Comparing Methods for Mapping Facial Expressions to Enhance Immersive Collaboration with Signs of Emotion

Natalie Hube*
Mercedes-Benz AG
University of Stuttgart

Oliver Lenz†
Technische Universität Dresden

Lars Engeln‡
Technische Universität Dresden

Rainer Groh§
Technische Universität Dresden

Michael Sedlmair¶
University of Stuttgart

ABSTRACT

We present a user study comparing a pre-evaluated mapping approach with a state-of-the-art direct mapping method of facial expressions for emotion judgment in an immersive setting. At its heart, the pre-evaluated approach leverages semiotics, a theory used in linguistic. In doing so, we want to compare pre-evaluation with an approach that seeks to directly map real facial expressions onto their virtual counterparts. To evaluate both approaches, we conduct a controlled lab study with 22 participants. The results show that users are significantly more accurate in judging virtual facial expressions with pre-evaluated mapping. Additionally, participants were slightly more confident when deciding on a presented emotion. We could not find any differences regarding potential Uncanny Valley effects. However, the pre-evaluated mapping shows potential to be more convenient in a conversational scenario.

Index Terms: Human-centered computing [Visualization]: Visualization system and tools—Visualization design and evaluation methods; Human-centered computing [Collaborative and social computing]: Collaborative and social computing systems and tools—

1 INTRODUCTION

Digital interpersonal communication is increasingly changing how people collaborate remotely. Video chats currently dominant in this area, though, they are limited compared to collaboration in augmented and virtual rooms with increasingly rich representation of human factors. Hereby, avatars enrich communication [6] as users' actions are mapped and can be equivalent to face-to-face dialogues [4]. Though this type of interaction increases complexity. The avatars must be humanized to the extent that the user is able to interpret facial expressions based on their initiated ability to recognize and evaluate the displayed non-verbal expression.

Our work is limited to interpersonal communication that is not conveyed through verbal language [22]. More specifically, we address facial expression of emotions. However, in immersive environments non-verbal cues are usually missing to fully understand underlying messages in discussions. Currently, the mapping of expressions to virtual faces is done primarily via a one-to-one transfer. This direct mapping of expressions may be oversubscribed and increases data transfer rates significantly. This can be a hindrance for satisfying immersive collaborations. Ongoing research mainly covers the question of how to depict every facet of a human in the immersive environment [1, 38] without taking aspects of communication theory into account. Nonetheless, avatars have a significant value for transferring non-verbal cues to increase social immersion and performance in general [2], and the lack of likewise leads to a state of confusion [37].

Therefore, we conduct a user study to compare a pre-evaluated mapping approach with a state-of-the art direct mapping method of facial landmarks to facilitate non-verbal emotion recognition by bringing real phenomena of analyzed facial expressions into the virtual world in an immersive setting. Thus, we developed a prototype to provide two different mapping methods. These two approaches mainly differ in the facial landmarks used for the animation of the avatar face. The emotions are based on the six basic emotions [12] and were applied using the Facial Action Coding System (FACS) [14] with regard to semiotics [8] for both mappings. We applied both mappings to a low-fidelity avatar to compare them in a controlled lab study with 22 participants. The avatar with the pre-evaluated emotion mapping achieved significantly better results in terms of task load and recognition rates.

In summary, our contributions are:

- Comparison of two implemented methods, an iconic direct and an indexical pre-evaluated mapping, to determine effects on users by conducting a controlled lab study.
- Presentation of the results of the user study to derive recommendations and thus, discuss limitations and future directions.

2 BACKGROUND & RELATED WORK

When working with facial expressions for non-verbal emotion recognition, semiotics are applied for describing elements and meanings in correspondence to possible abstractions. The theory of semiotics is described with its *syntactic*, *semantic*, and *pragmatic* function of signs [8, 27]. The syntax is the general structure of a subject. Here, the syntax are facial landmarks. Semantic describes the meaning of a subject, like the expression formed by the syntactic structure of the facial landmarks. The meaning of the expression leads to the perceived emotion. The pragmatic effect on a user is how the user thinks and feels about the subject.

The semantic can be differentiated in the categories *symbolic*, *indexical*, and *iconic* meaning [30]. One-to-one representation and direct mappings are iconic. References and hints to contextualize a meaning are *indexical*. Thereby, using all tracked facial landmarks for displaying an expression is referred as iconic, and only using a reduced set of features is referred as *indexical*, as reduced features may give a hint to the actual facial expression. The pre-evaluated mapping is indexical, as it hints to the emotion, but not to the actual facial expression. *Symbols* are like social conventions, which the user has to learn. In this context, symbolic is the highest abstraction like mapping to a visual, in which the emotion is expressed.

2.1 Semiotics in HCI

The theory of semiotics, a technique used in linguistic, also found its way in human-computer interaction [10, 28]. Sabine et al. [35] suggest to increase the feeling of being fully immersed using semiotics for visual communication. However, semiotics is also used in terms of social interaction and for augmented reality (AR) research [7, 21].

*e-mail: natalie.hube@daimler.com
†e-mail: oliver.lenz@tu-dresden.de
‡e-mail: lars.engeln@tu-dresden.de
§e-mail: rainer.groh@tu-dresden.de
¶e-mail:michael.sedlmair@visus.uni-stuttgart.de

Unlike these approaches, we explore the application of semiotics theory to describe non-verbal communication by decoupling the emotion from the actual facial expression.

2.2 Emotions in Immersive Collaboration

Certain forms of virtual avatar representation have an impact on the social entity [18, 20]. Specifically, behavioral information such as facial expressions and gaze have a unique role as the level of realism can be increased [34] to not fall into the Uncanny Valley [5]. Here, the Uncanny Valley describes the paradoxical effect that artificial human characters are recognized and accepted to a certain iconicity [26]. Thus, it is important to keep this effect in mind to not cause any discomfort by mapping emotions. Hence, we pursue a narrowed human representation. As depending on the context of an immersive experience, low-fidelity avatars additionally allow to increase disconnection from reality.

Overall, research [15, 24] indicates that non-verbal behaviors are essential in maintaining interpersonal collaboration to enable a natural way of interaction. Similar to Oh et al. [29], our approach to use semiotics to describe facial expressions to transfer emotions also infers an amplifying character as we seek to compare the recognizably of two mapping approaches using basic emotions.

2.3 Modulation of Facial Expressions

Approaches to modulate facial expressions [29] and natural behaviors [33] seek to increase social presence [36]. Systems with pre-evaluated approaches focus on identifying perceived emotions. Most research [11,32] uses machine learning and other approaches to automatically recognize facial emotions to increase accuracy in detecting the users state of mind or synthesis of 3D avatars based on the input material. Here, perception through a human being is only a means to an end, not the focus of most studies as they focus on computational recognition through a system. Hence, we especially use the results by Hart et al. [18] to implement an approach to pre-evaluate facial expressions to facilitate the user's emotion recognition.

3 Pre-Evaluated Mapping

Pre-evaluation allows to overcome two issues of direct mapping approaches. First, direct mapping might lead to oversubscribed facial expressions, bearing the risk of falling into the Uncanny Valley. Second, it leads to an increased data transmission size, as all recognized facial landmarks need to be transmitted. This aspect is particularly important for online multi-user applications due to higher latency.

To implement a pre-evaluated strategy, we use the results presented by Hart et al. [18]. Therefore, we devise a three-step pipeline (see Figure 2). First, we considered how facial expressions are converted into a data format that translates the facial landmarks of

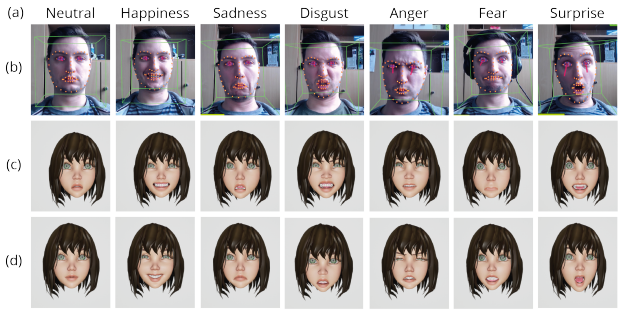

Figure 1: Mapping of emotions according to FACS. (a) Neutral baseline emotion and the six basic emotions by Ekman [12]. (b) Screenshot of recorded video stream of the facial expressions for each emotion. Animated avatar faces with (c) directly mapped and (d) pre-evaluated Action Units.

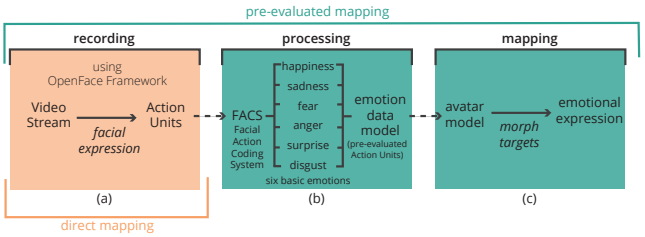

Figure 2: Pipeline for implemented approaches highlighted by color. (blue) pre-evaluated mapping. (orange) direct mapping. (a) Action Units extracted from video stream based on the facial expression. (c) An emotion is determined using rules based on FACS, resulting in an emotion data model. (d) The data model is used to animate the avatar using morph targets to adapt the emotional expression.

Ekman et al. [13]. Second, we aggregated emotions into meaningful modules. Here, we used the theory of semiotics. Finally, we map these extracted and aggregated facial expressions onto an avatar face. In the following, details on each of these steps are briefly provided.

3.1 Recording of Facial Expression

To display facial expressions on an avatar, facial feature points must be recognized and processed. First, we determine facial landmarks and translate them into a machine-readable representation of facial expression by analyzing live video material and extracting Action Units. Action Units are the smallest visually perceptible movements on a face. Thereby, almost any facial expression can be displayed through Action Units. To automatically recognize the Action Units from a facial expression from a given video stream, we use the OpenFace Framework [3], which combines methods of machine learning as well as other approaches to detect facial features.

3.2 Processing Emotions through FACS

To express facial landmarks computationally, we chose FACS [13], which describes expressions in different levels of detail by using Action Units. When using FACS alone, every single change in expression resembles an Action Unit, that has by itself no further meaning. FACS itself is descriptive and does not contain any emotional definitions, opposed to other systems [9, 16]. It is a common standard to categorize facial expression of emotions systematically [17]. Emotional description enables mapping only specific emotions, whereas FACS applied with semiotics are able to map more than plain emotions and transitions between animated emotions. Thus, allowing to define further ambiguous expressions.

An expression does not necessarily correspond to an emotion. Furthermore, a catalog of expressions has to be defined beforehand. Individual Action Units are of no meaning per se, but can be composed according to rules described by Ekman et al. [13]. They are ideal to form a meaningful syntactical structure. A large part of information content would be lost when creating refined attributes of semantics, since a facial expression conveys the intended meaning as a whole. With the description of facial expressions using semiotics and FACS, the following attributes result:

- *Syntactic* - a facial expression consists of several Action Units that are linked together using rules and patterns
- *Semantic* - a facial expression conveys a certain meaning, the understanding changes in a given context
- *Pragmatic* - a facial expression may lead to a context sensitive user reaction

In addition, semiotics are used to process emotions. The individual feature points are regarded as *syntactic*, the resulting expressions with a meaning through determined Action Units as *semantics* and the comprehension by the user of this expression as *pragmatic*.

3.3 Mapping Emotions onto Virtual Face

The last step describes the mapping of the identified emotions by means of pre-evaluated Action Units onto the virtual face of an avatar (see Figure 2 (e)). Using a lo-fidelity avatar is more suitable to prevent falling into the Uncanny Valley [25].

Based on our data model, we map the pre-evaluated Action Units to the corresponding morph targets in the underlying 3d model (see Figure 2 (d)). The morph targets correspond to the captured Action Units. Thus, the pre-evaluated emotions on the avatar always look identical, independent of the user's face and the intensity of its expressions. The visualization component itself was created using the Unreal Engine 4.22.3.

3.4 Bandwidth Analysis

To reduce bandwidth size in multi-user online collaboration, we seek to have a shorter string length when using our pre-evaluated method. The data model for the pre-evaluated mapping consists of six Action Units, whereas the state-of-the-art direct mapping has 17 Action Units. As the system is designed to pre-evaluate the facial expressions on the client side, only the evaluated Action Units would have to be sent to other clients.

While measuring the amount of data sent to the visualization component, we determined a size of 179 to 185 bytes for pre-evaluated mapping and a size of 463 to 469 bytes for direct mapping. Making the pre-evaluated mapping approach 2.5-times smaller regarding data transmission opposed to the directly mapped Action Units.

4 User Study

We conducted a controlled experiment in which we compare two conditions, pre-evaluated and direct mapping of facial expressions to identify emotions. Therefore, we examine the recognizability of emotions on a virtual avatar as well as the workload and comfort regarding the Uncanny Valley. A low-fidelity avatar was chosen to present the emotions. Our aim is to verify whether pre-evaluated mapping of facial expressions facilitates emotion recognition.

4.1 Participants

The study involved 22 participants (4 female). The mean age is 28 years, ranging from 22 to 38 years. 81% of the participants are related to computer science or media design. Participants have a self-estimated experience of 2.9 of 5 with virtual reality and 3.6 of 5 with computer games.

4.2 Setup

The setup consists of two components, a tracking component and a visualization component. The participant is equipped with an HTC Vive Pro. To avoid any bias from artificial, non-reproducible social or behavioral cues such as appearance, postures, facial or gaze displays, participants were put in an empty virtual room where the user sits in front of an avatar, that displays expressions (see Figure 3). The expressions are determined by the tracking component and mapped onto the avatar. We use prerecorded video material instead of a live stream of a real person, as we aim to increase comparability between both conditions and emotions to anticipate illumination issues, false recognition and performance-wise fps drops.

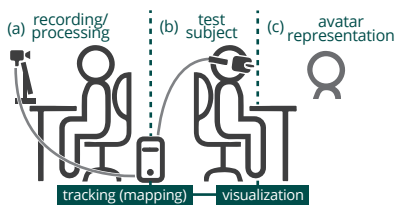

Figure 3: The overall setup includes three parts: (a) A prior recorded video material for the tracking component. (b) Study setup with the participant who is wearing a head-mounted display. (c) Avatar presentation within the immersive environment.

4.3 Tasks

We tasked participants with judging emotional expression of a virtual avatar. A total of 12 tasks were performed in a random order: for both conditions and 6 emotions. The two conditions differ as follows:

- **Pre-Evaluated** For the pre-evaluated mapping condition, we use our proposed method.
- **Direct** For the direct mapping condition, we followed a standard approach and directly mapped facial landmarks (see Figure 1, red dots) in form of Action Units onto the virtual face.

Six basic emotions (*happiness*, *sadness*, *fear*, *disgust*, *surprise*, and *anger*) were used to represent the emotions on the avatar face, as these are known and predictable by each human. Each emotion was shown in random order. Although, the six emotions were used for both conditions, participants did not know how often an emotion will be shown, therefore avoiding tactical choices. By limiting answer options, we aim to increase comparability as no synonyms appear.

4.4 Design and Variables

We used a within-subject design. The tasks were operated in two conditions (pre-evaluated mapping, direct mapping) in random seated order, while wearing a VR head-mounted display.

4.4.1 Independent Variables

We considered one independent variable: the mapping approach. The main independent variable distinguishes between pre-evaluated mapping and direct mapping. The study was conducted in a randomized procedure to exclude disturbances and to prevent systematic distortion, each participant answered a set of questions according to the demonstrated study setup (direct or pre-evaluated mapping).

4.4.2 Dependent Variables

The measured data includes categorical judgment of the recognized emotion and a subjective feedback using questionnaires. Our questionnaire includes a set of demographic questions, the Social Presence Module of the Game Experience Questionnaire (GEQ) [31], the NASA-TLX [19] and a set of questions regarding the emotion recognition. However, the key question was whether the pre-evaluated or direct mapping of facial landmarks is more suitable to determine emotions and to reduce discomfort regarding the Uncanny Valley. In addition, recognition and confidence rates were measured.

4.5 Hypotheses

Based on our experiment design, we have the following hypotheses:

- **H1** Recognizability of expressions is better with pre-evaluated mapping than with direct mapping (*pre-evaluated* > *direct*).
- **H2** Workload is lower using pre-evaluated mapping (*pre-evaluated* < *direct*).
- **H3** Comfort towards the avatar is higher with pre-evaluated mapping than with direct mapping (*pre-evaluated* > *direct*).

4.6 Procedure

Introduction Phase. First, we introduced the system and the experiment. Then participants filled out a demographic questionnaire. Before each condition, participants were allowed to adjust to the immersive environment as needed.

Batch repeated phase. The animated avatar face was presented to the participants in a randomized order for both conditions and 6 emotions. Participants judged each presented facial expression. The study leader then selects the given answers and proceeds with the next expression. Participants were allowed to re-watch the current

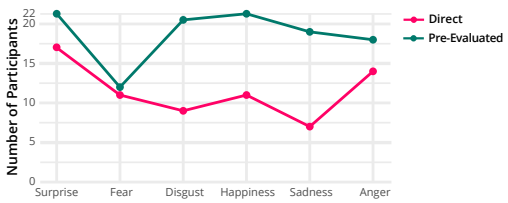

Figure 4: Recognition values for all emotions in both conditions. All emotions were better recognized in the pre-evaluated condition. The difference of the emotion fear is small, though.

expression, but had to judge the emotion before the next expression. After each batch of expressions, participants filled out the Social Presence Module, the NASA-TLX and rated feeling of comfort towards the avatar. There was a 5 minutes break between batches.

4.7 System Description

Our developed environment ran on a Windows 10 computer with an Intel Core i7 7700 CPU, 16 GB RAM and a NVIDIA GeForce 1060 GPU. The webcam is a Logitech C920 HD Pro offering a maximum resolution of 1920x1080 pixels and an automatic exposure compensation. The HTC VIVE Pro was used as the head-mounted display for the immersive environment.

5 Results

First, we report on the statistical results. We checked normality using Kolmogorov-Sminov test (KS-test). If the data obey normality, a pairwise t-test was used to analyze the Likert scale data. If not, we used a Wilcoxon signed-rank test (WSR-test). The results of our user study were statistically evaluated using *R* language.

5.1 Recognition Rate

After presenting an emotion, participants judged the emotion by choosing one of the six basic emotions. The recognition rate was determined by correctly identified emotions for each condition.

Participants received statistical significantly lower recognition rates for the direct mapping ($M = 3.14$, $SD = 1.04$) than for the pre-evaluated mapping condition ($M = 5.18$, $SD = 0.91$), for the overall sample group, $t(21) = -7.44$, $p < 0.001$ (see Figure 5 (a) and Figure 4). Additionally, to the recognition rate, we measured the confidence regarding the given answers using a Likert scale raging from 1 - 5 for each condition (see Figure 7).

5.2 Uncanny Valley

Comfort assessment related to the Uncanny Valley was performed on a 1-5 Likert scale (see Figure 5 (b)). As the normality after

Figure 5: Recognition and comfort rates based on 95% confidence interval for both conditions. (a) Value for average of correctly recognized emotions with a maximum of six. (b) Average comfort regarding the Uncanny Valley effect with a maximum of five *(1: very uncomfortable - 5: very comfortable)*.

Table 1: Results of t-test for each index of the Social Presence Module of GEQ with existing normality. *Negative Feelings* does not meet this requirement, thus, the WSR-test was calculated. No significant values could be measured.

	Mapping	Mean	SD	SE	Condition
Empathy	direct	1.82	0.4	0.09	$t(21) = -0.42$;
	pre	1.86	0.54	0.11	$p_2 = 0.682$
Behavioral Participation	direct	2.1	0.38	0.081	$t(21) = -0.24$;
	pre	2.11	0.45	0.09	$p_2 = 0.809$;
Negative Feelings	direct	1.1	0.16	0.03	$Z = -1.89$;
	pre	1.0	0.04	0.01	$p_2 = 0.063$

the KS test is not given, the Wilcoxon Signed-Ranks test (WSR-test) is used. The WSR-test indicated no statistically significantly difference between pre-evaluated mapping ($M = 3.37$) and direct mapping ($M = 3.14$, $Z = -1.078$, $p = 0.172$).

5.3 Social Presence Module

The Social Presence Module, surveys the participant's *empathy*, *negative feelings* and *behavioral participation* towards the avatar. Here, a significant difference could not be determined for *empathy* and *behavioral participation*. The WSR-test of *negative feelings* showed no significant differences either (see Table 1).

5.4 NASA-TLX

Subjective demands were checked individually, as these are plotted on a Likert scale of 0-100 (see Table 2). In our case, significance could be demonstrated for every requirement, except for effort and physical demand (see Figure 6). The TLX includes all subjective demands for calculating a total value. The calculation of the t-test resulted in a significantly higher total load for direct mapping. An influence of through demographics could not be determined.

6 Discussion

The results of the user study indicate various advantages of using pre-evaluated mapping of emotions on virtual avatars.

6.1 Emotion Judgment

In general, emotion recognition is good in the pre-evaluated mapping condition. The recognition of emotions *happiness* and *surprise* were correctly recognized by each participant (see Figure 4 and Figure 8), making them the most reliable in the overall comparison. The least recognized emotion is *fear* (see Figure 4). Nevertheless, a poor

Table 2: Results of t-test for each NASA-TLX demand with existing normality. The WSR-test was calculated for index *physical demand* as no normality could be determined. Green cells indicate statistically significant results.

	Mapping	Mean	SD	SE	Condition
Mental Demand	direct	54.5	23.59	5.03	$t(21) = 3.52$;
	pre	40.68	20.43	4.36	$p_2 = 0.002$
Physical Demand	direct	8.0	4.27	0.91	$Z = -0.632$;
	pre	8.64	6.93	1.48	$p_2 = 0.383$
Temporal Demand	direct	33.9	23.55	5.021	$t(21) = 2.43$;
	pre	25.46	17.25	3.68	$p_2 = 0.024$
Performance	direct	58.4	18.92	4.03	$t(21) = 4.79$;
	pre	33.64	19.35	4.12	$p_2 < 0.001$
Effort	direct	36.8	25.57	5.45	$t(21) = 1.09$;
	pre	31.59	20.49	4.37	$p_2 = 0.288$
Frustration	direct	46.8	25.51	5.44	$t(21) = 2.90$;
	pre	32.95	20.74	4.42	$p_2 = 0.008$
Task Load Index	direct	50.3	18.56	3.96	$t(21) = 4.213$;
	pre	34.86	15.94	3.39	$p_1 < 0.001$

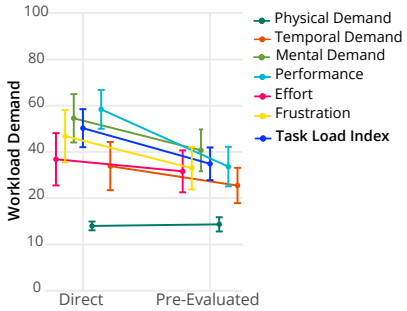

Figure 6: The NASA-TLX measures the perceived workload during the emotion judgment. The scale for each task is rated within a 100 points rage *(0: very low - 100: very high)*, thus, lower ratings are better. All ratings are combined to the TLX (blue). The plot shows the 95% confidence interval for each sub-scale of the questionnaire.

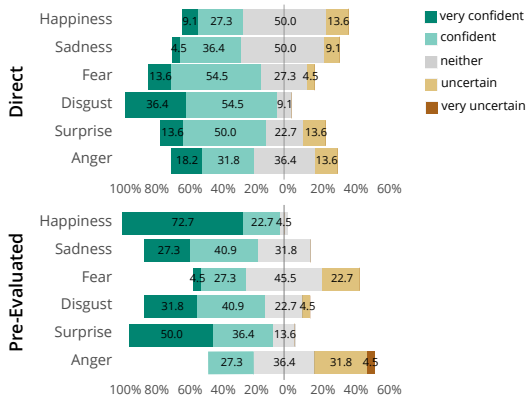

Figure 7: After judging an emotion, participants were asked how confident they were when answering. Results are shown in a diverging stacked bar chart. The answers were given using a Likert scale ranging from *1 (very uncertain) - 5 (very confident)*.

recognition can also be distinguished by the fact that a larger part of participants was less confident when deciding (see Figure 7). This could hint to a lack of context, as it may not be entirely clear whether an emotion is positively or negatively connoted.

6.2 Confidence in Emotion Judgment

With regard to the confidence level, we noticed that the majority used higher values when answering (see Figure 7). This finding does not relate to the low recognition results for the corresponding emotion (see Figure 4). For the direct mapping condition, 90.9% of participants were confident or very confident when judging the emotion *disgust*, but only 40.9% of participants judged the emotion correctly. The emotion *fear* has the same pattern. In the pre-evaluated mapping condition, we observed the opposite effect. The emotions *happiness* and *surprise* had high confidence in answering with each participant judging the emotion correctly (see Figure 8 (a) and Figure 7). Participants confidence assessment was more distinct according to the given answers in pre-evaluated mapping. In contrast to direct mapping, where the most confident assessment does not concur with the judged emotion, characterizing a more unsteady pattern.

6.3 Emotion Mix-Up

In the direct condition, emotions were more prone to be mixed-up. For example, the emotion of *disgust* was often mixed-up with *anger* by over 50% by participants (see Figure 8 (a)). When looking at Figure 8 (a) and Figure 7, there is no correlation between the mix-up

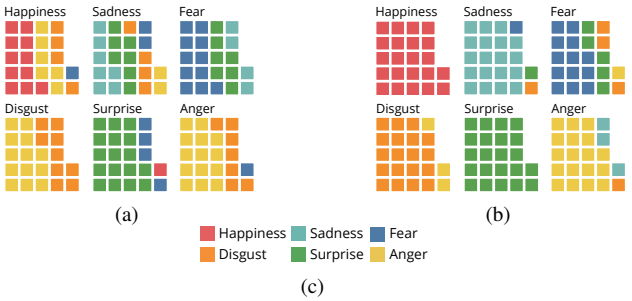

Figure 8: The waffle chart displays (c) specific answers given for (a) direct and (b) pre-evaluated mapping conditions for each participant, sorted by color.

and the confidence in the answers. However, Action Units for both emotions are similar.

In the pre-evaluated mapping condition, no emotion could be identified that is confused as often as in the direct mapping condition. Still, *sadness* and *disgust* are the least recognized emotions when using direct mapping. Marcos-Pablos et al. [23] had a similar discovery, sadness and disgust were named as the most difficult emotion to recognize in their findings. This mix-up could be due to a lack of context, that we purposely left out in the user study.

6.4 Perceived Workload & Comfort

Pre-evaluated mapping showed a significantly lower TLX (see Figure 6, blue) with significantly higher recognition rates (see Figure 4). Additionally, we subjectively observed that the average speed in judging emotions was considerable faster. Thus, we assert pre-evaluated mapping can facilitate emotion recognition in a conversational scenario. Regarding the Uncanny Valley, we could not identify a negative impact in our presented study. Only slight differences exist with regard to social presence of the avatar (see Section 5.3), thus, pre-evaluated mapping is unlikely to have any drawbacks.

6.5 Interpretation of Results

The quantitative results provide the following discussion points:

H1 This hypothesis can be confirmed by means of a significantly higher recognition rate. Likewise, the temporal demand (see Table 2) indicates faster judgment. Participants estimated the time required to solve a task for the direct mapping to be significantly higher.

H2 The calculation of the NASA-TLX confirms this hypothesis. The TLX for the pre-evaluated mapping is significantly lower. Furthermore, there is no difference in terms of effort (see Table 2).

H3 This hypothesis cannot be confirmed by analyzing the data. Meaning, there is no difference for the avatar representation in our presented scenario. The refuted hypothesis does not justify any disadvantage when applying direct mapping to a low-fidelity avatar.

6.6 Limitations

In our presented study, we omit user interaction as we do not want to influence emotion judgment depending on the type of interaction. According to Ekman et al. [13], the six basic emotions are recognizable without interaction or additional context. Regardless of this limitation, future studies with additional context would be interesting to conduct, as some emotions like *anger* or *disgust* might benefit from this extension. Expressions according to Morris' semiotic model [27] can take on special meanings within a context. However, comments from participants showed that context could help with judgment. Particularly for the emotion *surprise*, as this emotion can be both positively and negatively connoted. Again, this could be due

to the lack of context. Potential task bias should be decreased by showing more expressions than examined emotions.

7 Conclusion

We presented a user study, where we empirically compared two conditions, direct and pre-evaluated mapping, and gathered subjective feedback of the participants. Therefore, we briefly presented our implemented pipeline to map facial expressions.

Our findings showed that recognition rates and perceived workload in the pre-evaluated mapping condition achieved significantly better results. Additionally, participants estimated the time required to judge an emotion to be lower and had higher confidence values when judging emotions. We did not find any substantial differences in terms of the Uncanny Valley. However, we think that our approach provides interesting initial results to build upon. Our results suggest that for non-verbal facial landmarks in immersive collaborations a pre-evaluation pipeline could be used to represent emotions in a more natural way with lower bandwidth usage.

In future experiments, we want to amplify our setup to include more facial landmarks as well as other cues that are used by humans to communicate non-verbally. Further, we want to include context in a conversational scenario.

References

[1] M. H. Alkawaz, D. Mohamad, A. H. Basori, and T. Saba. Blend shape interpolation and facs for realistic avatar. *3D Research*, 6(1):6, 2015.

[2] S. A. Aseeri and V. Interrante. The influence of avatar representation and behavior on communication in social immersive virtual environments. In *Proc. IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, pp. 823–824, 2018.

[3] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency. OpenFace 2.0: Facial behavior analysis toolkit. In *Proc. ACM Conf. Human Factors in Computing Systems (CHI)*, pp. 59–66, 2018.

[4] G. Bente, S. Rüggenberg, and N. C. Krämer. Social presence and interpersonal trust in avatar-based, collaborative net-communications. In *Proceedings of the 7th Annual Int. Workshop on Presence*, 2004.

[5] G. Bernal and P. Maes. Emotional beasts: visually expressing emotions through avatars in vr. In *Proc. ACM Conf. Human Factors in Computing Systems (CHI)*, pp. 2395–2402, 2017.

[6] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th Annual Int. Workshop on Presence*, (January):1–9, 2001.

[7] B. Cameron, C. Sandor, and P. Mickan. Social semiotic analysis of the design space of augmented reality. In *IEEE Int. Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, 2011.

[8] P. Cobley. *Introducing semiotics: A graphic guide*. Icon Books Ltd, 2014.

[9] J. Cohn, T. Kanade, T. Moriyama, Z. Ambadar, J. Xiao, J. Gao, and H. Imamura. A comparative study of alternative facs coding algorithms. (CMU-RI-TR-02-06), 05 2002.

[10] C. S. de Souza, S. D. J. Barbosa, and R. O. Prates. A semiotic engineering approach to hci. In *Proc. ACM Conf. Human Factors in Computing Systems (CHI)*, CHI EA '01, p. 55–56. Association for Computing Machinery, New York, NY, USA, 2001.

[11] L. Deng, D. Yu, et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387, 2014.

[12] P. Ekman and W. Friesen. Facial Action Coding System. *Consulting Psychologist Press*, 1978.

[13] P. Ekman, W. V. Friesen, and J. C. Hager. Facial action coding system: The manual on cd rom. *A Human Face, Salt Lake City*, 2002.

[14] R. Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[15] U. X. Eligio, S. E. Ainsworth, and C. K. Crook. Emotion understanding and performance during computer-supported collaboration. *Computers in Human Behavior*, 28(6):2046–2054, 2012.

[16] J. Hager. A comparison of units for visually measuring facial actions. *Behavior Research Methods, Instruments, & Computers*, 1985.

[17] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods*, 200:237–256, 2011.

[18] J. D. Hart, T. Piumsomboon, L. Lawrence, G. A. Lee, R. T. Smith, and M. Billinghurst. Emotion sharing and augmentation in cooperative virtual reality games. In *Proc. of Symp. on Computer-Human Interaction in Play Companion Extended Abstracts*, 2018.

[19] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[20] P. Heidicker, E. Langbehn, and F. Steinicke. Influence of avatar appearance on presence in social vr. In *IEEE Symp. on 3D User Interfaces (3DUI)*, pp. 233–234, 2017.

[21] H. Iizuka, H. Ando, T. Maeda, and D. Marocco. Co-creativity of communication system on behavioral interaction. In *IEEE Int. Symp. on Robot and Human Interactive Communication*, pp. 298–303, 2012.

[22] M. L. Knapp, J. A. Hall, and T. G. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.

[23] S. Marcos-Pablos, E. González-Pablos, C. Martín-Lorenzo, L. A. Flores, J. Gómez-García-Bermejo, and E. Zalama. Virtual avatar for emotion recognition in patients with schizophrenia: A pilot study. *Frontiers in human neuroscience*, 10:421, 2016.

[24] G. Molinari, G. Chanel, M. Betrancourt, T. Pun, and C. Bozelle Giroud. Emotion feedback during computer-mediated collaboration: Effects on self-reported emotions and perceived interaction. In *To see the world and a grain of sand: Learning across levels of space, time, and scale: CSCL 2013 Conf. proceedings*, 2013.

[25] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley. *IEEE Robotics and Automation Magazine*, 19(2):98–100, 2012.

[26] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, pp. 98–100, 2012.

[27] C. W. Morris. Foundations of the theory of signs. In *Int. encyclopedia of unified science*, pp. 1–59. Chicago University Press, 1938.

[28] J.-L. Nespoulous, P. Perron, and A. R. Lecours. *The biological foundations of gesture: Motor and semiotic aspects*. Psychology Press, 2014.

[29] S. Y. Oh, J. Bailenson, N. Krämer, and B. Li. Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PloS one*, 11(9), 2016.

[30] C. S. Peirce. *A Syllabus of Certain Topics of Logic*. Alfred Mudge & Son, Boston, 1903.

[31] K. Poels, Y. A. W. de Kort, and W. A. IJsselsteijn. Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games. 2007.

[32] T. Randhavane, U. Bhattacharya, K. Kapsaskis, K. Gray, A. Bera, and D. Manocha. Learning perceived emotion using affective and deep features for mental health applications. In *Proc. IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*, pp. 395–399, 2019.

[33] D. Roth, P. Kullmann, G. Bente, D. Gall, and M. E. Latoschik. Effects of hybrid and synthetic social gaze in avatar-mediated interactions. In *Proc. IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*, pp. 103–108, 2018.

[34] D. Roth, J.-L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar realism and social interaction quality in virtual reality. In *Proc. IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, pp. 277–278, 2016.

[35] E. Sabine, B. Alexandra, and H. Wade. Evolution of discourses along human-computer interaction throughout computer game based teaching technology: The compentency building process of an individual. In *IEEE Int. Professional Communication Conf.*, pp. 1–8, 2012.

[36] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proc. ACM Conf. Human Factors in Computing Systems (CHI)*, pp. 1–12, 2018.

[37] J. Tanenbaum, M. S. El-Nasr, and M. Nixon. *Nonverbal communication in virtual worlds: Understanding and designing expressive characters*. ETC Press Pittsburgh, 2014.

[38] Z. Wang, Z. Liu, Z. Chen, H. Hu, and S. Lian. A neural virtual anchor synthesizer based on seq2seq and gan models. In *Proc. IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*, pp. 233–236, 2019.