Creating Generic Data-driven Face Rigs for Digital Actors

Volker Helzle
Animationsinstitut at Filmakademie Baden-Wuerttemberg
Ludwigsburg, Germany DE-71638
volker.helzle@filmakademie.de

Kai Goetz
Animationsinstitut at Filmakademie Baden-Wuerttemberg
Ludwigsburg, Germany DE-71638
kai.goetz@filmakademie.de

Diana Arellano
Animationsinstitut at Filmakademie Baden-Wuerttemberg
Ludwigsburg, Germany DE-71638
diana.arellano@filmakademie.de

ABSTRACT

The creation of faces for digital actors is a time-consuming and challenging process, especially when the character to recreate is a not living one, or looks very different to the original counterpart (e.g. aged). One crucial element is the facial rig, which can facilitate, or hinder the subsequent animation process. In this paper we present the Adaptable Facial Setup (AFS), an easy-to-use, semi-automatic rigging solution that uses a generic database of facial motion data to drive facial animation. To prove the reliability of our approach, and how it helps to overcome the challenges when creating historical characters, we created a digital version of Albert Einstein's head. The resultant animations created with our AFS facial rig and a blendshape rig were compared through a user perception evaluation, corroborating that our approach leads to more convincing and natural facial animation.

CCS CONCEPTS

• **Human-centered computing → User studies**; • **Computing methodologies → Animation**; **Motion capture**;

KEYWORDS

Rigging, facial animation, motion capture, perception, digital actor

ACM Reference format:
Volker Helzle, Kai Goetz, and Diana Arellano. 2017. Creating Generic Data-driven Face Rigs for Digital Actors. In *Proceedings of Data-Driven Animation Techniques (D2AT), Bangkok, Thailand, November 2017 (SIGGRAPH Asia Workshop'17),* 8 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The process of creating convincing digital human faces in motion picture and video games has progressed significantly in the last decade. Thereby two main strategies have emerged. On the one hand, progress in scanning technologies made it possible to create future-proof digital human datasets for existing individuals, resulting in actors being replaced by high-quality digital versions of themselves to a larger extent.

Systems like the Medusa Performance Capture [4], or the Light-Stage from the USC's Institute for Creative Technologies [9] allow to capture a human face shape and appearance at all possible extremes. Some examples of feature films where living actors were scanned to generate their digital faces are *Doctor Strange*, *Superman Returns*, *Spiderman 2*, among many others. Even eyes [5] and hair [19] could be captured individually.

SIGGRAPH Asia Workshop'17, November 2017, Bangkok, Thailand
2017. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

On the other hand, there are cases where more artistic approaches are necessary, especially when scanning methods cannot be used to digitally create or replicate a human face and its individual expressions. For instance, when the subject to create is a human-like creature with human expressions (e.g. Gollum in *Lord of the Rings*, Caesar in *War for the Planet of the Apes*, or the Na'vi in *Avatar*); or it is an aged version of an existing individual (e.g. Brad Pitt in *The Curious Case of Benjamin Button*) [14]; or it is a deceased person (e.g. Peter Cushing as Grand Moff Tarkin in *Rogue One*).

The approach presented in this paper, the Adaptable Facial Setup, follows the latter more artistic strategy. It is the core of a rigging procedure, based on FACS (Facial Action Coding System) [13], for a generic data-driven model that can be applied to any humanoid geometry. FACS describes all the facial muscle movements that can deconstruct and taxonomize human facial expressions by means of Action Units (AUs) [27]. We created our own dataset consisting of over 100 motion captured facial expressions clips. In order to obtain a generalized description of muscle group movements, we enhanced each individual performance by preserving its rich characteristic over the entire activation of the facial action.

The result is a rapid rigging approach that utilizes a dense data model describing an individual's facial movements during their entire activation phase, and not just a static extreme as in industry standard blendshape models. The proposed method is arranged in a toolset embedded in a standard DCC application that underwent numerous development iterations. It allows artistic adjustments to the dense data model to preserve its original characteristic and richness, while adapting to the desired facial action. This adaptation process can be used with any facial input, thus supporting the theory of the universality of human facial actions and emotions proposed by Ekman [12].

In the following we present previous works that have inspired our own. Then we explain the motivation for our system, as well as how it was implemented. A use case illustrates how we recreated a digital human. Finally, we will show the results of an informal perceptual study where we compare our results with the commonly used approach of linear interpolation between geometric representations of activated expressions, also referred to as *blendshapes*.

2 PREVIOUS WORK

The creation of synthetic faces is an open-field of research where scholars and industry have been working intensively during the past years. Realistic, human-like faces are the ones posing the most difficulties. A high level of control for deformations is needed in order to capture all the nuances of the human face. Several previous works have made great advances in facial rigging in order to create a powerful, capable of creating complex expressions, while being controllable and relatively simple to use.

2.1 Rig creation

Among the facial rigging methods we can find those based on skeletons and joints, physically-based muscle models, linear blendshapes, or combinations of them. For a better understanding of the advantages and disadvantages of these methods, see Li et al.[22].

One of the most accepted and used standards for facial rigging relies on the use of blendshapes. However, to obtain a better control of every area of the face, it is necessary to create a large number of shapes [28]. For instance, the facial animation of Gollum in the film *Lord of the Rings: The Two Towers* required 675 blend shapes, showing the complexity of a rig using this method [15]. Ritchie et al. [29] demonstrated in their book "The Art of Rigging" that it is possible to create such complex behavior and realistic facial expressions by using a smaller amount of shapes, or combining them with other animation techniques. However, it is intended for artists with very technical skills.

In an attempt to accelerate and facilitate the rig creation, researchers have been looking into techniques that made use of facial templates that can be adapted to different geometries, while providing the same control for the creation of facial movement. For instance, Li et al.[22] introduced a framework that automatically creates optimal blendshapes from a set of example poses of a digital face model. Dutreve et al. [11] presented an automatic process for re-targeting facial feature points between meshes with a different topology. The downside of these approaches is the need for a manually rigged character and all the shapes for key facial movements, used as the source for the feature transferring.

Unlike these approaches, our source rig has been automatically created from facial data obtained using motion capture, which describes facial movements during their full activation phase and not just one static shape.

2.2 FACS-based Databases

One of the sources to obtain large amount of facial data are "facial recognition databases". The most traditional ones consist mainly of still images of professional actors, or non-trained participants displaying either just a neutral expression, or emotional expressions. Some examples are the Karolinska Directed Emotional Faces [23], the Bosphorus database [32], or the color FERET Database [26].

Among the databases containing videos of facial action units is the MPI Facial Expression Database [20]. It contains 18,800 samples of video recordings of spontaneous and posed expressions of 19 individuals showing 55 expressions at 2 intensities. It also contains 3D facial scans of most participants. The target of this database are computer vision and affective computing applications. The Affectiva-MIT Facial Expression Dataset (AM-FED) is comprised of 242 webcam facial videos (168,359 frames) recorded in real world conditions, labeled for the presence of symmetrical and, similarly to our database, asymmetrical FACS action units, plus head movements, smile and other expressions [25].

The main difference, and advantage of our database, is that it captured the progression of the facial expressions from its onset to its offset. This is very important when it comes to the creation of the facial rig, where all involved areas of the face need to be precisely defined in order to achieve the desired target deformation.

2.3 Facial Rigs for Film and Game Productions

The challenges increase when the final product is intended for the film and game industry. Arghinenti [3] used a procedure similar to ours to create more than 30 different faces over 70 minutes of cutscenes for the AAA title KILLZONE3™. They started from a grid-like pattern of markers placed over the actor's face. The data, after having been tracked and stabilized, was loaded directly onto the rig. The FACS control was given by a muscle model on top of the rig that serves to drive animation in individual areas of the face.

A common method used in the entertainment industry is the creation of digital characters through scanned data. Seol et al. [33] combined the manual artistic input of an artist, who produces a generic facial *blendshape* rig, and the automatic capturing of FACS-compliant facial expressions from an actor, which are used for expression transferring. While our method does not require facial scans of the FACS AUs, the creation of displacement maps falls into the hands of an artist.

Commercial products have also tackled the issue of high-quality facial rigs. One effort was the now discontinued FaceRobot[1], a complete software solution for rigging and animating 3D faces easily, enabling studios to create, at great speed, life-like facial animation. Faceshift[2], now acquired by Apple, was a software that analyzed the facial movements of an actor and described them as a mixture of basic expressions, plus head orientation and eye gaze. This description was then used to animate virtual characters in movies and game productions. A more recent example of real-time facial rigs based (not entirely) on FACS is provided by the proprietary solution from 3Lateral[3]. Currently, they are enhancing their system by populating a database with scans from real people in order to "generate truly unique characters with appropriate facial gestures for their anatomy, while keeping the ability to use the same animation across the created population." [21]. Their technology was recently demonstrated in a VR installation called *Meet Mike* [34].

2.4 Data-drive Animation

To drive the animation of 3D characters, previous researchers have worked with video input, from which they extract feature points, or at best FACS AUs, that serve to control a facial rig, or a deformation model. Examples are the works of Mascaró et al. [24], who took the input data from videos of facial expressions, recorded while the participants played a computer game, and used it to control the animation of a FACS-compliant facial rig. Bhat et al. [6] used the input video, a neutral mesh of the actor and a set of blendshapes to solve for animation curves in the blendshapes and produce additional correctives, to improve the overall deformantion, at each frame. The advantage of this method is that the correctives are automatically created on top of the blendshapes. On the downside, these correctives are hard to modify and edit. Rizzo et al. [30] used the tracked feature points to create animations by volume morphing and appearance classification.

[1] http://altec.com.hk/altec_new/eng/Animation/Softimage_FaceRobot.html
[2] http://faceshift.com/studio/2015.2/introduction.html
[3] http://www.3lateral.com/

2.5 Digital Actors

Early examples of ground-breaking advancements in facial animation in digital actors are seen in the short-films "The Jester" (1999) [7] and "Young at Heart" (2000) [31].

One of the most notorious works in the creation of digital actors is the "Digital Emily" project. This work by Alexander et al. [1] represented a big step forward in the creation and animation of photo-realistic human characters. To achieve this, they used scans of an actress' face performing the different FACS action units. Then, they created blendshapes using small dots drawn on the actress' face to "coarsely warp a partial master mesh into the different expressions". To achieve an exact result, they added displacement maps obtained by placing each blendshape on top of its corresponding scan and calculating the difference between the two.

The main difference with our approach is that we used captured facial data to generate the basic building blocks of facial rigging, which in turn can be employed to animate entirely new synthetic performances. Some examples of games and short animated films where our approach have been implemented are: *James Bond Quantum Break* by Activision, *A Lost and Found Box of Human Sensation* by Lailaps Pictures [35], *The Gathering* by Sven Dreesbach [10][8], *Kinksi Revisited*[17] or *SARA* [2], demonstrating that it is a usable and state-of-the art methodology to follow.

3 ADAPTABLE FACIAL SETUP (AFS)

The approach we follow has been inspired by an earlier work that used a less comprehensive data model [18]. The *Facial Setup* (AFS), as we will refer to our approach and toolset, is an easy-to-use, semi-automatic rigging solution that uses a generic database of facial motion data to drive facial animation.

The basis of our data model consists of over 100 motion captured facial expressions clips. We further modified each individual expression clip performance to obtain a generalized description of muscle group movements while preserving its rich characteristic over the entire activation of the facial action.

The majority of the captured expressions were performed in accordance to the FACS (Facial Action Coding System) [13]. Our dataset also accounted for asymmetrical facial action units (AUs), which are better suited to replicate non-symmetrical movements in more detail than just using partial information of the symmetrical ones. A male actor in his early thirties performed all the AUs after having intensively studied all of them. His performance was also verified by a certified FACS scoring specialist, validating the content of the database. The facial actions were captured using 65 optical markers applied to the face of the actor. The distribution of the markers was intentionally chosen to account for as much facial muscle activation as possible. While performing, the actor was required to keep his head movement to a minimum, which was possible thanks to a neck rest support. Multiple takes of each muscle group activation were performed.

3.1 Motion Data Processing

In this process, the essentials of the facial movements are distilled to a standardized dataset describing the muscle activation through a set of 65 markers. We further refer to this data as motion data.

The first step in enhancing the data was to remove the global head movement (stabilisation). The actor wore a headband with multiple tracking markers. This worked for most cases where some facial actions, including the forehead, influenced the position of the headband. To compensate for the remaining head movement we carried out a stabilisation procedure as described in [16], where individual markers considerably reduced the global head movement.

The next step was to globally align all mocap markers with the neutral template. Applying standard values to the root of the markers did work in most cases for an initial match. However, some clips recorded towards the end of the session showed an increasing offset to the neutral template. This happened due to some markers being dismounted after extreme facial actions, and reapplied afterwards. To compensate this we manually moved the root of the markers so they match the neutral template markers with minimum offset. In some cases the facial action needed to be considered. For example, a lower face movement like *kiss* required optimal match with the mouth region rather than with the forehead.

Finally, we identified the performance clip with the highest quality. This was based on subjective attributes such as the evenness of motion from neutral to extreme, remaining global head movement, match to the neutral template, and unwanted accentuation in regions that do not belong to the AU as described in FACS. Once identified, the start and end of the motion segment were tagged. We then resampled the data on the neutral marker position to a unified activation range between neutral and expressive (full intensity).

4 FROM DATA TO DIGITAL HUMANS

4.1 Adaptation Process

Adaption describes the process of aligning the motion data to the desired humanoid face target geometry. It is performed through functions implemented in a tool GUI embedded in a standard DCC application for animation (Autodesk Maya). An *animation control* defines the geometry deformation, and it can be used later on by the animator to activate those deformations in the face rig.

There are no constraints for the face model geometry as long as it is human-like (mouth, nose and two eyes). The rigging process is initialized by applying a set of 65 deformation objects (joints grouped under Maya locator objects) on respective landmark positions of the target geometry. To facilitate the process, a reference of the template is provided. In a second step, all deformation objects are aligned globally and individually. This process requires thoughtfully adjusted influences of the deformation objects, also known as skin weights. As this is a crucial part, the toolset also offers a skin weights cloning process. It can reduce the time required to define the initial deformation areas, which use standard linear blend skinning of their surrounding area. In the next steps we connect the motion data to the deformation objects. The correlation to the initial template position is used to calculate an initial scale for the data. Afterwards, each individual animation control, corresponding to a motion data clip, is verified for its appearance. Finally, the rigging artist loads the associated adaptation information using the GUI, exposing the basic motion data, individually applied scale values, adaptation and rotation data (Figure 1).

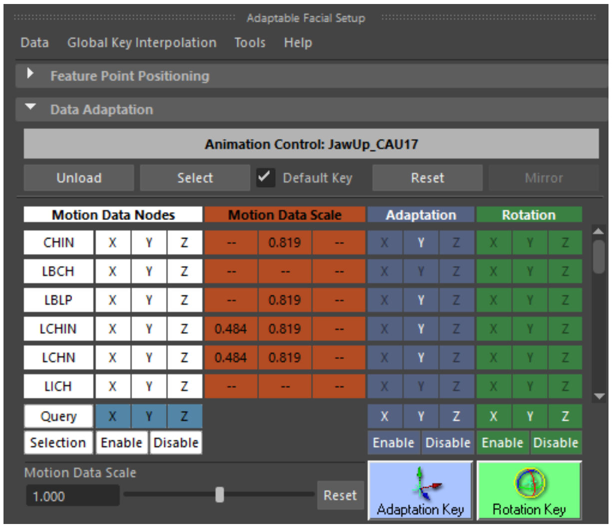

Figure 1: Adaptable Facial Setup GUI. Left to right: Motion data, Scale values, Adaptation data, and Rotation data.

4.1.1 Motion data. Motion data is the basic data associated to each animation control. It represents all involved locator objects for one animation control, each with a joint object carrying out the deformation below its hierarchy (Figure1: "Motion Data Nodes"). The total data consists of 12,969 curves for 117 animation controls.

4.1.2 Scale values. The scale values are directly associated to the motion data (Figure1: "Motion Data Scale"). They individually amplify/weaken specific regions by applying scale values. Involved locators and their respective axis can be selected with the help of selection masks.

4.1.3 Adaptation data. Adaption data allows to directly manipulate a locator and its associated deformation object (Figure1: "Adaptation"). New positions can be applied at any intensity level by creating *adaptation keys*. These keys are combined with the basic motion data and cause a local amplification/weakening of the locator/deformer at the given intensity.

4.1.4 Rotation data. As the motion data does not contain rotations this is an additional procedure to enhance the deformation with subtle rotations.

4.2 Advantages and Limitations

An advantage of AFS, in contrast to the static modeling of blendshapes, is the use of data adaptation of an existing motion template to create a detailed facial rig. All adaptation procedures are performed while constantly verifying the entire deformation of an animation control. For convenience we apply a standard linear animation to the animation control when loading its associated data to the tool GUI. By scrubbing the timeline we constantly verify the deformation associated to the animation control during its entire activation (from neutral to extreme).

The AFS provides a total of 117 animation controls. The basic set of controls can easily be extended as a mix of existing animation controls in combination with offset animation of individual locator (joint) objects. A baking procedure hereby creates a new set of motion data associated to the new custom animation control. The new control can use the same adaptation strategies as outlined above, providing a fast and flexible solution to extend the rig to animation artist requests. Figure 2 shows the AFS setup.

One might argue that in-between shapes could lead to similar results. This is true for the skilled artist that knows about muscle group activation in the human face. However, each update is a static modeling process with little intuitive feedback compared to our adaptation strategies. It is also worth noting that the time required to create a FACS compatible face rig with over 100 animation controls is reduced to a few days using the AFS approach.

On the other hand, using the chosen deformation method (Linear Blend Skinning) has its limitations. The method cannot account for finest vertex displacements or rigging/weighting imperfections that can occur with expressions in opposite extremes (e.g. lips stretch vs. lips pucker) To overcome those limitations we introduce a limited amount of Corrective Blendshapes (CBS) [4]. It is important to mention that these shapes should not be considered traditional blendshapes, as the basic movement is always caused by the AFS joint movement. In a complex rig with over 100 animation controls we use approximately 15 corrective shapes.

5 DIGITAL ACTORS IN DOCUMENTARY

To evaluate the capabilities of AFS in regard to photo-realistic human faces, we initiated a project in 2006 to recreate a deceased actor as an aged digital version of himself. The main idea was to integrate him into live action scenes that he originally played, appearing in several shots as an 80 years old man talking about his career and future plans [17].

The lack of real-life references for both modeling and animation required a strongly artistic and iterative workflow. Using the AFS setup, the animation was altered and fine-tuned until the last day of production, without the need for shot-specific scans or capture data. The duration of the project was two years and the results were very promising back then.

Recently, we decided to take up a new project to test the validity of the AFS and demonstrate how effective it is in creating high-quality rigs and producing new content from the first idea to delivery. All this even with a limited budget, time constraints and a core team consisting of two main artists and a technical director.

The selected use case investigates if historical characters can be recreated digitally. While actors in re-enactments resemble the original only to some degree, completely digital characters can become convincing counterparts down to the smallest wrinkle. However, the project does not aim at substituting the actors, but tries to find pipelines to conserve their unique performance, including voice and gestures to create natural looking and appealing digital actors.

For this use case, Albert Einstein served as a subject of investigation. His face is well-known and found its way into pop culture, although nobody has ever seen a colored image of him. Our goal is to create a video blog with several short episodes where Albert Einstein reports back every few weeks, commenting on recent events, and stating some of his famous quotes.

[4]https://animationsinstitut.de/en/research-development/tools/facial-animation-toolset/features/

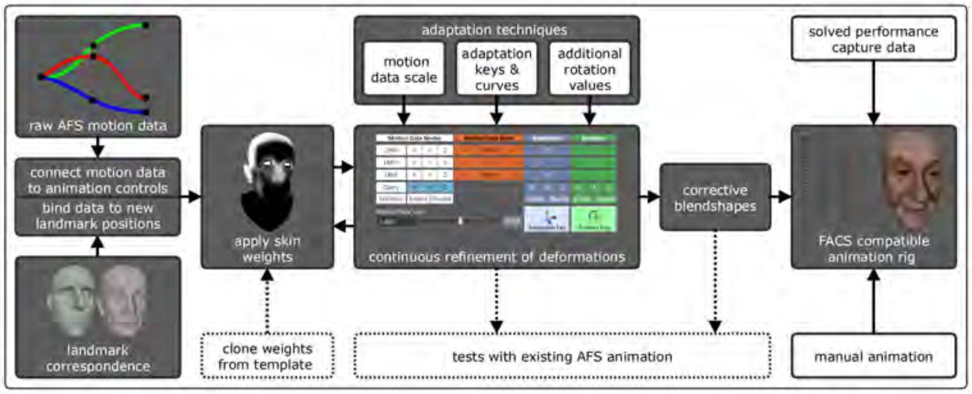

Figure 2: Schematic Setup Process of AFS.

5.1 Asset and Rig Creation

Before working digitally, different busts of Albert Einstein (clay, plastic (PVC) and silicone) were shaped by a professional sculptor. By scanning the PVC positive using photogrammetry, it was possible to quickly obtain a first medium quality 3D mesh as a starting point for a later digital sculpt. The scan of the silicone cast later served as a texture reference (Figure 3).

After tidying up the topology and correcting the overall shape, bringing the scan even closer to Albert Einstein's face, a modeler sculpted detailed wrinkles, and fine micro structures, such as pores. Simultaneously, the rig was set up. Applying the AFS tool a rigger was capable of creating a rough but fully functional facial rig within hours, including template animation, which allowed for first simulation and shading tests.

From then on the rig was continuously optimized, improving the skin weights and adjusting the maximum deflection of the individual facial joints driven by the animation control. Since AFS does not include an elaborated eye rig, an auto-rigger was used to rig the eyes, guaranteeing that the lids deform according to the eyeball and slide flawlessly over the eyeball geometry without intersection when blinking.

To enhance the realism of the character, we integrated a skull mesh and a muscle system that forced the skin to slide over the underlying geometry. The effect became most apparent in distinctive areas of the face such as the chin and the frontal bone.

Lastly, a curve-blending approach was used to simulate sticky lips. Although the setup consisting of various sub-rigs and muscle evaluation got quite complex in the end, the final rig performed well when being animated, offering close-to-real time feedback without the need for low-resolution proxy geometry.

The total duration of the 3D asset creation (basic geometry, animation rig, shader, dynamic displacements, facial hair and look tests) was approximately 77 days, i.e., almost 4 months.

5.2 Shot Production

The live action shoot consisted of filming an actor with a wig on location, such that he resembled Albert Einstein. Once the principal shoot of the plate material was completed, the footage was matchmoved and the facial performance tracked.

After the retargeting process, the resultant rough animation served as basis for a first slap comp. Already in this very first animation pass, it became obvious that the original facial expression needed to be altered in favor of a more appealing thus artistically motivated movement. For instance, blinks were added, facial expressions were replaced and the pace slightly changed. In the end, the animator decided to completely discard the animation derived from the tracking process as it was much more convenient to animate the face from scratch, keeping control over every single slider provided by the AFS, and being able to perform changes quickly when needed.

Concurrently, the 3D scene was shaded and lit using light probes from set. Light sources placed purposefully helped to harmonize live action footage and computer-generated imagery. The digital Einstein got merged with the backplate after rendering, applying soft masks to cover the transition to the actor's head. Color correction, depth of field and motion blur contributed to integrate the digital content and constituted the last steps in composing the final image (Figure 4).

5.3 Challenges in a Documentary Film

When creating photo-realistic digital versions of existing or deceased individuals no deviation from the reference is allowed. To make things even more challenging, everyone has a distinct image of a celebrity or a historical character like Albert Einstein in mind, although this image might not conform to reality.

While these challenges are inherent to digital actors in feature films as well, creating artificial characters for documentary poses special complexity, as the productions usually have to deal with minimum budgets, small teams and short production times. Under these conditions it does not appear reasonable to work with complex blendshape rigs including hundreds of scanned and polished shapes. Furthermore, it might be hard to justify extensive modification or even a rebuild of a rig requiring new facial scans when for example producing another episode of a series. AFS allows for fast adaption to new shots, without even touching the rig, at the same time producing high quality and convincing results. As soon as the plate has been shot, the artist can start animating right away.

Figure 3: Left to Right: clay sculpt, silicone cast, 3D scan, final mesh, rendering with textures.

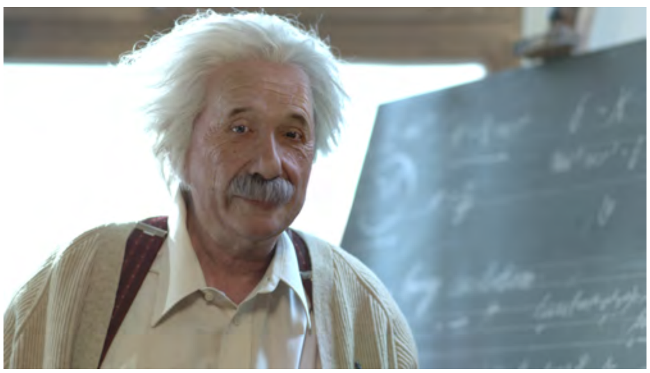

Figure 4: Digital face of Einstein composed with live action backplate.

6 USER EVALUATION

We conducted a perceptual evaluation in order to assess the visual quality of the generated rig and facial movements of the digital Albert Einstein. The goal was to demonstrate that our method (AFS) produces more natural and convincing facial movements with less labor effort compared to manually sculpted blendshapes.

6.1 Stimuli

Two sets of facial animation were created and presented to the participants in two parts. The first part (*Part 1*) consisted of two 10-seconds animations depicting single action units deflected to the maximum one after the other without any head movement or subtleties, such as blinks. The animations were created using both the blendshape rig (*Clip 1*) and AFS rig (*Clip 2*).

The second part (*Part 2*) consisted of two 3-seconds animations featuring several action units at once, including head movements as well. It constituted a more realistic use case of a talking face. As in the first part, they were created using both rig types: blendshapes (*Clip 1*) and AFS (*Clip 2*).

The blendshape rig was extrapolated from the AFS rig with no in-between shapes. This was the most effective way, as to create these 100+ shapes from scratch would have taken considerably longer.

6.2 Questionnaire

We created a questionnaire that could be filled either online or in paper-form. After introducing general data like age, gender and occupation, we asked the participants to rate their expertise in facial animation in a 5-point Likert scale (1: non-expert, 5: expert). Then, for each stimuli part (*Part 1* and *Part 2*), the participants were asked "*Which animation do you find more convincing?*". The possible answers were: *Clip 1*, *Clip 2*, *I see no difference*. They were also provided with some blank space to answer the question "*Why*".

6.3 Procedure

The evaluation took place in two contexts. They differed from each other in the audiovisual setup, way of administering the questionnaire and number of times the stimuli was replayed. In the following we will explain in detail how we proceeded with the evaluation.

6.3.1 Context 1: Evaluation and Results. In this context, the evaluation was carried out in a screening room at our own institution, where a 340 x 185 cm projection screen delivered the videos for both stimuli parts at full-HD resolution. The participants were provided with a paper questionnaire. The entire experiment took approximately 10 minutes.

A total of 22 participants (12 female, 10 male) between 18 and 54 years old took part in this evaluation. The majority of the participants (59.1%) were between 25-34 years old, and all of them had some relation to the film industry. 19 of them considered themselves as non-experts in facial animation, two considered having an intermediate experience, and only one recognized herself as expert.

For the stimuli Part 1, the participants were presented with the two videos in the following order: (a) animation using blendshape rig (*Clip 1*), shown twice, (b) animation using AFS rig (*Clip 2*), shown twice. The procedure was repeated three times. Afterwards, they filled in the part corresponding to Part 1 in the questionnaire.

When asked to compare the resulting facial movements produced using blendshapes and AFS, 59.1% of the participants agreed that the our method generated more convincing movements, while 27.3% perceived the version generated using blendshapes as more convincing. The remaining 13.6% did not notice any difference (Figure 5, Part 1). To test if there was a significant difference between the number or participants who chose the blendshapes version and those who chose the AFS version, we computed a difference of proportions with 95% confidence interval. The result is a difference of 0.318 with a confidence interval (-0.056, 0.692), meaning that the difference is not statistically significant. Nevertheless, when the participants were asked why the chose the AFS version as more convincing, some of the arguments included the perception of smoother skin movement, a softer (and therefore more realistic) transition from AU expression to neutral expression, and a more detailed deformation.

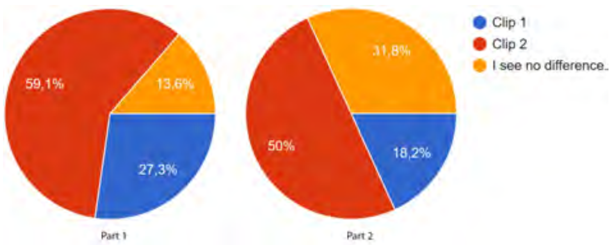

Figure 5: User Perception Evaluation: Context 1.

For the stimuli Part 2, again the participants were presented with two videos corresponding to: (a) first the animation using blendshape rig (*Clip 1*), shown three times (b) then the animation using the AFS rig (*Clip 2*), shown three times. Subsequently, they answered the questions corresponding to Part 2.

In total, 50% of the participants agreed that the facial movements produced with the AFS rig were more convincing than those produced with blendshapes. 31.8% noticed no difference between both animations (Figure 5, Part 2). Once again, when testing if there was a significant difference between the ratings of the participants, we obtained a difference of 0.182, with a 95% confidence interval of (-0.197, 0.561), meaning that the difference was not significant. This result was somehow expected given the complexity of the animation, the subtle differences between both clips and the influence of other elements like head rotation and lip movements (speech) in the perception of the animation.

6.3.2 Context 2: Evaluation and Results . In this context we made the questionnaire available online and indicated the links and credentials to access the videos with the stimuli. The goal was to reach a wider audience and to see if we could achieve statistically significance in our results.

A total of 18 participants (5 female, 13 male) between 18-44 years old took part in this evaluation. The majority of the participants (55.6%) were between 25-34 years old, and all of them had some relation with the film industry. However, as in the previous evaluation, most of them (13) considered themselves as non-experts in facial animation, four considered having an intermediate experience, and only one recognized herself as expert.

Although the evaluation could not be carried out in a controlled environment, the results strongly supported the outcome of the first evaluation.

For Part 1, when asked to compare the resulting facial movements produced using blendshapes and AFS, 72.2% of the participants agreed that our method generated more convincing movements, 11.1% perceived the blendshapes version more convincing, and the remaining 16.7% did not notice any difference (Figure 6, Part 1). To test if there was a significant difference between the scores for Clip 1 and Clip 2, we computed a difference of proportions with 95% confidence interval. The result was a difference of 0.611 with a confidence interval (0.289, 0.933), meaning that the difference is statistically significant. To the question why they chose the AFS version as more convincing, some of the arguments favoring our approach included the feeling of more skin sliding over bone/muscle, increased nose movement during intense expressions, better transitions, and less linear movement.

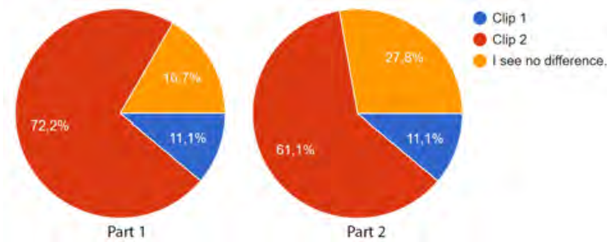

Figure 6: User Perception Evaluation: Context 2.

For the Part 2, 61.1% of the participants assessed the animation using the AFS rig as more convincing. Only 11.1% found the blendshape version as more convincing, and 27.8% noticed no difference between both animations (Figure 6, Part 2). When testing if there was a significant difference in the ratings between Clip 1 and Clip 2, we obtained a difference of 0.5, with a 95% confidence interval of (0.173, 0.827), meaning that in this group the difference was also statistically significant. Some arguments favoring the selection of the AFS version were an increased sense of liveliness, more details moving in the face, and that it felt more fluid probably because of an increased emphasis in the eyebrows and cheeks.

7 DISCUSSION

We presented the Adaptable Facial Setup (AFS), an approach that uses captured facial data to automatically generate the basic building blocks of facial rigging, which can be employed to animate entirely new synthetic performances. Thanks to an adaptation process, the non-linear characteristic of the facial movements can be preserved, allowing the transfer of movements to other characters with different topology, and speeding up the process in comparison to the creation of shapes. Our system has been integrated in a DCC application, where artists can work in a more or less straightforward manner.

To prove the suitability of our approach and to address the challenges in the creation of historical characters, we decided to recreate a digital version of the face and head of Albert Einstein. One of the main limitations was the little amount of colored footage and printed material from Einstein, making difficult to compare our results to a ground truth.

In order to better appraise the advantages of our rigging method (AFS) compared to more traditional methods like rigging through blendshapes, we ran an user perception evaluation. Then we could see whether the animations created with AFS were seen as more convincing than those created with a blendshape rig. The results proved that indeed, the animations created with the AFS rig were perceived as more natural, more expressive and with smoother transitions between facial movements.

The process described is still very manual in nature. This is something we considered important to maintain to give the artist enough freedom to work with the facial rigs without the constraints often implied by more automatic methods. The outcomes of this research project, including live action plates, matchmove data, textures and 3D scenes will be available to the public by the end of 2017 under Creative Commons license. The latest AFS version is free for academic and non-commercial use [5].

In the future, we want to investigate the effect of dynamic microstructure displacement when the skin stretches or compresses. Currently, a first integration has been done and we are evaluating the results. Moreover we are investigating possibilities to further automate the process of fitting the custom heads to the FACS model without losing artistic freedom. For example the initial data adaptation could be automated by enabling the tool to identify facial features and placing landmarks using deep learning. Also the adaption techniques might profit from predetermined sets of values, which could be loaded to automatically obtain a distinct range of motion and to achieve a natural look.

ACKNOWLEDGMENTS

This project is partly funded by the MFG Foundation in the scope of the *Karl-Steinbuch-Forschungsprogramm* (KFS) research program. Special thanks to the team at Filmakademie Baden-Württemberg: Leszek Plichta (digital sculpting), Lisa Ecker (rigging), Julian Oberbeck (real-time shading), Angela Jedek (animation), Jiayan Chen (facial hair creation), Simon Spielmann (camera), Andreas Schuster (boom), Claudia Frank (production design), Frederika Höllrigl (makeup), Holger Schönberger (lighting), and Juliane Voigtländer (casting). The cast was sculpted by Jan Ptassek and scanned at Disney Research Zurich by Thabo Beeler. The actor portraying Einstein was Ernst Konarek.

REFERENCES

[1] Oleg Alexander, Mike Rogers, William Lambeth, Matt Chiang, and Paul Debevec. 2009. Creating a Photoreal Digital Actor: The Digital Emily Project. In *IEEE European Conference on Visual Media Production (CVMP)*.
[2] Diana Arellano, Reinhold Rauh, Benjamin Krautheim, Marc Spicker, Ulrich Max Schaller, Volker Helzle, and Oliver Deussen. 2017. Interactive testbed for research in autism âĂŤ the SARA project. *Universal Access in the Information Society* (2017). https://doi.org/10.1007/s10209-016-0521-9
[3] Andrea Arghinenti. 2011. Animation Workflow in KILLZONE3&Trade;: A Fast Facial Retargeting System for Game Characters. In *ACM SIGGRAPH 2011 Talks (SIGGRAPH '11)*. ACM, NY, USA, 37:1–37:1.
[4] Thabo Beeler, Derek Bradley, Bernd Bickel, and Markus Gross. 2017. Medusa Performance Capture. (2017). Retrieved August 07, 2017 from http://medusa.disneyresearch.com/
[5] Pascal Bérard, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. Lightweight Eye Capture Using a Parametric Model. *ACM Trans. Graph.* 35, 4 (2016), 117:1–117:12. https://doi.org/10.1145/2897824.2925962
[6] Kiran S. Bhat, Rony Goldenthal, Yuting Ye, Ronald Mallet, and Michael Koperwas. 2013. High Fidelity Facial Animation Capture and Retargeting with Contours. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '13)*. NY, USA, 7–14.
[7] Paul Charette, Mark Sagar, Greg DeCamp, and Jessica Vallot. 1999. The Jester. In *ACM SIGGRAPH 99 Electronic Art and Animation Catalog (SIGGRAPH '99)*. ACM, NY, USA, 151–.
[8] Darren Cosker, Peter Eisert, and Volker Helzle. 2015. *Facial Capture and Animation in Visual Effects.* CRC Press, 305–315.

[5] https://animationsinstitut.de/en/research-development/tools/facial-animation-toolset/info/

[9] Paul Debevec. 2012. The Light Stages and Their Applications to Photoreal Digital Actors. In *SIGGRAPH Asia*. Singapore.
[10] Sven Dreesbach. [n. d.]. The Gathering. ([n. d.]). Retrieved August 10, 2017 from http://www.svendreesbach.com/the-gathering/
[11] Ludovic Dutreve, Alexandre Meyer, Veronica Orvalho, and Saida Bouakaz. 2010. Easy Rigging of Face by Automatic Registration and Transfer of Skinning Parameters. In *International Conference on Computer Vision and Graphics*. Springer, 333–341.
[12] Paul Ekman. 1999. *Facial Expressions.* John Wiley & Sons Ltd.
[13] Paul Ekman and Wallace V. Friesen. 1978. *The Facial Action Coding System.* Palo Alto, CA: Consulting Psychological Press.
[14] Tim Enstice. 2009. The Curious Case of Benjamin Button. In *ACM SIGGRAPH 2009 Computer Animation Festival (SIGGRAPH '09)*. ACM, New York, NY, USA, 36–37. https://doi.org/10.1145/1596685.1596715
[15] Joe Fordham (Ed.). 2003. Middle earth strikes back. *Cinefex* 92 (January 2003).
[16] Parag Havaldar. 2006. Sony Pictures Imageworks. In *ACM SIGGRAPH 2006 Courses*. ACM, NY, USA. https://doi.org/10.1145/1185657.1185845
[17] Volker Helzle. 2007. Kinski Revisited. In *ACM SIGGRAPH 2007 Computer Animation Festival (SIGGRAPH '07)*. ACM, NY, USA, 85–. https://doi.org/10.1145/1281740.1281815
[18] Volker Helzle, Christoph Biehn, Thomas Schlömer, and Florian Linner. 2004. Adaptable Setup for Performance Driven Facial Animation. In *ACM SIGGRAPH 2004 Sketches (SIGGRAPH '04)*. ACM, New York, USA, 54–.
[19] Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2014. Robust Hair Capture Using Simulated Examples. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2014)* 33, 4 (2014).
[20] Kathrin Kaulard, Douglas W. Cunningham, Heinrich H. Bülthoff, and Christian Wallraven. [n. d.]. The MPI Facial Expression Database - A Validated Database of Emotional and Conversational Facial Expressions. *PLOS One* ([n. d.]), 32:1–32:6.
[21] Daniel Kayser. 2017. GDC 2017: UE4 Animation and Physics Technical Showcase. (March 2017). Retrieved July 21, 2017 from https://www.unrealengine.com/en-US/blog/gdc-2017-ue4-animation-and-physics-technical-showcase
[22] Hao Li, Thibaut Weise, and Mark Pauly. 2010. Example-based Facial Rigging. *ACM Trans. Graph.* 29, 4 (2010), 32:1–32:6.
[23] Daniel Lundqvist, Anders Flykt, and Arne Öhman. 1998. *The Karolinska Directed Emotional Faces - KDEF.* Technical Report. Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.
[24] Miquel Mascaró, Francisco J. Serón, Francisco J. Perales, and Javier Varona. 2014. *Rigging and Data Capture for the Facial Animation of Virtual Actors.* Springer International Publishing, 170–179.
[25] Daniel McDuff, Rana El Kaliouby, Thibaud Senechal, May Amr, Jeffrey Cohn, and Rosalind Picard. 2013. Affectiva MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected "In the WildâĂŹâĂŹ. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.
[26] National Institute of Standards and Technology. 2011. color FERET Database. (2011). Retrieved July 24, 2017 from https://www.nist.gov/itl/iad/image-group/color-feret-database
[27] Jeffrey A. Okun and Susan Zwerman. 2014. *The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures.* CRC Press, 2014.
[28] Verónica Orvalho, Pedro Bastos, Frederic Parke, Bruno Oliveira, and Xenxo Alvarez. 2012. A Facial Rigging Survey. In *Proc Eurographics*, Vol. 32. 10–32.
[29] Kiaran Ritchie, Jake Callery, and Karim Biri. 2005. The Art of Rigging (A Definitive Guide to Character Technical Direction with Alias Maya, Volume 1). Vol. 1. CG ToolKit.
[30] Albert Rizzo, Ulrich Neumann, Reyes Enciso, Douglas Fidaleo, and J. Y. Noh. 2001. Performance-Driven Facial Animation: Basic Research on Human Judgments of Emotional State in Facial Avatars. *Cyberpsychology & Behavior* 4, 4 (2001).
[31] Mark Sagar. 2000. Young at heart. In *ACM SIGGRAPH 00 Electronic Theater (SIGGRAPH '00)*. ACM, NY, USA.
[32] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. 2008. Biometrics and Identity Management. Springer-Verlag, Berlin, Heidelberg, Chapter Bosphorus Database for 3D Face Analysis, 47–56.
[33] Yeongho Seol, Wan-Chun Ma, and J. P. Lewis. 2016. Creating an Actor-specific Facial Rig from Performance Capture. In *Proceedings of the 2016 Symposium on Digital Production (DigiPro '16)*. ACM, NY, USA, 13–17.
[34] Mike Seymour, Chris Evans, and Kim Libreri. 2017. Meet Mike: Epic Avatars. In *ACM SIGGRAPH 2017 VR Village*. ACM, NY, USA, 12:1–12:2. https://doi.org/10.1145/3089269.3089276
[35] Martin Wallner and Stefan Leuchtenberg. [n. d.]. A Lost and Found Box of Human Sensation. ([n. d.]). Retrieved August 10, 2017 from http://www.lailaps.eu/portfolio/gotcha/