Delft University of Technology

Knowing Me, Knowing AU
How Should We Design Agent-Mediated Mimicry?

Axelsson, Agnes Johanna; Chen, Weilun; Van Sinttruije, Deborah; Lefter, Iulia; Rook, Laurens; Jonker, Catholijn M.; Oertel, Catharine

DOI
10.1145/3715336.3735687

Licence
CC BY

Publication date
2025

Document Version
Final published version

Published in
DIS 2025 - Proceedings of the 2025 ACM Designing Interactive Systems Conference

Citation (APA)
Axelsson, A. J., Chen, W., Van Sinttruije, D., Lefter, I., Rook, L., Jonker, C. M., & Oertel, C. (2025). Knowing Me, Knowing AU: How Should We Design Agent-Mediated Mimicry? In N. J. Nunes, V. Nisi, I. Oakley, Q. Yang, & C. Zheng (Eds.), *DIS 2025 - Proceedings of the 2025 ACM Designing Interactive Systems Conference* (pp. 253-270). (DIS 2025 - Proceedings of the 2025 ACM Designing Interactive Systems Conference). Association for Computing Machinery (ACM). https://doi.org/10.1145/3715336.3735687

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Knowing Me, Knowing AU: How Should We Design Agent-Mediated Mimicry?

Agnes Johanna Axelsson
Delft Technical University
Department of Intelligent Systems
Delft, Netherlands
a.axelsson@tudelft.nl

Weilun Chen
Delft University of Technology
Intelligent Systems
Delft, Netherlands
weilunchen.cs@gmail.com

Deborah van Sinttruije
TU Delft
Intelligent Systems
Delft, Netherlands
d.vansinttruije@tudelft.nl

Iulia Lefter
Delft University of Technology
Multi Actor Systems
Delft, Netherlands
i.lefter@tudelft.nl

Laurens Rook
TU Delft
Delft, Netherlands
l.rook@tudelft.nl

Catholijn M Jonker
Delft University of Technology
Interactive Intelligence
Delft, Netherlands
C.M.Jonker@tudelft.nl

Catharine Oertel
TU Delft
Interactive Systems/Interactive Intelligence
Delft, Netherlands
c.r.m.m.oertel@tudelft.nl

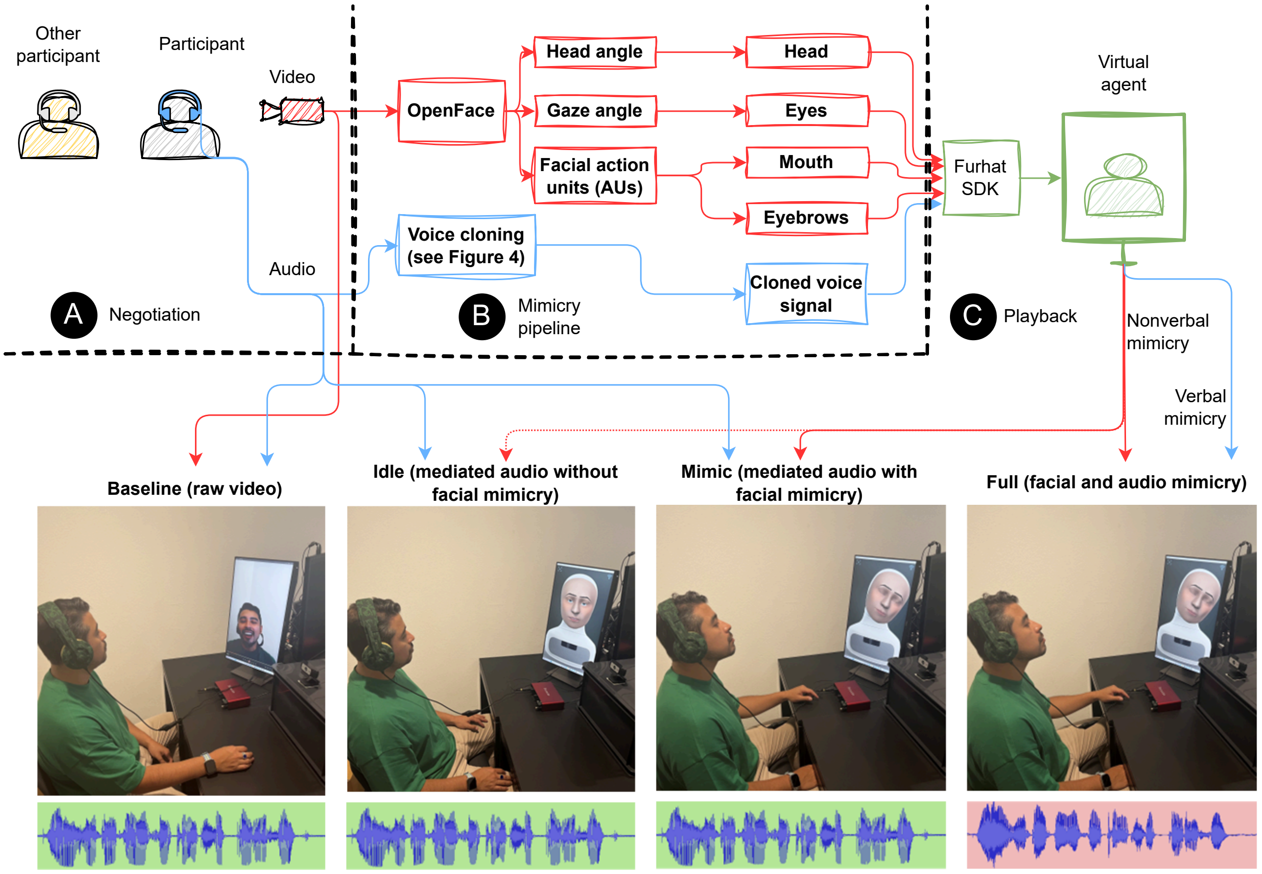

Figure 1: To explore agent-mediated mimicry, we set up a between-subject experiment, with some within-subject analysis. The first three steps of the experiment are illustrated here. In the negotiation (A), two participants perform a natural communication task, during which they are recorded. The mimicry pipeline (B) is applied to the recordings of participants from the negotiation. Participants' mimicked verbal and nonverbal behaviours are played back to them (C) in a manner that depends on the experimental condition (shown on the bottom of the figure). In the *baseline* condition, participants see the raw video without any processing. In the *idle* condition, they see an agent speaking back their audio to them, without any facial mimicry. The *mimic* condition has facial mimicry together with the original audio. The *full* condition has the same facial mimicry as *mimic*, but adds voice mimicry on top. We thus explore a range of mimicry.

Abstract

A lack of self-awareness of communicative behaviours can lead to disadvantages in important interactions. Video recordings as a tool for self-observation have been widely adopted to initiate behaviour change and reflection. Seeing oneself in a recording can lead to negative affect. Forcing an external perspective can lead to cognitive dissonance. Avatars and virtual agents have the advantage that they can copy a human's behaviour while potentially avoiding this dissonance. To explore the design space of mimicking agents, we set up a user study where a video baseline is compared to agent-mediated conditions ranging from idle non-verbal behaviour to complete mimicry of the voice and face. We show that participants gain increased self-awareness from seeing themselves mediated through the virtual agent. We further discuss qualitative observations for the future design of systems that aid in self-reflection, and particularly note that partial mimicry seems to be less appreciated than full mimicry.

This work is licensed under a Creative Commons Attribution 4.0 International License.
DIS '25, Funchal, Portugal
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1485-6/25/07
https://doi.org/10.1145/3715336.3735687

CCS Concepts

• **Computing methodologies** → *Speech recognition*; *Information extraction*; **Cognitive science**; Theory of mind.

Keywords

mimicry, negotiation, self-awareness, video self-confrontation, AUs

ACM Reference Format:
Agnes Johanna Axelsson, Weilun Chen, Deborah van Sinttruije, Iulia Lefter, Laurens Rook, Catholijn M Jonker, and Catharine Oertel. 2025. Knowing Me, Knowing AU: How Should We Design Agent-Mediated Mimicry?. In *Designing Interactive Systems Conference (DIS '25), July 05–09, 2025, Funchal, Portugal.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3715336.3735687

1 Introduction

Individuals are often unaware of their verbal and non-verbal behaviours in communication [11, 30]. Traditionally, a mirror or the playback function of a camera is used to render people aware of particularities in their behavioural repertoire. Considerable evidence exists that self-confrontation by means of mirror and playback interventions can raise someone's self-awareness of displayed emotions [1, 51] and non-verbal behaviours [14, 38, 59, 79] – insights that may eventually lead to behaviour change. Positive effects notwithstanding [14, 38, 59, 79], exposure to mirror and playback interventions could also be detrimental to a person's self-image and, for instance, cause anxiety [59] or decreased self-confidence [31, 39]. Adverse effects are especially documented among vulnerable groups, including people who are low in self-confidence [31]. That is, some people may experience exposure to direct and raw footage as too confrontational or uncomfortable, and this may not only apply to *seeing*, but also to *hearing* oneself [32, 33]. Such forms of sensory distress may be caused by the contrast between the way in which the voice is heard by others vs. how the individual perceives the voice herself [32, 81].

In this paper, we look at self-awareness of one's communicative behaviour when talking to another person. Admittedly, there are other types of self-awareness that can be affected by therapeutical interventions – for example, being self-aware of one's own job performance in nursing [65] or in business [73]. By looking at communicative behaviours, however, we are able to focus our design space to look at a field where communicative agents have an opportunity to help in a novel way.

Given the above, designing interactive systems aimed at increasing self-awareness by means of mirror or playback interventions is a difficult enterprise. There is a strong need for research and development of responsible interventions and applications that are capable of boosting someone's self-awareness, while being void of strong triggers of sensory distress. In the present study, we argue that embodied conversational agents are promising tools for doing so under experimentally controlled circumstances. Embodied conversational agents are interactive system applications, capable of communicating eye-gaze, mouth movements and smiles to the human user. In addition to this, they are capable of audio playback, which allows for morphing and changing how the agent speaks before exposing a human listener to the lip-synced audio [4]. Due to their highly customizable nature – i.e., the wide range of behavioural features that may be fully or partially played back to the human user – embodied conversational agents dramatically open up the range of possibilities to study the effectiveness of mirror and playback interventions for increasing self-awareness, and to explore the effectiveness of design alternatives.

Unlike traditional mirror and recording devices that confront someone with a full-fledged (unfiltered) version of themselves, embodied conversational agents allow for *partial mimicry* of someone's communicative behaviour, and mediated playback of selective aspects of a person's own behaviour. Because systems like this allow us full control of how much of an individual's behaviour we mimic and how we present it, they are a perfect test-bed for exploring the design space of non-verbal playback. We aim to explore how these designs can boost self-awareness; therefore, the present study explores how various degrees of **mimicry** of non-verbal and verbal behaviours impact reported self-awareness and behavioural changes in communication. This is a novel proposition, which, to date, has not been explicitly formulated or put to the test.

It is not a given that less mimicry is less confrontational in itself. Excluding certain features of the presented verbal and non-verbal features may not let our users draw the same conclusions from seeing it. As such, this paper has two research objectives. First, we seek to confirm that exposure to mediated (non-)verbal behaviours has a positive effect on reported self-awareness. Second, we explore the behavioural adjustments people make after receiving feedback about their own conduct. In particular, we are interested in investigating whether exposure to certain levels of (non-)verbal behavioural mimicry can be connected to specific behavioural adjustments by the people confronted with those forms of playback. The outcome of this exploration may have important ramifications for the development of future applications promoting self-awareness, based on mirror and playback interventions. Such interventions may be more efficient depending on playback of selected features, and the design implications of partial mimicry interventions for self-awareness will be addressed in the general discussion in Section 6.

2 Background

In this section, we will provide a review of the literature on self-awareness and exposure to (non-)verbal behaviour. We will first discuss traditional approaches, in which mirrors and video recordings are utilised for self-exposure. Then, we move on to prior work on self-awareness and the voice, which comes with issues relating to listening to one's own voice. Third, we highlight the promise of conversational agents as highly customizable interactive systems, capable of overcoming some of the pitfalls typically associated with traditional forms of mediated feedback. In the final section of the background, we point out how these insights are translated into an abstracted intervention of partial mimicry for the present study.

2.1 Self-awareness and non-verbal feedback

Promoting self-awareness by exposing individuals to their own verbal and non-verbal behaviours does not require an agent to be present. Historically, mirrors were used for such forms of confrontations with the self [43, 71]. In the last 70 years, researchers have primarily resorted to video recordings for providing mediated feedback on, and playback of, one's (non-)verbal manners and habits [14, 38, 59, 79]. Importantly, both mirror-based and video-based approaches simultaneously expose people to themselves and to their own communicative behaviours, notably *without separation*. This renders it challenging to disentangle effects stemming from exposure to particularities in the person's verbal and non-verbal communication, and those deriving from self-exposure per se. Among others, research shows that mirror-based self-exposure can negatively impact someone's expressivity [43], or lead to unhappiness in individuals who find it undesirable to clearly or passionately show their own emotions [71]. A common criticism of mirror-based pedagogy in dance practice, for example, is that it causes students to over-focus on what they see in the external mirror image, while they should also concentrate on how the dance movements feel to them [23].

Video self-confrontation (VSC) is a common technique to reinforce or change behaviour in education settings, because it allows for playback after the depicted behaviour has already taken place [59]. As such, VSC has a major advantage over physical mirrors, which can only reflect events as they are happening. Research shows that VSC interventions are more likely to boost self-awareness than mirror-based interventions [14, 38, 79]. Due to its playback functionality, VSC is useful as a reminder of behaviours that happened in the past, to prevent individuals from selectively remembering only certain parts of their behaviours, or from normalising them [61]. VSC may also enhance an individual's self-awareness with regard to non-verbal behaviours. This benefit to self-awareness is possible, because VSC is particularly suited for rendering people aware of previously unnoticed cues in their body language and behavioural tendencies [50]. This heightened awareness of unintended non-verbal behaviour can potentially help people control undesired motor behaviours (for an overview, see Manwaring and Kovach [50]).

On the other hand, the VSC technique may also have detrimental effects when used incorrectly or with people from specific populations and vulnerable groups. In general, one major downside of confronting people with videos that highlight certain aspects of their own behaviours is that the recordings may ruin someone's self-confidence [31, 39]. Research further shows that VSC can induce anxiety [59]. Negative reactions to seeing oneself in video may interact with cultural and gender factors – i.e., women are traditionally more sensitive to adverse effects of VSC than men, especially in the Western world [37]. Bailenson [9] identified a link between constantly being confronted with oneself, videoconferencing over the Covid-19 pandemic, and why this was so exhausting for many people.

2.2 Self-awareness and the voice

People do not only use their voice for verbal communication (words), but also to communicate nonverbal signals such as pitch, stress, and other more complicated communicative behaviours [82]. Humans learn from birth onwards to recognise their own voice [20, 45]. However, the difference between how sound conducts through bone and how sound conducts through air results in our voice sounding different to ourselves than it does to others [53]. Despite this, people have proven capable of recognising recordings of their own voice over those of others' voices [20]. This seems to suggest that we learn to recognise our voice not only by pitch, but also by features not warped by bone conductivity – like prosody, word choice or amplitude. The speech production systems in the brain that react to the pitch of one's own speech can compensate for small artificial pitch shifts, but will stop reacting for larger pitch shifts [41]. At least in neurotypical brains [6], these systems have an expectation for how the own voice should and can sound.

Self-awareness has traditionally been controlled for in psychological experiments by exposing participants to recordings of their own voice [21]. In a foundational experiment on self-awareness, Holzman and Rousey showed that adults tend to display a negative affective disturbance when listening to recordings of their own voice [32]. Participants in their study argued that the recording of their voice sounded unexpected and unfamiliar to them. Yet, the authors documented that this effect faded with exposure, and was most intense at the start of playing back the recording [32]. In a subsequent study, Holzman et al. [33] showed that the strong negative reaction came back after a break of three months, suggesting that those who react negatively to hearing their own voice cannot overcome the negative reaction without frequent exposure.

Research shows that facial expressions are more controllable than verbal expressions [86]. Zuckerman et al. [86] also argue that the voice reveals more of the speaker's internal state than their face. Correspondingly, children learn to decode emotional cues from voice before they learn to do so from the face [16]. Moreover, liars are more easily spotted by analysing their voice than by reliance on facial expressions [7]. If the voice is harder to control than the face, lasting change in the voice should thus be harder to achieve than lasting change in the use of the face.

2.3 Conversational agents and (self-) awareness

Conversational agents, as well as embodied conversational agents, are increasingly deployed across a wide range of domains to assist people in fulfilling their needs and preferences, and alleviating their problems [44]. Making people more aware of critical habits, patterns and behaviours is a crucial aspect of these applications. Examples include mediated assistance in behaviour change, such as offering decision support on dietary adjustments after a diabetes diagnosis [72], or supporting people to quit smoking [5]. Recent applications also include conversational agents that are designed to use mediated feedback on (non-)verbal behaviour to teach people to become better in emotion regulation [34] or supporting people in the autism spectrum to improve their social skills [75]. In the same way that VSC and mirrors can promote awareness of body posture, interacting with agents that position themselves in ways similar to the human can promote self-awareness in the human [17].

Previous work has shown that we generally prefer conversational agents that display non-verbal behaviours that resemble those we perform ourselves. These similarity effects hold for body language [49], facial expressions [58, 60] and speech features [74, 78]. The similarity effect is even documented for uncontrollable facial features like pupil size alignment between conversation partners as they agree with each other [64]. It follows that a preference of similar features does not necessarily imply that people are *aware* of their own (non-)verbal behaviours – the preference could be subconscious, even for features that are controllable. Indeed, people prefer avatars in virtual reality (VR) that mimic eye and mouth movements to those that do not, even if they can't tell exactly what the difference was between the two [40].

Conversational agent technologies are highly customizable, which dramatically increases the range of possibilities to study and design the impact of exposure to (non-)verbal behaviour on self-awareness. Customization essentially implies that a selective number of aspects of the self (rather than all of those aspects) may be included in the behavioural playback intervention. This is an important improvement over traditional mirror and VSC applications, which – out of technical necessity – are forced to rely on a 100% accurate representation of the (mirrored or recorded) self. Conversational agents, in contrast, allow for **mediation** or **abstraction** of non-verbal feedback – i.e., the filtering out of certain aspects of the self that may trigger undesired effects. Research shows that people, even when mediated non-verbal behaviour is not entirely accurate, still can competently over-express with the features that are present to compensate for those that are missing [42]. This means that mediation of non-verbal expression need not be perfect for effectiveness; users can learn to focus on those behaviours that are there, if they even notice the difference [40]. This has important ramifications for the design and implementation of behavioural interventions for the promotion of self-awareness that are based on partial (mediated) mimicry.

2.4 The present study

Motivated by the theoretically recognised link between self-awareness and exposure to one's (non-)verbal behaviour described above, and the possibilities for customization that are nowadays available in conversational agents, we developed a novel playback intervention, in which the amount of behavioural mimicry of verbal and non-verbal features could be varied. We apply facial mimicry as used by Kimmel et al. [40], Pourebadi and Riek [63], both of which applied facial features captured from a human face onto virtual agents with various methods of distortion. This mimicry is combined with self-awareness measures like those used by Choi et al. [17], who measured users' feelings of closeness to a remote partner via self-monitoring and perceived closeness. These factors combine to create a novel intervention.

Unlike earlier research that hid the fact that the agent was mimicking its user from that user, *we made it explicit* to our participants that the agent was reproducing their communicative behaviours. First, this circumvented issues regarding the nature, motivation and potential artificiality of (non-)verbal gestures observed or changed after exposure to the mimicking agent [66]. Second, this enabled us to treat our mimicry intervention as a mechanism for deliberate feedback delivery. In doing so, we sought to meet two research objectives: (1) replication of the positive association between playback interventions and self-awareness, and (2) exploring what behavioural adjustments people would make after being confronted with mimicry-based playback. A user study was developed, in which we put the degree of verbal and non-verbal behaviour a participant was exposed to under experimental control. This set-up enabled us to gain first insights into which type of mimicry may be most efficient in boosting self-awareness, which behavioural adjustments may be observed, and which specific personal features to which the human reacts negatively are filtered out in those cases.

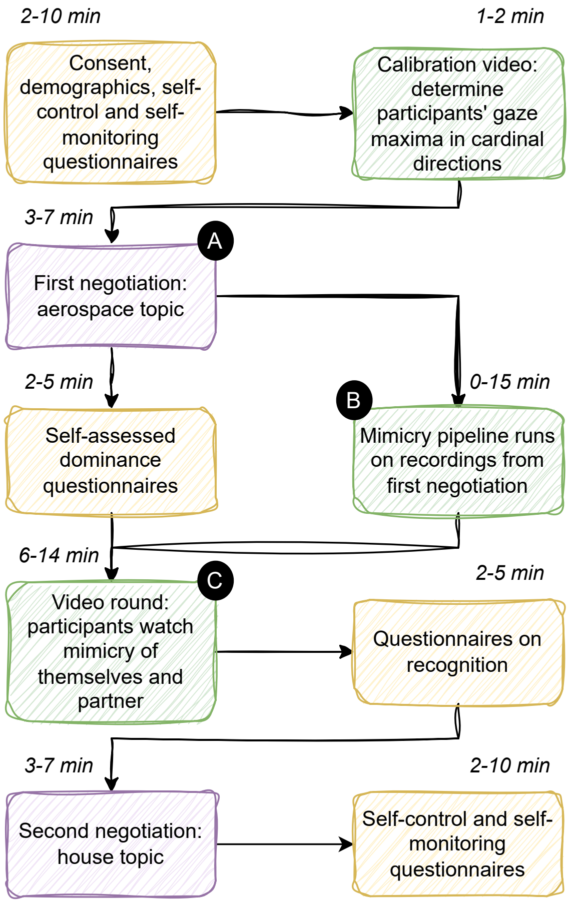
2-10 min
1-2 min
Consent, demographics, self-control and self-monitoring questionnaires
Calibration video: determine participants' gaze maxima in cardinal directions
3-7 min
A
First negotiation: aerospace topic
2-5 min
0-15 min
B
Self-assessed dominance questionnaires
Mimicry pipeline runs on recordings from first negotiation
6-14 min
2-5 min
C
Video round: participants watch mimicry of themselves and partner
Questionnaires on recognition
3-7 min
2-10 min
Second negotiation: house topic
Self-control and self-monitoring questionnaires

Figure 2: An illustration of the steps involved in our experiment and the approximate time taken for each step. The A, B and C labels mark the boxes that correspond to the steps seen in Figure 1.

3 Method

To explore the effect of agent-mediated mimicry on self-awareness, facial expressions and voice usage, we set up a between-subject experiment where participants performed two negotiations on predetermined topics in pairs. After our participants had performed the negotiation task once, but before they negotiated for the second time, they viewed an experimentally controlled recording of their face and voice during the negotiation. There were four conditions, each corresponding to a different level of mediation for the feedback videos the participants would see (see Section 3.2). Our participants then performed a second negotiation after having viewed the video of themselves and their participant. We experimentally measured the effect of having viewed the videos on the second negotiation. As detailed in Section 3.4, we relied on self-report measures filled out by the participants after each negotiation and on behavioural data (changes in face and voice activity) recorded during and compared between the two negotiations per participant.

It is important to emphasise that the negotiation activities served as a bogus task for our actual ambition to observe changes in reported self-awareness and (non-)verbal behaviour due to mimicry-based playback design. That is, the negotiations were supposed to distract participants from paying full attention to their own voice and facial expressions in communication. Using a communicative and meaningful task for this purpose is better for self-awareness after being exposed to VSC than irrelevant tasks like counting or reading text [22].

As part of local ethics procedures, participants had to be informed that they would negotiate and that they would be exposed to a recording, potentially agent-mediated, as part of the experiment. It is possible that participants were affected by knowing that the goal of the experiment was to explore self-awareness of facial and verbal expressions, and that this led them to then behaving in ways different to how they would normally act, distorting the results of our experiment. However, based on how cognitively demanding the negotiation task was, we are satisfied that participants would have had to spend their full attention on talking to their negotiation partner, and could thus not knowingly behave differently from how they would have otherwise.

3.1 Experimental procedure

Ethics approval for working with human subjects was granted by the HREC board of TU Delft with submission number 2412. At the start of the experiment, each pair of participants was led into the laboratory space and invited to take a seat in front of a camera, as illustrated in Figure 3. Participants first filled out a consent form and a questionnaire[1], which contained demographic items and two validated personality measures: the **self-control scale** [76] and the **revised self-monitoring scale** [46]. Self-control taps the ability to regulate one's inner responses (to control undesirable behaviours) [12]. Self-monitoring captures how people modify their self-presentation, and how sensitive they are to expressive behaviours of others [26].

Before the negotiation instructions were handed out, a short **calibration video** was recorded of each participant. In this video, which was typically around 10 seconds long, participants were asked to look forward and gaze in the cardinal and diagonal directions with their eyes only. This data was used to calibrate the gaze aversion of the Furhat agent in the conditions that used the agent. Next, the researcher handed out task instructions[2] for the **first negotiation**. If the participants had any questions after reading the instruction sheets, they were allowed to ask the researcher, who answered practical questions about the form of the negotiation. The first negotiation then started, and the researcher observed from behind a curtain. Negotiations lasted between 82.4 seconds and 495.0 seconds (8 minutes, 15 seconds), with a mean of 283.1 seconds (4 minutes, 43 seconds) and a standard deviation of 82.4 seconds.

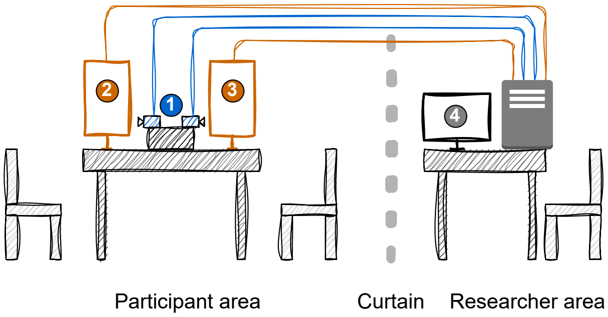

Figure 3: An illustration of the room where we set up our experiment. Two participants sat face-to-face at a table (to the left in the image). The researcher sat behind a curtain (to the right in the image). Ongoing recordings were shown on a monitor (4) so the researcher could confirm that the faces were being properly captured. Cameras (1) captured the participants' faces. Screens (2 & 3) displayed the feedback videos after the negotiation finished, and were empty during the negotiation task. Neither participant could see the other participant's screen.

After the first negotiation, participants were exposed to a **feedback video** of themselves and their opponent. Half of the participants watched their own video first and the one for their opponent second, and vice versa. The form of these two feedback videos depended on the condition, as described in Section 3.2. Participants were asked to state how well they recognised themselves – or their opponent – in the video they had seen, and to express their level of self-awareness after exposure to the feedback video. The **second negotiation** was activated upon completion. The second negotiation was similar to the first one, except that the roles were flipped (i.e., a participant who had played the investor before now became the investee in the negotiation, and vice versa). After the second negotiation task, participants were asked to fill out a final questionnaire (containing the self-control, self-monitoring and miscellaneous questions, see Section 3.4). After filling in these questionnaires, participants were compensated with a €10 gift card and allowed to leave.

3.1.1 *Setup.*

As illustrated in Figure 3, participants sat on chairs facing each other. A webcam was placed on a box in front of the participants, and the webcams were adjusted to capture the participants' faces at the start of the experiment. When the feedback videos were played back, the video played on monitors that were placed on the table in front of the participants. Participants could not see what was on the other participant's monitor, and wore headphones to prevent them from hearing the other participant's audio.

3.1.2 *Negotiation topics.*

The predetermined negotiation topics were loosely inspired on two sheets from the Harvard School of Law's Program on Negotiation (PON). The first negotiation was always on the topic of *Aerospace Investment*, with the second informed by *Bullard Houses*. One participant chosen at random would

[1] The raw questionnaires are included in the supplemental materials.
[2] These instructions are also included in the supplemental materials.

play the investor role for the first negotiation, switching to the other for the second negotiation. Note that the goal of the experiment was not to evaluate the success of the negotiation outcome.

3.2 Conditions

The experimental manipulation involved four conditions, described below. The conditions differed in how the feedback video was shown after the first negotiation was created.

Baseline No mimicry was used, and participants were simply shown the raw video recordings of themselves and their partner – effectively Video Self-Confrontation [59, cf.].

Idle After the first negotiation, participants were shown their own and their partner's audio played back via an agent, but the agent did not mimic facial expressions. However, the agent would play back the basic default Furhat facial expressions, like raising the eyebrows for emphasis, blinking, and performing lip-sync to the audio. It was thus not an entirely static face. This condition can be compared to a similar condition used by Riek et al. [66].

Mimic After the first negotiation, participants were shown an agent that mimicked facial expressions, but which used unmodified audio as its voice.

Full The full facial mimicry from the *mimic* condition was used, but the participant's voice and the partner's voice were also replaced by voice-cloned alternatives as described in Section 4.3.

The recorded videos of our participants negotiating with each other were put through the processing pipeline described in Section 4. Depending on experimental condition, different parts of the mimicry pipeline were disabled; the *baseline* condition had all parts of the pipeline disabled, while the *full* condition had all parts enabled.

3.3 Participants

A total of 128 participants were recruited. Their ages ranged from 18 to 38 years old (M = 25.5, SD = 3.93). Pairs of participants were randomly assigned to one of the four conditions. This resulted in 32 participants in the *mimic* condition (7 F, 25 M), 32 in the *idle* condition (12 F, 19 M, 1 NB), 32 in the *baseline* condition (19 F, 13 M) and 32 in the *full* condition (12 F, 20 M).

3.4 Measures

3.4.1 Personality measures.

Table 1 provides an overview of the bivariate associations between the levels of the experimental condition and participant scores on the two personality measures of **self-control and self-monitoring** as assessed before the experiment. The correlation matrix highlights the following three aspects: First, the conditions of the experiment (idle, mimic, full vs. baseline) are correlated with each other in the same direction – i.e., reflecting that their structural changes in mimicry playback are incremental rather than contrasting away from each other. Second, the self-monitoring scales (overall and their sub-scales) correlate positively with each other ($\rho = 0.245, p = .005$). In connection with the experimental conditions, only one significant negative bivariate correlation is observed: between idle (vs. baseline) and the self-monitoring sub-scale, which taps individual differences in sensitivity to express behaviour of others (Sensitivity, $\rho = -0.179, p = .044$). Third, trait self-control is negatively associated only with the full (vs. baseline) condition of the experiment ($\rho = -0.191, p = .031$). Because self-monitoring and self-control are hardly associated with the conditions of the experiment, it was decided not to include them in any of the follow-up analyses reported in Section 5. For the same reason, the self-monitoring and self-control measures assessed after the experiment were not included in follow-up analyses.

3.4.2 Facial expressions.

As stated in Section 1, we wondered how participants would alter their facial expressions after being exposed to a playback recording of (some of) their own communicative behaviours. To find out, we decided to abstract the facial action units (AUs) into combined emotions using the coefficients suggested by Yan et al. [84]. The linear combinations are approximations of Ekman's basic emotions [24] and a *neutral* category (see appendix). *Neutral* is thought to correlate with several ambiguous mouth expressions, which do not necessarily reflect a "poker face" [77]. As OpenFace did not generate AUs 16, 18, 27 or 28[3], we removed them from our equations. The resulting AU-to-emotion mappings were calculated frame-by-frame and averaged over the full video. This provided us with an average score for the extent to which each individual had expressed such specific facial movements during the respective negotiation rounds.

3.4.3 Voice activity.

In addition to facial expressions, we also wondered how participants would change their voice after exposure to mimicry-based playback. As the effects of hearing one's own voice are related to negative affect [32, 33], we chose to analyse the audio in terms of sentiment. We separated the participants' speech into segments of voice activity with *pyannote.audio* [15, 62]. Each individual segment was then classified for valence, arousal and dominance using the prosodic model from the *AffectToolbox* by Mertes et al. [55]. We chose to not consider the verbal content of the participants' utterances as the positivity and negativity of their word choices would be mixed up with the goals of the negotiation.

4 Implementation

We developed a framework for mimicking individuals' facial expressions and voice through a virtual agent. The steps of the full **mimicry pipeline** are visualised in Figure 1, with the voice mimicry split off into Figure 4.

4.1 Libraries and technology

OpenFace [10] was used to analyze the video footage of our participants' faces, and generate the data that was used to mimic their non-verbal behaviour with the help of a **Furhat** [4] virtual agent. Commands to tell the Furhat agent to make facial expressions or speak audio were sent through the Furhat real-time SDK. For voice mimicry, which was only used in the *full* condition, the **Coqui** library for Python was used, with the default **freevc24** [47] model for synthesis.

[3] Lower lip depressor, lip puckerer, mouth stretch and lip suck, respectively.

Table 1: Descriptive statistics, Spearman rho correlations, and reliability statistics. Cronbach α values for all Self-Monitoring and Self-Control scales, including their sub-scales, are presented in parentheses on the diagonal.

Index	Label	M	SD	1.	2.	3.	4.	5.	6.	7.
1.	Idle vs. Baseline	0.250	0.435	-						
2.	Mimic vs. Baseline	0.258	0.439	$\mathbf{-0.340^{***}}$	-					
3.	Full vs. Baseline	0.242	0.430	$\mathbf{-0.326^{***}}$	$\mathbf{-0.333^{***}}$	-				
4.	SM† (Modify, before)	3.114	0.651	-0.069	0.039	-0.018	(0.760)			
5.	SM† (Sensitive, before)	3.106	0.600	$\mathbf{-0.179^{*}}$	0.004	0.047	$\mathbf{0.351^{***}}$	(0.662)		
6.	SM† (Overall, before)	3.110	0.527	-0.148	0.004	0.041	$\mathbf{0.854^{***}}$	$\mathbf{0.761^{***}}$	(0.782)	
7.	SC* (Before)	3.078	0.519	-0.021	0.091	$\mathbf{-0.191^{*}}$	0.138	0.126	0.149	(0.772)

Note: $N = 128$; * $p < .05$ level, ** $p < .01$ level, *** $p < .001$ level; two-tailed. †: Self-Monitoring. ★: Self-Control.

4.2 Facial Feature Extraction

We used OpenFace [10] for facial landmark detection, Facial Action Unit (AU) extraction, head pose and eye gaze tracking. We captured features related to **head pose** by translating OpenFace's estimated head angles into head angles for the virtual Furhat agent. The same was done for **gaze angles**. The virtual Furhat agent also mimicked participants' **smiles**. We combined AUs 6, 12 and 25[4] to have the mimicking agent smile if AU 6 and AU 12 exceeded a threshold value (regardless of the value of the AU 25 feature), and to open its mouth without a smile if the AU 25 feature was activated on its own. **Eyebrow movements** were mapped by extracting AUs 1, 2, 4 and 9[5] and converting them to the respective features of the Furhat agent.

4.3 Voice cloning

We implemented voice cloning by using the *Coqui* Python library on the sound file recorded from the microphone. Five male voices and five female voices were chosen from the English Mozilla Common Voice corpus[6] and used as targets for the cloning. The voices were chosen to represent young, older as well as lower-pitched and higher-pitched samples for both female and male voices, while avoiding recordings that had background noise or static.

In order to choose a target voice for a participant, the *least* common voice from the same gender[7] as the participant was retrieved. We did this by encoding the participant's voice into a feature vector from the Python *Resemblyzer* library, which implements voice embeddings according to Wan et al. [80]. This vector was then compared to the feature embedding vectors of the candidate voices, and the furthest candidate by cosine distance was selected as the target. This process is illustrated in Figure 4. The least similar voice was chosen to expose the user to the least recognisable voice possible, avoiding self-exposure.

5 Results

5.1 Descriptive statistics: self-recognition

In order to contextualise the mixed model approaches that are used further on, we first wanted to confirm that participants had perceived that the mimicry worked at all. Figure 5 shows box plots for how answers to two questions – both asked in the questionnaire that followed the feedback video – were distributed in every condition.

On top, answers to the statement *"I recognized myself in the feedback video."* are shown. Participants in the *baseline* condition, who saw their own raw recorded video, predictably ranked the highest (M = 5.47, SD = 1.22). Not everyone in the *baseline* condition answered with strongest possible answer ("Strongly agree"), suggesting that recognition was interpreted as more than simply confirming whether the video had practically contained the participant. Predictably, the *mimic* condition (M = 5.28, SD = 1.02) was slightly worse than *baseline*, the *full* condition (M = 4.69, SD = 1.71) was less positive (reflecting the effect of voice mimicry), whereas the *idle* condition (M = 3.03, SD = 1.67) was rated worst. The latter outcome made much sense, given that it had not even attempted to mimic the participants' facial expressions.

The same patterns were observed for the statement *"The feedback video of my opponent is similar to my actual opponent during the negotiation."*, which is visualized at the bottom of Figure 5. Interestingly, the conditions mediated by an agent (M = 3.56, SD = 1.44 for *full*, M = 4.22, SD = 1.21 for *mimic*, M = 2.66, SD = 1.56 for *idle*) were ranked lower on average than those in the self-recognition question, while answers in the *baseline* condition (M = 5.47, SD = 1.52) were ranked higher. We interpreted this as an indication that participants had identified the mimicry they saw in their own and their negotiation partner's non-verbal and verbal expressions.

5.2 Analysis of change in self-awareness

In the questionnaires that were administered after each negotiation, participants filled out their agreement to the statements *"I was self-aware of my non-verbal expressions during the negotiation"* and *"I was self-aware of my expressed emotion during the negotiation"*. Consistent with our first research objective, we conducted repeated measures analyses of variance (RMANOVAs) to test whether the change between these two reported values of self-awareness had been significant in general and as a function of feedback through

[4] Cheeks raised, lip corners pulled and lips parted, respectively [25].

[5] Raising inner, raising outer, lowering brow, and wrinkling the nose, respectively [25].

[6] The voice samples we used as targets are available in the supplemental materials.

[7] For non-binary participants, instead of picking the five voices from the same gender as the participant, the system was set up to use the five closest voices to the participant by the same distance measure. No non-binary participants participated in the *full* condition, so this was not used in practice.

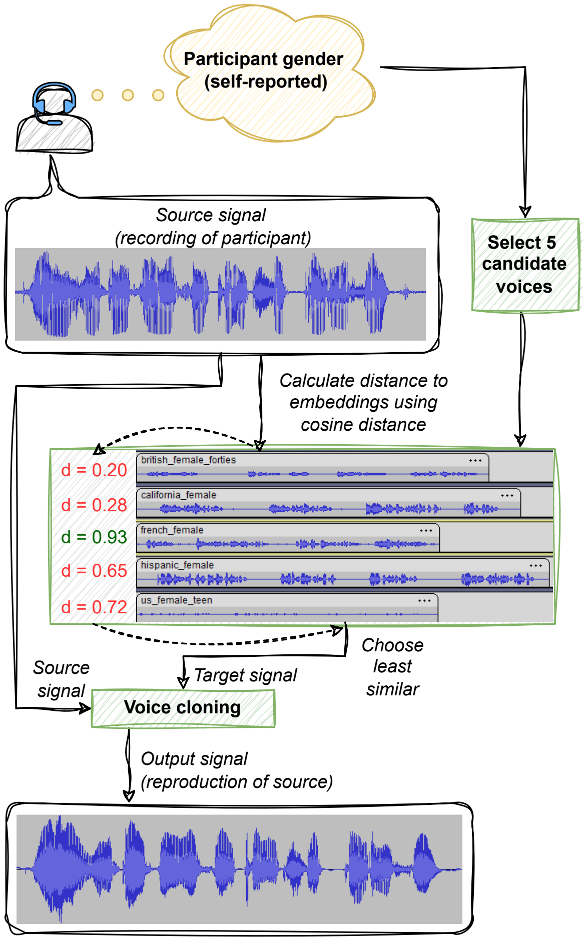

Figure 4: An illustration of the voice cloning process. The source signal is compared to five candidate voices by cosine distance. The least similar is chosen and used as a target for the cloning process, creating the sound wave seen on the bottom. The sound wave illustrated here contains one of the authors acting out the sentence "Yeah, I don't know if I can accept that offer" - note that the timing of syllables is identical in the cloned audio on the bottom and the original audio on top of the figure.

playback – i.e., as a function of the mimicry conditions that had been put under experimental control. Because the independent variable (condition) involved 4 categorical levels, dummy coding was applied. We computed three dummies, for which "*baseline*" (the control group in the experiment) served as the reference category, against which all other categories were compared. This was based on the commonly accepted procedures documented in Aiken et al. [3], Cohen [18].

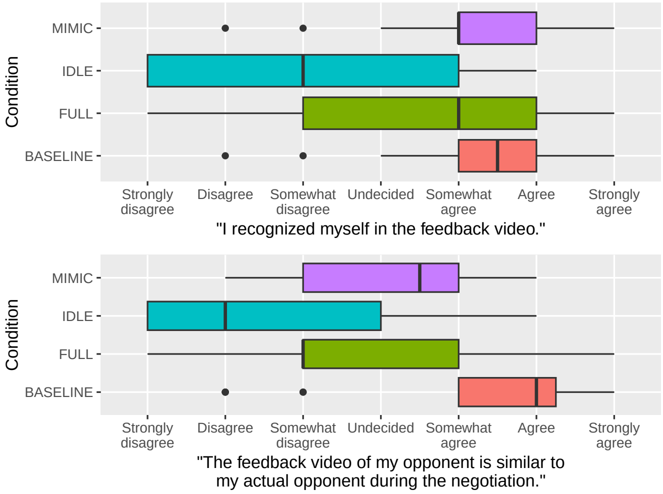

Figure 5: Top: Answers to *"I recognized myself in the feedback video"*, asked after the feedback video was shown to the participants. Participants can be seen to generally recognise themselves outside of the *idle* condition. Bottom: Answers to *"The feedback video of my opponent is similar to my actual opponent during the negotiation"*, also asked after the feedback video was shown. The same patterns as for self-recognition roughly hold, although the agreement is lower for all conditions except *baseline*, where agreement is slightly higher than for self-recognition. Both: For both the top and bottom graph in the figure, means and standard deviations are reported in Section 5.1.

Table 2 presents the within subject effects as well as the between subject effects for the reported changes in self-awareness averaged over the levels of the experimental conditions. We first discuss the output of the within-subject effects, which show a significant change in awareness both of nonverbals and of expressed emotions. Because of the repeated measures ANOVA structure, partial η^2 was used to assess the magnitude of these effects. Based on instructions in Cohen [19], Miles and Shevlin [56], we interpreted $\eta_p^2 < .06$ as "small", $\eta_p^2 \geq .06$ as "medium", and $\eta_p^2 \geq .14$ as a "large"effect size. Accordingly, the observed change in awareness of nonverbals was of medium effect size; the change in awareness of expressed emotions qualified as small effect size. Post hoc comparisons revealed a significant mean difference ($M = -0.984, SE = 0.184, t = -5.344, pbonf < .001, pholm < .001$) of awareness of nonverbals between the two measurement occasions (after the second negotiation minus after the first negotiation). The self-reported change in self-awareness was higher for nonverbal expressions after the second negotiation. Also for awareness of expressed emotions, a significant mean difference between the two measurement occasions was revealed ($M = -0.344, SE = 0.163, t = -2.107, pbonf < .036, pholm < .036$). Again, the change in awareness of expressed emotions was higher when measured after the second negotiation. These effects are illustrated in Figure 6.

Table 2 further displays these within-subjects effects as a function of experimental condition. For nonverbal expressions, the change was significant in all conditions. Relying again on the criteria listed in Cohen [19], Miles and Shevlin [56], the effect size for *full* vs. *baseline* was of medium magnitude, and the remaining effect sizes were small. Post hoc tests confirm significant changes in self-awareness for each possible combination of non-baseline condition vs. *baseline* and each measurement occasion (all $pholm < .05$). The exception is the change between *baseline* and *idle* within occasion 2, which is not significant. However, this specific post hoc comparison does not capture the hypothesized change in awareness *between* occasion 1 and 2, and thus can be ignored. Similarly, significant positive changes in self-awareness of nonverbals are observed for all comparisons of *mimic* vs. *baseline* over the two measurement occasions (all $pholm < .003$). Only the comparisons between *baseline* and *mimic* within occasion 1 and within occasion 2 are not significant. Again, these post hoc tests do not capture the change *between* the two occasions, and should be disregarded.

Finally, the RMANOVA shows a significant positive change in self-awareness of expressed emotions for *full* vs. *baseline* only, which had to be considered small in effect size. Post hoc tests suggest that this effect may be attributed to an increase in self-awareness of expressed emotions that took place for participants in the *full* condition over time ($M = -0.672, SE = 0.258, t = -2.605, ptukey = .048, pbonf < .058, pholm = .058$). Note that this simple effect is significant when Tukey's test is applied, but trending towards significance when other procedures (Bonferroni, Holm) are chosen. All other post hoc comparisons between full and baseline and measurement occasions are non-significant. This is illustrated in the bottom two plots of Figure 6.

5.3 Mixed model analysis of facial emotions

We used a multilevel generalized linear mixed model (GLMM) approach to model the effect of the mimicry-based playback conditions on the participants' usage of facial action units. We resorted to this class of multilevel techniques, because the facial action units were non-normally distributed variables – which implied that the regular RMANOVA procedure (for normally distributed variables) could not be applied. It should be noted that there is a strong connection between the two approaches: the GLMM can be understood as the non-parametric alternative to the RMANOVA procedure. Interestingly, GLMMs do not require empirical (logarithmic) transformation of the outcome variable; they contain a link function that may account for non-normal error distributions [35]. In our case, we defined the data by the identity link function. Because dummy coding applies to non-parametric multilevel applications [2], we used the same dummy variables as before. Negative dummy estimates represent a positive effect, and positive dummy estimates represent a negative effect.

Table 3 shows the estimates and fit statistics for the generalized linear mixed models of the experimental conditions on the 7 AU-to-Emotion mappings for anger, disgust, fear, happiness, surprise, sadness, and neutral that were derived from Yan et al. [84] and introduced in Section 3.4. The fit statistics (deviance, log likelihood, AIC and BIC) indicated good model fit of the fixed effects for most models, but also point to growing model misfit for some (happiness and neutral). Parametric bootstrapping (with 2000 samples) was applied to ensure that parameter estimates and standard errors were robust to outliers. The effect size per multilevel model is reported using the intraclass correlation ρ, which must be understood as tapping the proportion of variance explained at the multilevel component [35]. The intraclass correlation ρ is derived from the intercept and the variance of residuals at the random part: $\rho = \sigma^2_{\upsilon 0}$ / $(\sigma^2_{\upsilon 0} + \sigma^2_{\varepsilon})$. With the exception of fear and happiness, the AU-to-Emotion mappings explain large amounts of variance at the higher level (all > .30), which is considered high for this type of social science-inspired data (see Hox et al. [35, p. 14]).

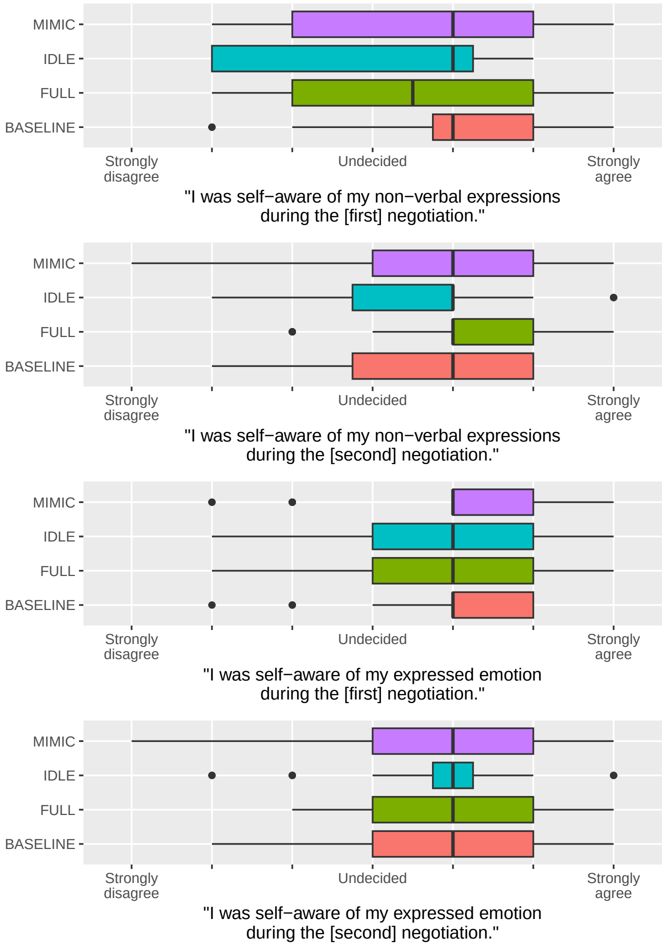

Figure 6: Top: Awareness of non-verbal expressions during the first negotiation. Second from top: The same for the second negotiation. Second from bottom: Awareness of expressed emotions during the first negotiation. Bottom: The same for the second negotiation. Note that this is between-subject, unlike Table 2.

Results show that the average strength of fear ($t = 3.089; \chi^2(1) = 9.491, p = .002, p(bootstrap) = .002$), happiness ($t = 2.865; \chi^2(1) = 8.169, p = .004, p(bootstrap) = .005$) and neutral emotion ($t =$

Table 2: Repeated Measures ANOVAs for change in awareness.
Top: The within-subject effects and between subject effects are presented of reported change in awareness of non-verbal expressions. Bottom: The within-subject effects and between subject effects are presented of reported change in awareness of expressed emotions.

	Sum of Squares	df	Mean Square	F	p	η_p^2
Change in Awareness of Non-Verbals						
(Within-Subject Effects)						
Non-Verbal (NV) Change	31.008	1	31.008	28.560	**< 0.001*****	0.102
Idle vs. Baseline x NV Change	7.563	1	7.563	6.966	**0.009****	0.027
Mimic vs. Baseline x NV Change	9.766	1	9.766	8.995	**0.003****	0.034
Full vs. Baseline x NV Change	25.000	1	25.000	23.027	**< 0.001*****	0.084
Residuals	273.594	252	1.086			
(Between Subject Effects)						
Idle vs. Baseline	7.562	1	7.562	2.580	0.109	0.010
Mimic vs. Baseline	0.141	1	0.141	0.048	0.827	1.904×10^{-4}
Full vs. Baseline	1.972×10^{-29}	1	1.972×10^{-29}	6.729×10^{-30}	1.000	2.670×10^{-32}
Residuals	738.594	252	2.931			
Change in Awareness of Expressed Emotions						
(Within-Subject Effects)						
Expressed Emotions (EE) Change	3.781	1	3.781	4.441	**0.036***	0.017
Idle vs. Baseline x EE Change	0.766	1	0.766	0.899	0.344	0.004
Mimic vs. Baseline x EE Change	1.562	1	1.562	1.835	0.177	0.007
Full vs. Baseline x EE Change	6.891	1	6.891	8.093	**0.005****	0.031
Residuals	214.562	252	0.851			
(Between Subject Effects)						
Idle vs. Baseline	0.391	1	0.391	0.178	0.674	7.045×10^{-4}
Mimic vs. Baseline	0.562	1	0.562	0.256	0.613	0.001
Full vs. Baseline	0.016	1	0.016	0.007	0.933	2.820×10^{-5}
Residuals	554.062	252	2.199			

Note. Type III Sum of Squares. $***$: $p < 0.001$; $**$: $p < 0.01$; $*$: $p < 0.05$.

$2.094; \chi^2(1) = 4.443, p = .035, p(bootstrap) = .038$) was significantly lower after receiving feedback in the *full* condition than in *baseline*. The average strength of sadness was higher after receiving feedback in *full* ($t = -2.959; \chi^2(1) = 8.735, p < .003, p(bootstrap) = .005$) than in *baseline*, but lower after receiving feedback in *mimic* ($t = 1.990; \chi^2(1) = 4.024, p = .045, p(bootstrap) = .047$). Less surprise was observed after receiving feedback in *mimic* than in *baseline* ($t = 2.851; \chi^2(1) = 8.128, p = .004, p(bootstrap) = .005$). Note that the latter effects derive from models with "participant" as the single grouping factor. The models for sadness and surprise are equivalent to a regular RMANOVA output (they represent fixed effects in the panel), but generated with non-parametric multilevel techniques [35].

5.4 Analysis of voice changes

As detailed in Section 3.4, we analyzed changes in the voice of our participants by classifying each segment of voice activity by sentiment. Each participant had a different number of voice clips to classify, depending on how often and for how long they spoke.

Based on previous research by Holzman and Rousey [32], Holzman et al. [33], we would expect to see lower *valence* in the second session than in the first for those participants who heard their own voice in the treatment (*baseline*, *idle* and *mimic* conditions). We compared each condition to the baseline to explore if this was the case in our data.

Table 4 presents the estimates and fit statistics for the mixed model of the experimental conditions on change in valence per measurement occasion. Results show a significant drop in valence for participants in *idle* ($t = -2.319; F(1, 95.047) = 5.376, p = .023$) and *mimic* ($t = -2.420; F(1, 114.965) = 5.855, p = .017$), compared to the *baseline* condition. No such drop in valence is observed in the *full* condition ($t = -1.236; F(1, 182.774) = 1.527, ns.$). The variance of residuals (σ_ε^2), intercepts (σ_{v0}^2), covariance of slopes and intercepts (σ_{v1}^2), and the variance of slopes (σ_{v2}^2) are all significant (all $p < .001$; and $p = .025$ for σ_{v2}^2), confirming the multilevel nature of these effects as nested in segment and measurement occasion for each participant. The associated intraclass correlations ρ_2 and ρ_3 for segment and measurement occasion are 0.500 for both of the nested levels.

Table 3: Generalized linear mixed model estimates for AU-to-Emotion mappings with fixed effects and random effects grouping factors.
A series of generalized linear mixed models were fitted for the fixed effects of the experimental conditions on 7 emotion mappings from Yan et al. [84] ("anger", "disgust", "fear", "happiness", "surprise", "sadness", and "neutral") with Gaussian family and identity link function. Model terms were tested with likelihood ratio tests. "Participant" and "occasion" were used as random effect grouping factors for all emotions, except for "sadness" and "surprise". For those emotions, the variance/correlation estimates were zero when "occasion" was considered. Their results are shown with "participant" as the single grouping factor. For each model, parameter estimates are provided with the associated standard error in parentheses.

	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral
Fixed Part							
(Intercept)	**0.782***** (0.046)	**0.947***** (0.064)	**0.797***** (0.053)	**1.165***** (0.108)	**0.859***** (0.050)	**0.858***** (0.032)	**0.100***** (0.009)
Idle vs. Baseline	0.020 (0.033)	0.029 (0.045)	−0.019 (0.032)	−0.053 (0.064)	0.062 (0.035)	0.035 (0.023)	0.002 (0.006)
Mimic vs. Baseline	0.030 (0.033)	0.039 (0.045)	0.008 (0.032)	−0.021 (0.064)	**0.070*** (0.035)	**0.065**** (0.023)	0.001 (0.006)
Full vs. Baseline	−0.018 (0.033)	−0.064 (0.045)	**0.099**** (0.032)	**0.182**** (0.064)	**−0.104**** (0.035)	0.015 (0.023)	**0.012*** (0.006)
Random Part							
σ^2_{ε}	0.017 (0.129)	0.022 (0.150)	6.021 (0.145)	1.054 (0.232)	0.054 (0.258)	0.032 (0.180)	7.262×10^{-4} (0.027)
$\sigma^2_{\upsilon 0}$	0.059 (0.244)	0.119 (0.345)	0.056 (0.236)	0.231 (0.481)	0.026 (0.162)	0.017 (0.130)	0.002 (0.043)
$\sigma^2_{\upsilon 1}$	1.309×10^{-5} (0.004)	1.112×10^{-4} (0.011)	0.001 (0.039)	0.007 (0.085)			6.882×10^{-6} (0.003)
Deviance	-35.080	83.190	-6.465	283.500	42.270	-39.950	-859.900
log Likelihood	17.540	-41.590	3.233	-141.700	-21.130	19.980	430.000
df	7.000	7.000	7.000	7.000	6.000	6.000	7.000
AIC	-21.08	97.190	7.535	297.500	54.270	-27.950	-845.900
BIC	-3.733	122.000	32.350	322.300	75.540	-6.683	-821.100
ρ	0.776	0.844	9.215×10^{-4}	0.180	0.325	0.347	0.734
nObs	256	256	256	256	256	256	256
nParticipants	128	128	128	128	128	128	128
nOccasions	2	2	2	2			2

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

5.5 Qualitative analysis of open-ended question on takeaways from mimicry

After the second negotiation, participants were given a chance to answer the question **Please explain briefly what you have learned regarding your non-verbal facial expressions**. Responses to this question were generally brief and not suitable for a full-scale thematic analysis, but we have presented select quotes from all four conditions below to illustrate how participants perceived the experiment.

5.5.1 Responses in the mimic and baseline conditions. In the *baseline* and *mimic* conditions, several participants expressed surprise that they had looked down to read the negotiation instructions in front of them more than they had recalled before seeing the feedback video.

- "looking down (to my paper) looks less confident. I should look more at my opponent." (participant in *mimic* condition)
- "I tend to look down a lot to read the instructions. Besides I keep a neutral/friendly mimic." (participant in *mimic* condition)

Others directly stated that they felt that looking down was a way to avoid eye contact with their partner:

- "I realized during the 1st experiment I did not make a lot eye-contacts with my opponent. And I realized it from the feedback video, So I tried to make more eye-contact in the 2nd one." (participant in *mimic* condition)
- "I avoid eye contact. I feel unsure. I need to read even during the interview [sic]." (participant in *baseline* condition)

Several participants in the *baseline* condition, who had seen their own unedited video footage, expressed surprise at seeing how much they smiled or laughed:

Table 4: Mixed model estimates for voice changes with fixed effects and random effects grouping factors.
A mixed model was fitted for the fixed effects of the experimental conditions on voice change (in valence) over time. Model terms were tested with likelihood ratio tests. Data from 2 participants and 168 observations were removed due to missing values. "Participant", "occasion" and "segment" were used as random effect grouping factors. Parameter estimates are provided with the associated standard error in parentheses.

	Valence
Fixed Part	
(Intercept)	−0.012
	(0.015)
Idle vs. Baseline (I-B)	0.036
	(0.021)
Mimic vs. Baseline (M-B)	−0.015
	(0.021)
Full vs. Baseline (F-B)	0.001
	(0.022)
I-B x Occasion	**−0.039***
	(0.017)
M-B x Occasion	**−0.044***
	(0.018)
F-B x Occasion	−0.023
	(0.019)
Random Part	
σ^2_{ε}	$\mathbf{0.046^{***}}$
	(0.001)
σ^2_{v0}	$\mathbf{0.005^{***}}$
	(0.001)
σ^2_{v1}	$\mathbf{0.005^{***}}$
	(0.001)
σ^2_{v2}	$\mathbf{2.086 \times 10^{-6*}}$
	(9.310×10^{-7})
Deviance	
log Likelihood	-1265.209
df	11.000
AIC	-1243.209
BIC	-1168.137
ρ_2	0.500
ρ_3	0.500
nObs	6800
nParticipants	126
nSegments	94
nOccasions	2

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

- "I laughed quite often during negotiation which may have some influence on the whole negotiation process. (maybe i should act more [unclear word])"
- "Looking people straight in the eyes seems to give me a more dominant/confident image which helps during the negotiation. But yes I smile too much. I can not control that."
- "I laugh more than I thought. More reading from paper. active interest by eye contact."
- "I tend to look happy and laugh during serious moments."

5.5.2 Unexpected responses in the full and idle conditions. We chose to look at answers where participants in the *full* condition responded by talking about their voice, and answers where participants in the *idle* condition talked about their face, as we were surprised that these responses occurred in the first place.

It appears that participants in the *full* condition had no problem thinking of the mimicked voice as their own, and even attributed features like volume or mumbling to themselves when hearing it in the mimicked voice:

- "To be honest, I think I was more aware of my verbal expressions, because I think I mumbled too much. The furhat showed minimal facial expressions, so I did not know how to adjust to that, and so I haven't. [...]"
- "After watching the recording I have learnt to control my non-verbal facial expressions (partially). And also I should learn to sound more secure in English."
- "I move my head too much, and speak in gaps. In future, I shall keep this in mind."

Similarly, some participants in the *idle* condition expressed that they had learned something about their facial expressions from viewing the feedback video. Although it was not communicated to the participants, the *idle* condition did not attempt to mimic facial expressions, so any lack of facial expressions in the feedback video was unrelated to the individual's facial expressions:

- "I smile a lot and raise my eyebrows, more than I thought I did."
- "I make less facial expressions than expected."

Other participants in the *idle* condition made specific comments about aspects of their facial expressions that they had reflected on and changed during the negotiations. Participants were thus able to change and reflect on their facial expressions even without having seen them in the mimicry. This can be compared to recent results by Kimmel et al. [40], in whose experiment less than half of the participants noticed that a virtual agent was showing facial expressions.

- "I do not have the same expressions as the one shown by furhat after the first negotiation. I have a lot of mimic expressions and my eyebrows move a lot."
- "I know I smile when I am feeling nervous/uncertain of what I am saying. Watching the furhat did not make me more aware of it, I just tried to smile less during the second part. [...]"
- "I look in every direction when I am uncomfortable."[8]

6 Discussion

In Section 1, we stated that we would (i) seek confirmation that exposure to mediated behaviours had an effect on self-awareness, and (ii) explore how participants would change their behaviour in response to the system.

[8] The agent in the *idle* condition did not avert its gaze from the center, so this must be a self-observation unrelated to the feedback video.

6.1 Confirmation that self-awareness did improve

In Section 5.2, we confirmed that engaging in repeated negotiations enhanced our participants' self-awareness. Participants expressed higher self-awareness of their non-verbal expressions after having participated in our experiment, regardless of the experimental condition, but **the effect was stronger in every mediated condition than the baseline**. This **confirms that self-awareness increased even in the mediated conditions**. Similar effects were observed for awareness of expressed emotions, but the results were less pronounced than those for awareness of non-verbal expressions. This suggests that the combination of mediated facial expressions and mediated, mimicked, voice led to the improved awareness of emotions, while facial mimicry on its own could not be confirmed to have such an effect.

These findings are encouraging, because they show that we can design systems that leverage the positive effects of exposure to the self, while mitigating the negative effects of self-exposure. Mediating communicative behaviour can retain the benefits to self-awareness that video self-confrontation can promote [59, 67], while also introducing a cognitive distance [70] to the observed behaviours. An external perspective, separated by time or distance, is useful for self-reflecting on mental illness [70], negative memories [8] or other people's experiences [27]. Mediation allows our users to take a similar view on their own communicative behaviours.

Mediation is appropriate for those scenarios where system designers aim to avoid negative reactions to the largest possible extent. In carefully created contexts where exposure to a negative aspect is part of the self-awareness that the designers want to promote (e.g. VSC for schizophrenia self-awareness, see Schandrin et al. [70]), mediation is probably not a good idea.

6.2 The nature of change in facial expressions

To analyse non-verbal behaviour, we investigated how participants utilised non-verbal expressions in the two negotiations. In the mixed model analysis presented in Table 3, we confirmed that various changes in the face occurred after participants viewed the feedback video, and that these changes were dependent on the experimental condition.

In the *mimic* condition, we see a decrease of expressed sadness and surprise with the face compared to *baseline*. We stress that facial analysis using linear combinations of action unit activations cannot be assumed to be an expression of an internal state. However, since our participants were using their facial expressions towards their partner in the negotiation task, this does capture a change in displayed facial expressions.

In the *full* condition, participants expressed more sadness and less fear, less happiness and less neutrality than in the *baseline* condition. In terms of emotion communicated to the negotiation partner, these changes in the face can be interpreted as an overall reduction of expressed *valence*. Participants learned through participating in the experiment – and being exposed to simultaneous voice and face mimicry – to express less intensity with their faces. This specific finding might well be due to the use-case of negotiation, where it might be advantageous to show less intense facial expressions. In future work, it would be interesting to investigate whether the same effect could also be observed across a range of different use cases.

In the *idle* condition, we see no significant differences compared to *baseline*, as expected based on this condition not revealing any facial expressions for the participants to change. In human-robot interaction, *emotional contagion* [29] can happen such that a human mimics the emotions displayed by a robot (or believed to be displayed by the robot) [85]. Such contagion is also connected to the expressivity and human-likeness of the agent [52]. Agents with less expressivity (like our *idle* condition) result in less change in the face – but in our case, *emotional contagion* is not an appropriate term, as the user is (a) viewing a representation of their own emotions and expressions, and (b) presumably looking for behaviours to *avoid* rather than mimic.

An advantage of video self-confrontation (VSC) over feedback from friends, therapists or colleagues is that the evidence is objective [67] – the behaviour has been caught on video and individuals can notice it themselves [59]. However, as stated by Perlberg [59], this feeling of objectivity assumes that the video footage is truly representative of the situation – that it is not capturing an angle that highlights or hides things, and that details in the surrounding context of what was happening around the footage are not lost [59]. Our mediated approach hides even more context from the individual than raw VSC. This can have pedagogical advantages – for example, a system can choose to mimic the parts of the face that are relevant to what the participant wants to see, and not other, distracting features. This makes it less objective than VSC, and VSC is already non-objective if taken away from the context of the footage. Agent-mediated mimicry is thus more efficient if one wants to selectively mimic parts of a behaviour to focus on it, or filter out unpleasant cues that take away from the point that the feedback is trying to make.

6.3 The nature of change in the voice

We looked at the valence of our participants' voice after being exposed to the mimicry. The results show that the valence of participants' voices was lower in the second negotiation than in the first only when comparing the *idle* and *mimic* conditions to the *baseline* condition. While this confirms that participants did change the use of their voice as a function of experimental treatment, the difference between conditions is not in line with what we would expect from prior research [32, 33].

One interpretation may be that participants preferred the artificial face when combined with an artificial voice, but reacted negatively to the combination of an artificial face (whether static, as in the *idle* condition, or mimicking, as in the *mimic* condition) and their natural voice. Perhaps, the mismatch between hearing ourselves while seeing someone else was unusual and potentially uncomfortable. This finding is essential for system designers to keep in mind when designing reflection evoking systems as it strongly as it implies that **partial mimicry is worse than no mimicry at all**. Prior research by Anolli and Ciceri [7], Zuckerman et al. [86] suggested that the face is more controllable than the voice and that emotional states estimated from the voice thus are a better indication of a person's internal state than from the face. The only condition where we saw both changes in the face and the affect of the voice was the *mimic* condition, where we saw a reduction of valence in the voice while the sadness and surprise that was expressed in the face was reduced compared to the *baseline* condition. The reduction of sadness and surprise in the face may have been attempts to mask emotions rather than reflections of an internal emotional change or state.

Technology has made it much easier to be exposed to one's own voice since the late 1960s, when Holzman and Rousey [32], Holzman et al. [33] showed that exposure to one's own voice had negative affective results. One explanation for the absence of negative effects in our *baseline* condition may, therefore, have been that our participants, mostly in their early-to-mid 20s, were not that bothered by hearing their own voice as they are exposed to it regularly – especially through modern video sharing tools like Snapchat or TikTok.

6.4 Limitations

Our work is based on the results of a single study, which may be considered a first limitation. In general, conclusions stemming from multiple studies are favoured over findings that derive from a single study. Even though this is very true for results that are based on p-values, it may be less of a problem for studies, such as this one, for which the bootstrapped results and effect sizes are reported. This is, because the magnitude of an effect, measured as an effect size, in one study is assumed to generalise to other populations [35]. Still, it remains to be seen to what extent the findings reported here hold in other settings. Second, and related, the participants in our study were undergraduate students. Undergraduates must be considered more "tech-savvy" than older populations. The possibility exists that older demographics would have interacted differently with the virtual agent, perhaps by experiencing technology-related psychological distress [28].

Third, there was a discrepancy between experimental conditions regarding the time it took for the mimicry-based playback video to be generated. This was due to the amount of behavioural mimicry that had to be generated, which was less in conditions with no or partial mimicry. Even though this may have had some impact on participants, we believe to have adequately controlled for it by comparing all three agent-mediated conditions via dummy coding against the baseline condition, in which participants had less of a wait. This makes us confident that our effects were, indeed, explained by variance in behavioural mimicry.

Fourth and finally, we relied on AU-to-emotion mappings to calculate differences in our participants' faces, which is standard practice in computer science. Yet, there is growing conceptual disagreement about the universality of emotions [68] in psychological science. Among others, an outwardly shown emotion may not represent an internally experienced emotion [69]. Notwithstanding this, we can safely claim that the facial expressions operationalised in the present study captured a physical change in the use of the facial muscles, and that these changes depended on the amount of mimicry induced in the playback intervention.

6.5 Future work

In recent years, multimodal interaction technologies have emerged with the purpose to alleviate the needs and problems of vulnerable groups in society via emotion sensing and recognition capacity [44]. Research could seek implementation of mimicry-based playback interventions in future multimodal interaction technology applications. Research shows, for instance, that people with social anxiety disorder tend to suffer from maladaptive emotion regulation under stressful circumstances characterised by social interactions with other people [83]. It makes a lot of sense for future work to begin designing new interactive systems, aimed at gradual improvement of the communication and social skills of such vulnerable members of society via agent-mediated playback interventions. More in general, research could focus on interactive training, coaching and personal development programs for under-represented demographics in society. Possible applications would range from behavioural mimicry-based negotiation training for young women in the early stages of career [13] to mid-life women who wish to overcome issues with self-representation in the physical or online world [57], to self-awareness training for individuals who stutter [48], or learning how one's communication patterns come across in stressful situations where communication modalities may be limited [54].

Another interesting avenue of future research is related to the appearance of the agent itself. Currently, a white face is the default face of most embodied conversational agents. Future work could explore the extent to which participants under study may feel represented by this default mode, or rather prefer some other neutral or non-human skin colour. One way of doing so would be to grant participants control over how they want the agent to represent them. Such research would have to solve the practical problem how to balance exposure to features that the participant finds recognisable with exposure to features that help the participant perceive facial expressions as if coming from someone else.

Finding the right balance will likely depend on the individual being mimicked, on the goal of the mimicry, and on whether the mimicry is intended only for the individual themselves, for the individual and others, or exclusively for others [36]. This is largely uncharted territory, but it should not remain unconsidered, given that features like skin colour are part of an individual's identity.

6.5.1 Choosing what (not) to mimic. In the present study, we confirmed that people consider themselves sufficiently represented by a mediating agent to achieve increased self-awareness and to encourage voice change. The exact nature of why an individual chooses to change something about their expression, or decides to pay attention to a specific feature in mimicry were beyond the scope of our research. Yet, it would be interesting to explore in future research **what** to mediate when doing any kind of translation from human communicative behaviour to an avatar. There is a real need to investigate this issue, given that people tend to display a tendency to perceive social cues where they may not exist, and to even overinterpret behaviours that do exist. In the present study, we observed this phenomenon among participants who claimed that they perceived facial expressions in the *idle* condition, while no such thing had been programmed into the agent. Likewise, social VR users cited in Kukshinov et al. [42] reportedly thought that all kinds of emotional communication were possible in VR, because the voice was sufficient. It is important to not depend too heavily on the participant when choosing what to mimic or not when reproducing that user's facial expressions or voice, but to arrive at a mediated common understanding instead of what they wish to avoid, what they wish to or what they should be exposed to.

7 Conclusion

In this study, we explored the impact of virtual agent-mediated playback on people's (non-)verbal behaviours. We found that exposure to agent-mediated mimicry increases people's awareness of the emotions and non-verbal particularities they display. We also observed that people tend to use this feedback to change the use of their face and voice. Mimicry-based playback interventions, therefore, seem to render people aware of potentially unwanted communicative behaviours and even give rise to behaviour change. Our work should be seen as a first step towards the design of highly customizable behavioural mimicry interventions for interactive training, coaching and personal development applications. Future design projects should take into consideration that agents can effectively invoke reflection on features users recognise from themselves, without the need to rely on negative feelings that stem from full self-exposure.

Acknowledgments

The authors would like to thank Luis Loaiza Pacheco for providing crucial assistance in running the *full* condition. We also thank Luis for posing in the teaser figure, Figure 1. The authors would also like to thank Senthil Chandrasegaran for valuable feedback on how to frame the work. AA was funded by *Technology Assisted Self-Management: Preventing Relapse and Crisis by the Severe Mentally Ill Themselves* (with project number KICH1.GZ03.21.002) of the research programme KIC-MISSIE 2021, in turn partly financed by the Dutch Research Council (NWO). This work was also in part funded by the Hybrid Intelligence Centre, funded through the Netherlands Organisation for Scientific Research (NWO), https://hybrid-intelligence-centre.nl (024.004.022).

References

[1] Robert S Adler, Benson Rosen, and Elliot M Silverstein. 1998. Emotions in negotiation: How to manage fear and anger. *Negotiation journal* 14, 2 (1998), 161–179.

[2] Leona S Aiken, Stephen A Mistler, Stefany Coxe, and Stephen G West. 2015. Analyzing count variables in individuals and groups: Single level and multilevel models. *Group Processes & Intergroup Relations* 18, 3 (2015), 290–314.

[3] Leona S Aiken, Stephen G West, and Raymond R Reno. 1991. *Multiple regression: Testing and interpreting interactions.* Sage Publications, Thousand Oaks, CA, USA.

[4] Samer al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction. In *Cognitive Behavioural Systems*, Anna Esposito, Antonietta M. Esposito, Alessandro Vinciarelli, Rüdiger Hoffmann, and Vincent C. Müller (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 114–130.

[5] Nele Albers, Mark A. Neerincx, Nadyne L. Aretz, Mahira Ali, Arsen Ekinci, and Willem-Paul Brinkman. 2023. Attitudes Toward a Virtual Smoking Cessation Coach: Relationship and Willingness to Continue. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 13832 LNCS (2023), 265 – 274. doi:10.1007/978-3-031-30933-5_17

[6] Andréia Rauber Ana P. Pinheiro, Neguine Rezaii and Margaret Niznikiewicz. 2016. Is this my voice or yours? The role of emotion and acoustic quality in self-other voice discrimination in schizophrenia. *Cognitive Neuropsychiatry* 21, 4 (2016), 335–353. doi:10.1080/13546805.2016.1208611 arXiv:https://doi.org/10.1080/13546805.2016.1208611 PMID: 27454152.

[7] Luigi Anolli and Rita Ciceri. 1997. The Voice of Deception: Vocal Strategies of Naive and Able Liars. *Journal of Nonverbal Behavior* 21, 4 (01 Dec 1997), 259–284. doi:10.1023/A:1024916214403

[8] Özlem Ayduk and Ethan Kross. 2010. From a distance: Implications of spontaneous self-distancing for adaptive self-reflection. *Journal of Personality and Social Psychology* 98, 5 (2010), 809–829. doi:10.1037/a0019205

[9] Jeremy N. Bailenson. 2021. Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior* 2, 1 (feb 23 2021), 5 pages. https://tmb.apaopen.org/pub/nonverbal-overload.

[10] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, IEEE, New York, USA, 59–66.

[11] Carol L. Barr and Robert E. Kleck. 1995. Self-Other Perception of the Intensity of Facial Expressions of Emotion: Do We Know What We Show? *Journal of Personality and Social Psychology* 68, 4 (1995), 608 – 618. doi:10.1037/0022-3514.68.4.608

[12] Roy F. Baumeister, Todd F. Heatherton, and Dianne M. Tice. 1994. *Losing control: How and why people fail at self-regulation.* Academic Press, San Diego, CA, US. xi, 307–xi, 307 pages.

[13] Katja Bouman, Iulia Lefter, Laurens Rook, Catharine Oertel, Catholijn Jonker, and Frances Brazier. 2022. The need for a female perspective in designing agent-based negotiation support. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, New York, NY, USA, 1–8.

[14] G Nicholas Braucht. 1970. Immediate effects of self-confrontation on the self-concept. *Journal of consulting and clinical psychology* 35, 1p1 (1970), 95.

[15] Hervé Bredin. 2023. pyannote.audio 2.1 speaker diarization pipeline: principle, benchmark, and recipe. In *Proc. INTERSPEECH 2023*. ISCA, Grenoble, France, 1983–1987. doi:10.21437/Interspeech.2023-105

[16] Albert J. Caron, Rose F. Caron, and Darla J. MacLean. 1988. Infant Discrimination of Naturalistic Emotional Expressions: The Role of Face and Voice. *Child Development* 59, 3 (1988), 604–616. http://www.jstor.org/stable/1130560

[17] Mina Choi, Rachel Kornfield, Leila Takayama, and Bilge Mutlu. 2017. Movement Matters: Effects of Motion and Mimicry on Perception of Similarity and Closeness in Robot-Mediated Communication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 325–335. doi:10.1145/3025453.3025734

[18] Jacob Cohen. 1968. Multiple regression as a general data-analytic system. *Psychological bulletin* 70 (1968), 426–443.

[19] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences.* Routledge, London, UK.

[20] Tatiana Conde, Óscar F. Gonçalves, and Ana P. Pinheiro. 2018. Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. *International Journal of Psychophysiology* 133 (2018), 66–78. doi:10.1016/j.ijpsycho.2018.08.007

[21] Edward Diener and Mark Wallbom. 1976. Effects of self-awareness on antinormative behavior. *Journal of Research in Personality* 10, 1 (1976), 107–111. doi:10.1016/0092-6566(76)90088-X

[22] H. O. Doerr and J. E. Carr. 1982. Videotape "Self-Confrontation Effect" as Function of Person Viewed and Task Performed. *Perceptual and Motor Skills* 54, 2 (1982), 419–433. doi:10.2466/pms.1982.54.2.419 arXiv:https://doi.org/10.2466/pms.1982.54.2.419

[23] Jessica Douglah. 2020. "Use the mirror now" – Demonstrating through a mirror in show dance classes. *Multimodal Communication* 9, 2 (2020), 32 pages. doi:10.1515/mc-2020-0002

[24] Paul Ekman. 1992. Are there basic emotions? *Psychological Review* 99, 3 (1992), 550–553. doi:10.1037/0033-295X.99.3.550

[25] Paul Ekman and Wallace V. Friesen. 1976. Measuring facial movement. *Environmental psychology and nonverbal behavior* 1, 1 (01 Sep 1976), 56–75. doi:10.1007/BF01115465

[26] Steven W Gangestad and Mark Snyder. 2000. Self-monitoring: Appraisal and reappraisal. *Psychological bulletin* 126, 4 (2000), 530.

[27] Adam Gerace, Andrew Day, Sharon Casey, and Philip Mohr. 2017. 'I Think, You Think': Understanding the Importance of Self-Reflection to the Taking of Another Person's Perspective. *Journal of Relationships Research* 8 (2017), 19. doi:10.1017/jrr.2017.8

[28] Nathalie Hauk, Anja S Göritz, and Stefan Krumm. 2019. The mediating role of coping behavior on the age-technostress relationship: A longitudinal multilevel mediation model. *PloS one* 14, 3 (2019), e0213349.

[29] Carolina Herrando and Efthymios Constantinides. 2021. Emotional Contagion: A Brief Overview and Future Directions. *Frontiers in Psychology* 12 (2021), 5 pages. doi:10.3389/fpsyg.2021.712606

[30] Ursula Hess, Sacha Sénécal, and Pascal Thibault. 2004. Do we know what we show? Individuals' perceptions of their own emotional reactions. *Cahiers de Psychologie Cognitive* 22, 2 (2004), 247 – 265. https://www.scopus.com/inward/record.uri?eid=2-s2.0-3042816684&partnerID=40&md5=141f501cdf76834addf16a42caada6b4

[31] JS Hinton and Michael W Kramer. 1998. The impact of self-directed videotape feedback on students' self-reported levels of communication competence and

apprehension. *Communication Education* 47, 2 (1998), 151–161.

[32] Philip S Holzman and Clyde Rousey. 1966. The voice as a percept. *Journal of Personality and Social Psychology* 4, 1 (1966), 79.

[33] Philip S Holzman, Clyde Rousey, and Charles Snyder. 1966. On listening to one's own voice: Effects on psychophysiological responses and free associations. *Journal of Personality and Social Psychology* 4, 4 (1966), 432.

[34] Katherine Hopman, Deborah Richards, and Melissa M. Norberg. 2023. A Digital Coach to Promote Emotion Regulation Skills. *Multimodal Technologies and Interaction* 7, 6 (2023), 18 pages. doi:10.3390/mti7060057

[35] Joop Hox, Mirjam Moerbeek, and Rens Van de Schoot. 2017. *Multilevel analysis: Techniques and applications.* Routledge, London, UK.

[36] Angel Hsing-Chi Hwang, John Oliver Siy, Renee Shelby, and Alison Lentz. 2024. In Whose Voice?: Examining AI Agent Representation of People in Social Interaction through Generative Speech. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 224–245. doi:10.1145/3643834.3661555

[37] Rick E. Ingram, Debra Cruet, Brenda R. Johnson, and Kathleen S. Wisnicki. 1988. Self-focused attention, gender, gender role, and vulnerability to negative affect. *Journal of Personality and Social Psychology* 55, 6 (1988), 967 – 978. https://search.ebscohost.com/login.aspx?direct=true&amp;db=pdh&amp;AN=1989-15524-001&amp;site=ehost-live

[38] David H Jonassen. 1979. Video-mediated objective self-awareness, self-perception, and locus of control. *Perceptual and Motor Skills* 48, 1 (1979), 255–265.

[39] Katherine A Karl and Jerry M Kopf. 1993. Guidelines for Using Videotaped Feedback Effectively. *Human resource development quarterly* 4, 3 (1993), 303–10.

[40] Simon Kimmel, Frederike Jung, Andrii Matviienko, Wilko Heuten, and Susanne Boll. 2023. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 429, 16 pages. doi:10.1145/3544548.3580707

[41] Oleg Korzyukov, Alexander Bronder, Yunseon Lee, Sona Patel, and Charles R. Larson. 2017. Bioelectrical brain effects of one's own voice identification in pitch of voice auditory feedback. *Neuropsychologia* 101 (2017), 106–114. doi:10.1016/j.neuropsychologia.2017.04.035

[42] Eugene Kukshinov, Daniel Harley, Kata Szita, Reza Hadi Mogavi, Cayley Macarthur, and Lennart E. Nacke. 2024. Disembodied, Asocial, and Unreal: How Users Reinterpret Designed Affordances of Social VR. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1914–1925. doi:10.1145/3643834.3661548

[43] John T Lanzetta, James J Biernat, and Robert E Kleck. 1982. Self-focused attention, facial behavior, autonomic arousal and the experience of emotion. *Motivation and Emotion* 6 (1982), 49–63.

[44] Iulia Lefter, Laurens Rook, and Theodora Chaspari. 2024. Editorial: Multimodal interaction technologies for mental well-being. 2 pages. doi:10.3389/fcomp.2024.1412727

[45] Maria Legerstee, Diane Anderson, and Alliza Schaffer. 1998. Five- and Eight-Month-Old Infants Recognize Their Faces and Voices as Familiar and Social Stimuli. *Child Development* 69, 1 (1998), 37–50. doi:10.1111/j.1467-8624.1998.tb06131.x arXiv:https://srcd.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8624.1998.tb06131.x

[46] Richard D Lennox and Raymond N Wolfe. 1984. Revision of the Self-Monitoring Scale. *Journal of Personality and Social Psychology* 46, 6 (1984), 1349–1364.

[47] Jingyi Li, Weiping Tu, and Li Xiao. 2022. FreeVC: Towards High-Quality Text-Free One-Shot Voice Conversion. arXiv:2210.15418 [cs.SD] https://arxiv.org/abs/2210.15418

[48] Jingjin Li, Shaomei Wu, and Gilly Leshed. 2024. Re-envisioning Remote Meetings: Co-designing Inclusive and Empowering Videoconferencing with People Who Stutter. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1926–1941. doi:10.1145/3643834.3661533

[49] Pengcheng Luo, Victor Ng-Thow-Hing, and Michael Neff. 2013. An Examination of Whether People Prefer Agents Whose Gestures Mimic Their Own. In *Intelligent Virtual Agents*, Ruth Aylett, Brigitte Krenn, Catherine Pelachaud, and Hiroshi Shimodaira (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 229–238.

[50] Melissa Manwaring and Kimberlee K Kovach. 2012. Using Video Recordings: A Mirror and a Window on Student Negotiation. *Assessing our students, assessing ourselves* (2012), 97–115.

[51] Theodore C Masters-Waage, Jared Nai, Jochen Reb, Samantha Sim, Jayanth Narayanan, and Noriko Tan. 2021. Going far together by being here now: Mindfulness increases cooperation in negotiations. *Organizational Behavior and Human Decision Processes* 167 (2021), 189–205.

[52] Tetsuya Matsui and Seiji Yamada. 2016. Emotional contagion between user and product recommendation virtual agent. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, New York, USA, 1172–1176. doi:10.1109/ROMAN.2016.7745257

[53] Dieter Maurer and Theodor Landis. 2009. Role of Bone Conduction in the Self-Perception of Speech. *Folia Phoniatrica et Logopaedica* 42, 5 (12 2009), 226–229. doi:10.1159/000266070 arXiv:https://karger.com/fpl/article-pdf/42/5/226/2799578/000266070.pdf

[54] Eleonora Mencarini and Tommaso Zambon. 2023. Becoming a Speleologist: Design Implications for Coordination in Wild Outdoor Environments. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 533, 12 pages. doi:10.1145/3544548.3581545

[55] Silvan Mertes, Dominik Schiller, Michael Dietz, Elisabeth André, and Florian Lingenfelser. 2024. The AffectToolbox: Affect Analysis for Everyone. arXiv:2402.15195 [cs.HC] https://arxiv.org/abs/2402.15195

[56] Jeremy Miles and Mark Shevlin. 2000. *Applying regression and correlation: A guide for students and researchers.* SAGE Publications Ltd, London, UK.

[57] Margaret E Morris, Daniela K Rosner, Paula S Nurius, and Hadar M Dolev. 2023. "I Don't Want to Hide Behind an Avatar": Self-Representation in Social VR Among Women in Midlife. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 537–546. doi:10.1145/3563657.3596129

[58] Takashi Numata, Hiroki Sato, Yasuhiro Asa, Takahiko Koike, Kohei Miyata, Eri Nakagawa, Motofumi Sumiya, and Norihiro Sadato. 2020. Achieving affective human–virtual agent communication by enabling virtual agents to imitate positive expressions. *Scientific Reports* 10, 1 (06 Apr 2020), 5977. doi:10.1038/s41598-020-62870-7

[59] Arye Perlberg. 1983. When professors confront themselves: Towards a theoretical conceptualization of video self-confrontation in higher education. *Higher education* 12 (1983), 633–663.

[60] Giulia Perugia, Maike Paetzel, and Ginevra Castellano. 2020. On the Role of Personality and Empathy in Human-Human, Human-Agent, and Human-Robot Mimicry. In *Social Robotics*, Alan R. Wagner, David Feil-Seifer, Kerstin S. Haring, Silvia Rossi, Thomas Williams, Hongsheng He, and Shuzhi Sam Ge (Eds.). Springer International Publishing, Cham, 120–131.

[61] Angèle Picco, Arjan Stuiver, Joost De Winter, and Dick De Waard. 2025. "Why were you speeding?": A self-confrontation study on awareness and reasons for speed behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour* 109 (2025), 421–438. doi:10.1016/j.trf.2024.12.015

[62] Alexis Plaquet and Hervé Bredin. 2023. Powerset multi-class cross entropy loss for neural speaker diarization. In *Proc. INTERSPEECH 2023.* ISCA, Grenoble, France, 3222–3226. doi:10.21437/Interspeech.2023-205

[63] Maryam Pourebadi and Laurel D. Riek. 2022. Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future. *ACM Trans. Comput. Healthcare* 3, 2, Article 23 (March 2022), 32 pages. doi:10.1145/3483598

[64] Eliska Prochazkova, Luisa Prochazkova, Michael Rojek Giffin, H. Steven Scholte, Carsten K. W. De Dreu, and Mariska E. Kret. 2018. Pupil mimicry promotes trust through the theory-of-mind network. *Proceedings of the National Academy of Sciences* 115, 31 (2018), E7265–E7274. doi:10.1073/pnas.1803916115 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1803916115

[65] Subia P. Rasheed, Ahtisham Younas, and Amara Sundus. 2019. Self-awareness in nursing: A scoping review. *Journal of Clinical Nursing* 28, 5-6 (2019), 762–774. doi:10.1111/jocn.14708 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jocn.14708

[66] Laurel D. Riek, Philip C. Paul, and Peter Robinson. 2010. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces* 3, 1 (01 Mar 2010), 99–108. doi:10.1007/s12193-009-0028-2

[67] Linda P Rouse. 1998. Perspectives On Video Self-confrontation. *Clinical Sociology Review* 16 (1998), 43–66.

[68] James A. Russell. 1995. Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin* 118, 3 (1995), 379–391. doi:10.1037/0033-2909.118.3.379

[69] Christian E. Salas, Darinka Radovic, and Oliver H. Turnbull. 2012. Inside-out: Comparing internally generated and externally generated basic emotions. *Emotion* 12, 3 (2012), 568–578. doi:10.1037/a0025811

[70] A. Schandrin, M.-C. Picot, G. Marin, M. André, J. Gardes, A. Léger, B. O'Donoghue, S. Raffard, M. Abbar, and D. Capdevielle. 2022. Video self-confrontation as a therapeutic tool in schizophrenia: A randomized parallel-arm single-blind trial. *Schizophrenia Research* 240 (2022), 103–112. doi:10.1016/j.schres.2021.12.016

[71] Paul J Silvia. 2002. Self-awareness and the regulation of emotional intensity. *Self and Identity* 1, 1 (2002), 3–10.

[72] Natalie Stein and Kevin Brooks. 2017. A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults. *JMIR Diabetes* 2, 2 (01 Nov 2017), e28. doi:10.2196/diabetes.8590

[73] Anna Sutton. 2016. Measuring the effects of self-awareness: Construction of the self-awareness outcomes questionnaire. *Europe's journal of psychology* 12, 4 (2016), 645.

[74] Noriko Suzuki, Yugo Takeuchi, Kazuo Ishii, and Michio Okada. 2003. Effects of echoic mimicry using hummed sounds on human–computer interaction. *Speech*

Communication 40, 4 (2003), 559–573. doi:10.1016/S0167-6393(02)00180-2

[75] Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2017. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLOS ONE* 12, 8 (08 2017), 1–15. doi:10.1371/journal.pone.0182151

[76] June P Tangney, Angie Luzio Boone, and Roy F Baumeister. 2018. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. In *Self-regulation and self-control*. Routledge, London, UK, 173–212.

[77] Leigh Thompson, Victoria Husted Medvec, Vanessa Seiden, and Shirli Kopelman. 2001. Poker face, smiley face, and rant 'n'rave: Myths and realities about emotion in negotiation. In *Blackwell handbook of social psychology: Group processes*. Blackwell Publishers, Malden, Massachusetts, USA, 139–163.

[78] Robin Ungruh, Susanne Schmidt, Nahal Norouzi, and Frank Steinicke. 2023. Insights From a Study on Subtle Mimicry in Human-Agent Interaction. In *ICAT-EGVE 2023 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, Jean-Marie Normand, Maki Sugimoto, and Veronica Sundstedt (Eds.). The Eurographics Association, Eindhoven, Netherlands, 8 pages. doi:10.2312/egve.20231313

[79] Garry R Walz and Joseph A Johnston. 1963. Counselors look at themselves on video tape. *Journal of Counseling Psychology* 10, 3 (1963), 232.

[80] Li Wan, Quan Wang, Alan Papir, and Ignacio Lopez Moreno. 2018. Generalized End-to-End Loss for Speaker Verification. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, New York, USA, 4879–4883. doi:10.1109/ICASSP.2018.8462665

[81] Alan J Weston and Clyde L Rousey. 1970. Voice confrontation in individuals with normal and defective speech patterns. *Perceptual and motor skills* 30, 1 (1970), 187–190.

[82] Morton Wiener, Shannon Devoe, Stuart Rubinow, and Jesse Geller. 1972. Nonverbal behavior and nonverbal communication. *Psychological Review* 79, 3 (1972), 185–214. doi:10.1037/h0032710

[83] Matthias J Wieser, Paul Pauli, Georg W Alpers, and Andreas Mühlberger. 2009. Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. *Journal of anxiety disorders* 23, 1 (2009), 93–103.

[84] Wen-Jing Yan, Qian-Nan Ruan, Xiaolan Fu, and Yu-Qi Sun. 2022. Perceived emotions and AU combinations in ambiguous facial expressions. *Pattern Recognition Letters* 164 (2022), 74–80. doi:10.1016/j.patrec.2022.10.018

[85] Wenjing Yang and Yunhui Xie. 2024. Can robots elicit empathy? The effects of social robots' appearance on emotional contagion. *Computers in Human Behavior: Artificial Humans* 2, 1 (2024), 100049. doi:10.1016/j.chbah.2024.100049

[86] Miron Zuckerman, Deborah T. Larrance, Nancy H. Spiegel, and Rafael Klorman. 1981. Controlling nonverbal displays: Facial expressions and tone of voice. *Journal of Experimental Social Psychology* 17, 5 (1981), 506–524. doi:10.1016/0022-1031(81)90037-8

A Linear combinations of AUs

We list the linear combinations from Yan et al. [84] below for reference. The AUs marked by † are not generated by OpenFace, and were thus always 0 in our system.

$$\begin{aligned}
\text{Anger} &= 0.4659 * \text{AU25} + 0.4337 * \text{AU9} \\
&+0.4236 * \text{AU10} + 0.3587 * \text{AU4} + 0.3459 * \text{AU16}^{\dagger} \\
\text{Disgust} &= 0.5964 * \text{AU10} + 0.5330 * \text{AU4} \\
&+0.2973 * \text{AU17} + 0.2527 * \text{AU9} + 0.2163 * \text{AU25} \\
\text{Fear} &= 0.5111 * \text{AU25} + 0.4033 * \text{AU12} \\
&+0.3729 * \text{AU27}^{\dagger} + 0.2995 * \text{AU16}^{\dagger} + 0.2852 * \text{AU1} \\
\text{Happiness} &= 0.7040 * \text{AU12} + 0.5143 * \text{AU25} \\
&+0.2491 * \text{AU27}^{\dagger} + 0.2032 * \text{AU6} + 0.1730 * \text{AU10} \\
\text{Sadness} &= 0.6723 * \text{AU4} + 0.3462 * \text{AU25} \\
&+0.2979 * \text{AU1} + 0.2859 * \text{AU17} + 0.2359 * \text{AU15} \\
\text{Surprise} &= 0.5926 * \text{AU25} + 0.4665 * \text{AU5} \\
&+0.3820 * \text{AU26} + 0.3490 * \text{AU1} + 0.3411 * \text{AU2} \\
\text{Neutral} &= 0.1022 * \text{AU14} + 0.0463 * \text{AU18}^{\dagger} \\
&+0.0446 * \text{AU23} + 0.0245 * \text{AU28}^{\dagger} + 0.0242 * \text{AU26}
\end{aligned}$$

The Action Units listed represent [25]:

AU1	Inner brow raiser
AU2	Outer brow raiser
AU4	Brow lowerer
AU5	Upper eyelid raiser
AU6	Cheek raiser
AU9	Nose wrinkler
AU10	Upper lip raiser
AU12	Lip corner puller
AU14	Dimpler
AU15	Lip corner depressor
AU16	Lower lip depressor
AU17	Chin raiser
AU18	Lip puckerer
AU23	Lip tightener
AU25	Lip parting
AU26	Jaw dropping
AU27	Mouth stretching
AU28	Lip sucking