visualizations of nonverbal behaviour during real-time video chat?

How can such visualizations of nonverbal behaviour help to train the culture-general skills of awareness and adaptation?

In order to address these research questions, I implemented a system integrating videoconferencing and face-tracking that is able to detect, track and compare nonverbal behaviours and then offer live visual feedback to the users. I designed interactive visualizations that aimed at helping the users being more self-aware of their behaviour and of their interlocutor's so that they can choose to adapt or not.

Together with my tutor I then conducted a pilot study to assess the comprehensibility of the designed visualizations and to get to know how to design better and more intuitive visualizations.

References

[1] R. W. Brislin and T. Yoshida, *Intercultural Communication Training: An Introduction*. SAGE Publications, 1994.

[2] W. Messner and N. Schäfer, "Advancing Intercultural Competencies for Global Collaboration," in *Globalization of Professional Services*, U. Bäumer, P. Kreutter, and W. Messner, Eds. Springer Berlin Heidelberg, 2012, pp. 189–201.

[3] K.-Y. Ng and P. C. Earley, "Culture + Intelligence Old Constructs, New Frontiers," *Group & Organization Management*, vol. 31, no. 1, pp. 4–19, Feb. 2006.

[4] H. Ai He, "Facilitating Intercultural Communication: The Role of Technology in Training Culture-General Skills." To be released.

[5] E. Ofek, S. T. Iqbal, and K. Strauss, "Reducing disruption from subtle information delivery during a conversation: mode and bandwidth investigation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 3111–3120.

[6] J. W. Neuliep, *Intercultural Communication: A Contextual Approach: A Contextual Approach*. SAGE Publications, 2014.

[7] L. Laroche, *Managing Cultural Diversity in Technical Professions*. Routledge, 2012.

[8] S. Ting-Toomey, *Communicating across cultures*. New York: Guilford Press, 1999.

[9] A. Mehrabian, *Nonverbal Communication*. Transaction Publishers, 1972.

[10] Cynthia Danisile Ntuli, "Intercultural Misunderstanding in South Africa: An Analysis of Nonverbal Communication Behaviour in Context." University of South Africa, South Africa.

[11] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3–4, pp. 169–200, May 1992.

[12] P. Ekman and W. Friesen, "Facial Action Coding System: A technique for the measurement of facial movements," *Consulting Psychologist*, vol. 2, 1978.

[13] P. Ekman and E. L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.

[14] E. T. Hall, *The hidden dimension (1st ed.)*, vol. xii. New York, NY, US: Doubleday & Co, 1966.

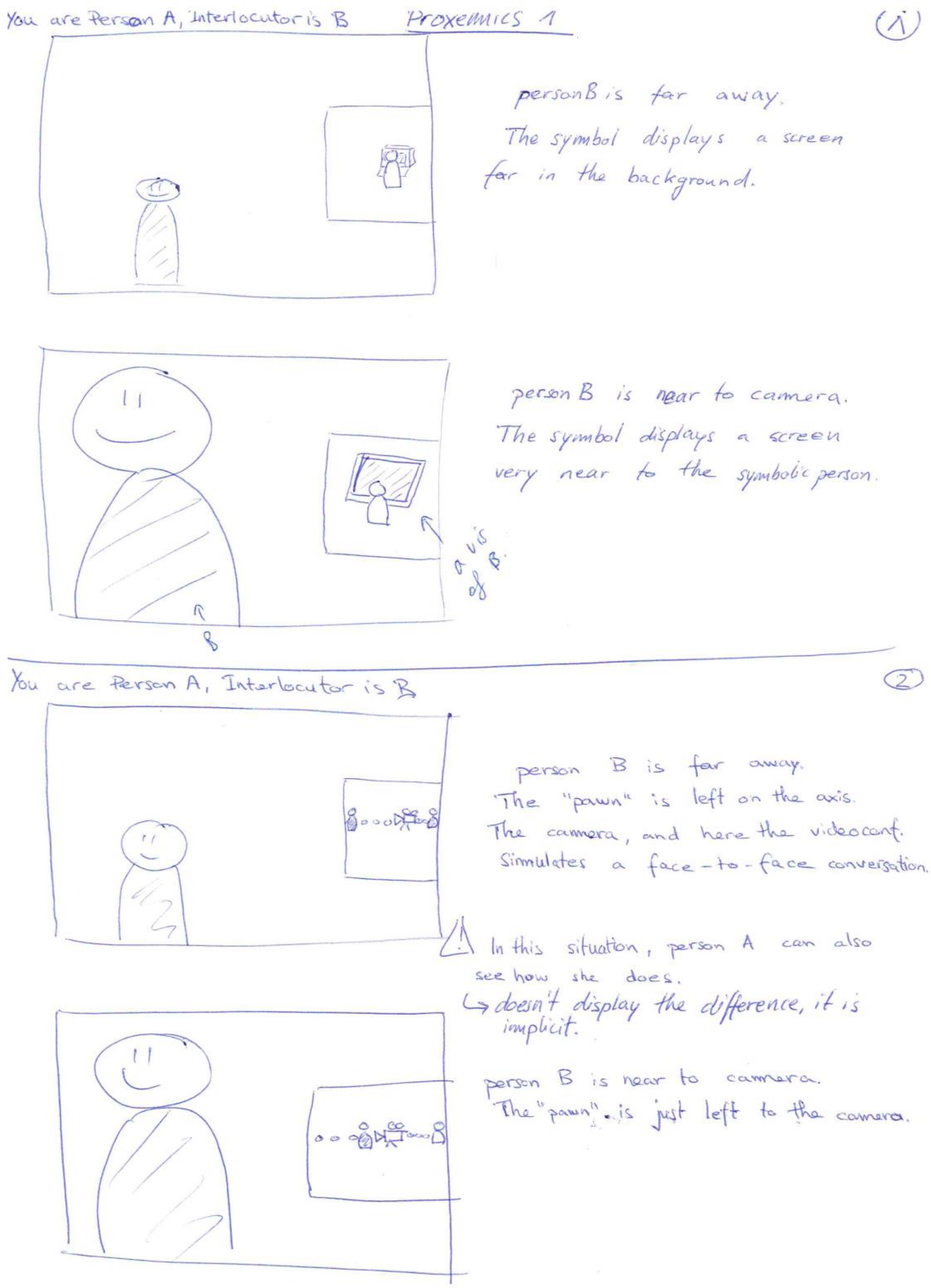
[15] E. T. Hall, "A System for the Notation of Proxemic Behavior1," *American Anthropologist*, vol. 65, no. 5, pp. 1003–1026, Oct. 1963.

[16] M. I. Tanveer, E. Lin, and M. (Ehsan) Hoque, "Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, New York, NY, USA, 2015, pp. 286–295.

Appendices

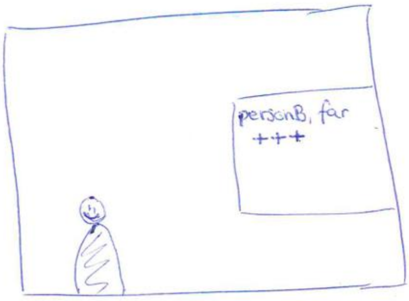
Appendix A – Visualization designs sketches

Initial sketches



Proxemics 2

(3)

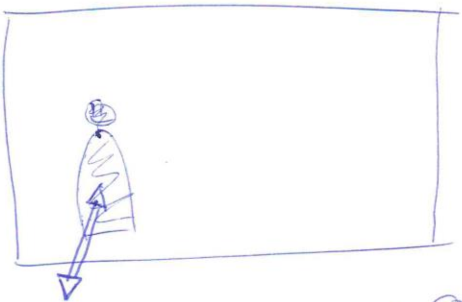


(?) Should we show the difference or only what person B does in case of a difference.
↳ only what person B does seems more intuitive.

text: Scale is much more, more, bit more, equal, bit less, less, much less

---

(4)



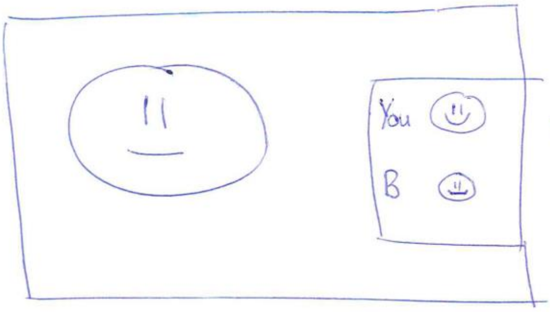
person B far
↳ in scene
3D-arrow:
long and narrow represent distance

→ (?) color?

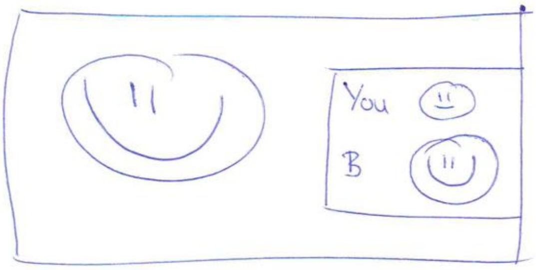
person B near
↳ 3D-arrow:
short and large represent closeness.

Smiling 2

(7)



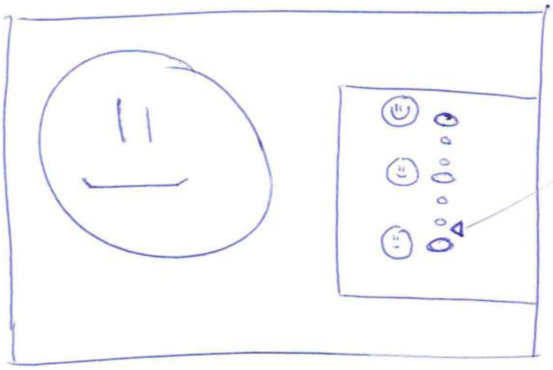
person B is smiling much less.
person B's smiling icon is smaller and less smiling.



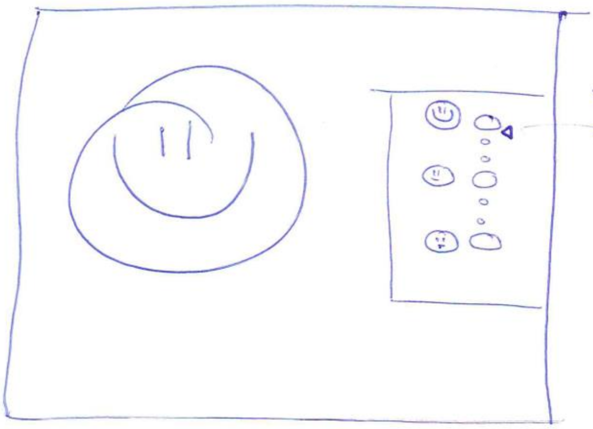
person B is smiling much more.
person B's smiling icon is bigger and more smiling

---

(8)



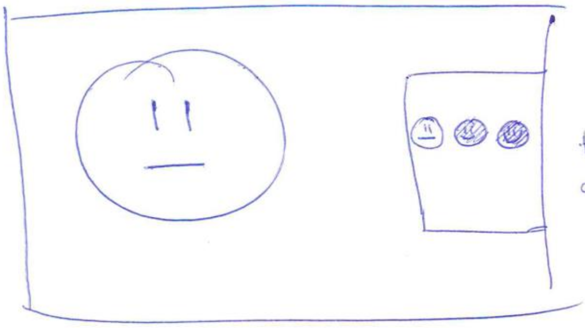
person B smiles much less.
The ◁ cursor is near to the inexpressive icon.



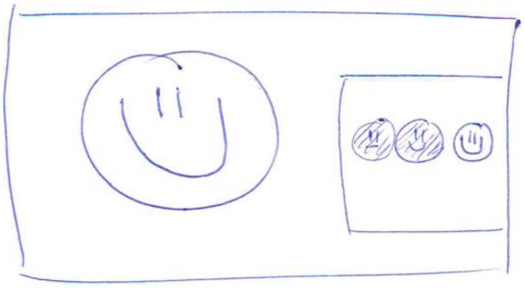
person B smiles much more.
The ◁ cursor is near to the smiling icon.

Smiling1

(5)

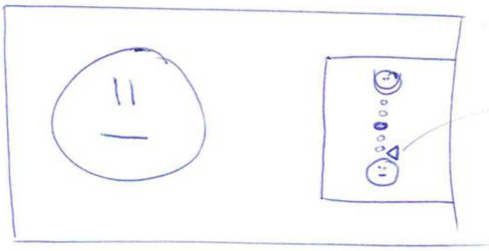


person B is smiling less.
the not-corresponding states are greyed-out.

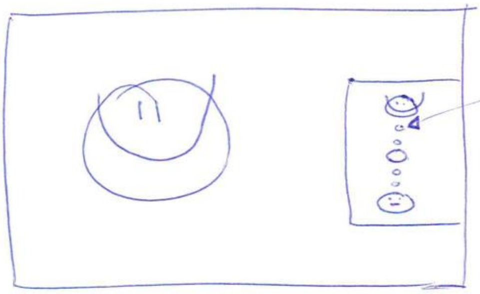


person B is smiling much more.

(6)



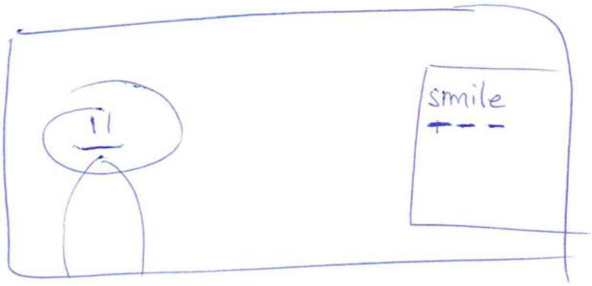
person B is smiling much less.
the ◀ cursor is near to the inexpressive smiley.



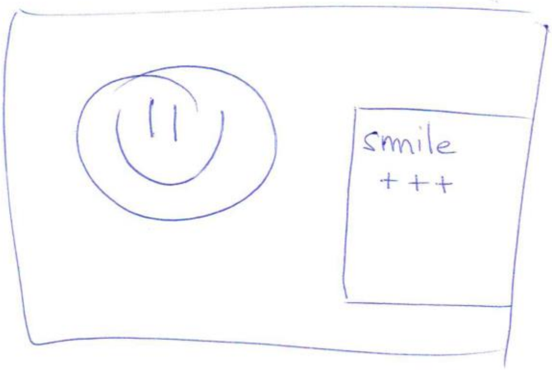
person B is smiling much more.

Smiling 3

(9)



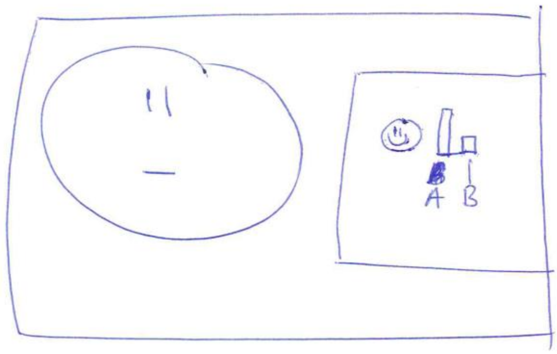
person B smiles much less



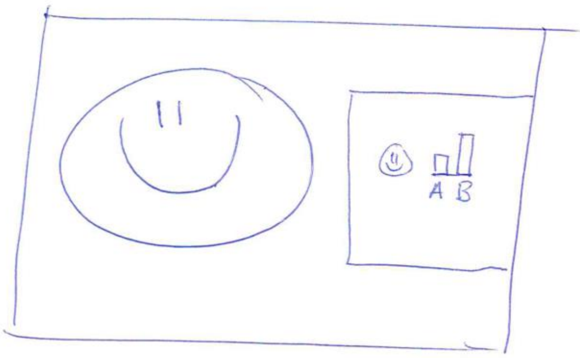
person B smiles much more.

---

(10)

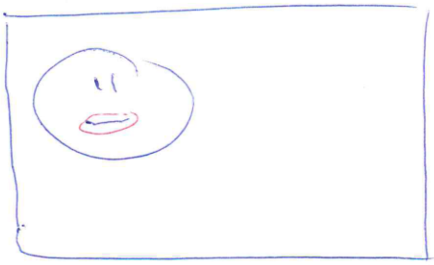


bar graphs show the smiling behaviour

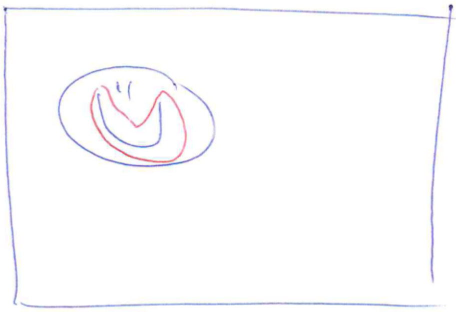


Smiling 4

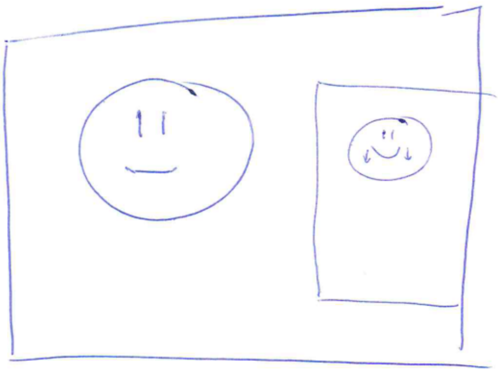
(11)



illuminates the mouth when there is a difference in behaviour.

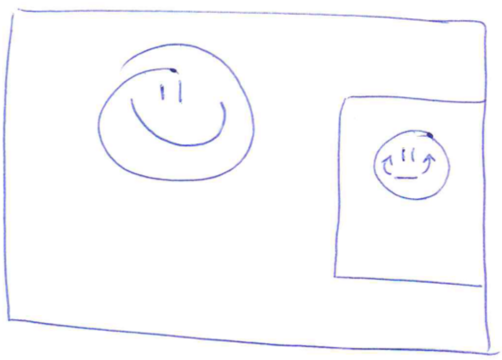


(12)



Person B smiles less.
An animation goes from smiling to inexpressive.

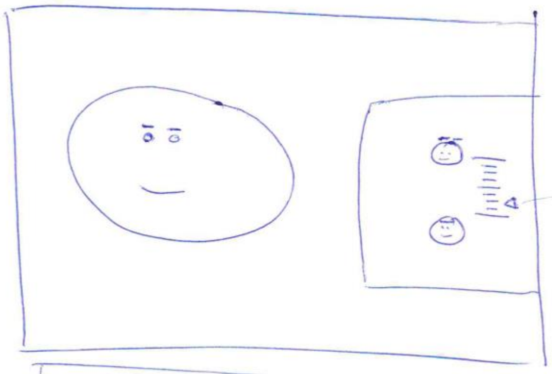
(⚠ this could be interpreted as prescriptive ?)



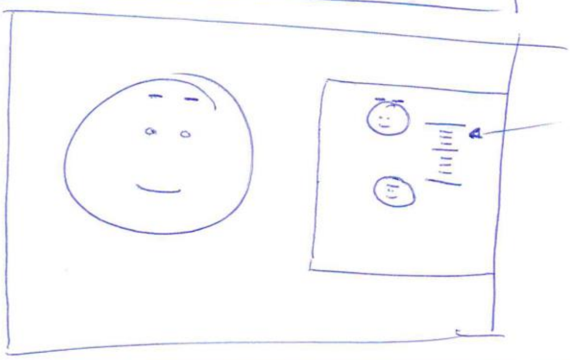
Person B smiles much more.
The animation goes from inexpressive to smiling.

Eyebrows 1

(13)

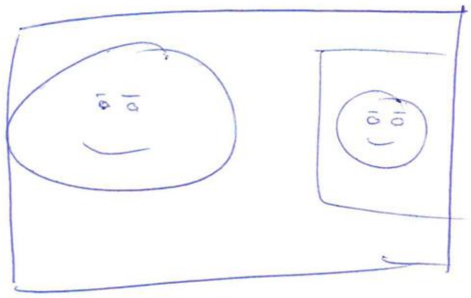


Person B does not raise eyebrows very much. The ◄ cursor is near to the inexpressive icon.

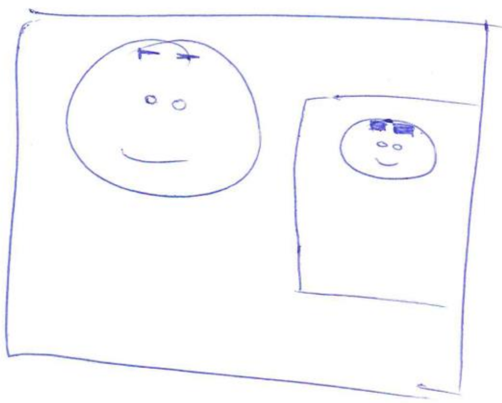


Person B raises eyebrows much more.

(14)



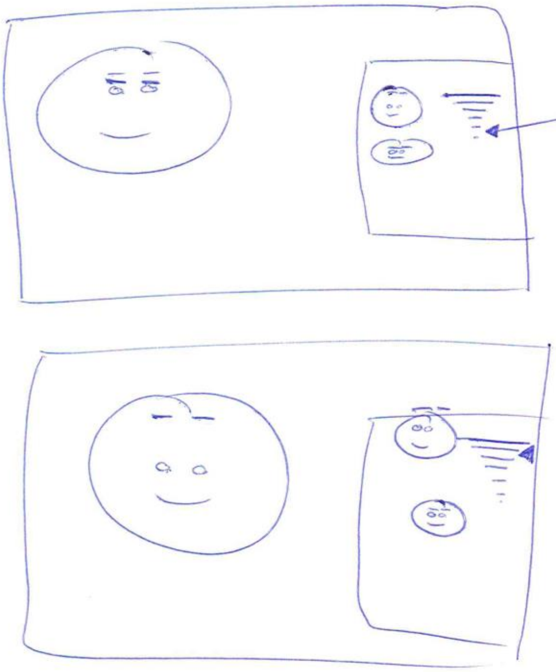
normal



Person B raises eyebrows much more
↳ the figure shows high and big eyebrows

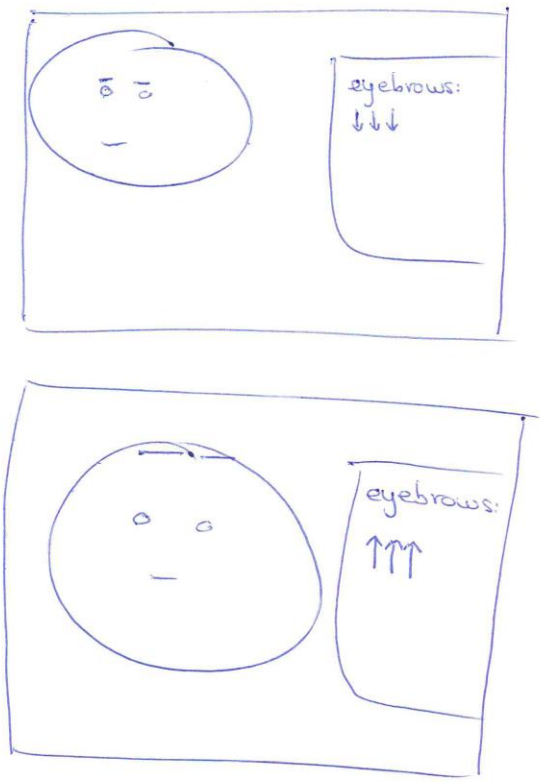
Eyebrows 2

15



Person B lowers eyebrows more
↳ the cursor is near to the low-eyebrows face

16

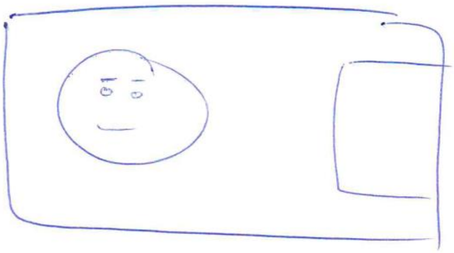


Person B lowers much more: text and arrows

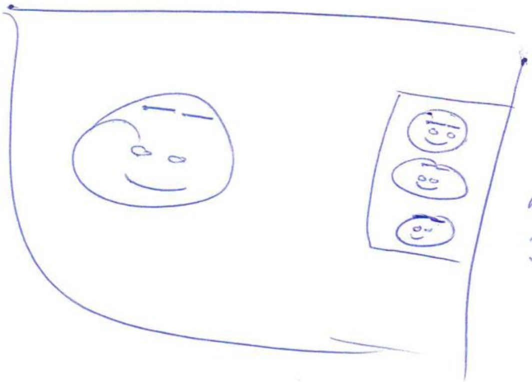
(here it can also be interpreted as a suggestion for mimicking)

Eyebrows 3

(17)



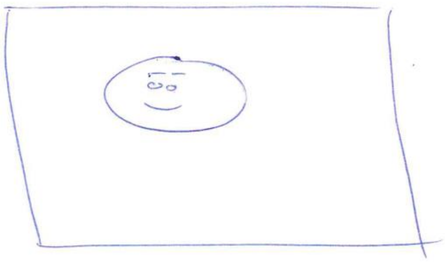
No difference: no display

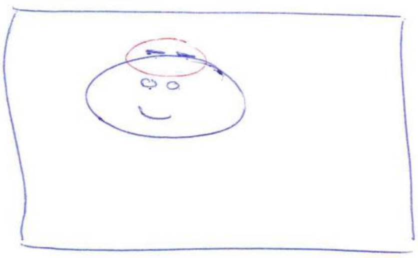


Person B raises eyebrows much more:
3 faces
(2 for more)
(1 for a bit more)

---

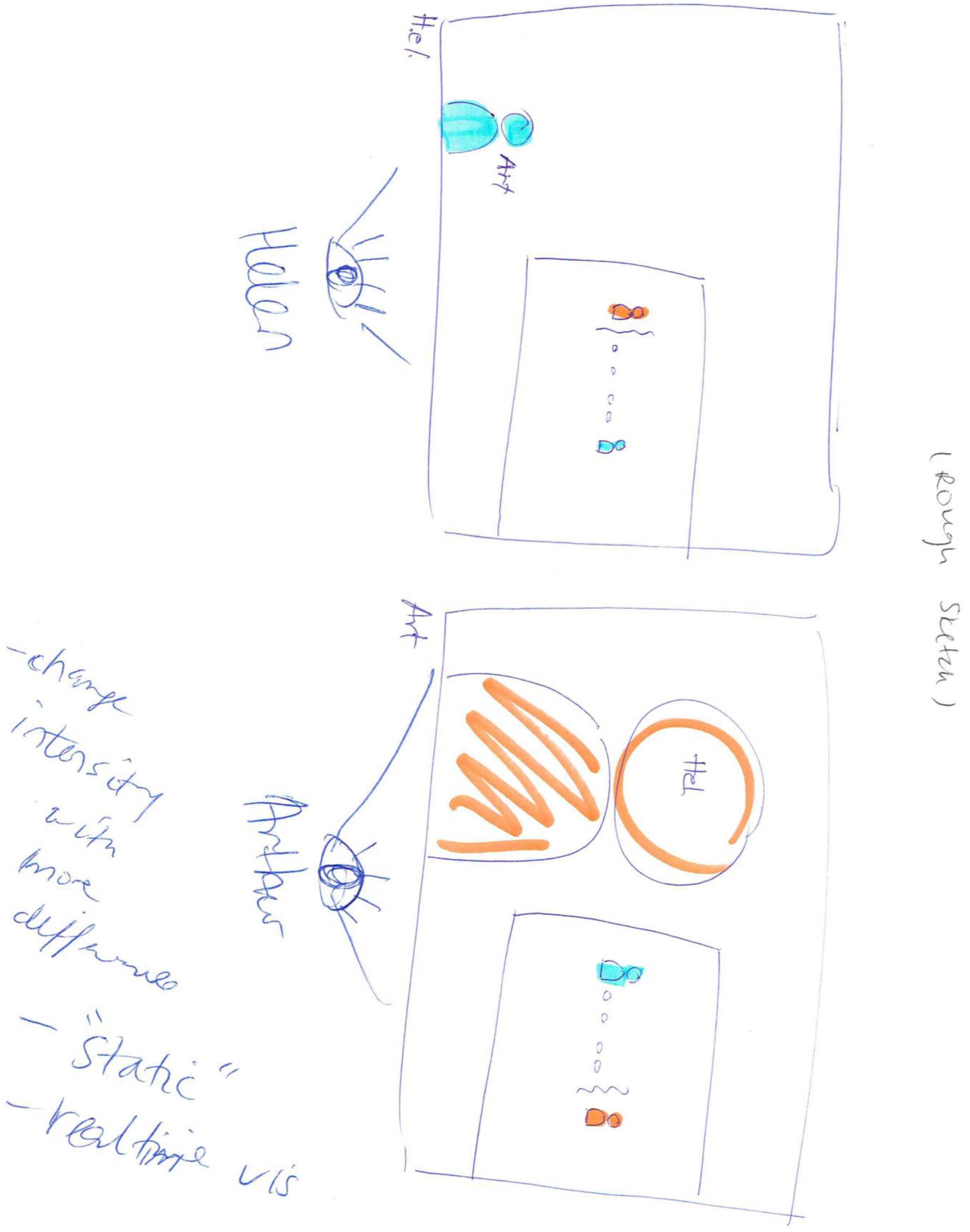
(18)

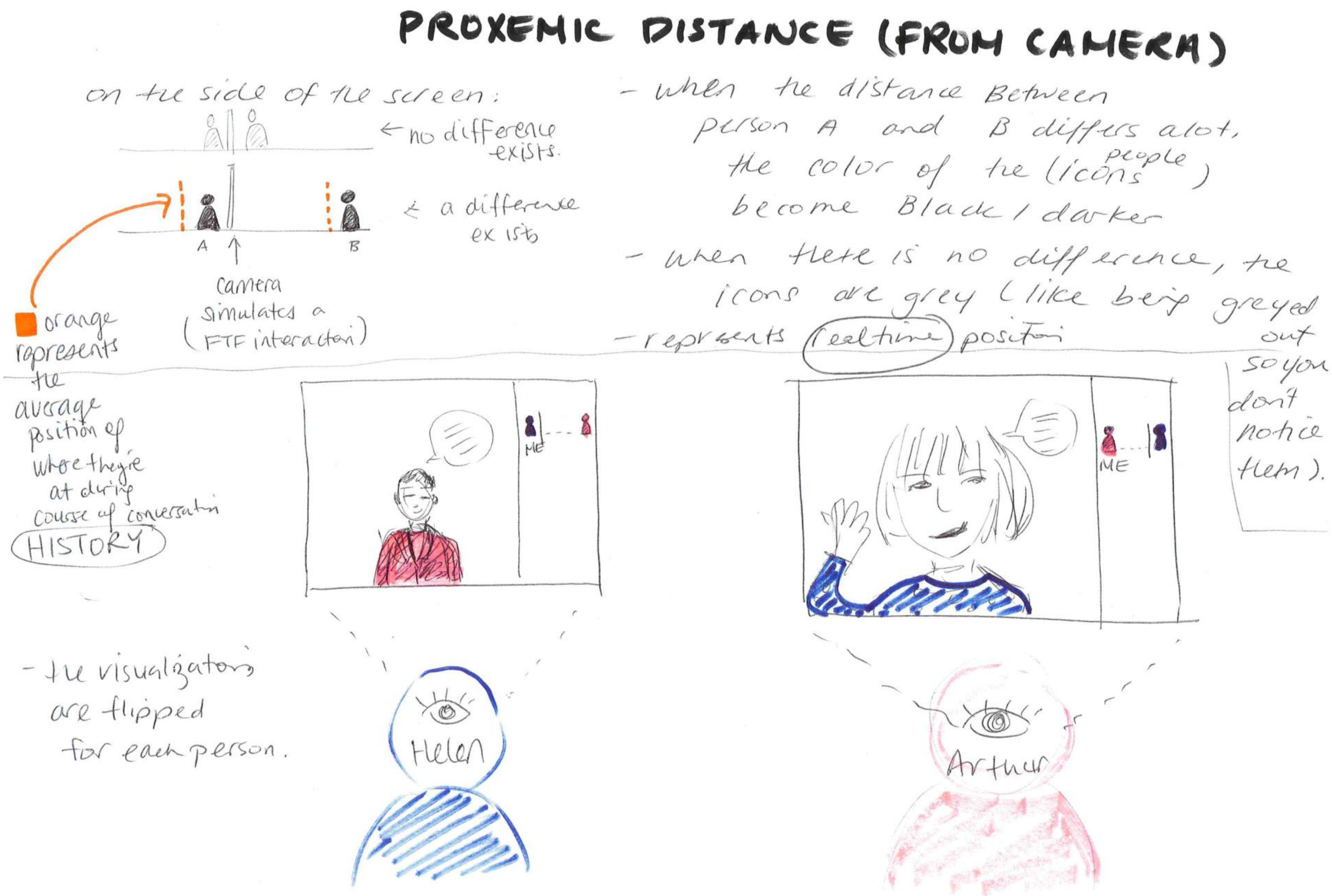




on scene:
Illuminate when there is a difference in behaviour.

Intermediate sketches

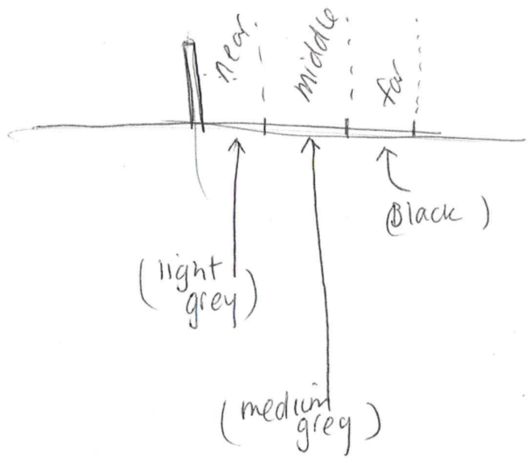



PROXEMIC DISTANCE (FROM CAMERA)
on the side of the screen:
← no difference exists.
← a difference exists
A
B
camera
simulates a
(FTF interaction)
orange represents the average position of where they're at during course of conversation
HISTORY
- when the distance between person A and B differs alot, the color of the (icons people) become Black / darker
- when there is no difference, the icons are grey (like being greyed out so you don't notice them).
- represents realtime position
ME
ME
- the visualizations are flipped for each person.
Helen
Arthur

PROXEMICS

REALTIME

- could calculate the intensity of color based on three distance ranges



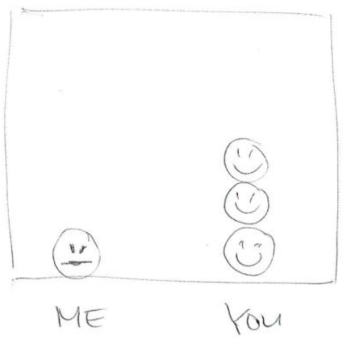
- REALTIME VIS.

- like a slider.
- as people move, ~~they~~ (the icons) move in real time
(so ppl. get feedback right away)

SMILING

- intensity ←  How to show intensity info as well?
- occurence

- show ratio of smiles occurence through Quantity of Heads?

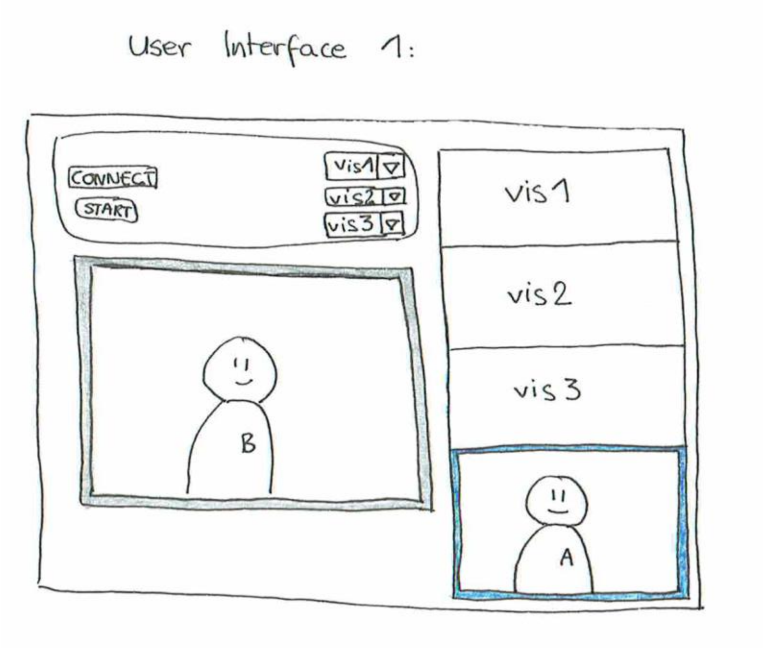


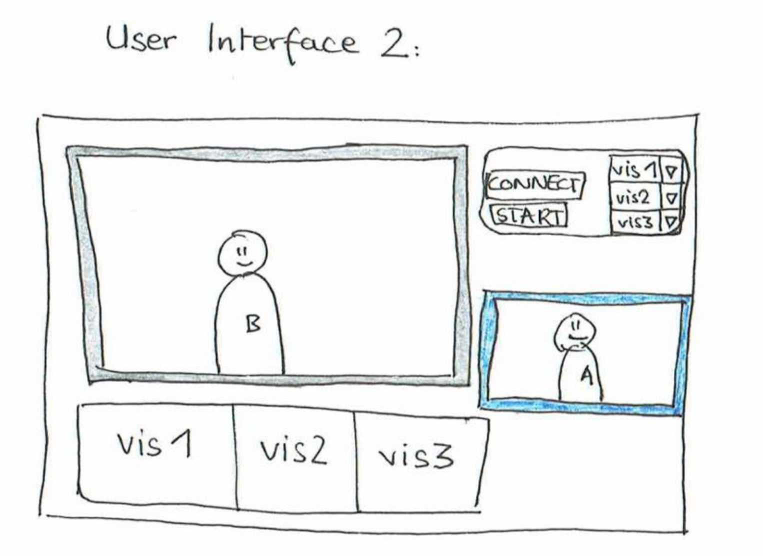
Quantity
- the actual # is not important but rather the ratio.

- Smiling not something you do while ~~[illegible]~~ talking, so realtime in differences of Smiling doesn't make much sense

Appendix B – User interface design

The following are two variations of a user interface for the videoconferencing tool I designed. User Interface 2 is the one I presented to the participants of the pilot study.





Appendix C – Pilot study consent form

Informed Consent Form

A pilot study to explore the design of video conferencing tools to support intercultural communication

The purpose of this study is to gain insight into how computer-supported communication tools can support *intercultural communication* – that is, communication between people of different cultural backgrounds. Specifically, we focus on non-verbal communication, such as distance to the camera, gestures, smiling, etc.
The goal of the video chat tool is to help support awareness of your own non-verbal behaviors in comparison to the non-verbal behaviors of other people.

What will I be asked to do?
If you choose to participate in this study, you will be asked to:

1) watch a video recording of a video chat conversation
2) talk to another person using this video chat tool; and
3) share your feedback in using this tool.
4) share feedback about the study design.

What information will be collected?
You will be asked demographic information about your occupation, cultural background, and your interactions with people from different cultural backgrounds. With your permission, the interview will be audio-recorded and we will take notes during your interview.

Are there risks to participating?
There is no risk to participate in this study, beyond the risks associated with normal everyday activity.

Participation in the study is voluntary and confidential. Your data will be anonymized. If it is ever shared with anyone outside of the research team, including any written publications or oral presentations based on this research, you will be identified only by a participant number (e.g. P12) or a pseudonym of your choosing.

You are free to withdraw your participation at any point during the study, without needing to provide any reasons. However, unless you request otherwise, any information you contribute up to the point at which you choose to withdraw will be retained and may be used in the study.

Uses of the interview data
All of your original data (notes, audio files, photos) will be saved on password-protected devices or locked in university filing cabinets at the University of Zurich. They will be stored for a maximum period of 5 years.

The data can be used and seen by researchers affiliated with this project. The results of this study may appear in both internal and external presentations and publications, as well as academic journals and conference proceedings. In all cases, your data will be anonymized.

University of Zurich[UZH]

Consent

By signing this form, you confirm the following statements:

- A researcher explained the study and the listed conditions to me.
- I had the opportunity to ask questions.
- I understood the answers and accept them.
- I am at least 18 years old.
- I had enough time to make the decision to participate.
- I agree to the participation.

In no way does signing this form waive your legal rights or release the investigators or involved institutions from their legal or professional responsibilities. You are free to withdraw from this research project at any time. Please feel free to ask for clarification or new information at any time during your participation.

___________________________________	_____________________________________
Participant's name (please print)	Researcher's name (please print)
___________________________________	_____________________________________
Location and date	Location and date
__________________________________	_____________________________________
Participant's signature	Researcher's signature

Questions or Concerns?

A copy of this consent form has been given to you to keep. The researcher has kept a copy of the consent form.   If you have further questions regarding our research, and/or your participation in this study, please contact:

Helen Ai He (primary contact)
helen.he@ifi.uzh.ch
University of Zurich

Jagoda Walny
jkwalny@ucalgary.ca
University of Calgary

Arthur Toenz
Arthur.toenz@uzh.ch
University of Zurich

Anthony Tang, Ph.D.
tonyt@ucalgary.ca
University of Calgary

Prof. Dr. Elaine M. Huang, Ph.D.
huang@ifi.uzh.ch
University of Zurich

Sheelagh Carpendale, Ph.D.
sheelagh@cpsc.ucalgary.ca
University of Calgary

Chat Wacharamanotham
chat@acm.org
University of Zurich

Appendix D – Pilot study pre-interview questionnaire

QUESTIONNAIRE

Name ________________________

What is your age bracket? Please circle one.

19 – 25 years old

26 – 35 years old

36 – 45 years old

46 – 55 years old

56 – 65 years old

Over 65

In what city and country were you born?

Other than your birth country, in which countries have you lived in and for how long? Approximately how old were you when you lived there?

Please tell me about your current (or most recent) occupation.

How often do you communicate in *Face-to-Face* with people from different cultural backgrounds? Please select one.

Daily

Weekly

Monthly

Other. Please explain. _______________

How often do you communicate in *Video-chat* with people from different cultural backgrounds? Please select one.

Daily

Weekly

Monthly

Other. Please explain. _______________

Appendix E – Videos used in the pilot study

The videos we presented to the participants of the pilot study are available at the following:
\\samba.ifi.uzh.ch\share\zpac\Research Projects\Helen - Culture research\Arthur Thesis 2015
Note that this is a shared folder internal to the ZPAC lab of the UZH IFI.

Appendix F – Code

The code for the implementation of the system presented in this thesis is located at
https://github.com/Toenza/RealSenseiConfFusion
Note that this is a private repository.