Phace: Physics-based Face Modeling and Animation

ALEXANDRU - EUGEN ICHIM, EPFL
PETR KADLEČEK, Charles University in Prague / University of Utah
LADISLAV KAVAN, University of Utah
MARK PAULY, EPFL

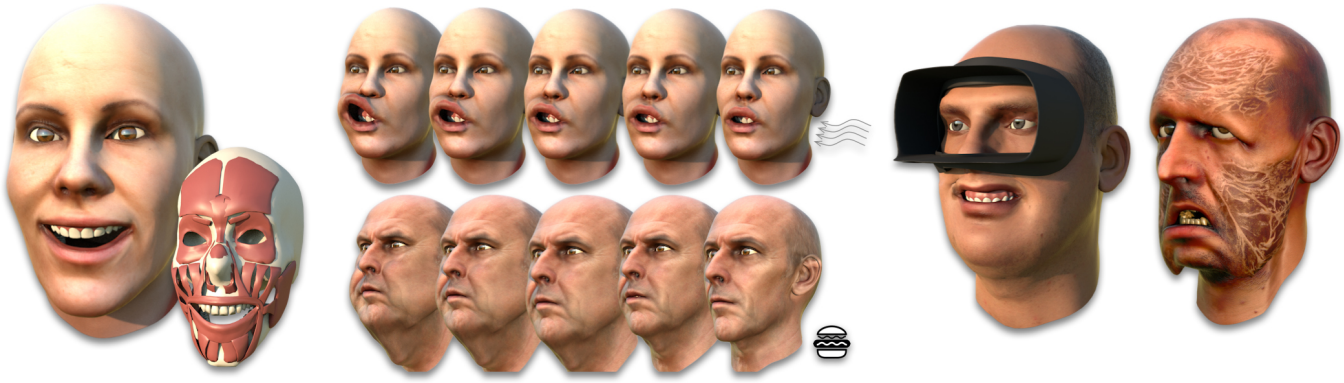


Fig. 1. Physics-based simulation facilitates a number of advanced effects for facial animation, such as applying wind forces, fattening and slimming of the face, wearing a VR headset, and even turning into a zombie.

We present a novel physics-based approach to facial animation. Contrary to commonly used generative methods, our solution computes facial expressions by minimizing a set of non-linear potential energies that model the physical interaction of passive flesh, active muscles, and rigid bone structures. By integrating collision and contact handling into the simulation, our algorithm avoids inconsistent poses commonly observed in generative methods such as blendshape rigs. A novel muscle activation model leads to a robust optimization that faithfully reproduces complex facial articulations. We show how person-specific simulation models can be built from a few expression scans with a minimal data acquisition process and an almost entirely automated processing pipeline. Our method supports temporal dynamics due to inertia or external forces, incorporates skin sliding to avoid unnatural stretching, and offers full control of the simulation parameters, which enables a variety of advanced animation effects. For example, slimming or fattening the face is achieved by simply scaling the volume of the soft tissue elements. We show a series of application demos, including artistic editing of the animation model, simulation of corrective facial surgery, or dynamic interaction with external forces and objects.

CCS Concepts: • **Computing methodologies** → **Physical simulation**;

Additional Key Words and Phrases: 3D avatar creation, facial animation, anatomical models, rigging

DOI: http://dx.doi.org/10.1145/3072959.3073664

ACM Reference format:
Alexandru - Eugen Ichim, Petr Kadleček, Ladislav Kavan, and Mark Pauly. 2017. Phace: Physics-based Face Modeling and Animation. *ACM Trans. Graph.* 36, 4, Article 153 (July 2017), 14 pages.
DOI: http://dx.doi.org/10.1145/3072959.3073664

1 INTRODUCTION

Accurate simulation of facial motion is of paramount importance in computer animation for feature films and games, but also in medical applications such as regenerative and plastic surgery. Realistic facial animation has seen significant progress in recent years, largely due to novel algorithms for face tracking and improvements in acquisition technology (Klehm et al. 2015; von der Pahlen et al. 2014).

High-end facial animations are most commonly produced using a sophisticated data capture procedure in combination with algorithmic and manual data processing. While video-realistic animations can be created in this manner, the production effort is significant and costly. A main reason is that complex physical interactions are difficult to recreate with the commonly employed reduced model representations. For example in blendshape rigs, collisions around the lip regions or inertial effects of the facial tissue are typically not accounted for. To remedy these shortcomings, artists often introduce hundreds of corrective shapes that need to be carefully sculpted and blended to achieve the desired effect in each specific animation sequence (Lewis et al. 2014).

Recent work (Barrielle et al. 2016; Ichim et al. 2016) proposes to avoid these shortcomings by augmenting the generative approach of blendshape animation with a simulation-based solution. A key

benefit of physics-based simulation is the ability to correctly handle collision and contact, both for internal contact of facial tissue or bones, as well as for collisions with external objects. In addition, secondary motion, such as inertial deformations or other time-dependent effects can easily be integrated into the optimization pipeline.

One major difficulty in simulation-based approaches is to achieve the required level of realism, which is particularly challenging for facial animation, due to the heightened human sensitivity for facial motion perception (Bruce and Young 1986). Accurate simulation requires building a detailed volumetric face model that faithfully represents the shape and dynamics of the captured subject. However, acquiring such a volumetric face model is challenging. Volumetric data produced by CT or MRI scanners is often difficult to convert into a simulation-ready representation. So far, successful pioneering methods required a significant amount of manual editing (Sifakis et al. 2005), which makes them difficult to deploy at scale.

We approach this problem by combining easy-to-obtain facial surface scans with a template model that integrates rigid bone structures, active muscle tissues and passive flesh, fat, and skin layers in a fully volumetric simulation model of the human face (see Figure 3). By scanning the subject in multiple facial poses, we obtain a representation of the geometry and expression dynamics of the acquired person. We then solve an inverse problem to estimate the activation parameters of the registered template rest pose in order to best reproduce the scanned expressions under activation.

We propose a novel muscle activation model in order to match the input scans more accurately. Unlike previous models that are constrained by fixed fiber directions, our model introduces additional degrees of freedom to support any deformation devoid of global rotation (since a muscle cannot rotate itself). This generalized model avoids the problem of relying on pre-determined fiber directions which are often inaccurate.

Subsequently, we can create new animations driven by muscle activations using a forward physics simulation that incorporates collision handling, volume preservation, inertia, and external forces such as wind forces or gravity. Muscle activations can be computed from a temporal sequence of blendshape weights, which enables straightforward integration into existing animation environments.

Contributions. The main technical contributions of our work are:

- a novel muscle activation model that offers more flexibility than standard fiber-based models,
- a physics-based simulation method that retains realism even with significant external forces or substantial modifications of the face geometry and tissue material properties,
- an inverse modeling optimization to adapt the simulation template to a series of expression scans of a specific person.

An important feature of physics-based approaches is that their parameters can be controlled to achieve the desired effects. In our case, the parameters include the stiffness of simulation elements, their rest shape volume, the static bone structure, or the muscle activation parameters. This detailed control facilitates numerous new applications that are difficult to achieve with existing methods. Examples we show in this paper include

- slimming and fattening of the face by adapting the volume of soft tissue,
- simulation of corrective facial surgery, such as orthognathic surgery to correct for jaw malformations,
- dynamic interaction with external forces (e.g. wind) and objects (e.g. VR headsets),
- artistic editing of facial expression dynamics by modifying tissue stiffness or muscle behavior.

Overview. Figure 2 provides a visual summary of our physics-based face modeling and animation approach. Central to our method is a face template model that combines volumetric and surface elements as shown in Figure 3. Physics-based optimization is performed on a tetrahedralized volumetric model composed of rigid bones and deformable tissue. The latter is further separated into *active* muscles, and *passive* flesh and skin. Muscles actively deform to drive the dynamic motion of the face model. In order to control the animation, we augment the volumetric template with a surface blendshape basis that represents the facial expression space. This also provide an interface to the surface scans used to build actor-specific simulation models.

The core algorithmic components of our method are the inverse and forward physics simulation modules. Inverse physics is used in a model building stage to create a simulation-ready anatomical face model of a specific person. As input to this preprocessing stage, we assume a set of surface scans that are first transformed to a user-specific blendshape model. An anatomy transfer step warps the volumetric template towards the neutral expression of the blendshape model. Subsequently, our inverse physics solver computes suitable muscle activations of the simulation model to best approximate each expression blendshape.

Given the person-specific simulation model and corresponding muscle activation patterns, we can apply forward physics simulation to compute dynamic face articulations. This animation stage takes as input a temporal series of blendshape weights that are mapped to per-frame muscle activations. External effects such as gravity or object collisions can be incorporated in the simulation to support a wide range of dynamic effects.

The rest of the paper is organized as follows: we first put our work in context by discussing related work in Section 2. In Section 3 we present our simulation template model. Then we introduce the forward and inverse physics simulation algorithms in Sections 4 and 5, respectively. Section 6 explains how these components are integrated into the model building and animation stages. In Section 7 we analyze the behavior of our method and provide comparisons to previous work. We show several application demos in Section 8, before concluding with a discussion of limitations and future work.

2 RELATED WORK

Data-driven methods. A significant body of work in facial animation is based on data-driven techniques. Multi-view stereo acquisition systems are used extensively to acquire detailed geometry and texture models, e.g., (Alexander et al. 2010; Amberg et al. 2007; Beeler et al. 2010). Avatar creation based on simple cell-phone camera acquisition was proposed by Ichim et al. (2015), while depth sensors are often used to create 3D avatars suitable for realtime

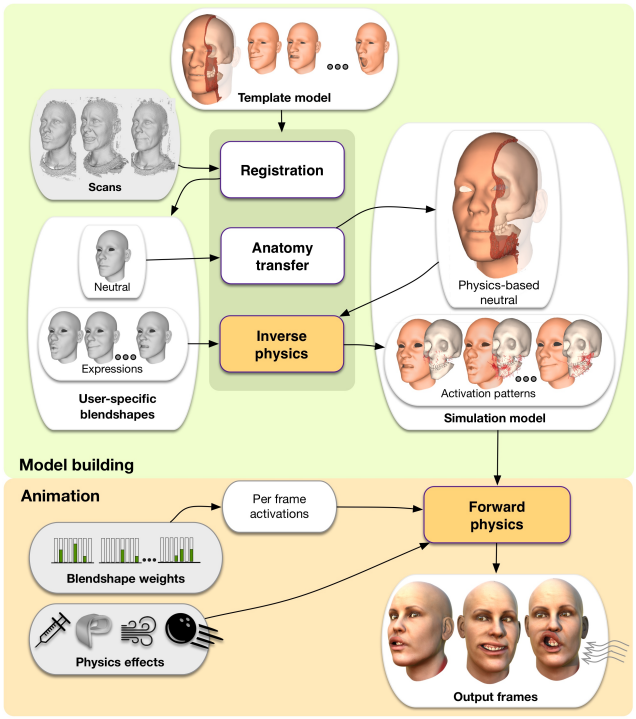


Fig. 2. Schematic workflow of our method.

tracking, e.g., (Bouaziz et al. 2013; Li et al. 2013; Weise et al. 2011). These methods typically rely on data-driven priors to guide the reconstruction process, in particular morphable face models (Blanz and Vetter 1999) or multi-linear (tensor) decomposition (Cao et al. 2014; Vlasic et al. 2005). They can be used in combination with upsampling methods, for example to add subject-specific details such as wrinkles (Bermano et al. 2014). Data-driven methods typically do not capture dynamic effects, even though some recent progress on this front has been made in the case of full-body animation (Pons-Moll et al. 2015).

Advanced acquisition. In recent years we have witnessed significant improvements in acquisition of facial performance and morphology, in particular detailed skin microstructure (Nagano et al. 2015), eyes (Bérard et al. 2016, 2014), eyelids (Bermano et al. 2015), hair (Hu et al. 2015), lips (Garrido et al. 2016) and teeth (Wu et al. 2016a). Modern methods can also capture medium-scale details (wrinkles) from monocular input in real-time (Cao et al. 2015). Anatomical constraints have proven useful in estimating the rigid transformation of the skull (rigid stabilization) (Beeler and Bradley 2014) and extracting detailed flesh deformations (Wu et al. 2016b).

Anatomical models. Early anatomical models were based on procedural models such as FFD (Chadwick et al. 1989). Procedural muscle models were also used in pioneering work on facial reconstruction (Kähler et al. 2003). Physics-based models of muscles and passive soft tissues were explored by Teran and colleagues (2003; 2005a; 2005b), later extended into a comprehensive biomechanical model of the upper body (Lee et al. 2009), and combined with fluid simulation to study swimming (Si et al. 2014).

Biomechanical modeling is a complex task and several software platforms support soft tissue simulation, such as Sofa (Allard et al. 2007), ArtiSynth (Lloyd et al. 2012), or FEBio (Maas et al. 2012). An important aspect of soft tissue modeling is the capture of material properties (Bickel et al. 2009) and their reproduction using modern fabrication methods such as 3D printing (Bickel et al. 2012).

Algorithmic and numerical aspects of soft tissue simulation continue to be a topic of active research; recently, Fan et al. (2014) proposed an Eulerian-on-Lagrangian method to simulate dynamic musculoskeletal systems, while Saito et al. (2015) applied Projective Dynamics (Bouaziz et al. 2014) to simulate hypertrophy or atrophy of the muscles or fat.

More recently, Kadlecek et al. (2016) studied the inverse problem of full-body modeling, inferring effects such as hypertrophy or atrophy of skeletal muscles from input 3D scans. Despite certain similarities to faces, a key difference is that full-body animation is characterized by muscles moving the bones, e.g., biceps moving the elbow. In facial animation, the skeletal articulation is limited to the jaw bone and facial expressions are created mainly by muscles pulling one another without any associated bone motion.

Physics-based face animation. The pioneering work of Sifakis et al. (2005) proposed a fully physics-based facial animation model, built from MRI and laser scan data of one specific subject. The key differences of our method are a more flexible muscle activation model combined with a more efficient inverse physics solver (Section 5). While Sifakis et al. (2005) also solve the inverse activation problem, their approach needs to invert a dense $n \times n$ matrix, where n is the number activation variables. This limits their method to using only low-dimensional activations, such as one activation per muscle, as opposed to our approach that allows for a richer high-dimensional activation model. A key benefit of our approach is that building the simulation model for different people is an almost entirely automatic process, as opposed to the substantial manual work required in previous work (Sifakis et al. 2005). Without the need of detailed parameter tuning, our approach also simplifies facial modifications such as slimming/fattening or geometric edits of the rigid bones.

The problem of scaling physics-based animation to different subjects has also been addressed in Cong et al. (2015). They propose a method that uses only the neutral expression to adapt an anatomical face model to different characters, including fictional ones. However, they do not attempt to closely match specific expressions of the target character as in our approach.

Cong and colleagues (2016) introduce "art-directed muscles", i.e., blendshape models applied to the muscles. This approach caters to experienced visual artists who appreciate direct control over their anatomical rigs. However, the art-directed muscles lack translation and rotation invariance which limits their ability to generalize, e.g., to significant facial modifications or large external forces inducing displacements of entire muscle groups. We propose a translation and rotation invariant muscle activation model and an automatic inverse physics procedure for inferring muscle activations from target expressions.

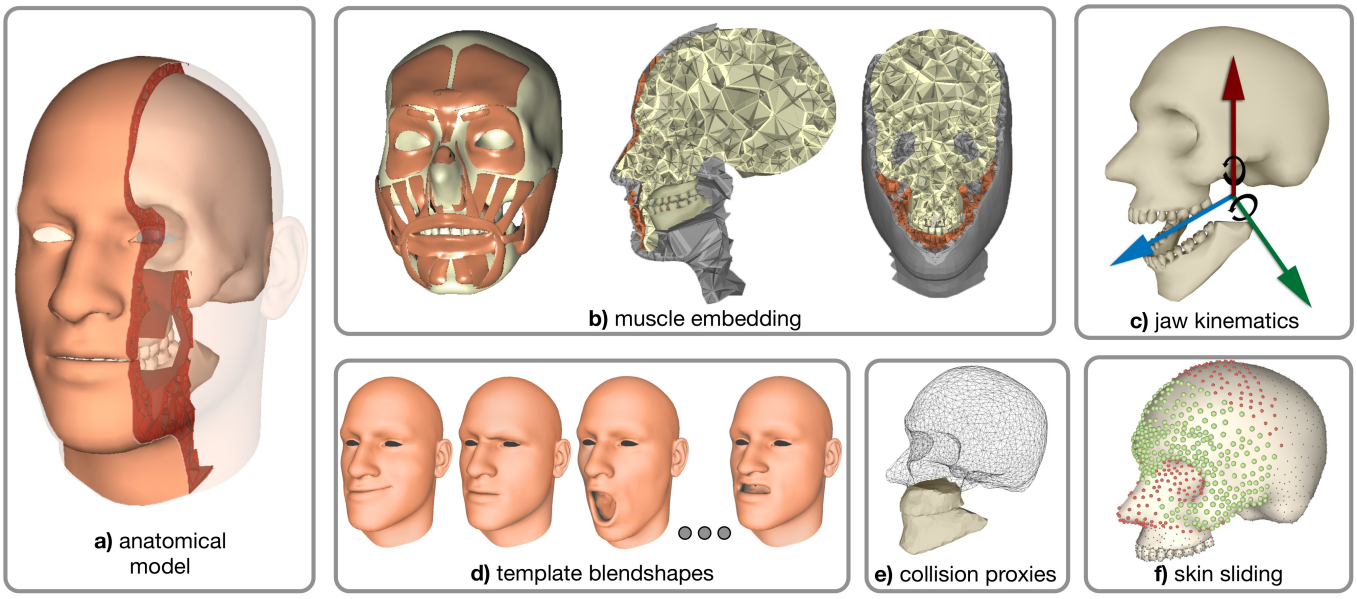


Fig. 3. Our template model consists of a volumetric representation of the tissue and bones (a), and a surface blendshape basis to represent the expression space (d). Muscles are embedded into a non-conforming tetrahedral mesh discretization (b). We explicitly model jaw kinematics with a 5 DoF joint (c) and utilize low-resolution geometry proxies for faster collision detection for the teeth region (e). Dynamic skin sliding is supported by introducing both sliding (green) and fixed (red) constraints for bone-tissue connections (f).

Alternative approaches to physics-based facial animation use mass-spring systems (Ma et al. 2012) or finite element modeling of the face as elastic thin shell (Barrielle et al. 2016). While these methods support certain types of physics-based effects, a surface-only approach does not correctly handle collisions or support volumetric face modifications, such as visualizing the outcome of facial surgery. Modeling interior tissue and bones is also important when the face is subjected to inertial or external forces that visibly expose the rigidity of the internal bone structure.

Volumetric blendshapes as proposed in Ichim et al. (2016) introduce energy terms attracting deformation gradients to their target values derived from input facial expressions of a given person. The volumetric blendshapes are translation invariant, but they lack rotation invariance, introducing similar artifacts as linear elasticity, especially in situations with large external forces (Figure 10). In this paper, we create a model compatible with traditional blendshape interfaces, but we push the anatomical realism further by utilizing a novel muscle activation model, separating active and passive soft tissue layers, and introducing sliding constraints to attach soft tissue to the bones. As a consequence, our model implements a variety of advanced animation effects and supports significant modifications of the face simulation model, which enables a number of new applications as demonstrated in Section 8.

3 TEMPLATE FACE MODEL

Our approach starts from a *generic face model* – an anatomical face template corresponding to an average human subject (see Figure 3). We created this model from a commercially available anatomical data set (Zygote 2016) that contains polygonal representations of the bones (the skull, the jaw, including teeth), skin (including a realistic model of the oral cavity), and 33 facial muscles. Using the winding-number method of Jacobson et al. (2013) we generate a tetrahedral mesh discretizing the soft tissue of the face. Our tetmesh conforms to the skin and the bones, but not to the muscles, because a conforming discretization of the numerous thin facial muscles would require prohibitively many elements. Instead, we use non-conforming discretization where every tetrahedron can represent multiple types of soft tissues. We distinguish between two types of soft tissues: *active* corresponds to muscles, while *passive* corresponds to subcutaneous fat, connective tissue and the skin, i.e., tissue that is not voluntarily activated by neural signals (Figure 3-b).

Up to the accuracy of the discretization, the active layer corresponds to the union of all facial muscles, while the passive layer forms the region between the active layer and the skin and fills in areas between the bones. Even though this model is not as accurate as modeling every muscle individually, it captures the key fact that the shape of the skin is affected by facial muscles only *indirectly*, i.e., the contracted muscles deform passive soft tissue, which consequently induces skin deformations.

Jaw kinematics. The relative motion of the jaw with respect to the skull contributes significantly to the final articulation of the face. The kinematics of the temporomandibular joint is non-trivial, consisting of both rotational and translational motion. In our model (see Figure 3-c), we define the major rotation axis (x-axis, corresponding to mouth opening) as the axis passing through the centers of the mandibular condyles. Halfway through the condyles, we define a perpendicular axis (y-axis) corresponding to vertical jaw rotation. The jaw does not normally rotate about the third orthogonal axis

(z-axis), but it can translate (slightly) in all three directions. This amounts to 5 DoFs for the jaw motion, expressed with respect to the skull, which is treated as a free rigid body (our model does not include the craniocervical junction). We concatenate the kinematic parameters of the jaw bone into a vector $\mathbf{b} \in \mathbb{R}^5$.

Template blendshapes. Given an anatomical model of the face, a natural control interface would be activation signals for all motor units. While biologically meaningful, such controls would not be user-friendly, because many motor units can affect a surface point in a complex, non-linear way. Instead, we augment our template model with a set of 48 blendshapes inspired by FACS (Ekman and Friesen 1977) that have been sculpted by an artist on our generic face model. These blendshapes are only defined on the skin as a basis for parametrizing the space of facial expressions. They provide no information about the internal deformations, which are handled by physics-based simulation (Section 4 and Section 5). This combination of surface blendshape basis and volumetric simulation model allows us to retain compatibility with commonly used blendshape controls, while offering the benefits of advanced physics-based simulation effects.

4 FORWARD PHYSICS

The goal of the forward physics algorithm is to compute the deformed soft tissue and resulting skin surface given bone kinematics and muscle activation parameters. We model the latter with a vector $\mathbf{a}$ (see "Active tissue" below) that represents the amount of activation (contraction) of all facial muscles. Even though in reality the jaw motion is controlled by muscle activations (in particular the masseter muscle) our model assumes the bones are directly controlled kinematically and the muscle activations are used only to create the facial expressions.

At the heart of our method is a physics-based model of soft tissue elasticity including muscle activation. We define this model using linear finite elements on our tet-mesh adapted for a given subject. Let $\mathbf{x}$ denote a vector stacking all degrees of freedom of the soft tissue, i.e., the 3D coordinates of all nodes.

Passive tissue. For passive tissue we define deformation energy

$$E_{\text{pass}}(\mathbf{x}) = \sum_i \min_{\mathbf{R}_i \in SO(3)} W_i^{\text{pass}} \mu \|\mathbf{F}_i(\mathbf{x}) - \mathbf{R}_i\|_F^2 + W_i^{\text{pass}} \lambda (\det(\mathbf{F}_i(\mathbf{x})) - 1)^2, \quad (1)$$

where the index i goes over all tets and $W_i^{\text{pass}} \geq 0$ denotes the volume of the i-th tetrahedron that is occupied by passive tissue, precomputed during template construction with Monte-Carlo sampling. The first term in Eq. 1 corresponds to the commonly used co-rotated elasticity (measure of deviation from rigid motion), while the second term models the resistance to changes of volume. $\mathbf{F}_i(\mathbf{x})$ denotes the deformation gradient, and $\mathbf{R}_i$ is an auxiliary rotation matrix used in the co-rotated model (Sifakis and Barbic 2012). μ and λ are material parameters that we set by default to $\mu = 1$ and $\lambda = 3$. We can change these parameters to achieve specific effects as discussed in Section 8.

Active tissue. For tets corresponding to the active layer (muscles), we propose a novel activation model. Previous muscle models typically assume a given direction of muscle fibers along which the muscle contracts (Lee et al. 2009; Teran et al. 2005a). While this corresponds to the biological structure of muscles, the problem is that the exact muscle fiber directions are in general not known. Medical imaging techniques such as diffusion tensor imaging are prohibitively expensive and time consuming, and the signal quality is limited.

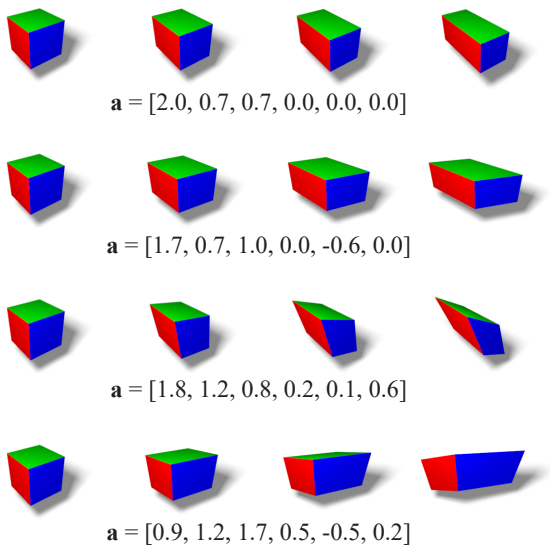


Fig. 4. Visualization of the capabilities of our 6-DoF activation model by squishing a cube, corresponding to a small sample of muscle tissue.

Previous work in graphics (Saito et al. 2015) applied ad-hoc muscle fiber approximations which worked well for major skeletal muscles (such as the biceps), but are not sufficiently accurate for the delicate facial muscles. Along with the exact location of muscle insertion points, tuning of these parameters to obtain realistic facial expressions is possible, but tedious (Sifakis et al. 2005). To circumvent these issues, we propose a different muscle activation model that does not require explicit knowledge of fiber directions, but relies on the elementary bio-mechanical fact that muscles can generate only internal forces. In other words, an isolated muscle is not capable of translating or rotating by itself (even though the muscle can of course be translated or rotated due to contact with the surrounding tissues).

The property that the muscle cannot translate itself is already guaranteed by the translation invariance of the deformation gradient operator $\mathbf{F}_i(\mathbf{x})$. Since a muscle tet should also not rotate itself, we require the activation to be a *symmetric* 3×3 matrix. Every symmetric matrix in $\mathbb{R}^{3\times3}$ has an eigendecomposition of the form $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}$, where $\mathbf{Q} \in SO(3)$ and $\mathbf{\Lambda} \in \mathbb{R}^{3\times3}$ is diagonal. Therefore, the symmetric activation matrix corresponds to non-uniform scaling ($\mathbf{\Lambda}$) in an arbitrary orthonormal coordinate system ($\mathbf{Q}$). In other words, the symmetric matrix represents pure distortion without any change of orientation (Shoemake and Duff 1992) (see Figure 4).

For each active tet, we define an activation vector $\mathbf{a}_i \in R^6$ and use a linear operator $\mathcal{S} : \mathbb{R}^6 \rightarrow \mathbb{R}^{3\times3}$ to generate the corresponding symmetric matrix $\mathcal{S}(\mathbf{a}_i) \in \mathbb{R}^{3\times3}$. Muscles, like most biological soft tissue, are approximately incompressible, which means that $\det(\mathcal{S}(\mathbf{a}_i)) = \det(\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}) = \det(\mathbf{\Lambda})$ should be close to 1. However, to compensate for discretization errors, we do not enforce $\det(\mathcal{S}(\mathbf{a}_i)) = 1$ strictly, but only as a soft constraint, as discussed in Section 5.

We use this activation model to define the deformation energy $E_{\text{act}}(\mathbf{x}, \mathbf{a})$ of active tissue, where $\mathbf{a}$ is a vector stacking the 6-dimensional activation parameters for all active tets. Specifically, we define:

$$E_{\text{act}}(\mathbf{x}, \mathbf{a}) = \sum_i \min_{\mathbf{R}_i \in SO(3)} W_i^{\text{act}} \mu \|\mathbf{F}_i(\mathbf{x}) - \mathbf{R}_i \mathcal{S}(\mathbf{a}_i)\|_F^2 + W_i^{\text{act}} \lambda (\det(\mathbf{F}_i(\mathbf{x})) - \det(\mathcal{S}(\mathbf{a}_i)))^2, \quad (2)$$

where the index i goes over all tets and $W_i^{\text{act}} \geq 0$ represents the volume of the i-th tet that corresponds to active tissue. Here the co-rotated term aims to find the rotation $\mathbf{R}_i$ that best aligns the deformation gradient $\mathbf{F}_i(\mathbf{x})$ with $\mathcal{S}(\mathbf{a}_i)$. The second term encourages the volume ratio of the deformed tet (i.e., $\det(\mathbf{F}_i(\mathbf{x}))$) to align with the volume ratio of the activation matrix $\det(\mathcal{S}(\mathbf{a}_i))$, which should be close to 1, i.e., volume conserving.

Bone attachments. Muscles are connected to the bones using a complex network of connective tissue, whose exact function is a matter of active research (Schleip et al. 2013). In animation, the visual importance of skin sliding is well recognized (Li et al. 2013). To distinguish areas where soft tissue is directly attached to the bones from areas where soft tissue slides over the bones, we create two types of constraints: 1) pin constraints and 2) sliding constraints. The pin constraints are straightforward to implement using Dirichlet boundary conditions. The sliding constraints are modeled as point-on-plane constraints on the tangent planes of the bone surfaces. We found this approximation to be sufficient even for curved regions, since the amount of sliding displacement is generally small.

Formally, we express both pin and sliding constraints using a function $\mathbf{c}(\mathbf{x}, \mathbf{b})$ that depends also on the kinematic parameters $\mathbf{b} \in \mathbb{R}^5$ of the jaw bone. All of the constraints are satisfied if and only if $\mathbf{c}(\mathbf{x}, \mathbf{b}) = 0$. We have manually distributed the pin and sliding constraint as shown in Figure 3-f. The constraint types were chosen to achieve realistic deformations. For example, for an eyebrow raise expression, the skin slides over the skull as illustrated in Figure 5.

Quasi-static solution. In this section we discuss how to compute the quasi-static solution of the forward physics simulation, deferring the discussion of dynamics to Section 6. Quasi-statics means calculating a steady state where all dynamic motion has settled. The quasi-static regime is useful in generating static expressions and is particularly important when solving for muscle activations from observed shapes, as discussed in Section 5. Finding the steady state can be formulated as the following optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & E_{\text{pass}}(\mathbf{x}) + E_{\text{act}}(\mathbf{x}, \mathbf{a}) + E_{\text{grav}}(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{c}(\mathbf{x}, \mathbf{b}) = 0, \mathbf{p}(\mathbf{x}) \geq 0, \end{aligned} \quad (3)$$

where $E_{\text{grav}}(\mathbf{x})$ represents a linear gravitational potential (i.e., the familiar *mgh*). The inequality constraints $\mathbf{p}(\mathbf{x})$ are used to resolve penetrations (collision response) as follows. When collision detection finds a surface vertex penetrating the volumetric face model (see below for more details), an inequality constraint is appended to $\mathbf{p}$. For each offending vertex we find its projection onto the surface and create a tangent plane at this point. The inequality constraint requires the vertex to be at the half-space opposite the volume.

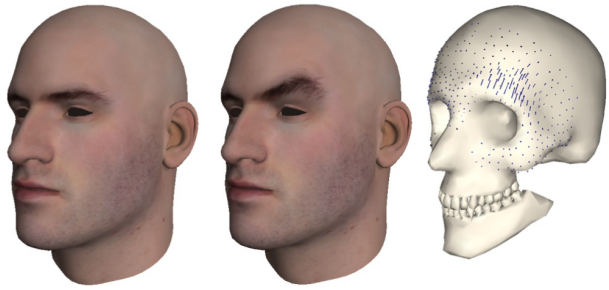


Fig. 5. An eyebrow raise expression uses the skin sliding feature of our model. The blue arrows show the displacement of the contact vertices between the cranium and the flesh.

We solve Eq. 3 by alternating between an interior point solver used to minimize Eq. 3 for fixed collision constraints $\mathbf{p}$, and collision detection to update $\mathbf{p}$. We have initially implemented a "homebrew" augmented Lagrangian solver, but ultimately decided to use the IPOPT package (Wächter and Biegler 2006), which has proven to be more robust and usually needs fewer iterations to converge.

Collisions - implementation details. We implemented collision handling between lips and the teeth and between the upper and lower lip, which are the areas most prone to inter-penetration. Because the geometry of the teeth is quite detailed (and we are not aspiring to simulate e.g. flossing where the detail would be necessary), we start by creating proxy collision shapes for the upper and lower teeth (see Figure 3-e). The upper and lower lip geometries are already sufficiently smooth and we do not need any special collision proxy. We detect collisions using AABB hierarchies built for the upper and lower teeth proxy geometries and the upper and lower lips. Since lips are deforming, we recompute the AABB hierarchies at run-time. We did not use techniques to avoid or amortize this recomputation costs as this was not a bottleneck in our implementation.

For each of the collision proxies and the lips, we also manually define a "projection region", i.e., subset of triangles where inter-penetrating vertices can be pushed to resolve collisions. Previous work considers all surface points as valid candidates for projection (McAdams et al. 2011). In our case this occasionally created problems such as resolving lip-teeth collisions by projecting the lip vertices behind the teeth, i.e., inside the mouth, which is rarely a plausible solution. Instead of more complicated continuous collision detection, we therefore disallowed these implausible projections. For each of the projection regions, we compute another AABB hierarchy that is used to find the closest point in the projection region, i.e., the location where an inter-penetrated vertex will be pushed in order to resolve the collision. A collision handling example during animation is shown in Figure 6.

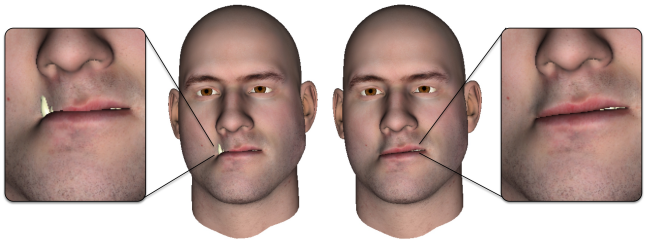


Fig. 6. Importance of collision handling. Without collisions, intersections between the teeth and the deformable tissue can occur (left). Our method correctly detects and handles the contact (right).

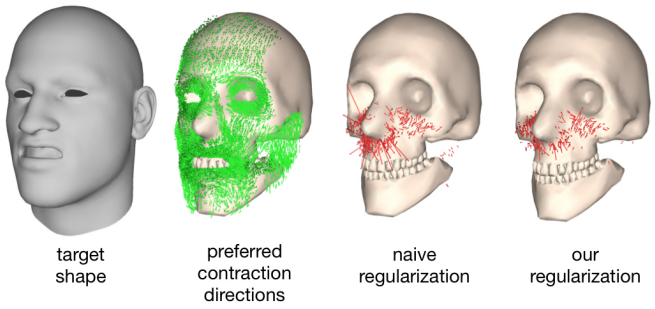


Fig. 7. Muscle activation regularization. Red lines indicate the direction and magnitude of the dominant muscle contraction, computed from the SVD of the activation matrix.

5 INVERSE PHYSICS

The previous section explains how to compute face articulations for given bone positions and muscle activations. In this section we discuss the inverse problem. For a given target shape of the skin, we want to compute the corresponding bone parameters $\mathbf{b}$ and muscle activations $\mathbf{a}$, which, when used in the forward simulation (Eq. 3), will produce a skin surface close to the input shape.

Optimization formulation. Let $\mathbf{t}$ denote the target vertex positions of the skin. The inverse modeling problem can be written as

$$\begin{aligned} \min_{\mathbf{x},\mathbf{a},\mathbf{b}} \quad & \|\mathbf{S}\mathbf{x} - \mathbf{T}\mathbf{t}\|^2 + R_{\text{act}}(\mathbf{a}) + R_{\text{sparse}}(\mathbf{a}) \\ \text{subj. to} \quad & \mathbf{c}(\mathbf{x},\mathbf{b}) = 0, \mathbf{p}(\mathbf{x}) \geq 0 \\ & \nabla_{\mathbf{x}} E_{\text{pass}}(\mathbf{x}) + \nabla_{\mathbf{x}} E_{\text{act}}(\mathbf{x},\mathbf{a}) + \nabla_{\mathbf{x}} E_{\text{grav}}(\mathbf{x}) = 0 \end{aligned} \tag{4}$$

where $R_{\text{act}}(\mathbf{a})$ and $R_{\text{sparse}}(\mathbf{a})$ are regularization terms discussed below. The objective term $\|\mathbf{S}\mathbf{x} - \mathbf{T}\mathbf{t}\|^2$ measures how close state $\mathbf{x}$ is to the target $\mathbf{t}$. The matrix $\mathbf{S}$ selects the simulation nodes corresponding to the skin surface. In addition $\mathbf{S}$ and $\mathbf{T}$ encode both position (point-to-point) and point-to-plane distance terms (Rusinkiewicz and Levoy 2001). The point-to-plane terms enable some amount of sliding (tangential motion) which is useful if we do not completely trust the correspondences represented by $\mathbf{t}$. The last vector equality constraint describes the condition of quasi-static equilibrium, i.e., the sum of all forces (gradients with respect to $\mathbf{x}$) is zero. Even though $\mathbf{x}$ is also an optimization variable, the desired output are the optimal values of muscle activations $\mathbf{a}$ and bone parameters $\mathbf{b}$.

Regularization. Without regularization, the optimization of Eq. 4 can lead to over-fitting and anatomically implausible activations $\mathbf{a}$. To provide an appropriate prior on activation patterns, we exploit the geometric structure of the muscles by estimating an approximate preferred contraction direction. Following Choi et al. (2013) we compute these directions by solving a Laplace equation and encode the corresponding orientations for the i-th tet as $\mathbf{Q}_i \in SO(3)$. The regularizing prior softly penalizes deviations in muscle contraction from the preferred direction and is defined as:

$$R_{\text{act}}(\mathbf{a}) = \sum_{i,m} f_{i,m} \left\| \mathbf{Q}_i^{\mathsf{T}} \begin{bmatrix} \gamma_{m(i)}^{-1} & 0 & 0 \\ 0 & \sqrt{\gamma_{m(i)}} & 0 \\ 0 & 0 & \sqrt{\gamma_{m(i)}} \end{bmatrix} \mathbf{Q}_i - \mathcal{S}(\mathbf{a}_i) \right\|^2$$

where i sums over all active tets and m over all muscles. Since our tet mesh does not conform to the muscles, some tets may be occupied only partially by a muscle, or be occupied by several muscles. We calculate the fraction $f_{i,m} \in [0,1]$ of tet i occupied by muscle m by Monte-Carlo sampling (if an active tet contains some amount of passive tissue, we get $\sum_m f_{i,m} < 1$ and the regularization strength is proportionally reduced as expected). The contraction parameters $\gamma_m \geq 0$ are auxiliary variables representing the contraction of muscle m. Recall that $\mathcal{S}(\mathbf{a})$ is our symmetric muscle activation matrix introduced in Section 4. Intuitively, $R_{\text{act}}(\mathbf{a})$ encourages all tets corresponding to a single muscle to contract in a uniform, volume preserving way (because $\gamma_{m(i)}^{-1}\sqrt{\gamma_{m(i)}}\sqrt{\gamma_{m(i)}} = 1$).

In addition to the muscle-activation regularization term R_{act}, we found it beneficial to also include the following term to promote sparse muscle activations:

$$R_{\text{sparse}}(\mathbf{a}) = \sum_m \sqrt{\sum_i f_{i,m} \|\mathbf{a}_i\|^2} \tag{5}$$

Specifically, this is a group sparsity term similar to L_1 regularization, but applied to entire groups – in our case, muscles. This term encourages all activations corresponding to one muscle to remain zero unless contributing significantly to the result. We introduced this term to avoid small spurious activations of remote muscles, which is justified when our target shapes $\mathbf{t}$ correspond to traditional FACS-type blendshape models which isolate individual action units. Figure 7 shows that compared to a naive L_2 regularization approach, our method leads to sparser activations that are better aligned with the geometric structure of the muscles.

Numerical solution. As in Section 4, we use interior-point methods (Wächter and Biegler 2006) to solve the constrained optimization problem in Eq. 4. Our implementation of the Hessian of the Lagrangian of Eq. 4 ignores third-order derivatives of E (pretends they are zero), amounting to the commonly used Gauss-Newton approximation of the Hessian (Bickel et al. 2012; Sifakis et al. 2005). We alternate the interior point solver with collision detection that determines the non-penetration constraints $\mathbf{p}$ as in Section 4.

Even though including the regularization term R_{act} could be directly incorporated into our optimization objective (adding γ_m as auxiliary variables), we found that this significantly increases the

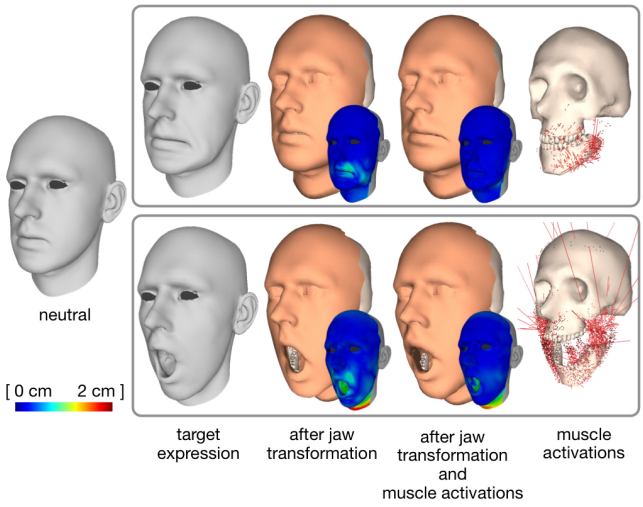


Fig. 8. Inverse physics finds jaw transformation and muscle activations that accurately reproduce the target blendshapes.

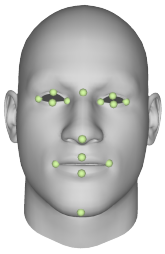
non-linearity of the problem and forces the non-linear solver to take many iterations, each making only slow progress towards the solution. To avoid this problem, we instead use a local-global approach (Sorkine and Alexa 2007). In the local step, the activations $\mathbf{a}$ are fixed and we compute optimal γ_m by finding roots of a 6-th order polynomial using the method of Brent (1971). In the global step, we call the interior point solver to optimize $\mathbf{a}$ for fixed γ_m, which is an easier optimization problem exhibiting fast convergence.

Figure 8 shows an example of an inverse physics solve for two blendshapes of a user-specific blendshape model, visualizing separately the effect of the jaw motion and the effect of muscle activations.

6 PHACE MODELING AND ANIMATION

In this section we explain how we integrate the optimization algorithms presented above into a complete system for creating and animating subject-specific face simulation models.

Model Building. We start by 3D scanning the face of our subject in neutral expression and about 5-10 additional premeditated facial expressions using a multi-view stereo setup as described in Ichim et al. (2016). Each of the scans is approximately aligned with the skin of our template model (Section 3) using rigid registration (plus uniform scale). Then we apply non-rigid ICP (Rusinkiewicz and Levoy 2001) to find dense correspondences between the template skin and the target scan, guided with a few manually chosen markers as shown in the inset. We denote the registered skin surfaces as $\mathbf{s}_{\text{neut}}$ for the neutral and $\mathbf{s}_k$ for k-th expression.



Next, we deform our volumetric template model such that its boundary (skin) aligns with $\mathbf{s}_{\text{neut}}$. This is accomplished with Anatomy Transfer (Dicko et al. 2013; Ichim et al. 2016). Note that during this process the generic face model can deform freely, i.e., the shape and/or volume of all cells can change, including the bones (in contrast to the deformation model considered in Section 4). We then use Example-Based Facial Rigging (Li et al. 2010) to convert the registered expressions $\mathbf{s}_k$ to subject-specific blendshapes $\mathbf{c}_j, j = 1, \ldots, 48$.

The processing steps so far essentially rely on existing methods to align the volumetric template to the neutral expression and to create the subject-specific blendshape model. We refer to the above cited papers for implementation details on these algorithms. After this geometric preprocessing, we now solve for activations $\mathbf{a}_j$ and jaw bone parameters $\mathbf{b}_j$ that correspond to each of the blendshapes $\mathbf{c}_j$ using the Inverse Physics optimization of Section 5.

Animation. To animate the created face model, we need to feed appropriate muscle activations and jaw bone parameters to the Forward Physics optimization of Section 4 for each animation frame. Given per-frame blendshape weights $\mathbf{w} = \{w_1, \ldots, w_{48}\}$, we compute muscle activations as $\mathbf{a} = \mathbf{a}_{\text{neut}} + \sum_j w_j(\mathbf{a}_j - \mathbf{a}_{\text{neut}})$, where $\mathbf{a}_{\text{neut}}$ corresponds to neutral activations, i.e., each activation $\mathcal{S}(\mathbf{a}_{j,i}) = \mathbf{I} \in \mathbb{R}^{3\times3}$. Linear blending of the activation parameters is justified because there is no rotational component in symmetric matrices (Shoemake and Duff 1992). Similarly, we compute the blended jaw kinematics parameters $\mathbf{b} = \sum_j w_j \mathbf{b}_j$. While blending of rotation angles is in general not recommended, we found that for the limited range of rotations of the jaw this simple scheme does not produce any visible artifacts.

Dynamics. Adding inertia corresponds to a minor change of Eq. 3. We use the popular backward Euler integration, which in its optimization form (Liu et al. 2013) corresponds to augmenting the objective of Eq. 3 with the term: $\frac{1}{2}\|\mathbf{x} - (\mathbf{x}_n + h\mathbf{v}_n)\|^2_{\mathbf{M}}$, where $\mathbf{x}_n$ and $\mathbf{v}_n$ are positions and velocities in the previous frame, $h > 0$ is the time step, and $\mathbf{M}$ is the mass matrix. We use a diagonal matrix $\mathbf{M}$ (mass lumping) with a soft tissue density of $1g/cm^3$. The minimizer $\mathbf{x}$ of Eq. 3 then becomes the new state $\mathbf{x}_{n+1}$ and the new velocity is $\mathbf{v}_{n+1} = (\mathbf{x}_{n+1} - \mathbf{x}_n)/h$. The main difference from the quasi-static solution is that the dynamic solution depends on the previous state $(\mathbf{x}_n, \mathbf{v}_n)$, i.e., we need to execute the time steps in sequence. To add non-conservative external forces, such as wind, we proceed as in Projective Dynamics (Bouaziz et al. 2014) and change the additional term to $\frac{1}{2}\|\mathbf{x} - (\mathbf{x}_n + h\mathbf{v}_n + h^2\mathbf{M}^{-1}\mathbf{f}_{\text{ext}})\|^2_{\mathbf{M}}$. Here $\mathbf{f}_{\text{ext}} \in \mathbb{R}^3$ is the external force vector, e.g., a wind force is a function of triangle normal, area, and wind direction.

Plasticity. To support effects such as fattening or slimming, we use a standard model of plastic deformations. Specifically, each total deformation gradient $\mathbf{F}_{\text{total}}(\mathbf{x})$ is assumed to be composed of an elastic deformation component and plastic deformation component, i.e., $\mathbf{F}_{\text{total}}(\mathbf{x}) = \mathbf{F}_{\text{elast}}(\mathbf{x})\mathbf{F}_{\text{plast}}$ or, equivalently, $\mathbf{F}_{\text{elast}}(\mathbf{x}) = \mathbf{F}_{\text{total}}(\mathbf{x})\mathbf{F}^{-1}_{\text{plast}}$. Note that $\mathbf{F}_{\text{plast}}$ does not depend on the current deformed state $\mathbf{x}$. The deformation gradient $\mathbf{F}_i(\mathbf{x})$ used in Eq. 1 and Eq. 2 corresponds to the elastic deformation component, because plasticity is a separate process, e.g., tissue growth, which is decoupled from elastic deformations. Therefore, the only modification we need to make to account for plasticity is to replace the $\mathbf{F}_i(\mathbf{x})$ in Eq. 1 and Eq. 2 by $\mathbf{F}_i(\mathbf{x})\mathbf{F}^{-1}_{\text{plast},i}$, where $\mathbf{F}_{\text{plast},i}$ describes the plastic deformation of the i-th tet. In our system, we use only uniform scaling, i.e., $\mathbf{F}_{\text{plast},i} = s_i\mathbf{I}$,

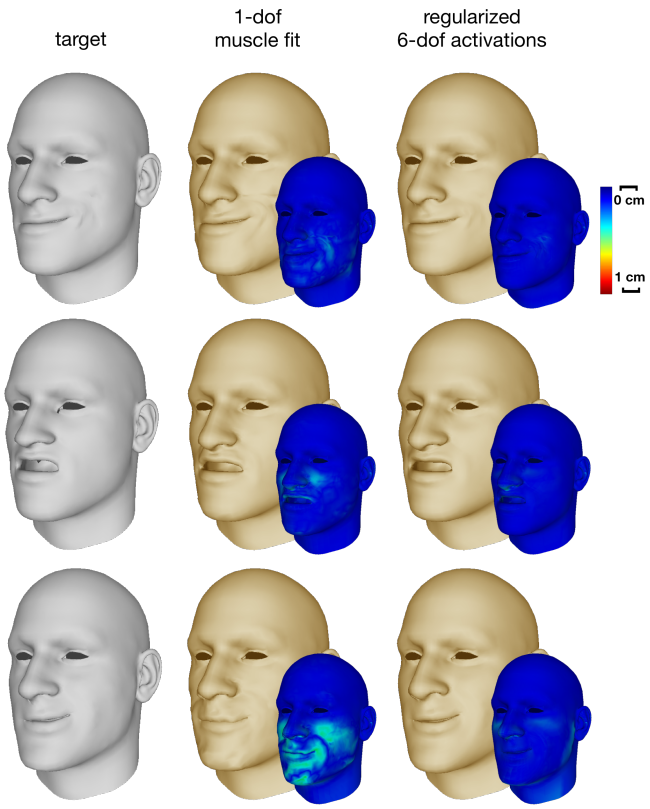


Fig. 9. Our 6-DoF muscle activation model (rigt) leads to more accurate reconstruction of the target expression (left) than previous 1-DoF fiber-aligned activations models (middle).

where $s_i > 0$ is a scaling coefficient (corresponding to growth for $s_i > 1$ and shrinking for $0 < s_i < 1$). The settings of the s_i parameters for each tet depend on the effect we wish to achieve as discussed in Section 8. Plasticity, as well as inertia and external forces are applied in forward physics only.

7 EVALUATION

Before showing application results of our method in Section 8, we evaluate the behavior of our optimization algorithms and provide comparisons to previous work.

Muscle activation model. As mentioned in Section 4, previous methods constrain the deformation along muscle fibre directions (Lee et al. 2009; Saito et al. 2015; Sifakis et al. 2005; Teran et al. 2005a). In our experiments we found that muscle fiber directions can be unreliable and lack the flexibility to accurately reproduce all facial expressions. This insight triggered the design of our more general activation model. In Figure 9 we compare the results of inverse physics with our method and the previous fiber-restricted model, where fiber directions are computed from our geometric muscle models using the method of Choi et al. (2013) (these directions are shown in Figure 7). The active tetrahedra of the 1-DOF muscle model act based on the following constraint energy:

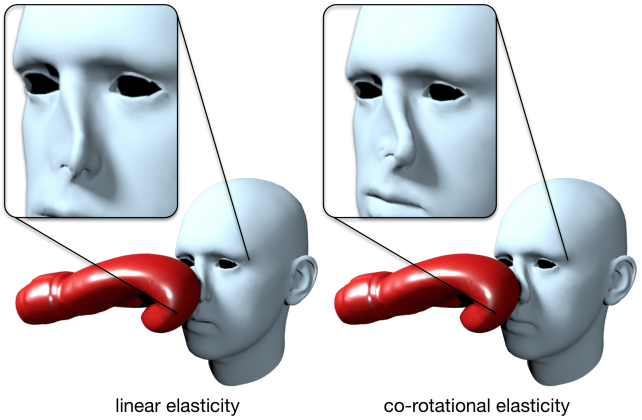


Fig. 10. A boxing punch to the nose results in artifacts with an elastic model lacking rotation invariance as in Ichim et al. (2016) (left). More realistic deformations are obtained with our rotation-invariant model (right).

$$E_{act\_1DOF}(\mathbf{x},\mathbf{a}) = W_i^{\text{act}}\mu\|\mathbf{F}_i(\mathbf{x}) - \mathbf{R}_i\mathbf{Q}^T\mathcal{S}_{1DOF}(\mathbf{a}_i)\mathbf{Q}\|_F^2 + W_i^{\text{act}}\lambda(\det(\mathbf{F}_i(\mathbf{x})) - \det(\mathcal{S}(\mathbf{a}_i)))^2,$$

where $\mathbf{Q}$ encodes the muscle fiber orientations and $\mathcal{S}_{1DOF}(\mathbf{a}_i) = diag(\mathbf{a}_i, 1, 1)$. As Figure 9 illustrates, muscle activations constrained to the fiber directions fail to closely match the desired target shape, while our activation model leads to a much more accurate reconstruction of the target expression.

Comparison to volumetric blendshapes. Defining a deformation model that is invariant under rigid motions is essential for correct tissue behavior. The volumetric blendshape approach of Ichim et al. (2016) lacks rotation invariance, which can lead to artifacts, e.g., when large rotation of the soft tissues are induced by external forces, such as the boxing punch shown in Figure 10. We propose rotation-invariant models for both passive and active soft tissue, leading to more realistic results. While we distinguish between passive and active tissue, previous work (Ichim et al. 2016) assumes that all soft tissue can activate. In addition, our approach includes a kinematic model for the jaw, whereas Ichim et al. (2016) only approximated the jaw by using a more stiff (but not exactly rigid) material. Finally, our method also allows for skin sliding, facilitating more realistic flesh deformations especially in areas such as the forehead (Figure 5).

Model adaptations. Our approach supports animating a character after significant modifications of the neutral pose (e.g. slimming/fattening, bone modifications, see Section 8) using the same muscle activation patterns. One might argue that the same effects could be obtained by using deformation transfer (Sumner and Popović 2004) on traditional linear animation models. For example, similar modifications as the ones we propose could be applied on the surface mesh of the neutral blendshape. Deformation transfer on all expression blendshapes will then yield new face rig that incorporates the desired changes. However, this approach has the significant drawback that the new blendshapes are not necessarily consistent with

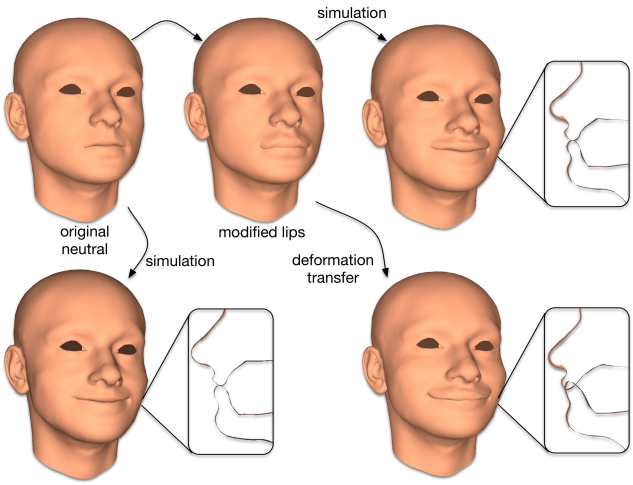


Fig. 11. Model adaptations such as increased lip volume are handled accurately in our approach, while deformation transfer (Sumner and Popović 2004) leads to self-intersections.

the same blendshape weights, e.g., self-intersections easily occur as shown in Figure 11.

In addition, direct transfer of modifications to the neutral pose cannot account for the complex force interactions in the elastic tissue. For example, when increasing the volume of the lips, the expression dynamics will change as a consequence of the changed stress distribution. Our indirect approach, that solves for the facial pose given muscle activations, can accommodate such scenarios and leads to more natural expressions.

Statistics. The interior point solver of the forward physics optimization requires on average 8 iterations per frame to converge. This takes approx. 22 seconds including the collision detection update on a consumer laptop with a 3.1 GHz Intel Core i7 processor and 16GB of main memory. The inverse problem needs approx. 15 iterations to compute the jaw transformation and muscle activations, averaging at about 3 minutes per target shape. The volumetric face template model of the passive flesh and active muscles used for the results presented in this paper has 8098 vertices and 35626 tetrahedra. The active muscle layer covers approx. 27% of the entire flesh. The surface mesh model of the entire skin has 6393 vertices and 12644 faces.

8 APPLICATION DEMOS

We present a series of application demos to highlight the versatility of our approach. A key benefit of our physics-based simulation is that we can modify the static and dynamic parameters of the model to achieve a number of advanced animation effects that would be difficult to obtain with purely generative geometric methods. Please also refer to the accompanying video to better appreciate the dynamics of the animations.

All animation examples were driven by a temporal sequence of blendshape weights obtained from the performance capture system of Weise et al. (2011). The tracking software also provides a rigid body transformation $\mathbf{T} \in SE(3)$ corresponding to the global rotation and translation of the head, as well as pitch and yaw for each of the eyeballs, which are parented to the head transformation $\mathbf{T}$.

Body mass index changes. Figure 12-a illustrates how an animated avatar can be modified to slim or fatten the person's face by adapting the plasticity scale for the soft tissue tets. As this adaptation alters the face geometry, simply re-animating the blendshape model would lead to unnatural expressions and visual artifacts caused by self-intersections. Our simulation approach avoids self-collisions and balances the stress distribution in the facial tissue while preserving the actuation forces, which leads to more plausible expressions and natural dynamics.

To create the scaling parameters $s_i > 0$, we start from a surface "fat map" painted by the user that specifies which areas of the face are more prone to fat accumulation. The values of the fat map are propagated into the volumetric tet-mesh by a diffusion process, similar to standard polygon-mesh diffusion flow ((Botsch et al. 2010), Chapter 4.2), but using the volumetric Laplacian instead of the surface Laplace-Beltrami. We apply forward Euler integration with time step and number of steps adjusted by the user in an interactive graphical tool to achieve the desired volumetric propagation effect. We used the same fat map for both characters in Figure 12-a, uniformly scaled to achieve slimming or fattening. To account for the increased fat content in the soft tissue, we lower the stiffness μ to $0.8, 0.5, 0.3$ for the three levels of fattening shown in Figure 12-a. For slimming, we keep the default stiffness $\mu = 1$.

Facial surgery. Figures 12-b and 12-c demonstrate potential applications in visualizing the possible outcomes of facial surgery, helping patients to choose between different versions of corrective or cosmetic procedures. Our method allows direct manipulation of the deformable soft tissue (e.g. lip fat injection) or the rigid bones (e.g. chin displacement). The simulation then provides a detailed visual preview of such interventions on the expression dynamics of the animated person. We modeled the lip fat injection using our plasticity model and diffusion tool, simulating the process of injecting filler material with a syringe through several points on the skin. To simulate the lower stiffness of the fat-like filler material, we decreased the soft tissue stiffness μ to 0.8 for the medium, and to 0.5 for the high lip volume effect (Figure 12-b). For the chin displacement we directly edited the bone using an interactive mesh modeling tool.

Inertia. Figure 13-a shows how our method incorporates inertial deformations in the dynamic simulation. Such secondary motion becomes particularly important in animations with strong accelerations, such as jumping, head shaking, or boxing.

Interaction with external forces and objects. Figure 13-b shows how an animation can be augmented with complex external force interactions produced by a dynamic wind field. Figure 13-d illustrates how a speech animation is affected when the subject is wearing a VR headset. Our contact resolution method adapts the face deformations to account for the collisions with the headset, creating non-linear bulging and wrinkling effects due to volume preservation of the facial tissue.

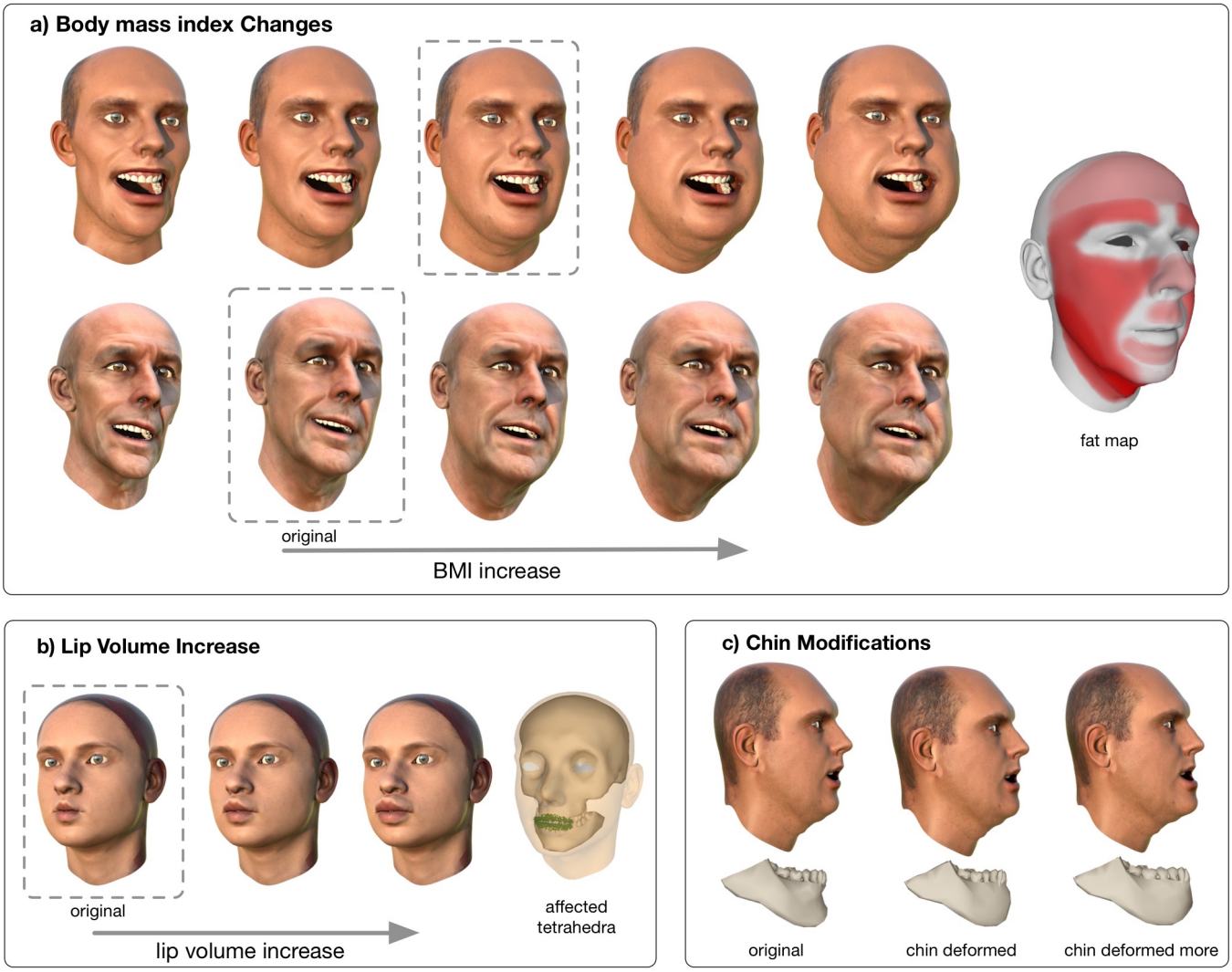


Fig. 12. Application Demos I: a) Body mass index changes and their impact on expressions. The original avatar is highlighted with dashed lines. More intense red in the fat map means more volume change of the corresponding face region. b) Results of lip injection, where affected tets are shown in green on the right. c) Effects of modifying the rigid bone structure of the chin.

Simulation of muscle paralysis. In Figure 13-c, we show how muscle activations can be modified to simulate Bell's palsy syndrome, where the affected person is unable to activate certain facial muscles. In this example, we marked the active muscles of the left half of the face to behave like passive tissue, which simulates the effect of partial facial paralysis.

Extreme face modifications. To push the limits of facial modifications, we created a virtual zombie character in Figure 13-e. We designed two texture maps to modulate the mass and stiffness (see Figure 13-e) and extrapolated their values into the volume using our diffusion tool. The idea was to increase the mass of the cheeks to create a flesh sagging effect, while increasing stiffness around the lips and the eyes to avoid excessive pulling of the flesh. The final μ values vary between $0.7 - 5.7$ and the density varies between $1 - 3g/cm^3$, achieving artistic "undead" effects.

9 LIMITATIONS AND FUTURE WORK

In our approach we rely solely on a generic volumetric template and a set of surface scans of the modeled person to derive the interior facial structure. This inherently limits the accuracy of our approach in terms of the true facial dynamics of the scanned actor. Getting access to the internal structure through volumetric scanning devices would allow building more faithful simulation models, but incurs a high acquisition cost. A potentially more practical approach for future work is to build a statistical model of the bone and tissue structures from a sufficiently large set of volumetric scans, similar

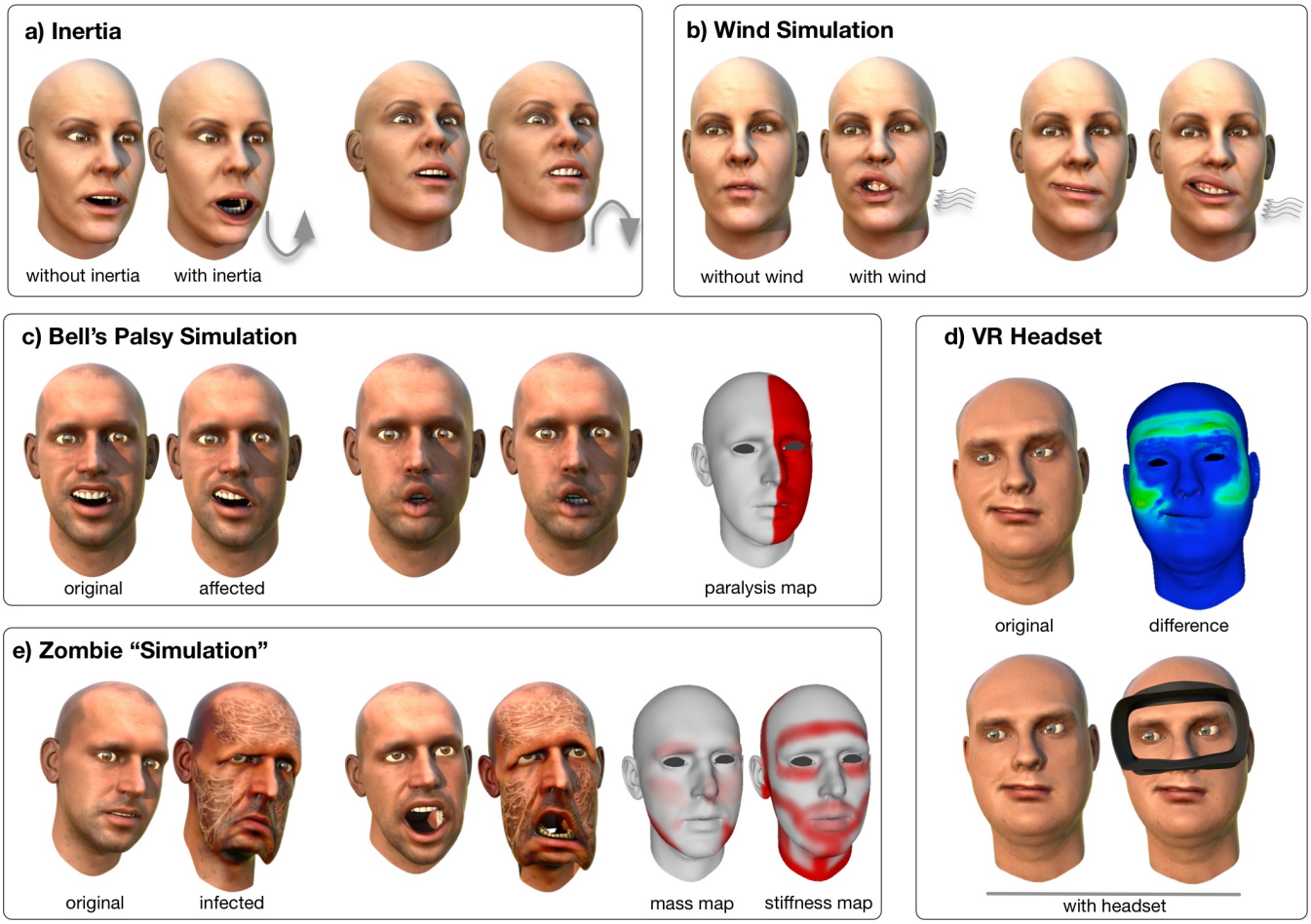


Fig. 13. a) Simulating inertia under sudden motion changes (e.g., jumping). b) Dynamic deformations in a wind force field. c) Simulating Bell's Palsy affecting half of the face of an actor. d) VR headset obstructing the full motion of expressions on the face. e) Artistic editing to create a zombie character by adapting the mass and stiffness distribution as indicated in the color-coded maps.

to the morphable face models that have been successfully applied for the skin surface (Blanz and Vetter 1999).

Detailed physical simulation is computationally involved and our method is currently not suitable for realtime animation. While computational efficiency was not the main focus of our work, we believe that significant speedups can be achieved, in particular by more explicitly exploiting spatial and temporal coherence. In the context of realtime animation, our approach could potentially be used to automatically create corrective shapes for a given blendshape basis in an offline process. How to select an optimal set of such correctives based on a given simulation is an interesting avenue for future research.

Our tet-mesh discretization is currently too coarse to correctly model small-scale effects such as skin wrinkles. However, increasing the resolution to the appropriate scale would lead to prohibitive computation times. Therefore, in future work, we want to explore ways to combine our simulation model with procedural or data-driven methods for wrinkle generation to further increase the visual realism of the animations.

Other avenues for future work include modeling and simulating hair, adding person-specific teeth models and a simulation of the tongue, and generalizing our model to full body simulations.

10 CONCLUSION

We propose a physics-based simulation approach to face animation that complements existing generative methods such as blendshapes. These purely geometric methods can produce artifacts such as self-intersections in facial poses that were not specifically considered during the modeling of the blendshape basis – ensuring consistency in all possible linear combinations quickly becomes intractable. Even more challenging is the correct handling of dynamic effects such as interactions with external objects or inertial deformations.

We advocate the use of physics-based simulation as a principled solution to these issues. Our experiments validate that this approach

leads to high-quality facial animations and facilitates new editing capabilities, e.g., by manipulating the face model's physical structure or its dynamic behavior. These advanced effects come at the cost of increased computational overhead. However, as computational power increases and algorithms are improved, we believe that the simulation-based approach to facial animation will become more and more viable in the future. This path certainly offers a rich set of opportunities for future research with applications not only in movies and games, but also in surgery simulation, interactive therapy, sports, or biomedical research.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant Numbers IIS-1617172 and IIS-1622360, the grant SVV-2017-260452 and GA UK 1524217. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also gratefully acknowledge the support of Activision.

REFERENCES

Oleg Alexander, Mike Rogers, William Lambeth, Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, and Paul Debevec. 2010. The digital Emily project: Achieving a photorealistic digital actor. *Computer Graphics and Applications, IEEE* 30, 4 (2010), 20–31.

Jérémie Allard, Stéphane Cotin, François Faure, Pierre-Jean Bensoussan, François Poyer, Christian Duriez, Hervé Delingette, and Laurent Grisoni. 2007. Sofa-an open source framework for medical simulation. In *MMVR 15-Medicine Meets Virtual Reality*, Vol. 125. IOP Press, 13–18.

Brian Amberg, Andrew Blake, Andrew Fitzgibbon, Sami Romdhani, and Thomas Vetter. 2007. Reconstructing high quality face-surfaces using model based stereo. In *ICCV 2007*. IEEE, 1–8.

Vincent Barrielle, Nicolas Stoiber, and Cedric Cagniart. 2016. Blendforces, a Dynamic Framework for Facial Animation. *Comp. Graph. Forum* (2016).

Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. 2010. High-Quality Single-Shot Capture of Facial Geometry. *ACM Trans. Graph.* 29, 3 (2010), 40:1–40:9.

Thabo Beeler and Derek Bradley. 2014. Rigid stabilization of facial expressions. *ACM Trans. Graph.* 33, 4 (2014), 44.

Pascal Bérard, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. Lightweight eye capture using a parametric model. *ACM Trans. Graph.* 35, 4 (2016), 117.

Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus H Gross. 2014. High-quality capture of eyes. *ACM Trans. Graph.* 33, 6 (2014), 223–1.

Amit Bermano, Thabo Beeler, Yeara Kozlov, Derek Bradley, Bernd Bickel, and Markus Gross. 2015. Detailed spatio-temporal reconstruction of eyelids. *ACM Trans. Graph.* 34, 4 (2015), 44.

Amit H Bermano, Derek Bradley, Thabo Beeler, Fabio Zund, Derek Nowrouzezahrai, Ilya Baran, Olga Sorkine-Hornung, Hanspeter Pfister, Robert W Sumner, Bernd Bickel, and others. 2014. Facial performance enhancement using dynamic shape space analysis. *ACM Trans. Graph.* 33, 2 (2014), 13.

Bernd Bickel, Moritz Bächer, Miguel A Otaduy, Wojciech Matusik, Hanspeter Pfister, and Markus Gross. 2009. Capture and modeling of non-linear heterogeneous soft tissue. *ACM Trans. Graph.* 28, 3 (2009).

Bernd Bickel, Peter Kaufmann, Mélina Skouras, Bernhard Thomaszewski, Derek Bradley, Thabo the, Phil Jackson, Steve Marschner, Wojciech Matusik, and Markus Gross. 2012. Physical face cloning. *ACM Trans. Graph.* 31, 4 (2012), 118.

Volker Blanz and Thomas Vetter. 1999. A morphable model for the synthesis of 3D faces. In *Proc. of the 26th annual conf. on Comp. graph. and interactive techniques*. 187–194.

Mario Botsch, Leif Kobbelt, Mark Pauly, Pierre Alliez, and Bruno Lévy. 2010. *Polygon mesh processing*. CRC press.

Sofien Bouaziz, Sebastian Martin, Tiantian Liu, Ladislav Kavan, and Mark Pauly. 2014. Projective dynamics: fusing constraint projections for fast simulation. *ACM Trans. Graph.* 33, 4 (2014), 154.

Sofien Bouaziz, Yangang Wang, and Mark Pauly. 2013. Online modeling for realtime facial animation. *ACM Trans. Graph.* 32, 4 (2013), 40.

Richard P. Brent. 1971. An algorithm with guaranteed convergence for finding a zero of a function. *Comput. J.* 14, 4 (1971).

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

Chen Cao, Derek Bradley, Kun Zhou, and Thabo Beeler. 2015. Real-time high-fidelity facial performance capture. *ACM Trans. Graph.* 34, 4 (2015), 46.

Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. 2014. Facewarehouse: a 3d facial expression database for visual computing. *Visualization and Computer Graphics, IEEE Transactions on* 20, 3 (2014), 413–425.

John E Chadwick, David R Haumann, and Richard E Parent. 1989. Layered construction for deformable animated characters. In *ACM Siggraph Computer Graphics*, Vol. 23. ACM, 243–252.

Hon Fai Choi and Silvia S Blemker. 2013. Skeletal muscle fascicle arrangements can be reconstructed using a laplacian vector field simulation. *PloS one* 8, 10 (2013), e77576.

Matthew Cong, Michael Bao, Kiran S Bhat, Ronald Fedkiw, and others. 2015. Fully automatic generation of anatomical face simulation models. In *Proc. of the EG/SIGGRAPH Symposium on Comp. Anim.* ACM, 175–183.

Matthew Cong, Kiran S Bhat, and Ronald Fedkiw. 2016. Art-directed muscle simulation for high-end facial animation. In *Proc. of the EG/SIGGRAPH Symposium on Comp. Anim.* 119–127.

Ali-Hamadi Dicko, Tiantian Liu, Benjamin Gilles, Ladislav Kavan, François Faure, Olivier Palombi, and Marie-Paule Cani. 2013. Anatomy transfer. *ACM Trans. Graph.* 32, 6 (2013), 188.

Paul Ekman and Wallace V Friesen. 1977. Facial action coding system. (1977).

Ye Fan, Joshua Litven, and Dinesh K Pai. 2014. Active volumetric musculoskeletal systems. *ACM Trans. Graph.* 33, 4 (2014), 152.

Pablo Garrido, Michael Zollhöfer, Chenglei Wu, Derek Bradley, Patrick Pérez, Thabo Beeler, and Christian Theobalt. 2016. Corrective 3D reconstruction of lips from monocular video. *ACM Trans. Graph.* 35, 6 (2016), 219.

Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2015. Single-view hair modeling using a hairstyle database. *ACM Trans. Graph.* 34, 4 (2015).

Alexandru Ichim, Ladislav Kavan, Merlin Nimier-David, and Mark Pauly. 2016. Building and Animating User-Specific Volumetric Face Rigs. In *Proc. of the EG/SIGGRAPH Symposium on Comp. Anim.*

Alexandru Eugen Ichim, the Bouaziz, and Mark Pauly. 2015. Dynamic 3D Avatar Creation from Hand-held Video Input. *ACM Trans. Graph.* (2015).

Alec Jacobson, Ladislav Kavan, and Olga Sorkine-Hornung. 2013. Robust inside-outside segmentation using generalized winding numbers. *ACM Trans. Graph.* 32, 4 (2013), 33.

Petr Kadlecek, Alexandru-Eugen Ichim, Tiantian Liu, Jaroslav Krivanek, and Ladislav Kavan. 2016. Reconstructing Personalized Anatomical Models for Physics-based Body Animation. *ACM Trans. Graph.* 35, 6 (2016).

Kolja Kähler, Jörg Haber, and Hans-Peter Seidel. 2003. Reanimating the dead: reconstruction of expressive faces from skull data. In *ACM Trans. Graph.*, Vol. 22. ACM, 554–561.

Oliver Klehm, Fabrice Rousselle, Marios Papas, Derek Bradley, Christophe Hery, Bernd Bickel, Wojciech Jarosz, and Thabo Beeler. 2015. Recent advances in facial appearance capture. In *Computer Graphics Forum*, Vol. 34. 709–733.

Sung-Hee Lee, Eftychios Sifakis, and Demetri Terzopoulos. 2009. Comprehensive biomechanical modeling and simulation of the upper body. *ACM Trans. Graph.* 28, 4 (2009), 99.

John P Lewis, Ken Anjyo, Taehyun Rhee, Mengjie Zhang, Frederic H Pighin, and Zhigang Deng. 2014. Practice and Theory of Blendshape Facial Models.. In *Eurographics (State of the Art Reports)*. 199–218.

Duo Li, Shinjiro Sueda, Debanga R Neog, and Dinesh K Pai. 2013. Thin skin elastodynamics. *ACM Trans. Graph.* 32, 4 (2013), 49.

Hao Li, Thibaut Weise, and Mark Pauly. 2010. Example-based facial rigging. In *ACM Trans. Graph.*, Vol. 29. ACM, 32.

Hao Li, Jihun Yu, Yuting Ye, and Chris Bregler. 2013. Realtime Facial Animation with On-the-fly Correctives. *ACM Trans. Graph.* 32, 4 (2013).

Tiantian Liu, Adam W. Bargteil, James F. O'Brien, and Ladislav Kavan. 2013. Fast Simulation of Mass-Spring Systems. *ACM Trans. Graph.* 32, 6 (2013), 209:1–7.

John E Lloyd, Ian Stavness, and Sidney Fels. 2012. ArtiSynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. In *Soft tissue biomechanical modeling for computer assisted surgery*. Springer, 355–394.

Wan-Chun Ma, Yi-Hua Wang, Graham Fyffe, Bing-Yu Chen, and Paul Debevec. 2012. A blendshape model that incorporates physical interaction. *Computer Animation and Virtual Worlds* 23, 3-4 (2012).

Steve A Maas, Benjamin J Ellis, Gerard A Ateshian, and Jeffrey A Weiss. 2012. FEBio: finite elements for biomechanics. *Journal of biomechanical engineering* 134, 1 (2012), 011005.

Aleka McAdams, Yongning Zhu, Andrew Selle, Mark Empey, Rasmus Tamstorf, Joseph Teran, and Eftychios Sifakis. 2011. Efficient elasticity for character skinning with contact and collisions. In *ACM Trans. Graph.*, Vol. 30. 37.

Koki Nagano, Graham Fyffe, Oleg Alexander, Jernej Barbic, Hao Li, Abhijeet Ghosh, and Paul Debevec. 2015. Skin microstructure deformation with displacement map convolution. (2015).

Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J Black. 2015. Dyna: A model of dynamic human shape in motion. *ACM Trans. Graph.* 34, 4 (2015), 120.

Szymon Rusinkiewicz and Marc Levoy. 2001. Efficient variants of the ICP algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*. IEEE, 145–152.

Shunsuke Saito, Zi-Ye Zhou, and Ladislav Kavan. 2015. Computational Bodybuilding: Anatomically-based Modeling of Human Bodies. *ACM Trans. Graph.* 34, 4 (2015).

Robert Schleip, Thomas W Findley, Leon Chaitow, and Peter Huijing. 2013. *Fascia: the tensional network of the human body: the science and clinical applications in manual and movement therapy*. Elsevier Health Sciences.

Ken Shoemake and Tom Duff. 1992. Matrix animation and polar decomposition. In *Proceedings of the conference on Graphics interface*, Vol. 92. Citeseer, 258–264.

Weiguang Si, Sung-Hee Lee, Eftychios Sifakis, and Demetri Terzopoulos. 2014. Realistic biomechanical simulation and control of human swimming. *ACM Trans. Graph.* 34, 1 (2014), 10.

Eftychios Sifakis and Jernej Barbic. 2012. FEM simulation of 3D deformable solids: a practitioner's guide to theory, discretization and model reduction. In *ACM SIGGRAPH 2012 Courses*. 20.

Eftychios Sifakis, Igor Neverov, and Ronald Fedkiw. 2005. Automatic determination of facial muscle activations from sparse motion capture marker data. In *ACM Trans. Graph.*, Vol. 24. 417–425.

Olga Sorkine and Marc Alexa. 2007. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, Vol. 4.

Robert W Sumner and Jovan Popović. 2004. Deformation transfer for triangle meshes. In *ACM Trans. Graph.*, Vol. 23. 399–405.

Joseph Teran, Sylvia Blemker, V Hing, and Ronald Fedkiw. 2003. Finite volume methods for the simulation of skeletal muscle. In *Proc. of the EG/SIGGRAPH Symposium on Comp. Anim.* Eurographics Association, 68–74.

Joseph Teran, Eftychios Sifakis, Silvia S Blemker, Victor Ng-Thow-Hing, Cynthia Lau, and Ronald Fedkiw. 2005a. Creating and simulating skeletal muscle from the visible human data set. *Visualization and Computer Graphics, IEEE Transactions on* 11, 3 (2005), 317–328.

Joseph Teran, Eftychios Sifakis, Geoffrey Irving, and Ronald Fedkiw. 2005b. Robust quasistatic finite elements and flesh simulation. In *Proc. of the EG/SIGGRAPH Symposium on Comp. Anim.* ACM.

Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. 2005. Face transfer with multilinear models. In *ACM Trans. Graph.*, Vol. 24. ACM, 426–433.

Javier von der Pahlen, Jorge Jimenez, Etienne Danvoye, Paul Debevec, Graham Fyffe, and Oleg Alexander. 2014. Digital Ira and Beyond: Creating Real-time Photoreal Digital Actors. In *ACM SIGGRAPH 2014 Courses (SIGGRAPH '14)*. Article 1, 384 pages. DOI: https://doi.org/10.1145/2614028.2615407

Andreas Wächter and Lorenz T Biegler. 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106, 1 (2006).

Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. 2011. Realtime performance-based facial animation. In *ACM Trans. Graph.*, Vol. 30. ACM, 77.

Chenglei Wu, Derek Bradley, Pablo Garrido, Michael Zollhöfer, Christian Theobalt, Markus Gross, and Thabo Beeler. 2016a. Model-based Teeth Reconstruction. *ACM Trans. Graph.* 35, 6 (2016).

Chenglei Wu, Derek Bradley, Markus Gross, and Thabo Beeler. 2016b. An anatomically-constrained local deformation model for monocular face capture. *ACM Trans. Graph.* 35, 4 (2016), 115.

Zygote. 2016. Zygote Body. (2016). https://zygotebody.com [Online; accessed 28-Dec-2016].