Investigating perceptually-based models to predict importance of facial blendshapes

Anonymous Author(s)

ABSTRACT

Blendshape facial rigs are used extensively in the industry [32] for facial animation of virtual humans. However, storing and manipulating large numbers of facial meshes (blendshapes) is costly in terms of memory and computation for real-time applications. Blendshape rigs are comprised of sets of semantically-meaningful expressions, which govern how expressive the character will be, often based on Action Units from the Facial Action Coding System [20]. However, the relative perceptual importance of blendshapes has not yet been investigated. Research in Psychology and Neuroscience has shown that our brains process faces differently than other objects [4, 21, 29] so we postulate that the perception of facial expressions will be feature-dependent rather than based purely on the amount of movement required to make the expression. Therefore, we believe that perception of blendshape visibility will not be reliably predicted by numerical calculations of the difference between the expression and the neutral mesh. In this paper, we explore the noticeability of blendshapes under different activation levels, and present two new perceptually-based models to predict perceptual importance of blendshapes. The models predict visibility based on commonly-used geometry and image-based metrics.

CCS CONCEPTS

• **Applied computing** → **Psychology**; • **Computing methodologies** → **Mesh geometry models**; • **Mathematics of computing** → *Equational models.*

KEYWORDS

virtual character, animation, perception, linear model

ACM Reference Format:

Anonymous Author(s). 2020. Investigating perceptually-based models to predict importance of facial blendshapes. In *Proceedings of SAP '20: ACM Symposium on Applied Perception (SAP '20).* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In Psychology literature, the perception of human faces is a much studied area of research. In terms of virtual characters, expressions are generally created using blendshape rigs [32] based on FACS action units [20], however these rigs are computationally expensive for real-time applications. The question of importance of blendshapes is therefore of great interest to computer games and other real-time applications, with the aim of reducing the number of blendshapes needed for animating a rig [14], or prioritising which blendshapes to include in expressions for example-based blendshape rig creation algorithms [7, 8]. Additionally, algorithms that create or alter facial geometry are usually evaluated against ground-truth facial meshes using standard geometry error metrics (e.g., [8]), however, we postulate that standard error-metrics may not be sufficient to determine how perceptually different the results are to the ground-truth.

Due to the nature of facial perception, and how it is a special form of perception that humans are particularly attuned to [4, 21, 29], we expect that differences in perception of facial action units will not align with the magnitude of displacement on the mesh caused by the expression. We hypothesise that small displacements in salient regions (e.g., eyelids) will be more perceptually noticeable than larger displacements in less salient regions (e.g., puffing of cheeks), which will not be accurately reflected by the standard geometric and image error metrics. We also expect that due to social conditioning, sex and race will affect the perception of facial action units. It appears that female and male faces are observed differently, because the type and expressivity of particular emotions were found to be sex specific [22, 25, 43]. In addition, people perceive faces of their own race differently than other races [34, 54]

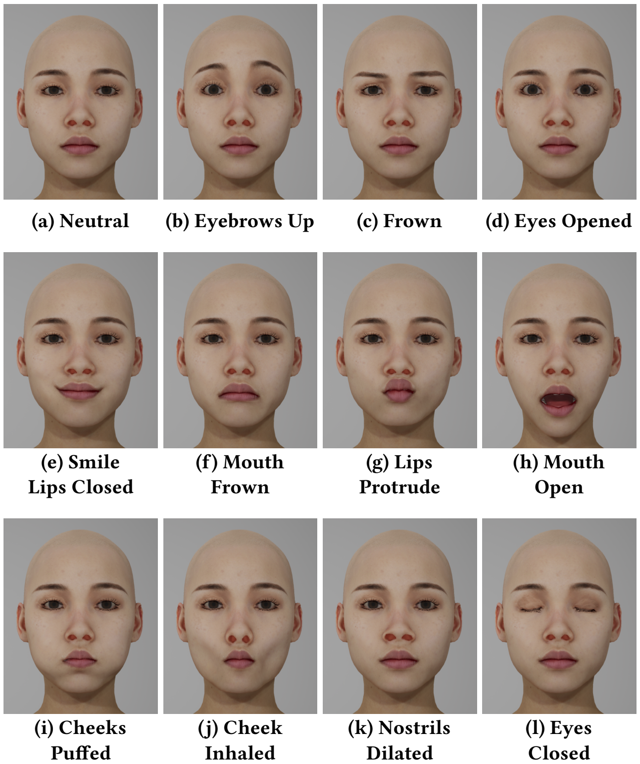
In this paper, we investigate the perceptual impact of a carefully selected range of expressive action units at varying activation levels across a number of characters of different race and sex. We then compare our qualitative perceptual results to quantitative metrics in order to determine whether the perceptual effect can be predicted easily. Geometric and image-based error metrics for triangle meshes are traditionally used for predicting mesh errors such as watermarking, simplification or lossy compression. However, we aim to determine if our question of perceived action unit importance can be predicted by simply calculating the *error* between the neutral pose and the expression blendshape, using common image and geometry error metrics. We investigate RMS, Hausdorff distance, and triangle difference for our geometric errors, as described in Section 6.1, and MSE and SSIM for our image metrics, as described in Section 6.2. We then perform linear regression analysis to determine if facial expression importance can be predicted using simple error metrics, or if a new perceptual metric specific to facial expressions is required.

We address a number of questions, such as:

- Are certain facial action units more perceptually noticeable than others?
- Are the same facial action units consistently noticeable across faces of different sex and race?
- Can we predict the saliency of facial action units using numerical metrics and which are better between geometry or image-based error metrics?

SAP '20, September 17–18, 2020, Washington DC

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn



(a) Neutral (b) Eyebrows Up (c) Frown (d) Eyes Opened

(e) Smile Lips Closed (f) Mouth Frown (g) Lips Protrude (h) Mouth Open

(i) Cheeks Puffed (j) Cheek Inhaled (k) Nostrils Dilated (l) Eyes Closed

Figure 1: The set of blendshapes used in our main experiment, shown on the Asian Female character at full activation (1.0).

Figure 2: The Mouth Open blendshape shown at the 5 activation levels used in our experiments, shown on the White Male character.

Our results could be used for optimisation of blendshape rigs through blendshape reduction. Blendshape rigs are memory intensive as they require large amounts of geometry to be stored in a blendshape matrix in order to display real-time animations. This also makes them computationally expensive due to the matrix-vector multiplication required [45]. By identifying and removing blendshapes of lower visual saliency, which equates to simply removing rows from the blendshape matrix, we can save both memory and computation required.

2 RELATED WORK

Error metrics have traditionally been used to assess dissimilarity between ground-truth geometry and geometry after undergoing simplification, watermarking, or lossy compression, with the goal of avoiding perceptible differences. The types of metrics used are *view-dependent* and *view-independent*, or image-based and model-based (see overview by Corsini et al. [13]). We are interested if these metrics can used in face-geometry perception.

Face perception is an interesting area of study, as humans have been shown to perceive faces in a different way to regular perception [4, 21, 29]. As well, the different areas of the face have been shown to be important in terms of speech and emotion perception [1, 3]. A great deal of research is ongoing in the areas of face recognition, detection, memory, the other-race effect and the effect of experience on face perception, critical features for recognition, and social evaluation of faces [41].

It has been shown that people perceive faces of their own race differently to faces of other races, with studies showing an own-race recognition memory advantage [34], as well as an own-race encoding advantage [54]. One explanation for this phenomenon is that people have more exposure to people of their own race, and there is evidence that experience can mitigate these other-group effects even if the experience is acquired during adulthood [9]. There is also a neurological basis for perceptual differences of faces based on both shape and pigment [2]. Social conditioning appears to play a role in face perception of different sexes as well. There are sex differences in the readiness to express certain emotions - males tend to more readily express anger [43], while females more frequently express fear and sadness [22]. Therefore, a female expression of anger can actually be perceived as more intense than a male expressing the same intensity of anger, due to the violation of viewers' expectations [25, 26]. For these reasons, we include a diverse set of characters in our experiment, ranging in race and sex, to generalise our results.

Many perceptual studies have been conducted with artificial data, i.e., static, usually black and white, and often posed photos [1]. With advanced game engines and highly realistic facial rigs, we can conduct perceptual experiments with control over the minute details of how stimuli are displayed. Using virtual characters as stimuli gives us a lot more control over the exact expressions being shown, which areas of the face are activated or not, the texture and lighting of the scene, and the replication of motion across characters. This allows us to do interesting experiments involving changing the level of activation of face or body motion [27] or creating artificial animations with varying face area onset orders [51].

The current state of the art for high quality facial animation is blendshape animation [32]. A blendshape is a mesh representing a certain shape, typically a simple movement like an eye blink or mouth open shape for facial blendshape. Animation is achieved by linearly combining a number of these blendshapes with the neutral face to create an expression. There is currently no consensus on what blendshapes a rig should contain, with the decision being left entirely to the artist. One solution is to use the Action Units from the Facial Action Coding System [19]. In theory, FACS breaks down facial expressions to their most basic components, making it a useful guideline for blendshape creation.

Blendshapes can be costly to create, however, they can be transferred from a template rig containing the desired shapes to a target character rig using Deformation Transfer [50]. The quality and personalisation of these blendshapes can be improved by providing examples of the target character face [33]. Similar to the question of which blendshapes should be included in a rig, there is no consensus on which examples should be provided to best improve a rig. Initial perceptual research has been done in this area [7] as well as a first attempt at creating an example suggestion algorithm [8]. Another method for personalising rigs is to use an actor's performance to train an existing set of blendshapes to better match the actor's face [36].

Optimisation of blendshape animation can be done in a few ways. Reducing mesh complexity is one method [23], however this causes correspondence issues between shapes. The animation itself can be optimised by passing blendshapes [35] and animation [16] to the GPU, and using GPGPU methods [14]. The most relevant optimisation method for this paper would be blendshape reduction, either removing blendshapes from a rig or from an animation. Naturally, this would reduce the expressivity of a rig and reduce the quality of animations, so identifying salient blendshapes as we attempt to do in this work is important. One area in particular where this optimisation method is applicable is optimisation for level of detail, where distance obscures the detail of the face so reduced quality is less perceptible.

Visual saliency of meshes is an interesting topic that leads to applications such as optimal viewpoint selection and mesh simplification. [30, 47]. This would be another route of investigation into the relationship between 3D geometry and face perception. Identifying the salient areas of the face, as well as optimal viewpoints, using mesh saliency methods and investigating those through perceptual experiments may give further insight into the special properties of face perception.

The recognition of Action Units from FACS using computer vision has been explored using facial component models, with AUs being recognised with greater than 95% accuracy [52]. Computer recognition of AUs is interesting to our work as it allows us to see the similarities and differences between human perception and computer vision. Most AUs were recognised correctly, with incorrect recognition attributing to either an additional similar AU being recognised (e.g. both Inner and Outer Brow Raiser being recognised when only one was present), or a similar AU being incorrectly recognised (e.g. Jaw Drop being recognised instead of Lips Part). It is noted that one of the pairs of AUs that were confused, Cheek Raiser and Lid Tightener, are confused by humans as well [12]. Recognition of AUs, as well as automatic recognition of intensity of AUs, has also been accomplished using pure deep learning methods, although with lower accuracy [46].

Neural network approaches for facial recognition require large amounts of data. While databases are available that contain real, in-the-wild data, it has been shown that generated databases can be used to train networks with a high degree of recognition accuracy [57]. Our stimuli creation pipeline may be of interest to such applications due to the high level of controllability.

Not all emotions are perceived equally. Happiness is most quickly recognised and least often confused with other emotions [6, 31, 42], while angry faces are more easily detected within a crowd [10]. There are also different areas of the face associated with recognition of each emotion.

For each emotional expression, specific parts of the expression appear to be more important for the classification of an emotion [48]. Since particular areas of the face are important for the recognition of emotion, different action units could potentially be more salient than others. The evidence supporting this suggests that areas, specialised for the perception of action units, exist in the brain (region pSTS). This could indicate that action units are a necessary precursor to categorization of emotion [49]. In addition, particular action units are responsible for a correct recognition of an emotion [56]: for happiness, this is the lip corner puller and parting of lips; for disgust, the most important are the raising and plucking of the lip. For fear, surprise, anger and sadness the regions around the eyes have the highest weights, with the lid raiser (exposing the sclera of the eyes) important for fear, and the lid tightener significantly most important for anger. Brows are important for sadness and both eyes and mouth contribute significantly to the recognition of surprise. There were also studies which used the information about individual action units to generate synthetic expressions. A gradual activation of specific action units resulted in detection of an expression [58]. Reverse engineering expressions based on perceptual relevance helped with improved facial recognition in artificial faces [11]. There is enough evidence to suggest that action units alone have a perceptually significant impact on emotion categorisation. However, it is unknown if certain action units are more salient than others because they are associated with a particular emotional expression.

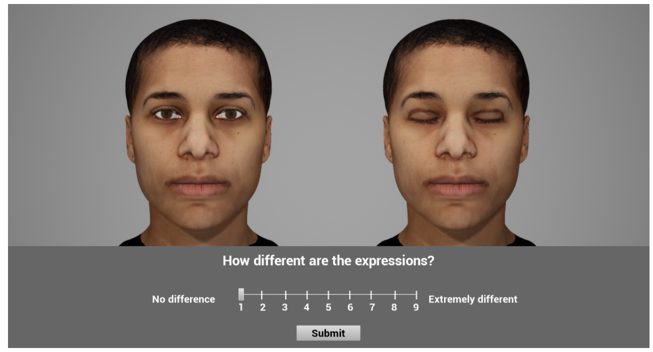


Figure 3: The experiment system in UE4 showing the character we used for training, which was also the template model used to create the experiment character rigs.

While the mouth is understandably a significantly attended to area due to its importance for emotional expression and communication [17, 40] and its size relative to other facial features, the eyes and eyebrows can also be considered highly important despite their considerably smaller size. Eyebrows are integral for emotional and conversational signals [18], and can alter the perception of the eyes [37], however they are important in their own right for face recognition [44] and not just in relation to how they change the perception of eyes.

3 STIMULI CREATION

We explored acquiring a range of high-resolution full-head meshes with semantically-matching AUs and diversity of facial features from open-source databases. However, to our knowledge, no such set exists, therefore we created our own data-set.

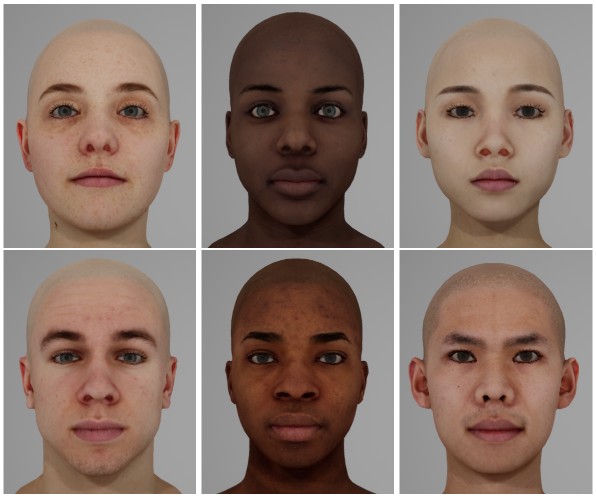


Figure 4: Neutral faces of the characters used in our experiment. Left: white, middle: black, right: asian faces. Top row shows the female faces, while the bottom row shows the male faces.

3.1 Model Acquisition

We first acquired a high-end photogrammetry-scanned *template model*, created by Eisko[1], a leading Digital Double company. The character had over 200 blendshapes, inspired by the FACS system [20] with additional shapes for emotion and speech (Figure 3). Our *experiment characters* were a set of 6 neutral faces (Figure 4) created utilising high resolution scan data, from 3D Scan Store[2].

One of the goals of this experiment was to obtain results that could be generalisable across different character faces, therefore we attempted to create a diverse set of stimuli by including 2 characters of each Asian, Black, and White race. Within each race group, there was 1 female and 1 male character.

3.2 Blendshape Transfer

In order to obtain a range of expressions for each of our experiment characters, we used the Russian 3D Scanner[3] Wrap 3.4 to transfer the topology of our template model to each of the neutral characters, using some feature points as guidance so that the semantics of the topology remained the same. We then used this wrapped mesh to warp the blendshapes of our template model to the experiment characters, thereby creating 6 new character rigs with equal topology and blendshapes. These characters can be seen in Figure 4. We chose not to include any hair on the characters as we are exclusively interested in facial features and wanted to avoid distracting elements.

3.3 Action Unit Selection

We carefully chose 11 blendshapes from the character's set of 200 for the experiment (see Table 1 and Figure 1). The AUs chosen are those which are important for emotion (AUs 2, 4, 5, 12, 15, 26, 38 [39, 56]), speech (AUs 18, 26 [38]), and those necessary for realistic and natural motion (AU 43 [28]). The cheeks have also been found to be important for facial recognition [5], so in order to fully cover potentially important features we also included cheek AUs 34 and 35. We also attempted to include opposite movements in each area, e.g. smile and frown.

[1]https://www.eisko.com/
[2]https://www.3dscanstore.com/3d-head-models/
[3]https://www.russian3dscanner.com/

AU	AU Name	Area
2	Eyebrows Up	Eyebrows
4	Frown	Eyebrows
5	Eyes Opened	Eyes
12	Smile Lips Closed	Mouth
15	Mouth Frown	Mouth
18	Lips Protrude	Mouth
26	Mouth Open	Mouth
34	Cheeks Puffed	Cheek
35	Cheek Inhaled	Cheek
38	Nostrils Dilated	Nose
43	Eyes Closed	Eyes

Table 1: Action Units included in the experiment.

3.4 Activation Levels

We are interested in whether the increase in onset of an AU linearly affects its perceptual importance, or whether there is a point at which the AU becomes more noticeable. For this reason, we investigate each AU at a number of different levels of activation. For each of these expressions, we show 5 activation levels: 0.2, 0.4, 0.6, 0.9, 1.0, (see Figure 2), with 1.0 being the maximum activation of that expression performed by the actor during the scanning process. In terms of blendshapes, this is simply a linear interpolation from the neutral face to the blendshape, with 1.0 being the fully activated expression (e.g. eyes fully closed) and each intermediate step being a transition from neutral to that expression, e.g. 0.4 of the eyes closed expression would be eyes almost half closed.

4 EXPERIMENT DESIGN

For each trial of the experiment, we displayed the Neutral expression on the left and the stimulus on the right, and asked the participants to answer "How different are the expressions?" using a slider. The slider ranged from 1 defined as "No Difference" to 9 defined as "Extremely Different". Participants were aware that the left image was always neutral. After each trial, a 1 second focus cross was displayed. We chose the Likert scale instead of a two-alternative forced-choice paradigm, in order to determine the relative saliency of AUs and activation levels, rather than simply whether the activation levels were noticed or not.

At the beginning of the experiment, participants conducted a training session, where they completed 11 trials showing the full activated blendshapes on the template character, which was not used in the main experiment (Figure 3). The idea of the training session was to calibrate participants to the most extreme examples of each AU.

Three hundred and sixty trials were shown to participants in random order, 12 AUs (including Neutral) x 5 activation levels × 6 characters. To avoid the experiment becoming too long, we used only one repetition of each character.

4.1 Participants

Twenty participants volunteered for the experiment (3 female, 16 male, 1 prefer not to answer; 8 were in the age range 18-27, 10 in 28-37, and 2 in 38-47). All reported medium or high familiarity with computer graphics and video games. As the characters we showed in this experiment varied in race, and there is a perceptual effect of one's own race and perception of other races [34, 54], we asked the participants to disclose their race (5 Asian, 13 White, 0 Black, 2 Other). Due to the fact that this was an in-laboratory experiment, recent restrictions related to the COVID-19 pandemic meant that we were unable to recruit a larger or more diverse sample of participants.

5 PERCEPTUAL EXPERIMENT RESULTS

We ran a 4-way repeated measures ANOVA on the Perceptual Difference results with the within factors Sex, Race, Action Unit, and Activation Level. Due to the imbalance between participant race and sex groups, we did not include these between-groups factors in the analysis. The ANOVA results can be seen in Table 2. We ran post-hoc analysis using Tukey's HSD tests throughout.

Factor	F(DFn, DFd) = F-value	p-value	η_p^2
Sex	F(1, 19) = 1.727	0.2	0.08
Race	F(2, 38) = 4.192	0.02*	0.18
Action Unit	F*(2.93, 55.58) = 123.8	0.00*	0.86
Activation	F*(1.21, 22.90) = 158.2	0.00*	0.89
Sex-Race	F(2,38) = 7.826	0.001*	0.29
Sex-AU	F(11, 209) = 2.99	0.001*	0.14
Race-AU	F(22, 418) = 6.885	0.00*	0.27
Sex-Activation	F(4, 76) = 2.887	0.03*	0.13
Race-Activation	F(8, 152) = 1.581	0.14	0.08
AU-Activation	F(44, 836) = 19.29	0.00*	0.50
Sex-Race-AU	F*(6.73, 127.86) = 5.301	0.00*	0.22
Sex-Race-Activation	F(8, 152) = 2.031	0.046*	0.10
Sex-AU-Activation	F(44, 836) = 0.979	0.5	0.05
Race-AU-Activation	F(88, 1672) = 1.592	0.001*	0.07
Sex-Race-AU-Activation	F(88, 1672) = 1.68	0.00*	0.08

Table 2: ANOVA interactions with dependent variable "Difference" from the perceptual results. (AU = Action Unit, * represents significant p-values, F* stand for Greenhouse-Geisser correction for violations of sphericity). Effects sizes are reported in the last column (η_p^2).

5.1 Character Sex & Race

There was no main effect of the sex of the character. There were some smaller interactions showing some individual differences in the models, but no interesting trends.

We found a main effect of character Race, where shape differences were less perceptible for Black characters overall than for Asian characters ($p < 0.02$). An interaction between Race and Sex gave further insight that shape differences were more perceptible for the Asian Female character than other characters except for the White Male ($p < 0.03$ for all). There was an interaction between Race and Activation Level, which showed the Frown and Cheeks

Figure 5: The Cheeks Puffed and Frown AUs shown at full activation on the Asian Female and Black Male characters.

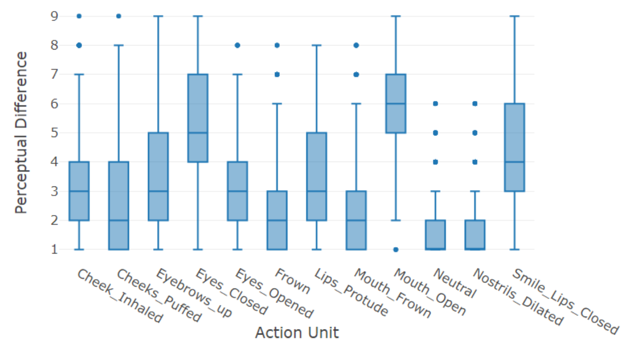


Figure 6: Main effect of AU from our experiment.

AU Name	Difference Rating
Mouth Open	5.97
Eyes Closed	5.2
Smile Lips Closed	4.18
Eyebrows Up	3.56
Lips Protude	3.55
Cheek Inhaled	3.24
Eyes Opened	3.15
Cheeks Puffed	2.77
Mouth Frown	2.56
Frown	2.22
Nostrils Dilated	1.78
Neutral	1.42

Table 3: The AUs ordered by average perceptual difference.

Puffed ($p < 0.02$) were the main AUs affected. This implies that differences in the cheek and frown expressions were less perceptible on Black characters (see Figure 5).

5.2 Activation Level

A main effect of Activation Level showed a significant increase in perceived differences as the activation increased, as expected.

There was no difference across all characters and AUs at the lowest Activation Level of 0.2. However, some characters were rated as relatively more different at higher Activation Levels. Specifically, Asian Female at 0.6 Activation Level was rated similarly than the AUs of some characters at the 0.8 level.

5.3 Action Units

Mouth Open, Eyes Closed, and Smile Lips Closed appeared to have a higher perceptual effect since the perceived differences were significantly higher when compared to all other AUs ($p < 0.02$). Nostrils Dilated had the smallest effect since it was not significantly different from the Neutral. See Figure 6 and Table 3.

Further interactions showed that Mouth Open was significantly more different than most other shapes ($p < 0.005$). Eyes Closed were also prominent on some characters, while Nostrils Dilated and Frown were not different from Neutral, for some characters.

Mouth Frown was the only AU to be rated significantly differently between the sexes ($p < 0.05$), with the female characters being rated as more different. This could potentially be related to the inverse effect of gender stereotyping increasing saliency of unexpected emotions seen in previous work (i.e., that females are perceived as more angry than males) [25]. We also found interactions with Race, as well as interactions with Race and Sex (see Table 2 and Appendix Figure 11). While we observed many significant differences from post-hoc tests, we did not observe any meaningful patterns.

6 ERROR METRICS

We now move to investigating the relationship between numerical error metrics and perception. We calculate each metric for each Activation Level of each AU, for each character. Each metric is calculated between the neutral face and the activated AU. For reference, we display these errors averaged across characters in the Appendix Tables 9 and 10.

6.1 Geometric Error Metrics

Root-Mean-Square. We calculate the RMS error between two meshes by getting the sum across all vertices $n \in N$ of the square root of the average of the square of each (x, y, z) component of each delta vertex $\delta \mathbf{v}_n$, i.e. the difference between that vertex position in the blendshape mesh and the same vertex in the neutral mesh, as described in Equation 1. A heatmap visualisation of RMS error can be see in Figure 7.

$$\delta_{RMS} = \sum_{n=1}^{N} \sqrt{\frac{1}{3} \delta \mathbf{v}_n^T \delta \mathbf{v}_n} = \frac{1}{\sqrt{3}} \sum_{n=1}^{N} \|\delta \mathbf{v}_n\| \tag{1}$$

Hausdorff. We calculate the bi-directional Hausdorff distance between two meshes of vertex sets N and M using the equation described in Equation 2 where $d(n, m)$ is the distance between two vertices.

$$d_H = \max \left\{ \sup_{n \in N} \inf_{m \in M} d(n, m), \sup_{m \in M} \inf_{n \in N} d(n, m) \right\} \tag{2}$$

Triangle Difference. We define the triangle difference δ_t of triangle $\mathbf{uvw}$ between two meshes A and B of equal connectivity in Equation 3.

$$\delta_t = \max \left\{ \frac{\mathbf{vu}_A}{\mathbf{vu}_B}, \frac{\mathbf{vu}_B}{\mathbf{vu}_A} \right\} + \max \left\{ \frac{\mathbf{vw}_A}{\mathbf{vw}_B}, \frac{\mathbf{vw}_B}{\mathbf{vw}_A} \right\} - 2 \tag{3}$$

Model	Gaussian	Poisson
RMS*AU	24878	24688 (Id)
RMS*AU+Race:Sex	24853	**24673** (Id)
RMS*AU*Sex*Race	24810	24749 (Id)
Hausdorff*AU*Sex*Race	24898	24814 (Id)
TriangleDifference*AU*Sex*Race	24819	24751 (Id)
SSIM*AU	26287	25591 (Id)
SSIM*AU*Sex*Race	24841	**24758** (log)
MSE*AU*Sex*Race	24839	24776 (Id)
Activation*AU*Sex*Race	24820	24759 (Id)

Table 4: Model comparison with AIC↓ to explain the perceived difference. Best link function reported for Poisson between Identity (Id), log and sqrt. Only Id link function has been tested with Gaussian distribution.

6.2 Image Error Metrics

To calculate our image metric results, we took screenshots of each stimulus during the experiment and cropped out a large amount of the empty space surrounding each head. An example of the crop can be seen in Figure 4. MSE and SSIM were calculated using scikit-image [53].

Mean-Square-Error. We calculate MSE by getting the per-pixel average error between images A and B, where N is the total number of pixels in the image, and $\mathbf{x}_n^A$ is the n^{th} pixel of image A.

$$\delta_{MSE} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n^A - \mathbf{x}_n^B \tag{4}$$

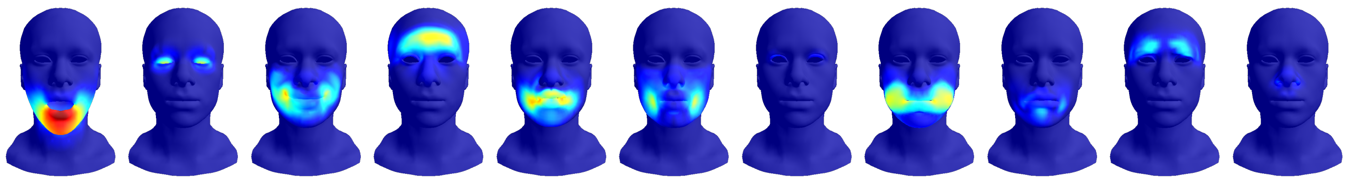
Structural Similarity Index Metric. SSIM is calculated as defined by Wang et al. [55] and using the default suggested parameters. It is designed to model the response of the human vision system and should correlate better to our perceptual results than MSE. As this metric measures similarity rather than difference, we sometimes invert this metric for better comparison with our other metrics. In these cases, we refer to it as 1-SSIM, as SSIM returns a value between 0 and 1 so to invert we simply subtract from 1.

6.3 Model Fit

In order to find the best model to describe the relationship between perception and error metrics, several Generalised Linear Models were tested and compared using Aikaike Information Criterion [15]. Poisson was compared to Gaussian distribution and it was found that Poisson distribution captures the discrete nature of the perceived difference better (all AICs with Poisson distribution are lower than their corresponding model with Gaussian distribution (Table 4)).

In the geometry domain, RMS is the selected as it is a popular metric to measure mesh deformation (Section 6.3.1). We note that RMS also achieves the lowest AIC in comparison to Hausdorff and Triangle Difference (Table 4) indicating it is the best predictor among the geometry metrics.

Similarly, in the image domain SSIM is chosen to create a model for predicting the perceived difference (Section 6.3.2). All models tested in Table 4 are acceptable for our dataset, as validated using their deviance (Poisson distribution) with χ^2 test [15].



(a) Mouth Open (b) Eyes Closed (c) Smile Lips Closed (d) Eyebrow Up (e) Lips Protrude (f) Cheek Inhaled (g) Eyes Opened (h) Cheeks Puffed (i) Mouth Frown (j) Frown (k) Nostrils Dilated

Figure 7: A heatmap visualisation of RMS error on the set of AUs used in our main experiment, shown on the template character at full activation (1.0). The AUs are shown in order of Perceptual Difference, from highest to lowest. No displacement is dark blue, while the largest displacement is shown in dark red in (as seen in (a) Mouth Open). The colours in order of displacement are: dark blue, blue, cyan, yellow, red, dark red.

6.3.1 Perceived difference w.r.t. geometry metrics. Table 5 shows the ANOVA when fitting a linear regression to explain the Perceived Difference (response variable) with explanatory variables RMS, AU and the 6 virtual characters used (captured with variables Sex and Race). As can be seen by the high values for Sum Sq., most of the perceived difference is explained using RMS and AU with their interactions (variable highlighted in green Table 5).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
RMS	1	11125.30	11125.30	6174.53	0.00
AU	11	5114.03	464.91	258.02	0.00
Sex	1	3.56	3.55	1.97	0.16
Race	2	24.91	12.45	6.91	0.001
RMS:AU	10	2103.48	210.35	116.74	0.00
RMS:Sex	1	7.98	7.98	4.43	0.035
AU:Sex	11	30.06	2.73	1.52	0.118
RMS:Race	2	21.93	10.96	6.09	0.002
AU:Race	22	132.01	6.00	3.33	0.00
Sex:Race	2	36.30	18.15	10.07	0.00
RMS:AU:Sex	10	10.55	1.05	0.59	0.827
RMS:AU:Race	20	63.22	3.16	1.75	0.02
RMS:Sex:Race	2	6.87	3.43	1.91	0.149
AU:Sex:Race	22	140.94	6.41	3.56	0.00
RMS:AU:Sex:Race	20	58.88	2.94	1.63	0.037
Residuals	7062	12724.35	1.80	NA	NA

Table 5: ANOVA interactions with dependent variable "Difference" and within factors RMS, Sex, Race and AU.

6.3.2 Perceived difference w.r.t. image metric. Table 6 shows the ANOVA when using the image metric SSIM, AUs and the 6 virtual characters used (captured with variables Sex and Race and shown in Figure 4). Based on a further ANOVA on MSE, we found the SSIM (Sum Sq. 6135) is systematically better than MSE (Sum Sq. 4736) in explaining the perceived difference, hence SSIM is the chosen image metric as a predictor to the perceived difference.

Interestingly, we found that SSIM as an image metric (captured in a 2D projective space) is not as powerful as the geometry metric RMS (measuring the deformation in 3D) for explaining the perceived difference (c.f. Sum Sq. first rows in Table 5 and 6). This is interesting, as our participants viewed the stimuli as a 2D projection, however their recorded perceived difference is better explained by geometric metrics computed from 3D meshes. A potential explanation may be that because faces are very familiar objects, a 3D representation is automatically imagined or inferred by participants when viewing 2D facial images. Despite this, having a model fitted using image metrics can be useful for prediction of perceived difference when geometry metrics are not available (e.g., for facial photograph comparisons).

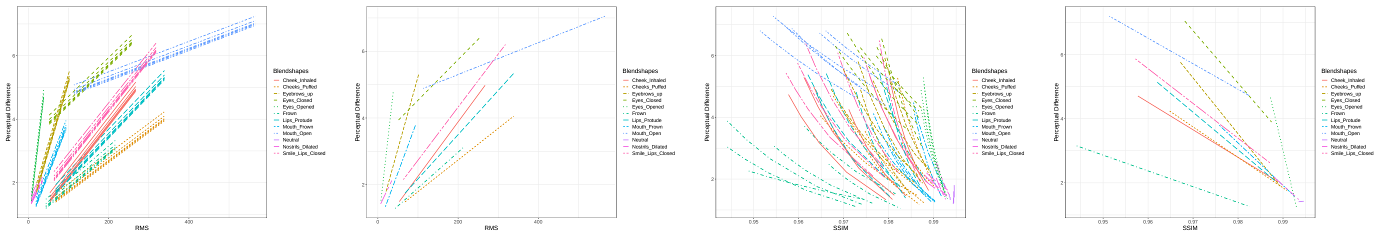
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SSIM	1	6135.88	6135.88	3393.49	0.00
AU	11	8500.78	772.8	427.4	0.00
Sex	1	204.14	204.14	112.9	0.00
Race	2	615.25	307.62	170.13	0.00
SSIM:AU	11	924.91	84.08	46.5	0.00
SSIM:Sex	1	23.57	23.57	13.03	0.00
AU:Sex	11	405.87	36.9	20.41	0.00
SSIM:Race	2	113.56	56.78	31.4	0.00
AU:Race	22	484.94	22.04	12.19	0.00
Sex:Race	2	645.41	322.7	178.47	0.00
SSIM:AU:Sex	11	97.8	8.89	4.92	0.00
SSIM:AU:Race	22	148.01	6.78	3.72	0.00
SSIM:Sex:Race	2	148.00	74.00	40.93	0.00
AU:Sex:Race	22	297.97	13.54	7.49	0.00
SSIM:AU:Sex:Race	22	100.09	4.55	2.52	0.00
Residuals	7056	12758.19241	1.808134	NA	NA

Table 6: ANOVA interactions with dependent variable "Difference" and within factors SSIM, Sex, Race and AU.

7 DISCUSSION

In this paper, we presented the first experiment on perceptibility of facial action units, and the relationship with numerical metrics describing the displacements. Our main contribution is our perceptual model, which will provide a starting point for the development of a universal perceptual error metric suitable for human faces. We will make available a GitHub repository with all of the data, and the R-code for the fit of the models, allowing others to build on our data investigating a larger range of faces, viewpoints, and facial action units.

Our other contribution is the results of our experiments which answer our questions from before. Firstly, we found that some facial action units were more perceptually noticeable than others, and provide a table showing the order of importance (Table 3). This perceptual ordering will be useful for developers for tasks



(a) RMS*AU+Race:Sex (b) RMS*AU (c) SSIM*AU*Race*Sex (d) SSIM*AU

Figure 8: Model-fit for perceived difference using geometry metric RMS (a and b) and image metric SSIM (c and d) as per models listed in Table 4. The 6 virtual characters behave in a similar fashion when using RMS (a) and are well captured with the simpler model (b) corresponding to the average model fit across the 6 virtual characters for each AU.

that require an order of blendshapes, such as level-of-detail blendshape reduction methods [14], or example creation for blendshape transfer [8]. By identifying and removing blendshapes of lower visual saliency, which equates to simply removing rows from the blendshape matrix, we can save both memory and computation.

We noted that diversity is missing from much of the psychology and computer vision research on recognition and perception of faces. Therefore, we included Asian, Black, and White characters with various skin tones. In general, there were no large differences at a per-Race or per-Sex level, implying that our results were generally consistent across characters. However, we did find an effect of Race in our ANOVA analysis (see Section 5.1), which showed that certain expressions were less perceptible on our Black characters. This result may have been due to our predominantly European and Asian participant pool, indicating that differences in perception of Black characters could be caused by the other-race effect [34, 54]. However, we also found that mouth and eye expressions were perceived similarly across race, indicating that the effect was not strong. A more diverse participant pool would be interesting to test for future experiments. The effect of Race may also be caused by the chosen background and lighting of the characters, which could be biased towards a clearer representation of lighter skinned characters, in particular the Asian Female character whose AUs were most noticeable.

We also hypothesized that male and female faces would be observed differently, but did not find much evidence for this, except that the Mouth Frown AU was more noticed on the female than on the male faces. We believe this could be related to the inverse effect of gender stereotyping increasing saliency of unexpected emotions, in this case the Mouth Frown could have been perceived as anger. This hypothesis will be investigated further in future work with a more balanced sampling of female and male participants.

Finally, we wanted to determine if standard geometry and image-based error metrics such as RMS and SSIM could predict the perceptibility of facial action units when compared to the neutral face. We did not find this to be the case, as our statistical models showed that each AU had a different relationship with the metrics (see different slopes in Figure 8). Also, we found that eye AUs (Eyes Closed and Eyes Opened) were rated highly in terms of perceptual difference (Table 3) despite their low error metric values (Appendix Table 7), showing that humans are relatively more sensitive to eye expressions than other areas of the face. Additionally, Frown was one of the least perceptually different AUs, however it had medium-level geometric error values compared to other AUs, and had either the highest or second-highest error using image-based metrics. In Figure 7, we provide a heatmap visualisation of RMS error of the action units we tested compared to the neutral mesh. Action Units are ordered from left to right from the most to the least noticeable (based on the Table 3). This shows more intuitively how small the RMS error is for certain highly salient action units such as Eyes Closed. These results show the need for a perceptually AU-based error metric for describing facial geometry alterations.

Interestingly, we found that image metrics were worse at predicting perceived differences than geometry metrics, even though the viewers only viewed the 3D geometry from a single viewpoint (i.e., they were not allowed to interact with the geometry). This implies that humans have a strong ability to infer 3D shape of faces from a 2D image, and that the pixel-based differences in the images do not capture these differences as well as 3D geometry comparisons. This is unlikely to hold true for different viewpoints besides the front view, but will be interesting to investigate in future work.

8 FUTURE WORK

It is well documented that face perception is a special type of perception that humans are particularly attuned to [4, 21, 29]. Using inverted faces [21, 24] is an effective way to analyze face-specific effects, which we will investigate in future work.

As this is an initial study into investigating the perceived importance of facial AUs, we limited our study to static expressions of single AUs. Naturally, perception of animated faces with combined expressions would be more complicated, particularly since specific AUs are important for the perception of emotions (e.g., AU 7 Lid Tightener for anger [56]). It might be the case that even if these AUs are not perceptually important according to our approach, removing them from a rig might alter the interpretation of emotion of a virtual human, which we will study in future work.

Our work may be helpful for optimising marker placement for animation, as those markers which are driving more salient areas of the face could be prioritised. This would also require investigation into animated stimuli first, however.

An interesting avenue to investigate will be using eye-tracking to determine the AUs that elicit greater visual attention and contrast this with the results of our study.

REFERENCES

[1] Ralph Adolphs. 2006. Perception and emotion: How we recognize facial expressions. *Current directions in psychological science* 15, 5 (2006), 222–226.

[2] Benjamin Balas and Charles A Nelson. 2010. The role of face shape and pigmentation in other-race face perception: An electrophysiological study. *Neuropsychologia* 48, 2 (2010), 498–506.

[3] Vicki Bruce and Andy Young. 1986. Understanding face recognition. *British journal of psychology* 77, 3 (1986), 305–327.

[4] Vicki Bruce and Andy Young. 2013. *Face perception.* Psychology Press.

[5] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces.* 205–211.

[6] Manuel G Calvo and Lauri Nummenmaa. 2008. Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology: General* 137, 3 (2008), 471.

[7] Emma Carrigan, Ludovic Hoyet, Rachel McDonnell, and Quentin Avril. 2018. A preliminary investigation into the impact of training for example-based facial blendshape creation.

[8] Emma Carrigan, Eduard Zell, and Rachel McDonnell. 2020. Expression Packing: As-Few-As-Possible Training Expressions for Blendshape Transfer. In *Proceedings of the 41st Annual European Association for Computer Graphics Conference (in press).* Eurographics Association.

[9] Viola Macchi Cassia, Marta Picozzi, Dana Kuefner, and Monica Casati. 2009. Short article: Why mix-ups don't happen in the nursery: Evidence for an experience-based interpretation of the other-age effect. *Quarterly Journal of Experimental Psychology* 62, 6 (2009), 1099–1107.

[10] Francesco Ceccarini and Corrado Caudek. 2013. Anger superiority effect: The importance of dynamic emotional facial expressions. *Visual Cognition* 21, 4 (2013), 498–540.

[11] Chaona Chen, Oliver GB Garrod, Jiayu Zhan, Jonas Beskow, Philippe G Schyns, and Rachael E Jack. 2018. Reverse Engineering Psychologically Valid Facial Expressions of Emotion into Social Robots. In *Int. Conf. on Automatic Face & Gesture Recognition.* 448–452.

[12] Jeffrey F Cohn, Adena J Zlochower, James Lien, and Takeo Kanade. 1999. Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology* 36, 1 (1999), 35–43.

[13] Massimiliano Corsini, Mohamed-Chaker Larabi, Guillaume Lavoué, Oldřich Petřík, Libor Váša, and Kai Wang. 2013. Perceptual metrics for static and dynamic triangle meshes. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 101–125.

[14] Timothy Costigan, Anton Gerdelan, Emma Carrigan, and Rachel McDonnell. 2016. Improving blendshape performance for crowds with GPU and GPGPU techniques. In *Proceedings of the 9th International Conference on Motion in Games.* ACM, 73–78.

[15] A. J. Dobson and A. G. Barnett. 2008. *An Introduction to Generalized Linear Models.* CRC Press, Third Edition.

[16] Bryan Dudash. 2007. Skinned instancing. *NVidia white paper* (2007).

[17] Hedwig Eisenbarth and Georg W Alpers. 2011. Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion* 11, 4 (2011), 860.

[18] Paul Ekman. 1979. About brows. *Emotional and conversational signals in: von Cranach, M./Foppa, K. ua (Hrsg.)(1979): Human Ethology, Cambridge* (1979).

[19] Paul Ekman and Wallace V Friesen. 1978. *Facial action coding system.* Consulting Psychologists Press, Stanford University.

[20] Paul Ekman and Wallace V. Friesen. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press.

[21] Martha J Farah, Kevin D Wilson, Maxwell Drain, and James N Tanaka. 1998. What is" special" about face perception? *Psychological review* 105, 3 (1998), 482.

[22] Agneta H Fischer, Patricia M Rodriguez Mosquera, Annelies EM Van Vianen, and Antony SR Manstead. 2004. Gender and culture differences in emotion. *Emotion* 4, 1 (2004), 87.

[23] Michael Garland and Paul S Heckbert. 1997. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques.* 209–216.

[24] James V Haxby, Leslie G Ungerleider, Vincent P Clark, Jennifer L Schouten, Elizabeth A Hoffman, and Alex Martin. 1999. The effect of face inversion on activity in human neural systems for face and object perception. *Neuron* 22, 1 (1999), 189–199.

[25] Ursula Hess, Reginald B Adams Jr, and Robert E Kleck. 2004. Facial appearance, gender, and emotion expression. *Emotion* 4, 4 (2004), 378.

[26] Ursula Hess, Sylvie Blairy, and Robert E Kleck. 1997. The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior* 21, 4 (1997), 241–257.

[27] Jessica Hodgins, Sophie Jörg, Carol O'Sullivan, Sang Il Park, and Moshe Mahler. 2010. The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception (TAP)* 7, 4 (2010), 22.

[28] Laurent Itti, Nitin Dhavale, and Frederic Pighin. 2003. Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation VI*, Vol. 5200. International Society for Optics and Photonics, 64–78.

[29] Nancy Kanwisher, Josh McDermott, and Marvin M Chun. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience* 17, 11 (1997), 4302–4311.

[30] Chang Ha Lee, Amitabh Varshney, and David W Jacobs. 2005. Mesh saliency. In *ACM SIGGRAPH 2005 Papers.* 659–666.

[31] Jukka M Leppänen and Jari K Hietanen. 2004. Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological research* 69, 1-2 (2004), 22–29.

[32] John P Lewis, Ken Anjyo, Taehyun Rhee, Mengjie Zhang, Frederic H Pighin, and Zhigang Deng. 2014. Practice and Theory of Blendshape Facial Models. *Eurographics (State of the Art Reports)* 1, 8 (2014), 2.

[33] Hao Li, Thibaut Weise, and Mark Pauly. 2010. Example-based facial rigging. In *Acm transactions on graphics (tog)*, Vol. 29. ACM, 32.

[34] D Stephen Lindsay, Philip C Jack, and Marcus A Christian. 1991. Other-race face perception. *Journal of applied psychology* 76, 4 (1991), 587.

[35] Tristan Lorach. 2007. Directx 10 blend shapes: Breaking the limits. *GPU Gems* 3 (2007), 53–67.

[36] Wan-Chun Ma, Mathieu Lamarre, Etienne Danvoye, Chongyang Ma, Manny Ko, Javier von der Pahlen, and Cyrus A Wilson. 2016. Semantically-aware blendshape rigs from facial performance measurements. In *SIGGRAPH ASIA 2016 Technical Briefs.* 1–4.

[37] Soyogu Matsushita, Kazunori Morikawa, Saya Mitsuzane, and Haruna Yamanami. 2015. Eye shape illusions induced by eyebrow positions. *Perception* 44, 5 (2015), 529–540.

[38] Zibo Meng, Shizhong Han, Ping Liu, and Yan Tong. 2018. Improving speech related facial action unit recognition by audiovisual information fusion. *IEEE transactions on cybernetics* 49, 9 (2018), 3293–3306.

[39] Marcello Mortillaro, Marc Mehu, and Klaus R Scherer. 2011. Subtly different positive emotions can be distinguished by their facial expressions. *Social Psychological and Personality Science* 2, 3 (2011), 262–271.

[40] Manfred Nusseck, Douglas W Cunningham, Christian Wallraven, and Heinrich H Bülthoff. 2008. The contribution of different facial regions to the recognition of conversational expressions. *Journal of vision* 8, 8 (2008), 1–1.

[41] Ipek Oruc, Benjamin Balas, and Michael S Landy. 2019. Face perception: A brief journey through recent discoveries and current directions.

[42] Romina Palermo and Max Coltheart. 2004. Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers* 36, 4 (2004), 634–638.

[43] E Ashby Plant, Janet Shibley Hyde, Dacher Keltner, and Patricia G Devine. 2000. The gender stereotyping of emotions. *Psychology of Women Quarterly* 24, 1 (2000), 81–92.

[44] Javid Sadr, Izzat Jarudi, and Pawan Sinha. 2003. The role of eyebrows in face recognition. *Perception* 32, 3 (2003), 285–293.

[45] Jaewoo Seo, Geoffrey Irving, JP Lewis, and Junyong Noh. 2011. Compression and direct manipulation of complex blendshape models. *ACM Transactions on Graphics (TOG)* 30, 6 (2011), 1–10.

[46] Zhiwen Shao, Zhilei Liu, Jianfei Cai, Yunsheng Wu, and Lizhuang Ma. 2019. Facial action unit detection using attention and relation learning. *IEEE Transactions on Affective Computing* (2019).

[47] Philip Shilane and Thomas Funkhouser. 2007. Distinctive regions of 3D surfaces. *ACM Transactions on Graphics (TOG)* 26, 2 (2007), 7–es.

[48] Marie L Smith, Garrison W Cottrell, FrédéAric Gosselin, and Philippe G Schyns. 2005. Transmitting and decoding facial expressions. *Psychological science* 16, 3 (2005), 184–189.

[49] Ramprakash Srinivasan, Julie D Golomb, and Aleix M Martinez. 2016. A neural basis of facial action recognition in humans. *Journal of Neuroscience* 36, 16 (2016), 4434–4442.

[50] Robert W Sumner and Jovan Popović. 2004. Deformation transfer for triangle meshes. *ACM Transactions on graphics (TOG)* 23, 3 (2004), 399–405.

[51] Marie-Hélène Tessier, Chloé Gingras, Nicolas Robitaille, and Philip L Jackson. 2019. Toward dynamic pain expressions in avatars: perceived realism and pain level of different action unit orders. *Computers in Human Behavior* 96 (2019), 95–109.

[52] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence* 23, 2 (2001), 97–115.

[53] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. 2014. scikit-image: image processing in Python. *PeerJ* 2 (6 2014), e453. https://doi.org/10.7717/peerj.453

[54] Pamela M Walker and James W Tanaka. 2003. An encoding advantage for own-race versus other-race faces. *Perception* 32, 9 (2003), 1117–1125.

[55] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions*

on image processing 13, 4 (2004), 600–612.
[56] Martin Wegrzyn, Maria Vogt, Berna Kireclioglu, Julia Schneider, and Johanna Kissler. 2017. Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PloS one* 12, 5 (2017).
[57] Tian Xu, Oliver Garrod, Steven H Scholte, Robin Ince, and Philippe G Schyns. 2018. Using psychophysical methods to understand mechanisms of face identification in a deep neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1976–1984.
[58] Hui Yu, Oliver GB Garrod, and Philippe G Schyns. 2012. Perception-driven facial expression synthesis. *Computers & Graphics* 36, 3 (2012), 152–162.

APPENDIX

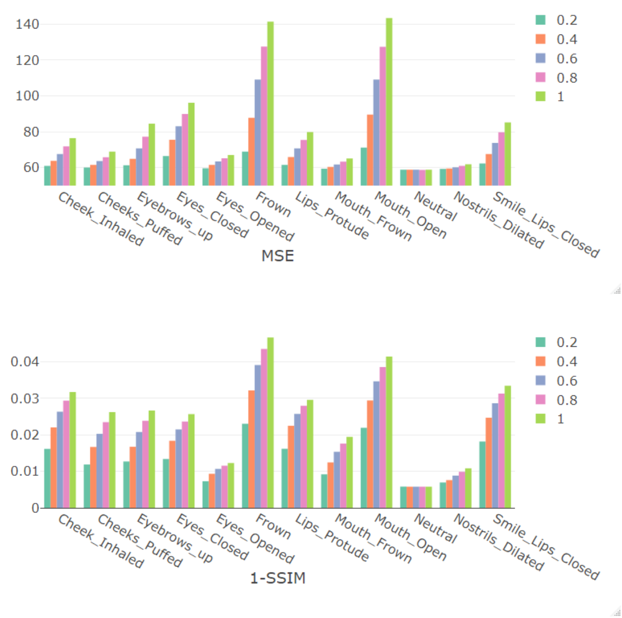


Figure 10: Image errors averaged across characters.

Figure 9: Geometry errors averaged across characters.

Order	RMS	Hausdorff	Tri. Diff	MSE	1-SSIM	Perceptual
1	Mouth Open	Mouth Open	Lips Protrude	Mouth Open	Frown	Mouth Open
2	Lips Protrude	Cheeks Puffed	Eyes Closed	Frown	Mouth Open	Eyes Closed
3	Cheeks Puffed	Cheek Inhaled	Smile Lips Closed	Eyes Closed	Smile Lips Closed	Smile Lips Closed
4	Smile Lips Closed	Lips Protrude	Cheek Inhaled	Smile Lips Closed	Cheek Inhaled	Eyebrows Up
5	Cheek Inhaled	Eyes Closed	Cheeks Puffed	Eyebrows Up	Lips Protrude	Lips Protude
6	Eyes Closed	Smile Lips Closed	Frown	Lips Protrude	Eyes Closed	Cheek Inhaled
7	Frown	Frown	Mouth Open	Cheeks Puffed	Eyebrows Up	Eyes Opened
8	Eyebrows Up	Eyebrows Up	Eyebrows Up	Cheek Inhaled	Cheeks Puffed	Cheeks Puffed
9	Mouth Frown	Mouth Frown	Mouth Frown	Eyes Opened	Mouth Frown	Mouth Frown
10	Eyes Opened	Nostrils Dilated	Eyes Opened	Mouth Frown	Eyes Opened	Frown
11	Nostrils Dilated	Eyes Opened	Nostrils Dilated	Nostrils Dilated	Nostrils Dilated	Nostrils Dilated

Table 7: Our AUs ordered by each of the different error metrics and the perceptual results, averaged across all characters and Activation Levels.

Figure 11: Interaction of AU, Sex, and Race from our experiment.