Perception of Linear and Nonlinear Motion Properties using a FACS Validated 3D Facial Model

Darren Cosker*
Department of Computer Science
University of Bath

Eva Krumhuber†
Swiss Center for Affective Sciences
University of Geneva

Adrian Hilton‡
Centre for Vision, Speech and Signal Processing
University of Surrey

Abstract

In this paper we present the first Facial Action Coding System (FACS) valid model to be based on dynamic 3D scans of human faces for use in graphics and psychological research. The model consists of FACS Action Unit (AU) based parameters and has been independently validated by FACS experts. Using this model, we explore the perceptual differences between linear facial motions – represented by a linear blend shape approach – and real facial motions that have been synthesized through the 3D facial model. Through numerical measures and visualizations, we show that this latter type of motion is geometrically nonlinear in terms of its vertices. In experiments, we explore the perceptual benefits of nonlinear motion for different AUs. Our results are insightful for designers of animation systems both in the entertainment industry and in scientific research. They reveal a significant overall benefit to using captured nonlinear geometric vertex motion over linear blend shape motion. However, our findings suggest that not all motions need to be animated nonlinearly. The advantage may depend on the type of facial action being produced and the phase of the movement.

CR Categories: I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism—Animation; H.1.2 [Models and Principles]: User/Machine Systems—Human Information Processing J.4 [Computer Application]: Social and Behavioural Sciences—Psychology

Keywords: FACS, 3D Dynamic Facial Capture and Animation

1 Introduction

Human realistic facial animation remains a challenging goal in computer graphics and animation research. The attainment of this depends on a number of factors, including both static and dynamic realism. Recent advances in skin rendering [Weyrich et al. 2006] have pushed static realism to new levels. Furthermore, advances in stereo capture now allow us to observe dense 3D dynamic facial surface changes in real-time. This provides substantially more information regarding high resolution dynamic facial movement than optical motion-capture alone. However, the achievement of human dynamic realism does not lie exclusively in the domain of computer graphics and animation. There is also a strong dependency on what we can learn from psychophysical and perceptual experimentation with faces. For example, subtle variations in expression timing [Krumhuber et al. 2007] and even a slight dampening of dynamic motion [Theobald et al. 2009] have been shown to strongly influence how expressions and individuals are perceived.

Research using dynamic facial expressions in computer science and psychology is largely focused on facial models with control parameters based on the Facial Action Coding System (FACS) [Ekman et al. 2002]. FACS provides detailed descriptions of 44 facial actions – termed Action Units (AUs) – which attempt to encompass the basic set of distinct facial movements capable by a face. This gives researchers a standard for coding, quantifying and communicating results. Given such a model experimenters may manipulate these parameters to measure the perceptual effect of AU variations in a controlled manner. FACS is also used extensively as a basis for animation systems in video games and movies [Sagar 2006; Duncan 2009]. Hence, it has a major role to play in both facial animation and perception research.

In this paper we present a 3D facial model for use in graphics and psychological research exploring the perception of facial expression dynamics. Our model has been independently validated by certified FACS experts to ensure that each of its AUs are accurately portrayed. This makes our model the first FACS valid animation model based on either static or dynamic 3D scans of human faces. The model consists of AU parameters which control facial changes as observed using dense 3D stereo surface scanning (see Figure 1). As the intensity of a parameter gets increased, the surface and texture of the facial model moves and transforms in a highly realistic manner. The data set used to train the model is recorded from an individual performing a range of different AUs. We believe that such a model is indispensable for both computer graphics and psychological research, since knowledge of complex facial behaviors in different social and interactive contexts can help to create more realistic and emotive facial animations [Krumhuber and Tamarit 2010; Roesch et al. 2010].

Using this model, we explore the perceptual differences between linear AU facial motions – represented by a linear blend shape approach – and the same AU facial movements as observed from a real face. These real movements are those captured using the dynamic 3D facial scanner, and are re-synthesized through the facial model. Such a perceptual study is important since linear blend shape animation approaches are highly popular in animation, whereas the natural facial movements during expressions are typically geometrically nonlinear as opposed to linear. This geometric nonlinearity is well known, and physical wrinkling and elasticity models are often employed to approximate the motion [Bickel et al. 2007]. Using our 3D dynamic scanner, we capture those geometric nonlinearities and then recreate them in our model.

The comparison of geometric linearity versus nonlinearity in the face raises several questions. For example, is there a perceptual advantage of nonlinear animation types over linear ones? If not,

*e-mail: D.P.Cosker@cs.bath.ac.uk
†e-mail: Eva.Krumhuber@unige.ch
‡e-mail: A.Hilton@surrey.ac.uk

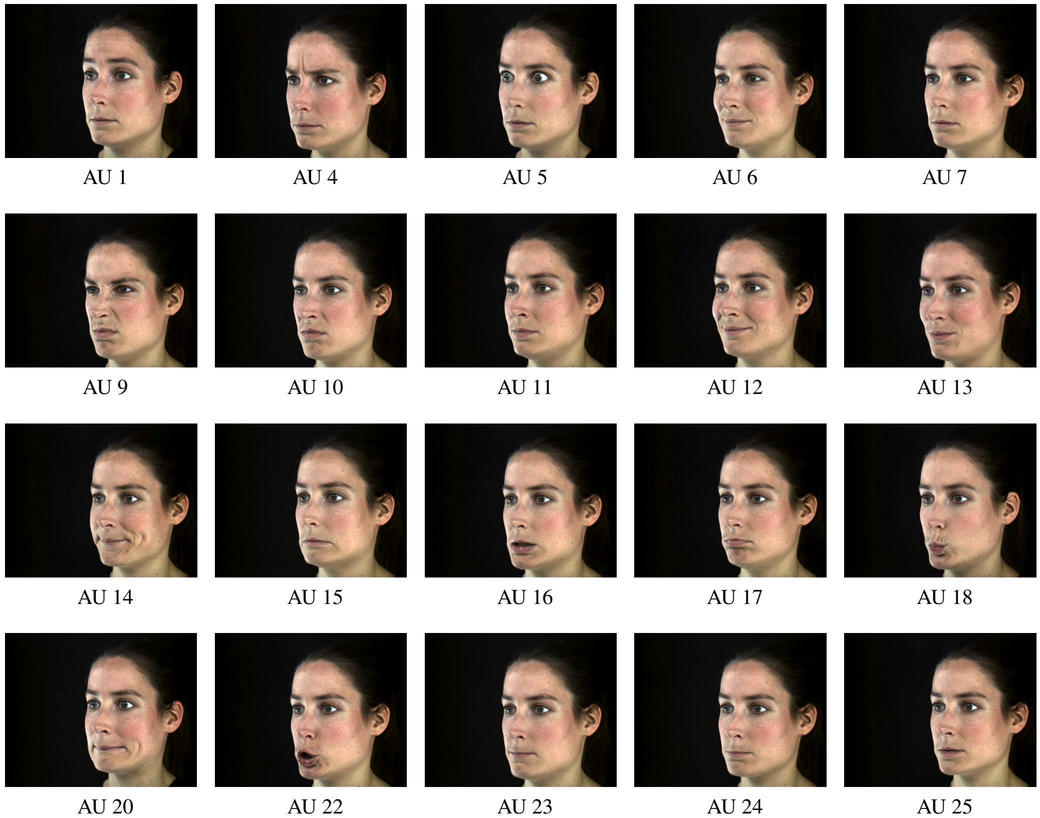


Figure 1: *Images from captured AU sequences as acquired from one of the color cameras on the 3D stereo capture system. The peak AU image of each sequence is shown.*

then linear blend shape models may be entirely adequate for facial animation purposes. However, it may also be the case that certain facial actions come across as being less natural if linear movements are used. We use FACS as the basis to explore the perception of these motion types since it provides the best systematic description of individual facial movements. To our knowledge, this is the first experiment of its kind to address the question of whether linear geometric facial motions provided by a blend shape model are as perceptually similar as nonlinear geometric (i.e. naturally deforming) motions of the same actions. In our experiments, we also account for linear and nonlinear *temporal* motion, by matching the *ease-in* and *ease-out* of the linear geometric motion in as natural a way as possible (see Sections 4 and 5).

We next describe related work (Section 2) and then our animation system (Sections 3 and 4). Design of our experiment is outlined in Section 5, and our results are presented in Section 6. In Sections 7 and 8 we then discuss the implications of our results and conclude.

2 Related Work

3D Facial Animation Since the early work of Parke [Parke and Waters 1996] advances in performance capture and control techniques have continued to advance the field of facial animation. FACS has become a popular reference for creating facial animation systems in movies [Sagar 2006]. Curio *et al* [Curio et al. 2007] presented a realistic 3D model based on FACS for the systematic investigation of moving faces. AUs were acquired using an ABW 3D stereo capture device, and a framework was presented for animating this model using motion-capture. However, since the AUs were not validated by FACS experts it is unclear how valid the model would be in situations where the accurate production of an AU is essential [Krumhuber and Tamarit 2010; Roesch et al. 2010]. Moreover, the employed motion capture approach only provides movement information for sparse 3D facial positions and does not capture dense surface deformation detail. Optical motion-capture based on sparse markers[Williams 1990] also has the drawback that wrinkle detail is not captured, and this has motivated work on artificially re-inserting approximated wrinkle detail [Bickel et al. 2007].

3D stereo capture approaches can overcome many of the limitations of sparse optical motion capture. Zhang *et al* [Zhang et al. 2004] developed a real-time 3D stereo surface acquisition technique and used this to capture both nonlinear geometric surface facial deformations as well as color texture during a performance. The color information observed contains expression wrinkle detail too subtle to capture during the stereo acquisition process. Ma *et al* [Ma et al. 2008] more recently developed a photometric stereo based capture system capable of observing nonlinear changes in skin pores and fine-scale wrinkles.

Unlike previous work using FACS [Curio et al. 2007; Sagar 2006], we employ certified FACS coders [1] to ensure that AUs performed in our captured data are valid representations. This makes the model fully reliable and valid for psychophysical experimentation. We also capture the AU movements using real-time stereo 3D capture technology, and create facial parameters that produce realistic nonlinear movements when being manipulated. The underlying model

[1]Certified FACS coders are experts in analyzing facial movements in terms of their component AUs. Becoming certified requires substantial training followed by final examination.

is akin to a 3D morphable model [Blanz and Vetter 1999] However, due to the real time nature of the scanning device, statistical information is available for the full dynamic evolution of each expression both in terms of its geometric and temporal information.

3D Faces and Dynamic Motion in Perception Research Jiang *et al* [Jiang et al.] and Knappmeyer *et al* [Knappmeyer et al. 2003] both utilized the morphable model of Blanz and Vetter to explore the role of 3D information in representations of familiar faces and the use of facial motion and form when processing identity. The motion-capture AU based model of Breidt *et al* [Breidt et al. 2003] has been used in several studies investigating expression recognition given different combinations of local expressions [Griesser et al. 2007] as well as the effects of different rendering conditions on expression perception [Wallraven et al. 2008].

Experiments based on the subtle manipulation of facial dynamics have revealed interesting effects. By manipulating the dynamic properties of smile expressions in a trust game Krumhuber *et al* [Krumhuber et al. 2007] showed a significant difference in the way people cooperated with a realistic avatar. Similarly, Cunningham and Wallraven [Cunningham and Wallraven 2009] demonstrated the importance of dynamic motion in the perception and recognition of facial expressions.

While methods for the parameterization and animation of 3D stereo data have begun to emerge, perceptual experimentation with such data is still in its infancy. Zhang *et al* [Zhang et al. 2004] briefly commented on how the perception of dynamic movement improved by using nonlinear temporal manipulation as opposed to linear interpolation but conducted no empirical tests. In a study by Wallraven *et al* [Wallraven et al. 2008], a comparison on the recognition of linear versus non-linear expressions using different facial model representations was carried out. However, only the play-back of captured dynamic 3D data was utilized, and no parameterization of the data was developed such that different nonlinear temporal aspects could be manipulated later on. In addition, the expressions tested were holistic, as opposed to AU based.

3 FACS Data Capture

Facial poses recorded from a FACS expert were used for building our model. In the following we present results with respect to the capture of 20 FACS validated AUs and their subsequent modelling. FACS validation of our recorded data and the output of our animation model (see Section 4) was performed blindly using two independent FACS coders.

Our acquisition device consisted of a 3DMD active stereo 3D capture system [3DM]. This uses a projected infra-red speckle pattern to calculate stereo correspondence and produces an accurate 3D surface reconstruction. The system has a capture rate of 60Hz, and therefore provides smooth temporal acquisition of fast facial movements. Each raw facial surface scan contains approximately 30K vertices along with 1280 x 1024 pixel color UV maps. Table 1 lists the captured AUs and provides notes on their validation as performed on the output of the animation model. Note that matching AU target and validation codings indicate that the animation modeling pipeline (Section 3) is faithfully preserving the AU and not corrupting its representation. Figure 1 shows images of AU peaks from each recorded sequence as captured by one of the color acquisition cameras in the stereo capture system. Note that while validation was only applied to the peak image, we ensured that no other AU interference occurred during the capture.

AU Target	FACS Name	AU Validation
0	Neutral	0
1	Inner Brow Raiser	1
4	Brow Lowerer	4^a
5	Upper Lid Raiser	5
6	Cheek Raiser	6
7	Lid Tightener	7
9	Nose Wrinkler	9^b
10	Upper Lip Raiser	10
11	Nasolabial Furrow Deepener	11^a
12	Lip Corner Puller	12
13	Cheek Puffer	13
14	Dimpler	14^c
15	Lip Corner Depressor	15
16	Lower Lip Depressor	16
17	Chin Raiser	17
18	Lip Pucker	18
20	Lip Stretcher	20
22	Lip Funneler	22
23	Lip Tightener	23
24	Lip Pressor	24
25	Lips Part	25

Table 1: *Captured and Validated Action Units (AUs). Note: [a] traces of AU38 (Nostril Dilate), [b] traces of AU17 (Chin Raiser) at the peak, [c] traces of AU12 (Lip Corner Pull) at the peak. AU 17 is a common concomitant of an intensely posed AU9 and does not interfere with its appearance. In AU14 the lip corners are pulled inwards and often also angle up, making it difficult to distinguish it from the smile expression (AU12). Generally, the term* trace *refers to the lowest intensity level of an AU and to very slight – sometimes barely noticable – activity of the facial action.*

4 Dynamic 3D Facial Modelling Pipeline

We have developed a pipeline that creates a highly detailed statistical 3D facial model from acquired raw 3D surface and UV map data. In this Section, we describe how we constructed the model.

Pre-processing We first apply a de-noising filter [Sun et al. 2007] to the 3D sequence in order to remove surface noise. We then create cylindrically unwrapped UV maps for each sequence.

Optical Flow Tracking We select a single mesh from a neutral facial pose to act as our *canonical mesh* [Parke and Waters 1996], i.e. a mesh with a known vertex number and topology. We then select a set of 54 landmarks on the canonical UV map relating to features such as mouth and eye corners and freckles. We next use a point tracker utilizing phase-based optical flow [Guatama and Hulle 2002] to track similar landmark positions through the UV sequences for each AU.

Registration Using correspondences from our tracked points we next align each UV map with the canonical UV map. We then calculate new modified positions for the canonical 3D mesh by: (1) calculating the barycentric coordinate of a canonical UV texture coordinate with respect to the surrounding UV coordinates in the registered mesh, (2) updating the corresponding canonical mesh 3D vertex coordinate using the surrounding 3D coordinates in the registered mesh. This process is repeated for each scan in each AU sequence, resulting in the entire data sequence having the same topology and number of vertices. Finally, rigid head movement is removed by applying ICP registration to the entire data set.

4.1 Morphable Modelling and AU Parameterization

Our initial pipeline registers all UV images to a common *shape free* space [Blanz and Vetter 1999], meaning that we may use the same number of pixels to represent the face area in each UV image. Texture (i.e. color) information is therefore represented using column vectors $\mathbf{t}$. The set of all UV textures across all

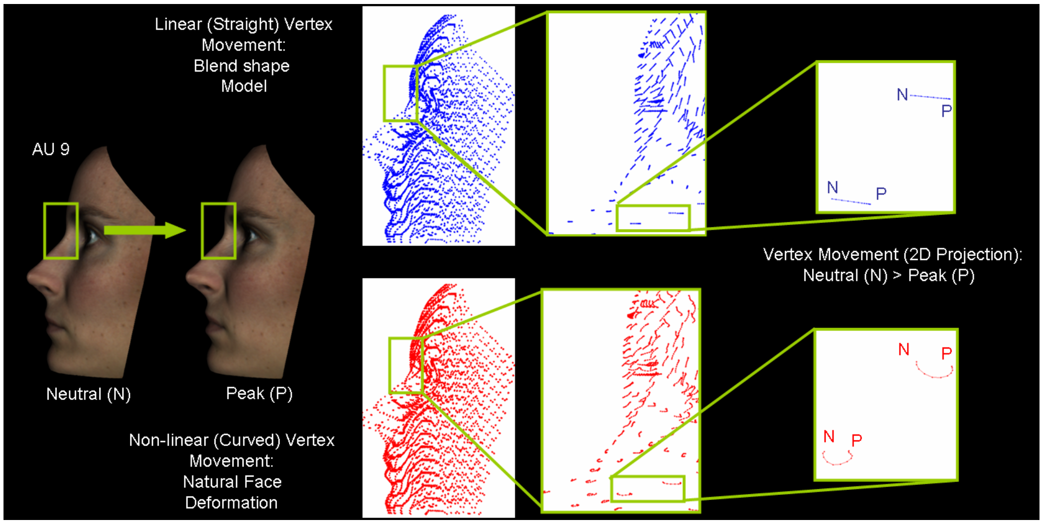


Figure 2: *Linear vs nonlinear geometric vertex motion for AU 9. Note that the vertices (subsampled) follow a curve in the nonlinear animations (recorded from a real facial performance) and a straight line in the linear animations (created using a blend shape model).*

AUs is represented as $\mathbf{T}^{AU}$. The sequence of UV maps corresponding to a specific AU is represented as $\mathbf{T}^{AU,i}$, where $i \in \{1, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 22, 23, 24, 25\}$, i.e. $\mathbf{T}^{AU,4}$ is the sequence of UV texture maps for AU 4.

Facial surface mesh data may also be represented using similar notation. The initial pipeline provides each 3D mesh with the same number of vertices and the same topology. A column vector of 3D vertex data for a single mesh is therefore represented as $\mathbf{s} = [x_1, y_1, z_1, \ldots, x_N, y_N, z_V]$, where N is the number of vertices in the mesh. The set of all 3D meshes across all AUs is represented as $\mathbf{S}^{AU}$. The sequence of 3D meshes corresponding to a specific AU is represented as $\mathbf{S}^{AU,4}$, where $i \in \{1, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 22, 23, 24, 25\}$, i.e. $\mathbf{S}^{AU,4}$ is the sequence of 3D meshes for AU 1.

In its raw form $\mathbf{T}_{AU}$ requires a great deal of memory for storage, especially since full color information is represented in the matrix. We therefore use Principle Component Analysis (PCA) to compress both the raw texture and shape data. This allows the dynamic AU sequences to be stored in a far more efficient manner. The process we follow is similar to that of 3D morphable model construction [Blanz and Vetter 1999]. PCA performs a basis transformation to a new orthogonal coordinate system represented by the eigenvectors $\mathbf{V}_s$ and $\mathbf{V}_t$ of the covariance matrices of the data sets $\mathbf{S}^{AU}$ and $\mathbf{T}^{AU}$. Given these matrices, any vector $\mathbf{s}$ or $\mathbf{g}$ may be represented as a linear weighted combination of the eigenvectors. Defining these weights as $\mathbf{b}_s$ and $\mathbf{b}_t$ for shape and texture respectively, we may write this linear combination as

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{V}_s \mathbf{b}_s \qquad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{V}_t \mathbf{b}_t \tag{1}$$

where $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ are the mean shape and texture vectors respectively.The eigenvectors form the columns of the matrices $\mathbf{V}_s$ and $\mathbf{V}_t$, and are organized in descending order according to their eigenvalues, i.e. the proportion of total variation in the data set. It is typically the case that the first few eigenvectors in each matrix represent the most significant proportion of data variation, which allows a large number of the columns of $\mathbf{V}_s$ and $\mathbf{V}_t$ to be removed. This means that the sizes of $\mathbf{b}_s$ and $\mathbf{b}_t$ can be made significantly smaller than $\mathbf{s}$ and $\mathbf{v}$. Each shape and texture vector is therefore converted to reduced set of weights using

$$\mathbf{b}_s = \mathbf{V}_s^T(\mathbf{s} - \bar{\mathbf{s}}) \qquad \mathbf{b}_t = \mathbf{V}_t^T(\mathbf{t} - \bar{\mathbf{t}}) \tag{2}$$

The raw sequences of mesh and texture data may now be stored in a far more compact manner. Similarly, animation can now also be carried out more efficiently by just manipulating the values $\mathbf{b}_s$ and $\mathbf{b}_t$. New vectors $\mathbf{s}$ and $\mathbf{v}$ may be created from new sequences of weights by simply projecting back onto the eigenvectors as in (1).

Using this compact representation, we store AU sequences in their weighted forms $\mathbf{b}_s^{AU,i}$ and $\mathbf{b}_t^{AU,i}$, where $\mathbf{b}_s^{AU,i}(j)$ and $\mathbf{b}_t^{AU,i}(j)$ are the weights for shape and texture at time *j* for AU *i*.

5 Evaluation of Linear vs Nonlinear AU Motion

We wished to assess the perceptual difference between AUs synthesized using a linear blend shape model with appropriate ease-in and ease-out (i.e. acceleration and deceleration) [Parke and Waters 1996] against a model which is replaying facial movement parameters from the same AUs observed from real facial motion (e.g. the weighted AU sequences defined in Section 4.1 and estimated from (2)). A model which stores or encodes geometric vertex displacements for each frame of a dynamic sequence will be more expensive in terms of computation and memory than a linear blend shape model. Therefore, a comparison of the two model types from a perceptual point of view has crucial implications when designing a facial model.

First, we ask an important question: How different is continuously observed facial motion from motion synthesized using a linear blend shape model? As we shall see the comparison is largely one of linear versus nonlinear geometric vertex motion.

5.1 Definitions and Measurement of Facial Motion

When the face deforms, it typically does so in a geometrically (or spatially) nonlinear manner [Bickel et al. 2007]. To demonstrate this, Figure 2 shows the geometric vertex motion of AU 9 (Nose Wrinkler) as re-created by a linear blend shape model, and as observed from a real face (and after re-synthesis through our model). In the linear animation (blue vertices on top row) the vertices of each facial expression move in a straight line from the neutral expression to the peak expression of the AU. In the nonlinear animation – based on continuously observed facial motion (red vertices

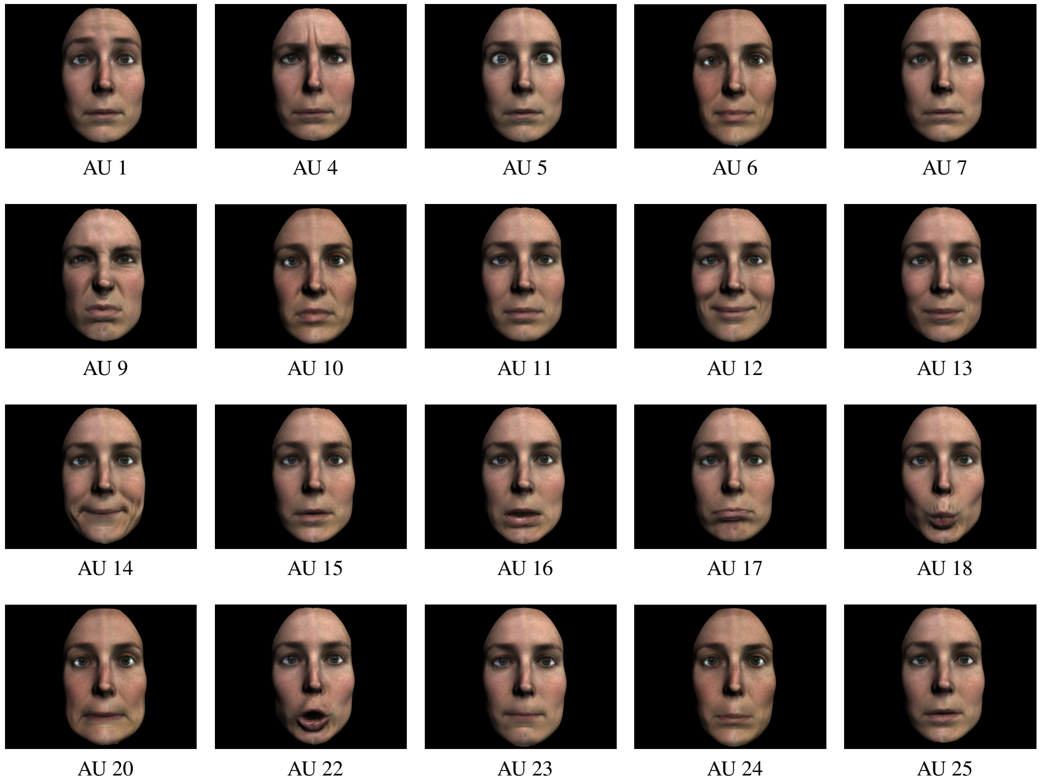


Figure 3: *Synthesized 3D faces at the* peak *AU position. Transition towards the peak is different depending on whether the linear blend shape model or playback of the captured dynamic movement (nonlinear) is used.*

on bottom row) – the vertices do not move in a straight line, but rather in a curve (red vertices on bottom row). It is the perceptual effect of these different animation types that we seek to compare in this study [2].

For our experiment nonlinear AU animations were created by re-synthesizing animations from the weights $\mathbf{b}_s^{AU,i}(j)$ and $\mathbf{b}_t^{AU,i}(j)$ for each AU. We also wished to assess whether the perception of motion types was different for the onset and offset. This was to determine whether contraction of the facial muscles (during onset) caused perceptually more or less geometrically linear motion than during relaxation stage (during offset). We therefore produced different animations for the *onset* of an AU and the *offset* of an AU. Onset AU animations ran from the neutral expression to the apex of the AU. Offset animations ran from the point where the expression begins to relax back to neutral, down to the neutral rest pose.

In terms of indices j of the parameter weights for shape and texture, onsets were defined as $j = 1, \dots P_{On}^{AU,i}$, and offsets as $j = P_{Off}^{AU,i}, \dots 1$, where $P_{On}^{AU,i}$ and $P_{Off}^{AU,i}$ are the indices of the last onset frame and the first offset frame respectively. Note that the duration of each AU differed in terms of overall length and onset/offset length. Linear animations were generated using $\mathbf{b}_s^{AU,i}(1)$ and $\mathbf{b}_s^{AU,i}(P_{On}^{AU,i})$ for the onset, and $\mathbf{b}_s^{AU,i}(P_{Off}^{AU,i})$ and $\mathbf{b}_s^{AU,i}(1)$ for the offset – with in-between frames synthesized using linear interpolation. Here, the proportion of mixing was defined using ease-in and ease-out curves. The shape of these curves was learned using a fitting process which matched the velocity of the linear animations as closely as possible to the nonlinear ones.

[2]Note that a degree of nonlinearity also exists in our linear animations. This is in terms of its *temporal* motion, or velocity, which is controlled by the ease-in and ease-out curves.

One important question to consider is *how nonlinear are the animations*? As far as the authors are aware, no previous work exists on attempting to numerically quantify the nonlinearity or naturalness of facial motion, making ours the first exploration of the issue. Here, we attempt to define a measure of nonlinearity here based on the overall spatial vertex distance between a linear and a nonlinear animation. Based on the assumption that vertices in a nonlinear animation will deviate from the corresponding linear ones (e.g. in Figure 2), this would give an indication of how much an animation deviates from a linear path, thereby also providing a basis of comparison between different nonlinear animations. Table 2 shows the total distance between each linear and nonlinear AU. This is calculated by summing each nonlinear vertices euclidean distance from its linear counterpart for each frame. The total value is then normalized using the overall maximum AU value for the given time phase. From inspection of the table, the onset movement of AU 20 deviated most from its linear counter part, while on the offset phase the most deviation was in AU 16.

By observing expressions frame by frame, we can also visualize differences in the facial changes between linear and nonlinear sequences. In a linear model we would expect changes to occur across the entire face at the same rate. However, in a captured facial sequence (nonlinear) these changes should occur at different places at different times. Figure 4 visualizes vertex displacement for AU 9 (Nose Wrinkler) at different stages during onset (neutral to peak expression). Note that whereas changes are globally uniform in the linear blend shape sequence (i.e. they occur simultaneously across the entire face at the same rate of change), they are non-uniform in the captured nonlinear sequence (appearing in different areas at different rates of change). This is compelling since it highlights potential deficiencies of the linear blend shape model with respect to the representation of complex facial deformations during an expression.

AU	1	4	5	6	7
Onset	0.816	0.078	0.002	0.183	0.286
Offset	0.900	0.159	0.526	0.882	0.100
AU	**9**	**10**	**11**	**12**	**13**
Onset	0.218	0.054	0.003	0.446	0.233
Offset	0.437	0.162	0.030	0.239	0.278
AU	**14**	**15**	**16**	**17**	**18**
Onset	0.208	0.160	0.407	0.098	0.204
Offset	0.279	0.140	1.000	0.386	0.415
AU	**20**	**22**	**23**	**24**	**25**
Onset	1.000	0.558	0.455	0.420	0.726
Offset	0.545	0.469	0.099	0.113	0.094

Table 2: *Overall Euclidean distance (normalized) between linear and captured (nonlinear) animations.*

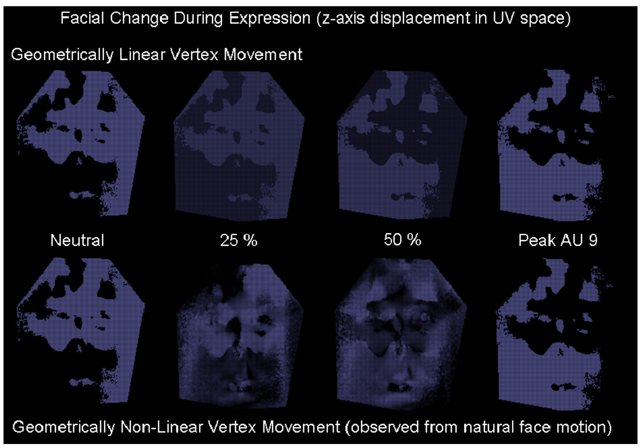


Figure 4: *Facial changes during linear (blend shape model) and nonlinear (captured) sequences.*

Linear sequences were of the same temporal length as their corresponding nonlinear AU sequences. Synthesized sequences of shape and texture data were rendered in 3D Studio Max. The rendering view was created using a fixed camera placed directly in front of the face model. After being rendered out as sequences of BMP images in 3D Studio Max, animations were converted into Quicktime movie format at a resolution of 640 x 480 pixels using MPEG-4 compression (which upon visible inspection did not add any detrimental compression artifacts). Figure 3 shows peak facial images for each AU sequence.

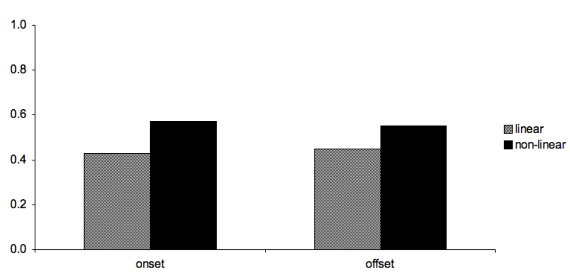


Figure 5: *Mean overall linear vs nonlinear preference frequencies for onset and offset phases(normalized)*

5.2 Experimental Design

To allow a stringent test of motion perception, a forced choice task was used. Participants were presented with linear and nonlinear AU animations side-by-side, and were asked to indicate which of the two clips had the more natural motion. The order of AU presentation was randomized for each participant, as was the side upon which the linear or nonlinear animation was presented. In total, there were 40 pairs of animations (20 AUs at the onset and 20 AUs at the offset phase). In order to account for the frequency of single AUs in everyday life, participants were shown before each animation real videos of another independent study showing the respective individual AU. This way, no constraint was laid upon participants familiarity with the actual movement in real life.

When watching the pairs of animations, participants were instructed to focus on the facial movement and to make a distinction based on which facial movement looked more realistic and natural (i.e., the one that moves in the most human like manner and is the least robotic looking). After making their decision, participants had to indicate for each pair of animation how confident they felt about their choice on a 5-point Likert scale (1 not confident at all, 5- very confident). A total of 15 voluntary subjects (10 men, 5 women, aged between 22 and 32 years) participated in the study. The display resolution used for the experiment was 1280 x 1024 pixels, and participants sat approximately 60 cm away from the screen.

6 Results

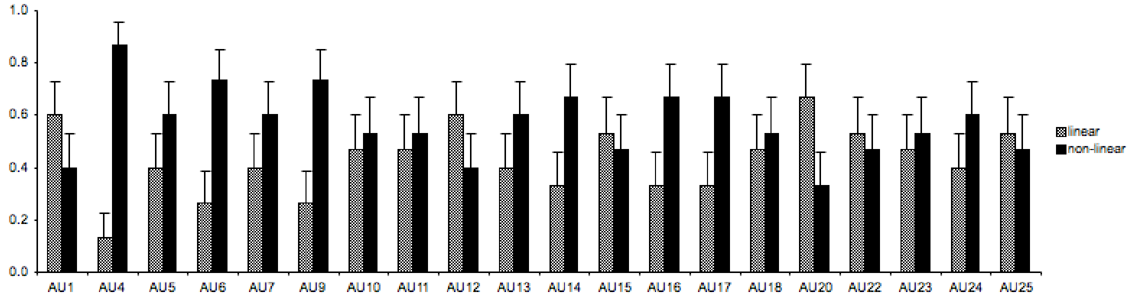
Choice of Animation To test whether there was an overall effect of motion (linear, nonlinear) and time phase (onset, offset) across all 20 AUs on participants'choice, we conducted a multivariate analysis of variance (MANOVA). Results showed that there was a significant difference in participants'choice of animation when rating their naturalness, $F(1, 14) = 783.94, p < .05$ [3]. When presented with pairs of linear and nonlinear animations, participants more often selected nonlinear animations $(M = 22.47, SD = 5.08)$ over linear animations $(M = 17.53, SD = 5.08)$ as being the more natural animation (see Figure 5). The effect of time phase was non-significant, $F(1, 14) = 0.39, p > .05$, suggesting that participants preference for nonlinear animations applied equally to both time phases (onset and offset). Figure 6 shows that for the majority of AUs at onset and offset phase, participants chose nonlinear animations to be the more natural. For AU4, AU9, and AU17 the differences in preference frequency for nonlinear animations over linear animations were found to be significant. Individual AU significance levels are as follows: $AU4, F(1, 14) = 38.50, p < .001$; $AU9, F(1, 14) = 7.72, p < .05$; $AU17, F(1, 14) = 4.42, p < .05$.

Confidence Ratings Results showed that participants'confidence ratings about their choice of animation neither varied as a function of the type of motion (linear, nonlinear), $F(1, 14) = 0.05, p > .05$, nor time phase (onset, offset), $F(1, 14) = 0.88, p > .05$. Overall, confidence ratings were the same for linear $(M = 2.42, SD = 0.77)$ and nonlinear animations $(M = 2.38, SD = 0.66)$. Thus, despite participants making different choices of motion type, they indicated similar levels of confidence about their judgments.

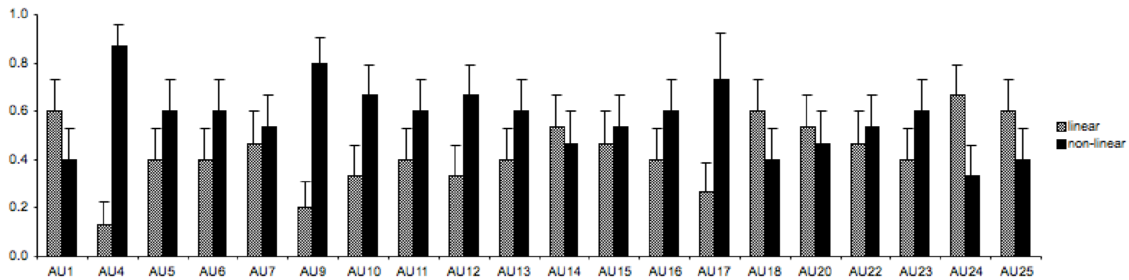
7 Discussion

The results showed an overall preference of the nonlinear animation over the linear animation. The evidence supports the hypothesis

[3] The same results were obtained when using only a subset of emotion-specific AUs (i.e. AU 1, 4, 5, 6, 7, 9, 12, 15, 23): motion $F(6, 9) = 4.36, p < .05$; time phase, $F(6, 9) = 1.15, p < .05$



(a) Onset motion linear/nonlinear preference frequency for each AU (normalized)



(b) Offset motion linear/nonlinear preference frequency for each AU (normalized)

Figure 6: *Mean percent preference rates of linear and nonlinear animations of 20 Action Units (AU). Error bars represent standard errors of the means.*

that animations with motion parameters based on nonlinear geometric movement recorded from a real face are judged more natural looking than those with linear geometric movement.

Certain nonlinear AU animations were very strongly preferred over linear ones, e.g. AU 4, 9 and 17. These preferences were also found to be significant. Each of these movements involves complex wrinkling and elastic deformation of the skin surface – a behavior which is inherently nonlinear geometrically, as demonstrated in Bickel *et al* [Bickel et al. 2007]. It therefore follows that the perceptual realism of skin wrinkling would certainly benefit from nonlinearly deforming surfaces in facial models. However, this is not to say that linear movement is not entirely undesirable in terms of their perceived motion characteristics. In this study there was a tendency for some linear AUs (i.e., AU 1, 10, 20, and 25) to be rated in both onset and offset phase as more natural than their nonlinear counterparts. For other AUs (i.e., AU 15, 22) participants found it difficult to differentiate between linear and nonlinear animations, resulting in similar preference rates for both types of motion. When designing an animation system this information may be important since it suggests that some AUs can still be based on linear morphing without compromising naturalness. The fact that the nonlinearity measures for AUs 4, 9 and 17 in Table 2 did not appear correlated with the strong nonlinear animation type preference is worth investigating. It is likely that the measure may have to be extended to account for the distribution of nonlinear variation across the face.

Interestingly, there was no significant difference in participants confidence ratings when choosing between the different motions (linear vs. nonlinear). This is also supported by the fact that each participant commented after the experiment that distinguishing between the two animations was rather difficult. Therefore, given the subtle nature of animation type, the findings suggest an implicit cognitive process. Although participants were unable to report any conscious feelings such as their confidence their behavior revealed an overall preference for nonlinear animations. Similar findings are well-known in psychological literature on subliminal priming in which conscious feelings can become decoupled from processes underlying behavioral reactions [Morsella 2008; Winkielman and Berridge 2004]. The subtle and also fleeting nature of facial movements in animations may therefore have reliable effects on people's preferences without being accessible to direct introspection.

Note that our decision to use FACS as the basis for facial animation in our experiments was that it provides the clearest description of facial movement available to researchers. It therefore allows replication of our experiment with different facial models and performers, and can be used independently by researchers in various fields. In this experiment we focused only on single AUs as this is the principle adopted by many facial model designers – whether creating models entirely artistically or capturing facial movement from actors using dynamic 3D systems such as MOVA [Duncan 2009]. Thus, knowing which AUs could benefit from nonlinear geometric movement could help an animator to focus realism on certain facial controllers. However, we acknowledge that a comparison of linear and nonlinear motion types in the perception of more complex expressions (e.g. anger, fear) would be worthwhile, and is a direction for future work.

8 Conclusions

In this paper we have presented a 3D facial model with FACS based AU parameters for use in graphics, animation and psychology research. Our model is based on AU data validated by certified FACS coders, and is therefore suitable for use in a range of perceptual experiments. Using this model, we have investigated the perceptual differences of facial motion observed from a real person versus linear motion, and have found a perceptual benefit towards displaying natural nonlinear geometric movements.

Future Work The question of a numerical measure of nonlinearity and expression naturalness is intriguing, and one we have only just started to investigate. Given a suitable measure and a correlation with perceived realism, it may be possible to predict how natural an animation will be judged even before it is rendered. We believe that the measure employed in Table 2 may be improved by also taking into account the position and spread of nonlinear changes across the face. For example, smaller nonlinear changes distributed across the face may be deemed more natural than large nonlinear changes localized on a single part of the face. Moreover, the influence of texture nonlinearity and its effect on perception (in our experiments and in general) is an interesting issue and remains to be determined. Calculating optical flow between textures would provide the necessary vector information to achieve this. Each of these questions are yet to be addressed by researchers, and there is large scope for future examination.

Another continuation of our work aims at the inclusion of several facial models in order to show generalization of effects across identity. Ultimately, we may hope to learn a set of key-facial movements that should be prioritized in terms of their realism when designing a facial model in order to give the greatest perceptual naturalness.

Acknowledgements

Thanks to everyone at CVSSP who helped with dynamic 3D facial capture, including Doris Peham, Gwenda Simons, Kornelia Gentsch and Michaela Rohr. Thanks to the Royal Academy of Engineering and EPSRC for funding this work.

References

3dmd. http://www.3dmd.com.

BICKEL, B., BOTSCH, M., ANGST, R., MATUSIK, W., OTADUY, M., PFISTER, H., AND GROSS, M. 2007. Multi-scale capture of facial geometry and motion. *ACM Trans. Graph. 26*, 3, 33.

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Proc. of ACM Siggraph*.

BREIDT, M., WALLRAVEN, C., CUNNINGHAM, D. W., AND BULTHOFF, H. H. 2003. Facial animation based on 3d scans and motion capture. In *ACM SIGGRAPH Sketches and Applications*.

CUNNINGHAM, D., AND WALLRAVEN, C. 2009. Dynamic information for the recognition of conversational expressions. *Journal of Vision 9*, 1 – 17.

CURIO, C., BREIDT, M., KLEINER, M., VUONG, Q., GIESS, M., AND BULTHOFF, H. 2007. Semantic 3d motion retargetting for facial animation. In *Proc. of ACM APGV*, 59–64.

DUNCAN, J. 2009. The unusual birth of benjamin button. *Cinefex*.

EKMAN, P., FRIESEN, W., AND HAGER, J. 2002. *Facial Action Coding System: Second Edition*. Salt Lake City: Research Nexus eBook.

GRIESSER, R., CUNNINGHAM, D., WALLRAVEN, C., AND BÜLTHOFF, H. 2007. Psychophysical investigation of facial expressions using computer animated faces. In *ACM APGV*, 11–18.

GUATAMA, T., AND HULLE, M. V. 2002. A phase-based approach to the estimation of optical flow using spatial filtering. *IEEE Trans. Neutral Networks 13*, 5.

JIANG, F., BLANZ, V., AND O'TOOLE, A. Three-dimensional information in face representations revealed by identity aftereffects. *Psychological Science 20*, 3.

KNAPPMEYER, B., THORNTON, I., AND BUTHOFF, H. 2003. The use of facial motion and facial form during the processing of identity. *Vision Research 43*, 18, 1921–1936.

KRUMHUBER, E., AND TAMARIT, L. 2010. Facsgen beta version: Development and validation of a new set of action units. *CISA Annual Research Forum*.

KRUMHUBER, E., MANSTEAD, A., COSKER, D., ROSIN, P. L., AND MARSHALL, A. D. 2007. Facial dynamics as indicators of trustworthiness and cooperative behaviour. *Emotion 7*, 4, 730–735.

MA, W., JONES, A., CHIANG, J., HAWKINS, T., FREDERIKSEN, S., PEERS, P., VUKOVIC, M., OUHYOUNG, M., AND DEBEVEC, P. 2008. Facial performance synthesis using deformation-driven polynomial displacement maps. *ACM Trans. Graph. 27*, 5, 1–10.

MORSELLA, J. B. . E. 2008. The unconscious mind. *Perspectives on Psychological Science 3*, 73–79.

PARKE, F. I., AND WATERS, K. 1996. *Computer Facial Animation*. A. K. Peters.

ROESCH, E., TAMARIT, L., REVERET, L., GRANDJEAN, L., SANDER, D., AND SCHERER, D. 2010. Facsgen: A tool to synthesise emotional facial expressions through systematic manipulation of facial action units. *Journal of Nonverbal Behaviour (in press)*.

SAGAR, M. 2006. Facial performance capture and expressive translation for king kong. In *ACM SIGGRAPH 2006 Sketches*, 26.

SUN, X., ROSIN, P. L., MARTIN, R., AND LANGBEIN, F. 2007. Fast and effective feature-preserving mesh denoising. *IEEE Trans. on Visualization and Computer Graphics 13*, 5, 925–938.

THEOBALD, B., MATTHEWS, I., MANGINI, M., SPIES, J., BRICK, T., COHN, J., AND BOKER, S. 2009. Mapping and manipulating facial expression. *Journal of Language and Speech 52*, 2, 369–386.

WALLRAVEN, C., BREIDT, M., CUNNINGHAM, D., AND BULTHOFF, H. 2008. Evaluating the perceptual realism of animated facial expressions. *ACM Trans. Appl. Percept. 4*, 4, 1–20.

WEYRICH, T., MATUSIK, W., PFISTER, H., BICKEL, B., DONNER, C., TU, C., MCANDLESS, J., LEE, J., NGAN, A., JENSEN, H. W., AND GROSS, M. 2006. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph. 35*, 3, 1013 – 1024.

WILLIAMS, L. 1990. Performance driven facial animation. *Computer Graphics 24*, 4, 235 – 242.

WINKIELMAN, P., AND BERRIDGE, K. 2004. Unconscious emotion. *Current Directions in Psychological Science 13*, 120–123.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph. 23*, 3, 548–558.