

# Automatic Determination of Facial Muscle Activations from Sparse Motion Capture Marker Data

Eftychios Sifakis\*  
Stanford University  
Intel Corporation

Igor Neverov†  
Stanford University

Ronald Fedkiw\*  
Stanford University  
Industrial Light + Magic

## Abstract

We built an anatomically accurate model of facial musculature, passive tissue and underlying skeletal structure using volumetric data acquired from a living male subject. The tissues are endowed with a highly nonlinear constitutive model including controllable anisotropic muscle activations based on fiber directions. Detailed models of this sort can be difficult to animate requiring complex coordinated stimulation of the underlying musculature. We propose a solution to this problem automatically determining muscle activations that track a sparse set of surface landmarks, e.g. acquired from motion capture marker data. Since the resulting animation is obtained via a three dimensional nonlinear finite element method, we obtain visually plausible and anatomically correct deformations with spatial and temporal coherence that provides robustness against outliers in the motion capture data. Moreover, the obtained muscle activations can be used in a robust simulation framework including contact and collision of the face with external objects.

**CR Categories:** I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Physically based modeling;

**Keywords:** facial animation, muscles, finite element method

## 1 Introduction

Facial modeling and animation, enabled by recent advances in technology, is a vital new area in high demand. While this is especially true in the entertainment industry (e.g. [Borshukov et al. 2003]), it is also quite popular elsewhere including applications to lip reading and surgical planning. For example, [Koch et al. 1998] pointed out the utility of synthesizing expressions on a post-surgical face to determine the effects of the surgical modifications.

Starting with data from the visible human data set [U.S. National Library of Medicine 1994], we used the techniques proposed in [Teran et al. 2005b] to construct a highly detailed anatomically accurate model of the head and neck region. This includes a triangulated surface for each bone, a tetrahedralized volume and a B-spline fiber field representation for each muscle, and a single tetrahedral mesh for all the soft tissue. Then we morphed this anatomically accurate model to fit data obtained from both laser and MRI scans of a living subject constructing new meshes where necessary.

Animating such a complex model can be rather difficult, so we propose using three dimensional sparse motion capture marker data

\*e-mail: {sifakis,fedkiw}@cs.stanford.edu

†e-mail: igor@graphics.stanford.edu

Copyright © 2005 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail [permissions@acm.org](mailto:permissions@acm.org).  
© 2005 ACM 0730-0301/05/0700-0417 \$5.00

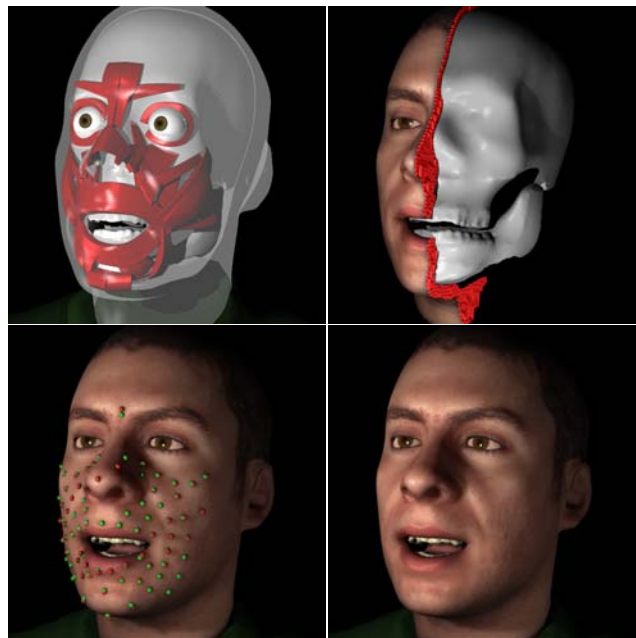


Figure 1: Facial expression created by the action of 32 transversely isotropic muscles (top left) and simulated on a quasistatic finite element tetrahedral mesh (top right). Muscle activations and bone kinematics are automatically estimated to match motion capture markers (bottom left) giving rise to the final synthesized expression (bottom right). The original markers are colored red, and the marker positions resulting from our simulation are depicted in green.

(see e.g. [Williams 1990; Guenter et al. 1998]) to automatically determine muscle activations. [Terzopoulos and Waters 1993] took a similar approach early on estimating muscle actuation parameters based on the position of facial features tracked by snakes. Later, [Morishima et al. 1998] contracted both individual and combinations of muscles in order to learn patterns, and then used two dimensional marker positions or optical flow as input for a neural network which estimated muscle contraction parameters. Both of these approaches aim to match a two dimensional projected image as opposed to our goal of matching the full three dimensional shape of the face.

A control-theoretic approach was used to estimate muscle contractions that match optical flow input in [Essa et al. 1996; Essa and Pentland 1997]. Although this approach might work for more detailed anatomical models, they only considered a two dimensional finite element model for skin along with a simple muscle model for actuation. [Basu et al. 1998b; Basu et al. 1998a] proposed avoiding the internal anatomy altogether constructing a two dimensional quasistatic finite element model for the lips that was used to minimize the strain as select nodes tracked motion data. This was done to *train* the lips, and PCA was used to reduce the subsequent degrees of freedom to about ten. This reduction mimics the fact that the actual degrees of freedom correspond to the muscles which are conveniently already in the *proper* lower dimensional space (as opposed to that obtained from PCA). Finally, they track lip motion

automatically determining the parameters of their model that match the motion data using a steepest descent iterative solver.

More recently, [Choe et al. 2001] used a two dimensional linear quasistatic finite element model of the skin surface, along with underlying muscles that apply forces to the surface mesh based on linear activations. Their lack of anatomical structure led to the use of heuristic correction forces near the mouth. Given marker data, they used a steepest descent method to calculate the muscle activations that best track the data including penalty forces to constrain muscle activations to the physical regime. They point out that their model is linear greatly simplifying this problem. This formulation suffered from both a lack of anatomical accuracy and a lack of nonlinearity, and [Choe and Ko 2001] pointed out that it could produce unnatural artifacts. Thus, they modified this procedure allowing the artist to sculpt basis elements to be combined linearly, and used the active set method to solve a constrained quadratic program to obtain the muscle activations (or weights). Typically, the basis elements need to be resculpted a number of times to obtain satisfactory results.

A major benefit of our approach is that we strive for anatomical accuracy which gives biomechanical meaning to our activations, and thus makes the problem tractable in the sense that a person's face *is* driven by muscle activations of this sort. Moreover, we take a biomechanically accurate fully nonlinear approach to both the constitutive model and the finite element method, which complicates the solution process but provides behavior not captured by approaches that linearly blend basis functions. We automatically solve for not only muscle activations, but the head position and jaw articulation parameters as well. Moreover, our approach uses a sophisticated simulation framework allowing us to place the animated face in complex environments including arbitrary contact and collision with external objects in the scene. During this interaction, the extracted muscle activation controls can still be applied providing a realistic combination of muscle contraction and external stimuli. Furthermore, if the collision events give rise to ballistic phenomena, we can readily replace the quasistatic simulation with a fully dynamic one while retaining the extracted muscle activation values.

## 2 Related Work

Three dimensional facial animation began with [Parke 1972] (see [Parke and Waters 1996] for a review). [Platt and Badler 1981] built a face model using masses and springs including forces generated by muscles, and made use of the Facial Action Coding System (FACS) [Ekman and Friesen 1978]. Other early work included [Waters 1987; Magnenat-Thalmann et al. 1988; Kalra et al. 1992]. [Lee et al. 1995] constructed an anatomically motivated facial model based on scanned data, and endowed it with a mass spring system driven by muscle contractions. [Waters and Frisbie 1995] built a muscle model for speech animation, stressing that muscles animate faces and that it is more productive to focus on a muscle model rather than a surface model. A number of authors have used finite element simulations in the context of facial surgery, e.g. [Pieper et al. 1992; Koch et al. 1996; Keeve et al. 1996; Roth et al. 1998] (see also, [Teschner et al. 2000]).

[DeCarlo et al. 1998] used variational modeling and face anthropometry techniques to construct smooth face models. [Pighin et al. 1998] used a number of photographs to fit a three dimensional template mesh to a given facial pose, and then obtained animations by blending different poses. [Pighin et al. 1999] used this technique to fit a face model to each frame of a video sequence estimating the pose for subsequent analysis, and [Joshi et al. 2003] automatically segments the face into smaller regions for blending. See also, [Zhang et al. 2003]. Starting from a database of face scans, [Blanz and Vetter 1999] derive a vector space representation of shapes and

textures such that any linear combination of examples gives a reasonable result. This framework has been used to transfer animations from one individual to another [Blanz et al. 2003], for face identification [Blanz and Vetter 2003], and to exchange a face from one image to another [Blanz et al. 2004]. [Kahler et al. 2001] built a mass spring model of a face and skull with a muscle model to drive the deformation, [Kahler et al. 2002] proposed a method for morphing this model to other faces, and [Kahler et al. 2003] extended this approach to forensic analysis.

Other work includes facial animation based on audio or text input [Cassell et al. 1994; Bregler et al. 1997; Brand 1999; Cassell et al. 2001; Ezzat et al. 2002; Cao et al. 2003; Cao et al. 2004], wrinkle formation [Wu et al. 1999], eye motion [Lee et al. 2002], and facial motion transfer [Noh and Neumann 2001; Pyun et al. 2003; Na and Jung 2004; Sumner and Popović 2004]. [Byun and Badler 2002] modified the MPEG-4 Facial Animation Parameters (FAPs) [Moving Picture Experts Group 1998] to add expressiveness, [Kshirsagar and Magnenat-Thalmann 2003] used PCA to deform the mouth during speech, and [Chai et al. 2003] used facial tracking to drive animations from a motion capture database. [Wang et al. 2004] used a multiresolution deformable mesh to track facial motion, and a low dimensional embedding technique to learn expression style. [Zhang et al. 2004] proposed a method for tracking facial animation fitting a template mesh to the data, and then used linear combinations of basis shapes to create an inverse kinematics system that allows one to create expressions by dragging surface points directly.

## 3 Anatomical Model

Our model building effort started with the volumetric data from the visible human data set [U.S. National Library of Medicine 1994]. As in [Teran et al. 2005b], we constructed level sets for each tissue and used them to create a triangulated surface for each bone and a tetrahedralized volume for each muscle. Since many of the muscles in the face are quite thin and thus not amenable to robust and efficient tetrahedral mesh simulation, we took an embedded approach to muscle modeling. First, a single tetrahedral flesh mesh was created to represent *all* the soft tissue in the face, and then we calculated the fraction of overlap between each muscle and each tetrahedron of the flesh mesh storing that fraction locally in the tetrahedron. We also create fiber fields for each muscle and store a single vector direction *per muscle* in each tetrahedron with a nonzero overlap. One graduate student spent 6 months constructing this template face and muscle model from the visible human data, but with existing tools this could be accomplished in 2 weeks.

Subsequently, we obtained laser and MRI scans of a living subject. The laser scans gave a high-fidelity likeness of the subject, and we wanted to adhere to them closely. The MRI scan was of much lower quality presenting only an approximate guideline, and we needed to reuse the bone, muscle, and to a lesser extent flesh geometry from our visible human template model. To that end we developed a set of point correspondences between the two models and morphed the geometry from the first using radial basis functions. This morphed geometry required further manual editing to satisfy considerations of aesthetics and general anatomical knowledge. Once we had geometry for the surface of the flesh volume, we again used the meshing algorithm to create a high-quality tetrahedral mesh for the face flesh. To summarize, our model consists of a rigid articulated cranium and jaw with about 30 thousand surface triangles, flesh in the form of a tetrahedral mesh with about 850 thousand tetrahedra out of which 370 thousand (in the front part of the face) are simulated, Dirichlet boundary conditions corresponding to bone attachments, and an embedded representation of 32 muscles. This subject specific model was constructed in 2 months by 5 undergraduate students, but would only take a single person a few days with existing

tools. For example, rebuilding the facial tetrahedral flesh takes only a few hours. A cross-section of the simulation volume is illustrated in figure 2.

In addition to the main functionality, a number of auxiliary features were considered. To provide realistic collisions of lips against the underlying rigid structure, we incorporated scans of teeth molds into the cranium and jaw. To achieve more realistic muscle action, we independently scaled the strength of each embedded muscle based on the amplitude and plausibility of their flexion. This biomechanically corresponds to adjusting the thickness of muscles, which we could not reliably infer from the MRI scan. Finally, we added eyes, teeth, shoulders, realistic rendering, etc.

## 4 Finite Element Method

The flesh mesh is governed by a Mooney-Rivlin constitutive model for the deviatoric deformation augmented by a volumetric pressure term for quasi-incompressibility. Tetrahedra which contain facial muscles have an additional anisotropic response for each muscle, which consists of both passive and active components scaled by the volume fraction. See [Teran et al. 2003; Teran et al. 2005b] for more details. The definition of nodal forces can be summarized as

$$\mathbf{f}(\mathbf{x}, \mathbf{a}) = \mathbf{f}_0(\mathbf{x}) + \sum_{i=1}^M a_i \mathbf{f}_i(\mathbf{x}) \quad (1)$$

were  $\mathbf{f}$  and  $\mathbf{x}$  denote the forces and positions of *all* nodes in the simulation mesh, and  $\mathbf{a} = (a_1, a_2, \dots, a_M)^T$  is the vector of activations of all  $M$  muscles.  $\mathbf{f}_0$  corresponds to the elastic material response of the flesh including the passive anisotropic component present in muscle regions. Each force component  $\mathbf{f}_i$  corresponds to the contribution of a fully activated muscle and is weighted by the corresponding current muscle activation level with  $a_i \in [0, 1]$ . The  $\mathbf{f}_i$  depend on the spatial configuration  $\mathbf{x}$  alone, making the total force an affine function of muscle activations. This linear dependence of force on activation is a fundamental property of the force-length curve of [Zajac 1989] that provides a useful simplification to our control framework.

We use a quasistatic simulation scheme where each input of muscle activations and skeletal configuration is directly mapped to the steady state expression it gives rise to. Such an assumption is fundamental to our control strategy, since it enables facial expressions to be defined as functions of the input control parameters without any dependence on the deformation history. We stress that this hypothesis is adopted only in the context of our optimization process to automatically determine muscle activations, and inertial effects can later be included in extracted expressions via a full dynamic simulation utilizing the same muscle control parameters.

Given a set of muscle activation parameters,  $\mathbf{a}$ , and appropriate boundary conditions, we substitute these into equation (1) and solve the resulting nonlinear equation  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ . The boundary conditions are derived from the position of the cranium and jaw (see section 8), and we abstractly encode this state with a vector  $\mathbf{b}$ . Solving this equation leads to an equilibrium configuration for the mesh,  $\mathbf{X}(\mathbf{a}, \mathbf{b})$ . These steady state positions are defined implicitly with the aid of equation (1) as

$$\mathbf{f}(\mathbf{X}(\mathbf{a}, \mathbf{b}), \mathbf{a}) = \mathbf{0}. \quad (2)$$

Note that  $\mathbf{b}$  does not explicitly appear in the definition of the finite element forces, but instead fully determines the value of some constrained nodes in the simulation mesh, which we denote  $\mathbf{X}^C(\mathbf{b})$ . Equation (2) can therefore be considered an implicit definition for the quasistatic positions of the unconstrained set of mesh nodes, denoted by  $\mathbf{X}^U(\mathbf{a}, \mathbf{b})$ .

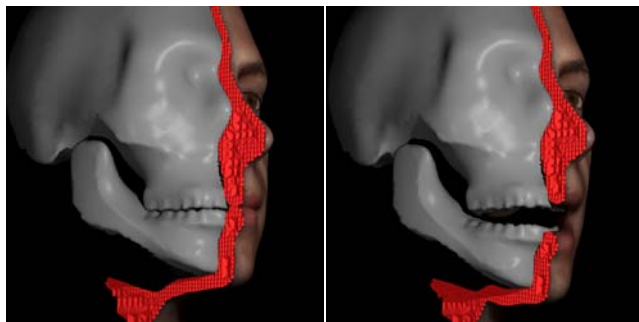


Figure 2: Illustration of the finite element flesh mesh, and the importance of colliding this mesh with the rigid bodies and itself.

We solve equation (2) with a Newton-Raphson iterative solver. At each step, the finite element forces are linearized around the current estimate  $\mathbf{X}_k$  as  $\mathbf{f}(\mathbf{X}_k + \delta\mathbf{X}) \approx \mathbf{f}(\mathbf{X}_k) + \partial\mathbf{f}/\partial\mathbf{x}|_{\mathbf{X}_k} \delta\mathbf{X}$ . Then we compute the displacement  $\delta\mathbf{X}$  that would restore the linearized equilibrium  $-\partial\mathbf{f}/\partial\mathbf{x}|_{\mathbf{X}_k} \delta\mathbf{X} = \mathbf{f}(\mathbf{X}_k)$ , and define the next iterate as  $\mathbf{X}_{k+1} = \mathbf{X}_k + \delta\mathbf{X}$ . Unfortunately,  $\partial\mathbf{f}/\partial\mathbf{x}|_{\mathbf{X}_k}$  is often indefinite leading to significant computational cost when solving for  $\delta\mathbf{X}$ . Thus, we utilize the enabling technology proposed in [Teran et al. 2005a] that allows a fast conjugate gradient solver to be used to find  $\delta\mathbf{X}$ . Moreover, the method proposed in [Teran et al. 2005a] allows for element inversion during the quasistatics solve speeding up our Newton-Raphson iteration by a significant amount. Overall, the convergence is particularly fast yielding an admissible solution to the nonlinear equilibrium problem within a few Newton-Raphson iterations even for drastic changes of activation levels. Although other solvers could be used, and solving equation (2) can be considered a “black box” as far as our method is concerned, [Teran et al. 2005a] makes our estimation of muscle activations practical as opposed to just doable. Mesh collisions between the lips or the lips and the teeth or gums are handled with the penalty force formulation of [Teran et al. 2005a]. See also [Heidelberger et al. 2004], [Teschner et al. 2003], etc. for more on collision handling.

## 5 Optimization Framework

We group all the muscle activations and kinematic parameters into a single set of controls  $\mathbf{c} = (\mathbf{a}, \mathbf{b})$  writing the equilibrium positions as  $\mathbf{X}(\mathbf{c})$ . The input to our model consists of a sparse set of motion capture marker data, but markerless techniques or animator key framing could alternatively be used as long as the final inputs are converted to target locations for points on the surface mesh. In the rest pose, we find the surface triangle closest to each marker and compute the barycentric coordinates for the marker rest position. If the marker does not lie on the surface mesh, we subtract its vector offset from *all* the data for that marker so that it does lie on the surface mesh (and should continue to as it is animated). Given values for the control parameters, the vector of all our embedded landmark positions is given by  $\mathbf{X}^L(\mathbf{c}) = \mathbf{W}\mathbf{X}(\mathbf{c})$  where  $\mathbf{W}$  is a sparse matrix of barycentric weights. Our goal is then to find the set of controls that minimize the distance between our landmark positions  $\mathbf{X}^L(\mathbf{c})$  and the motion capture marker data target positions  $\mathbf{X}^T$ , i.e.

$$\mathbf{c}_{\text{opt}}(\mathbf{X}^T) = \arg \min_{\mathbf{c} \in \mathbf{C}_0} \|\mathbf{X}^L(\mathbf{c}) - \mathbf{X}^T\|$$

where  $\mathbf{c}_{\text{opt}}(\mathbf{X}^T)$  stresses that the optimal set of controls is a function of the target positions. Here,  $\mathbf{C}_0$  is the *feasible* set of control configurations restricting the muscle activations (to the interval  $[0, 1]$ ) as well as the positioning and articulation of the head and jaw. Although any geometric or statistical norm could be used, we use the

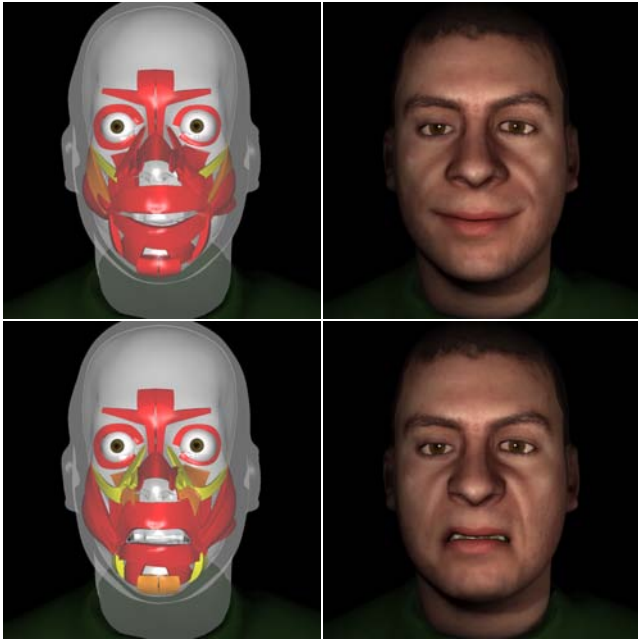


Figure 3: Synthetic expressions created by manual specification of activations. Yellow denotes fully active, and red is fully inactive.

Euclidean norm which leads to a nonlinear least squares optimization problem. The nonlinearity is from the dependence of  $\mathbf{X}^L(\mathbf{c})$  on  $\mathbf{X}(\mathbf{c})$  which is a complex nonlinear map defined implicitly by equation (2).

A standard Newton iterative approach to minimizing the functional  $\phi(\mathbf{c}) = \|\mathbf{X}^L(\mathbf{c}) - \mathbf{X}^T\|_2^2$  consists of replacing  $\phi(\mathbf{c})$  by its quadratic Taylor expansion about the current guess  $\mathbf{c}_k$ , i.e.  $\phi(\mathbf{c}_k + \delta\mathbf{c}) \approx \phi(\mathbf{c}_k) + \delta\mathbf{c}^T \nabla\phi(\mathbf{c}_k) + \frac{1}{2}\delta\mathbf{c}^T H_\phi(\mathbf{c}_k)\delta\mathbf{c}$  where  $H_\phi(\mathbf{c}_k) = 2\mathbf{J}_k^T \mathbf{J}_k + 2\mathbf{W}^T (\mathbf{X}^L(\mathbf{c}_k) - \mathbf{X}^T) : \partial^2\mathbf{X}/\partial\mathbf{c}^2|_{\mathbf{c}_k}$ ,  $\mathbf{J}_k = \mathbf{W} \partial\mathbf{X}/\partial\mathbf{c}|_{\mathbf{c}_k}$  and  $\delta\mathbf{c} = \mathbf{c} - \mathbf{c}_k$ . Then the quadratic approximation is minimized by solving  $-H_\phi(\mathbf{c}_k)\delta\mathbf{c} = \nabla\phi(\mathbf{c}_k)$  to find the next iterate  $\mathbf{c}_{k+1} = \mathbf{c}_k + \delta\mathbf{c}$ . In section 6 we illustrate how the Jacobian,  $\partial\mathbf{X}/\partial\mathbf{c}$ , of the quasistatic configuration can be computed using an efficient and reliable process. However, computation of  $\partial^2\mathbf{X}/\partial\mathbf{c}^2$  is particularly expensive as well as error prone unless very stringent accuracy requirements on the computation of both the quasistatic solution and its Jacobian are satisfied. In light of this, we propose an alternative optimization technique linearizing about  $\mathbf{c}_k$  to obtain

$$\mathbf{X}(\mathbf{c}) \approx \mathbf{X}(\mathbf{c}_k) + \partial\mathbf{X}/\partial\mathbf{c}|_{\mathbf{c}_k} \delta\mathbf{c} \quad (3)$$

which can be substituted into  $\phi(\mathbf{c})$  to obtain  $\hat{\phi}(\mathbf{c}) = \|\mathbf{X}^L(\mathbf{c}_k) + \mathbf{J}_k \delta\mathbf{c} - \mathbf{X}^T\|_2^2$ .  $\hat{\phi}(\mathbf{c})$  is minimized by the least squares solution of the linear system  $-\mathbf{J}_k \delta\mathbf{c} \cong \mathbf{X}^L(\mathbf{c}_k) - \mathbf{X}^T$ , and the normal equations approach to this requires solving  $-\mathbf{J}_k^T \mathbf{J}_k \delta\mathbf{c} = \mathbf{J}_k^T (\mathbf{X}^L(\mathbf{c}_k) - \mathbf{X}^T)$  or  $-2\mathbf{J}_k^T \mathbf{J}_k \delta\mathbf{c} = \nabla\phi(\mathbf{c}_k)$ . Notably, this final equation is equivalent to the Newton approach where the Hessian has been approximated by only its first term, removing the problematic  $\partial^2\mathbf{X}/\partial\mathbf{c}^2$  term. This is known as the Gauss-Newton approach (see e.g. [Gill et al. 1981]).

When  $\mathbf{X}^T$  is physically attainable, the Hessian is well approximated by its first term in the vicinity of the optimal value  $\mathbf{c}_{\text{opt}}$ . We have found this to be a very frequent case due to the expressive ability of our simulation model, especially in the context of expression tracking where the estimated control parameters at each frame constitute a very good initial guess to those at the next frame. However, large changes in activations or boundary conditions can make the solution to equation (2) unreliable causing the Gauss-Newton step to be

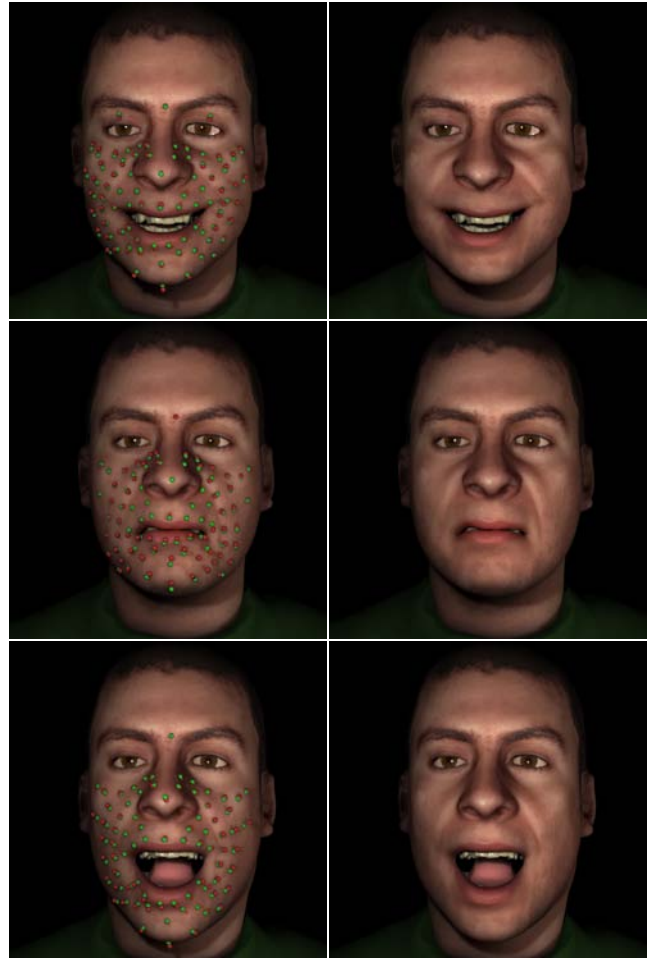


Figure 4: Expressions estimated from motion capture data, along with both the captured and simulated markers for comparison.

suboptimal. This can also happen when low quality input causes  $\mathbf{X}^T$  to be distant from the physically attainable configuration manifold. In order to safeguard against this suboptimal behavior, we use  $\delta\mathbf{c} = \mathbf{c}_{k+1} - \mathbf{c}_k$  as a *search direction* and minimize  $\phi(\mathbf{c})$  along the line segment connecting  $\mathbf{c}_k$  and  $\mathbf{c}_{k+1}$ . Since  $\phi(\mathbf{c})$  seemed to be unimodal in the vast majority of test cases, we used golden section search. Using linear interpolation to estimate the quasistatic configuration at internal points of the line segment provided a particularly good initial guess to the quasistatic solver making it typically converge in a single Newton-Raphson iteration for each golden section refinement. Given that no computation of Jacobians is necessary during the explicit line search, we found the incorporation of this process to incur only about a 10% performance overhead. As far as overall performance is concerned, remote initial guesses typically converge within an absolute maximum of 4-5 Gauss-Newton steps, while reasonable quality inputs typically led to convergence in a single step (notably with the full Gauss-Newton step to  $\mathbf{c}_{k+1}$ ).

## 6 Jacobian Computation

Our optimization framework relies on the ability to compute both the equilibrium positions,  $\mathbf{X}^* = \mathbf{X}(\mathbf{c}^*)$ , and the Jacobians,  $\partial\mathbf{X}/\partial\mathbf{c}|_{\mathbf{c}^*}$ , for a given control configuration  $\mathbf{c}^* = (\mathbf{a}^*, \mathbf{b}^*)$ . The first of these is readily computed by solving equation (2), and the results of this can be used to (nontrivially) compute the Jacobians as well. To do this, we rewrite equation (2) to explicitly highlight the



Figure 5: Tracking of a narration sequence. The captured markers are colored red, and the results of our simulation are depicted in green. Note that there is a one to one correspondence between the red and green markers, but some markers are occluding others in the figure.

dependence on both constrained and unconstrained nodes

$$\mathbf{f}(\mathbf{X}^C(\mathbf{b}), \mathbf{X}^U(\mathbf{a}, \mathbf{b}), \mathbf{a}) = \mathbf{0} \quad (4)$$

stressing that there is still only one equation for each unconstrained node, since the net force at constrained nodes is trivially zero. Differentiating equation (4) with respect to an activation parameter  $a_i$  yields  $\partial \mathbf{f} / \partial \mathbf{x}^U \Big|_{\mathbf{X}^*, \mathbf{a}^*} \partial \mathbf{X}^U / \partial a_i \Big|_{\mathbf{a}^*, \mathbf{b}^*} + \partial \mathbf{f} / \partial a_i \Big|_{\mathbf{X}^*, \mathbf{a}^*} = \mathbf{0}$  or

$$-\partial \mathbf{f} / \partial \mathbf{x}^U \Big|_{\mathbf{X}^*, \mathbf{a}^*} \partial \mathbf{X}^U / \partial a_i \Big|_{\mathbf{a}^*, \mathbf{b}^*} = \mathbf{f}_i(\mathbf{X}^*) \quad (5)$$

where  $\mathbf{f}_i$  was defined in equation (1) as the active force induced by a unit activation of the  $i$ -th muscle. This is a linear system of equations for the unknown partial derivatives  $\partial \mathbf{X}^U / \partial a_i$ . Additionally note that the activations have no effect on the constrained boundary nodes, and thus  $\partial \mathbf{X}^C / \partial a_i$  is identically zero.

Some of the kinematic controls,  $\mathbf{b}$ , such as the base frame of reference for the position of the cranium, do not affect the strain of the deformable model. The Jacobian of the quasistatic positions with respect to such controls can be determined analytically since a rigid transformation of their associated boundary conditions simply induces the same rigid body transformation for the entire simulation mesh. Other kinematic parameters, such as those that articulate the jaw, nonrigidly change the quasistatic configuration. Differentiating equation (4) with respect to such a kinematic parameter  $b_i$  and rearranging gives

$$-\partial \mathbf{f} / \partial \mathbf{x}^U \Big|_{\mathbf{X}^*, \mathbf{a}^*} \partial \mathbf{X}^U / \partial b_i \Big|_{\mathbf{a}^*, \mathbf{b}^*} = \partial \mathbf{f} / \partial \mathbf{x}^C \Big|_{\mathbf{X}^*, \mathbf{a}^*} \partial \mathbf{X}^C / \partial b_i \Big|_{\mathbf{b}^*} \quad (6)$$

which is a linear system of equations for the unknown  $\partial \mathbf{X}^U / \partial b_i$ . Note that  $\partial \mathbf{X}^C / \partial b_i$  is analytically known from the definition of the kinematic parameters. Also,  $\partial \mathbf{f} / \partial \mathbf{x}^C$  is the stiffness of the forces on the unconstrained nodes with respect to the positions of the boundary conditions and is also analytically known from the definition of the finite element forces in our model. The entire right hand side of equation (6) can be interpreted as the linearized differential of the forces on the unconstrained nodes resulting from a displacement by  $\partial \mathbf{X}^C / \partial b_i$  of only the boundary conditions.

Both equations (5) and (6) require solving a linear system with the coefficient matrix  $-\partial \mathbf{f} / \partial \mathbf{x}^U$  which is the same coefficient matrix used in section 4, and thus the same solution techniques can be applied. Moreover, the coefficient matrix is symmetric positive definite near an equilibrium configuration, and thus special treatment is required only for element inversion (not definiteness). The Jacobians need to be computed before each Gauss-Newton step amounting to 44 applications of the conjugate gradient solver (for 32 activations and 12 kinematics parameters). In a sequence of expression tracking, an excellent initial guess to the conjugate gradient solver consists of using the Jacobians from the previous frame rotated according to the cranium motion. In fact, this allows us to update

all the Jacobians at a cost approximately equal to that of a single quasistatic solve.

## 7 Muscle Activation Constraints

In order to restrict the optimization process to the allowable parameter set  $\mathbf{C}_0$ , we augment  $\hat{\phi}(\mathbf{c})$  with a weighted penalty term  $\rho \phi_p(\mathbf{c})$  which consists of piecewise quadratic penalty terms of the form  $(\min\{0, a_i, 1 - a_i\})^2$  for each activation. These  $C^2$  continuous functions vanish within the allowable parameter space  $\mathbf{C}_0$ , and penalize the optimization functional when the control parameters drift away from  $\mathbf{C}_0$ . We typically initialize  $\rho$  with a value that is smaller than  $\phi(\mathbf{c}_{\text{opt}})$  (for example the variance of the localization error of the motion capture system), and then progressively increase its value in multiplicative increments of 5% until it is  $10^6$  times larger than the current value of  $\phi(\mathbf{c})$ . This drives the activations to within a maximum distance of  $10^{-4}$  of the allowed interval  $[0, 1]$ . At the maximum value of  $\rho$ , the contribution of  $\rho \phi_p(\mathbf{c})$  to the overall optimization functional is typically on the order of .1% implying that our minimization effort is properly focused on the constrained minimum of the proximity error  $\phi(\mathbf{c})$ .

In each step of the Gauss-Newton approach, we replace  $\phi(\mathbf{c})$  with  $\hat{\phi}(\mathbf{c})$  and apply the standard Newton approach to obtain  $-H_{\hat{\phi}}(\mathbf{c}_k) \delta \mathbf{c} = \nabla \hat{\phi}(\mathbf{c}_k)$  or  $-2\mathbf{J}_k^T \mathbf{J}_k \delta \mathbf{c} = \nabla \hat{\phi}(\mathbf{c}_k)$ . In our penalty term formulation,  $\mathbf{c}_{k+1}$  is obtained as the limit of the solutions  $\mathbf{c}_k^i$  to a nested sequence of unconstrained optimization problems minimizing  $\hat{\phi}(\mathbf{c}) + \rho_i \phi_p(\mathbf{c})$  for an increasing procession of weights  $\rho_i$ , i.e.  $\mathbf{c}_{k+1} = \mathbf{c}_k^N$  where the maximum  $\rho$  is  $\rho_N$ . Each step of this nested iteration is given by

$$-\left(2\mathbf{J}_k^T \mathbf{J}_k + \rho_i \partial^2 \phi_p / \partial \mathbf{c}^2 \Big|_{\mathbf{c}_k^i}\right) \delta \mathbf{c}^i = \nabla \hat{\phi}(\mathbf{c}_k^i) + \rho_i \nabla \phi_p(\mathbf{c}_k^i) \quad (7)$$

where  $\delta \mathbf{c}^i = \mathbf{c}_k^{i+1} - \mathbf{c}_k^i$ . Note that all derivatives of  $\phi_p$  can be computed analytically. Each iteration involves the solution of a low dimensional system (32 dimensions for the activations, but 56 total when the 24 kinematics parameters are included as outlined in section 8), and thus the overall computational cost is practically negligible.

## 8 Kinematics and Jaw Articulation

The kinematic parameters determine the placement of the cranium and mandible, and thus the position of specific mesh nodes of the interior flesh surface that have been rigidly attached to these bones. The frame of reference for the cranium determines the position and orientation of the entire head, while the frame of reference of the jaw is specified relative to the cranium and is subject to anatomical constraints that limit its relative placement and degrees of freedom.

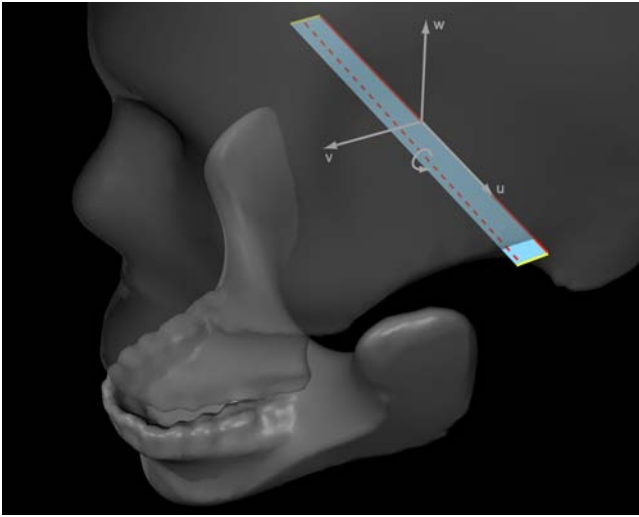


Figure 6: The mandible rotates around an axis (dashed red line) whose endpoints are allowed to move asymmetrically along two parallel segments (yellow lines) at the sides of the cranium.

In order to define each frame of reference, a displacement vector for the origin of each system must be supplied together with a descriptor of the orientation. Typical descriptions of orientation are poor choices for our optimization framework, since equation (3) indicates that we need to linearize the quasistatic configuration with respect to the control parameters. Large linearized rotations induce significant erroneous nonrigid distortion leading to a poor approximation of the rotation slowing convergence of the Gauss-Newton iteration especially for a remote initial guess. Thus, we propose an atypical penalty term formulation.

We begin by describing the rigid body frame by a general *affine* transform, i.e. the frames that describe the cranium and mandible are  $(\mathbf{M}_c, \mathbf{t}_c)$  and  $(\mathbf{M}_m, \mathbf{t}_m)$  where  $\mathbf{M}$  is *any* matrix (not necessarily orthogonal) and  $\mathbf{t}$  is a translation. Then the vector of kinematic controls,  $\mathbf{b}$ , consists of the 24 coefficients specifying these two affine transforms. Under this parameterization the mapping from the coefficients of the global frame of reference  $(\mathbf{M}_c, \mathbf{t}_c)$  to the positions of *all* nodes in the flesh mesh is *linear*, when the matrix  $\mathbf{M}_c$  is restricted to the set of rotation matrices. This implies that the linearization of equation (3) can capture exactly all rigid body transformations of the landmarks without any geometric distortion. Orthogonality of  $\mathbf{M}_c$  and  $\mathbf{M}_m$  (implying rigidity of the corresponding linear transform) can be enforced using the penalty term  $\phi_{\text{rigid}}(\mathbf{M}) = \|\mathbf{M}^T \mathbf{M} - \mathbf{I}\|_F^2$  with  $F$  representing the Frobenius norm. Note that this does not penalize the rotation, since a polar decomposition of  $\mathbf{M} = \mathbf{Q}\mathbf{S}$  into a rotation plus a symmetric matrix yields  $\mathbf{M}^T \mathbf{M} = \mathbf{S}^T \mathbf{Q}^T \mathbf{Q} \mathbf{S} = \mathbf{S}^T \mathbf{S}$  which removes the rotation. Thus it only penalizes the symmetric nonrigid deformation to be the identity matrix, thus removing it. Under the progressive stiffening schedule for  $\rho$  described in section 7, this penalty term keeps the singular values of the affine matrices within  $10^{-5}$  of unity. In order to ensure convexity of this penalty term we should project its Hessian to its positive definite component either through explicit eigenanalysis or through the process proposed in [Teran et al. 2005a]. Furthermore, we only need to solve equation (6) for the jaw, as the partial derivative of the quasistatic positions with respect to a component of the cranium affine transform can be analytically computed as  $\partial \mathbf{X}^j / \partial b_i = (\partial \mathbf{M}_c / \partial b_i) \mathbf{X}^j + \partial \mathbf{t}_c / \partial b_i$  for the quasistatic position of *any* node  $\mathbf{X}^j$  assuming orthogonality of  $\mathbf{M}_c$ .

We model the joint between the cranium and the mandible by a three degree of freedom articulation system as depicted in figure 6. During opening of the mouth, the lower jaw rotates around a hor-

izontal axis passing through the mandibular condyles, which are located at the rear extreme of the jawbone and are free to slide a short distance along the temporal bone of the cranium. We model the allowable trajectories of the condyles with two parallel line segments. The condyles can slide symmetrically or asymmetrically along their designated tracks; the latter effectively results in rotation of the mandible about a vertical axis. We formalize these constraints by requiring the horizontal axis of rotation to always lie on the plane defined by the two sliding tracks and restricting the midpoint of the two condyles to positions on that plane that are equidistant from the two side tracks.

To provide algebraic descriptions for these anatomical constraints, we equip the jaw with three characteristic normalized vectors defining the geometry of the temporomandibular joint in its rest configuration (fully closed with horizontally aligned dentures). In the reference frame of the cranium,  $\mathbf{u}$  points from the right to the left condyle,  $\mathbf{v}$  is parallel to the sliding tracks of the condyles directed from back to front, and  $\mathbf{w} = \mathbf{v} \times \mathbf{u}$ . Labeling the initial location of the midpoint of the two condyles as  $\mathbf{m}$ , the algebraic constraints for anatomical validity are

$$\begin{aligned} \psi_1(\mathbf{M}_m, \mathbf{t}_m) &= \mathbf{w}^T \mathbf{M}_m \mathbf{u} = 0 \\ \psi_2(\mathbf{M}_m, \mathbf{t}_m) &= \mathbf{w}^T (\mathbf{M}_m \mathbf{m} + \mathbf{t}_m - \mathbf{m}) = 0 \\ \psi_3(\mathbf{M}_m, \mathbf{t}_m) &= \mathbf{u}^T (\mathbf{M}_m \mathbf{m} + \mathbf{t}_m - \mathbf{m}) = 0 \\ \psi_4(\mathbf{M}_m, \mathbf{t}_m) &= \mathbf{v}^T [\mathbf{M}_m (\mathbf{m} - (l/2)\mathbf{u}) + \mathbf{t}_m - \mathbf{m}] \in [0, d] \\ \psi_5(\mathbf{M}_m, \mathbf{t}_m) &= \mathbf{v}^T [\mathbf{M}_m (\mathbf{m} + (l/2)\mathbf{u}) + \mathbf{t}_m - \mathbf{m}] \in [0, d] \\ \psi_6(\mathbf{M}_m, \mathbf{t}_m) &= \mathbf{w}^T \mathbf{M}_m \mathbf{v} \in [-s, 0] \end{aligned}$$

where  $l$  is the distance between the two condyles,  $d$  is the length of the sliding tracks, and  $s$  is the sine of the maximum opening angle of the mouth.  $\psi_1$  forces the horizontal rotation axis to be parallel to the plane of the sliding tracks,  $\psi_2$  forces the midpoint to reside on the same plane, and  $\psi_3$  forces it to be equidistant from the two sliding tracks. The additional constraints keep the three remaining degrees of freedom within their allowable range.  $\psi_4$  and  $\psi_5$  constrain the left and right condyle on their sliding tracks, and  $\psi_6$  regulates the opening angle. Finally, the kinematic validity penalty term is

$$\begin{aligned} \phi_{\text{kin}}(\mathbf{b}) &= \phi_{\text{rigid}}(\mathbf{M}_c) + \phi_{\text{rigid}}(\mathbf{M}_m) \\ &\quad + \psi_1^2 + \psi_2^2 + \psi_3^2 + \min\{0, \psi_4, d - \psi_4\}^2 \\ &\quad + \min\{0, \psi_5, d - \psi_5\}^2 + \min\{0, -\psi_6, s + \psi_6\}^2 \end{aligned}$$

noting that all the piecewise quadratic terms based on  $\psi_1$  to  $\psi_6$  are convex functions, and therefore no adjustment of their Hessian is necessary. All the terms in  $\phi_{\text{kin}}$  are included in  $\phi_p$  and handled as in section 7.

## 9 Examples

We evaluate our system by estimating muscle activations and kinematic parameters from a set of test motion capture sequences. These include a 33 second long narration sequence (figure 5) and several individual examples of pronounced expressions typically 2-3 seconds long (figure 4), which can be compared with expressions obtained by manual specification of muscle activation levels (see figure 3). Our single mocap session used 79 markers, and we focused them on mouth and jaw movement as opposed to the forehead and eyes. The motion capture input was processed at a frame rate of 60 frames per second. With the possible exception of the first frame of each capture sequence, we typically used a single Gauss-Newton step followed by a golden section line search for estimating the muscle activations and kinematic parameters at each subsequent frame. The average processing time for our simulation model of

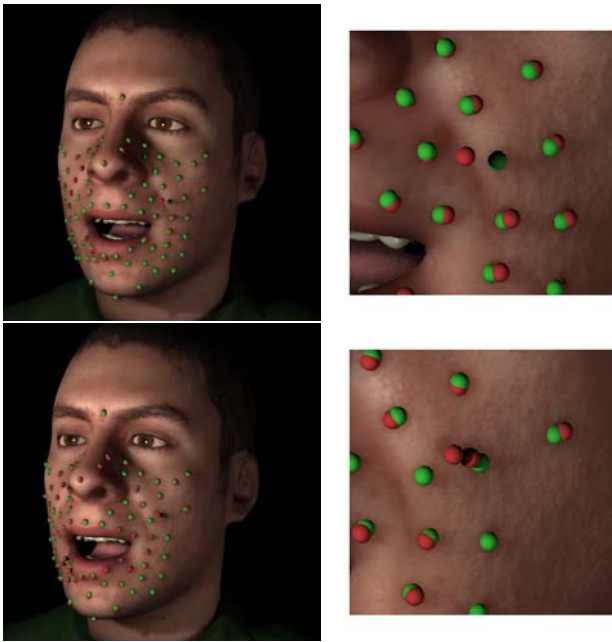


Figure 7: Outliers of noisy motion capture markers are handled robustly without incurring spurious local deformation. An enlarged view of a poor quality captured marker (colored red) is shown to the right of each figure.

370K tetrahedral elements and 32 transversely isotropic muscles was 8 minutes per frame on a single Xeon 3.06Ghz CPU, which includes 10 quasistatic solves for the chosen search depth of the line search with full collision handling in addition to 44 linear solves for the update of the control jacobians by application of equations (5) and (6). Using linear interpolation for the initial guess given to the quasistatic solver during each golden section search refinement and using the transform of the global frame of reference to precondition the initial guesses for both the quasistatic positions and their jacobians proved to be the most important performance optimizations. The cost of computing the Gauss-Newton step itself, once the linearization of equation (3) had been updated, was less than one second per frame.

Between successive frames, the activations can change by as much as 20%-40% and the kinematics can experience rotations of 3-4 degrees for both the global frame of reference and that for the jaw. We stress that our approach is trivially parallelizable, as a result of our quasistatic formulation. At the expense of estimating the first of a sequence of expressions from a suboptimal guess (which typically requires 3-5 Gauss-Newton iterations), processing a long sequence of motion capture frames can be partitioned arbitrarily.

An important aspect of our approach is that the search for the optimal match for the motion capture markers is performed over the space of physically attainable configurations, as parameterized by the muscle activations. This results in robust handling of noisy input data or motion outliers as illustrated in figure 7, since their non-physical component is discarded through the optimization process. Free form deformation and shape based animation schemes do not exhibit this property, and unfiltered motion outliers incur nonphysical localized deformation.

Once the muscle activations and kinematic parameters for an input sequence have been computed, we can address a number of post-processing tasks using the extracted, physically based animation parameters. Interpolation between expressions can be performed in the muscle activation space with an automatic guarantee that the



Figure 8: Accentuated expressions created by scaling of the force-length curve (unscaled, doubled and quadrupled from left to right).

interpolated expressions are physically valid and attainable (figure 10). Furthermore, an expression can be exaggerated or deemphasized by multiplying the muscle activations by a scaling factor, and clamping the result within the valid activation range  $[0, 1]$ . One can also exaggerate an expression (or sequence) beyond the physically attainable limits. It would be inadvisable to do so by extending the activation values beyond 1, since the force-length relationship is undefined for such values and heuristic extrapolations can be problematic. Instead, we scale the entire force-length curve effectively scaling the overall anisotropic behavior of the muscle while still maintaining plausible behavior as the muscle activation varies over the interval  $[0, 1]$  (see figure 8). The same effect would be difficult to achieve using blending techniques as pointed out in [Zhang et al. 2004].

Our estimation process is based on a quasistatic simulation of the face which disregards inertial phenomena. The quasistatic hypothesis is consistent with the empirical fact that humans tend to avoid sudden ballistic motion of their head (e.g. this is how boxers knock each other out). When we compared our quasistatic simulation to a fully dynamic one using biologically realistic material parameters and the estimated activations and kinematic controls, the differences were unnoticeable. We had to loosen up the material parameters to get noticeable ballistic effects, e.g. softening the cartilage in the nose (as shown in the accompanying video). Even for highly dynamic motion such as a person jogging, one could still capture the actor's performance quasistatically and add the dynamics as a post process. Finally, external elements can be introduced into a quasistatic or dynamic simulation that uses the extracted parameters. For example, figure 9 illustrates a quasistatic simulation of the face interacting with a kinematic sphere.



Figure 9: Interaction of the face with an external colliding object.



Figure 10: Interpolation between two motion captured expressions in the space of activations and rigid body kinematics.

## 10 Conclusions and Future Work

We have presented an anatomically accurate face model controlled by muscle activations and kinematic bone degrees of freedom. A novel algorithm was developed to automatically compute control values that track sparse motion capture marker input. Once the controls are reconstructed, the model can be subjected to interaction with external objects, used in a dynamic simulation to capture ballistic motion, expressions can be edited in the activation space by combining multiple existing segments or making manual adjustments, etc. We are currently building an even more accurate face model from a more powerful MRI scanner and a laser scan of a highly detailed cast of the face. We are also working to obtain improved motion capture data including data for the forehead and eye region, as well as more detailed mouth and lip tracking (placing markers *on* the lips as opposed to only around them). An obvious extension would be to generalize the control estimation framework to accept markerless input data. In fact, we are currently undertaking a project that determines the muscle activations associated with the articulation of phonemes, and this will require more detailed lip motion and muscle data.

There are many application areas that we can now address including, for example, the ability to learn patient-specific muscle activations that can be used to predict the effect that surgical modifications will have on expression. A natural but highly promising research direction would be to estimate not just the controls, but also the model parameters including bone and flesh structure, material constitutive parameters, muscle locations and shapes in the rest state, etc. This would allow us to correct anatomical modeling errors and make the model more predictive in a data driven fashion. Finally, it would be interesting to analyze a large number of facial expressions, for example deriving correlations between muscle activations. In this vein, we are also working on validating our muscle activation results using electromyography.

## Acknowledgements

Research supported in part by an ONR YIP award and a PECASE award (ONR N00014-01-1-0620), a Packard Foundation Fellowship, a Sloan Research Fellowship, ONR N00014-03-1-0071, ONR N00014-02-1-0720, ARO DAAD19-03-1-0331, NSF IIS-0326388, NSF ITR-0205671 and NIH U54-GM072970. E.S. was supported in part by a Stanford Graduate Fellowship. We would like to thank Dinesh Pai and Paul Kry for the motion capture data (acquired with a Vicon system), Garry Gold for the MRI data, CyberWare for the laser scan, Jiayi Chong and Michael Turitzin for modeling and rendering, Sergey Koltakov for help assembling the final figures and videos, and Mike Houston, Christos Kozyrakis, Mark Horowitz, Bill Dally and Vijay Pande for computing resources. Finally, we would like to thank Demetri Terzopoulos, Lance Williams and Dinesh Pai for many valuable discussions!

## References

- BASU, S., OLIVER, N., AND PENTLAND, A. 1998. 3D lip shapes from video: a combined physical-statistical model. *Speech Communication* 26, 131–148.
- BASU, S., OLIVER, N., AND PENTLAND, A. 1998. 3D modeling and tracking of human lip motions. IEEE Computer Society, 337–343.
- BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3D faces. In *Proc. of ACM SIGGRAPH*, ACM Press, 187–194.
- BLANZ, V., AND VETTER, T. 2003. Face recognition based on fitting a 3D morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 29, 9, 1063–1074.
- BLANZ, V., BASSO, C., POGGIO, T., AND VETTER, T. 2003. Reanimating faces in images and video. In *Proc. of Eurographics*, vol. 22.
- BLANZ, V., SCHERBAUM, K., VETTER, T., AND SEIDEL, H. P. 2004. Exchanging faces in images. In *Proc. of Eurographics*, vol. 23.
- BORSHUKOV, G., PIPONI, D., LARSEN, O., LEWIS, J. P., AND TEMPELAAR-LIETZ, C. 2003. Universal Capture - image-based facial animation for “The Matrix Reloaded”. In *ACM SIGGRAPH 2003 Sketches & Applications*, ACM Press, 1–1.
- BRAND, M. 1999. Voice puppetry. In *Proc. of ACM SIGGRAPH*, 21–28.
- BREGLER, C., COVELL, M., AND SLANEY, M. 1997. Video Rewrite: driving visual speech with audio. In *Proc. of ACM SIGGRAPH*, 353–360.
- BYUN, M., AND BADLER, N. I. 2002. FacEMOTE: Qualitative parametric modifiers for facial animations. In *Proc. of ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, ACM Press, 65–71.
- CAO, Y., FALOUTSOS, P., AND PIGHIN, F. 2003. Unsupervised learning for speech motion editing. In *Proc. of the ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 225–231.
- CAO, Y., FALOUTSOS, P., KÖHLER, E., AND PIGHIN, F. 2004. Real-time speech motion synthesis from recorded motions. In *Proc. of 2003 ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 347–355.
- CASSELL, J., PELACHAUD, C., BADLER, N., STEEDMAN, M., ACHORN, B., BECKET, T., DOUBILLE, B., PREVOST, S., AND STONE, M. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proc. of ACM SIGGRAPH*, ACM Press, 413–420.
- CASSELL, J., VILHJÁLMSSON, H. H., AND BICKMORE, T. 2001. BEAT: the Behavior Expression Animation Toolkit. In *Proc. of ACM SIGGRAPH*, 477–486.
- CHAI, J., XIAO, J., AND HODGINS, J. 2003. Vision-based control of 3D facial animation. In *Proc. of ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 193–206.
- CHOE, B., AND KO, H.-S. 2001. Analysis and synthesis of facial expressions with hand-generated muscle actuation basis. In *Proc. of Comput. Anim.*, 12–19.
- CHOE, B., LEE, H., AND KO, H.-S. 2001. Performance-driven muscle-based facial animation. *J. Vis. and Comput. Anim.* 12, 67–79.
- DECARLO, D., METAXAS, D., AND STONE, M. 1998. An anthropometric face model using variational techniques. In *Proc. of ACM SIGGRAPH*, ACM Press, 67–74.



- EKMAN, P., AND FRIESEN, W. V. 1978. *Facial Action Coding System*. Consulting Psychologist Press, Palo Alto.
- ESSA, I., AND PENTLAND, A. 1997. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19, 7, 757–763.
- ESSA, I., BASU, S., DARRELL, T., AND PENTLAND, A. 1996. Modeling, tracking and interactive animation of faces and heads using input from video. In *Proc. of Computer Animation*, IEEE Computer Society, 68–79.
- EZZAT, T., GEIGER, G., AND POGGIO, T. 2002. Trainable videorealistic speech animation. In *ACM Transactions on Graphics*, ACM Press, vol. 21, 388–398.
- GILL, P. E., MURRAY, W., AND WRIGHT, M. H. 1981. *Practical Optimization*. Academic Press, San Diego, USA.
- GUENTER, B., GRIMM, C., WOOD, D., MALVAR, H., AND PIGHIN, F. 1998. Making faces. In *Proc. ACM SIGGRAPH*, ACM Press, 55–66.
- HEIDELBERGER, B., TESCHNER, M., KEISER, R., MÜLLER, M., AND GROSS, M. 2004. Consistent penetration depth estimation for deformable collision response. In *Proc. of Vision, Modeling, Visualization (VMV)*, 339–346.
- JOSHI, P., TIEN, W. C., DESBRUN, M., AND PIGHIN, F. 2003. Learning controls for blend shape based realistic facial animation. In *Proc. ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 365–373.
- KAHLER, K., HABER, J., AND SEIDEL, H.-P. 2001. Geometry-based muscle modeling for facial animation. In *Proc. of Graphics Interface*, 37–46.
- KAHLER, K., HABER, J., YAMAUCHI, H., AND SEIDEL, H.-P. 2002. Head shop: Generating animated head models with anatomical structure. In *Proc. of ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 55–63.
- KAHLER, K., HABER, J., AND SEIDEL, H.-P. 2003. Reanimating the dead: Reconstruction of expressive faces from skull data. In *ACM Trans. on Graphics*, vol. 22, 554–561.
- KALRA, P., MANGILI, A., MAGNETAT-THALMANN, N., AND THALMANN, D. 1992. Simulation of facial muscle actions based on rational free form deformations. In *Proc. of Eurographics*, 59–69.
- KEEVE, E., GIROD, S., PFEIFLE, P., AND GIROD, B. 1996. Anatomy-based facial tissue modeling using the finite element method. In *Proc. of Visualization*, 21–28.
- KOCH, R. M., GROSS, M. H., CARLS, F. R., VON BUREN, D. F., FANKHAUSER, G., AND PARISH, Y. I. H. 1996. Simulating facial surgery using finite element models. *Computer Graphics* 30, Annual Conference Series, 421–428.
- KOCH, R., GROSS, M., AND BOSSHARD, A. 1998. Emotion editing using finite elements. *Proceedings of Eurographics 1998* 17, 3.
- KSHIRSAGAR, S., AND MAGNETAT-THALMANN, N. 2003. Visyllable based speech animation. In *Proc. of Eurographics*, vol. 22.
- LEE, Y., TERZOPOULOS, D., AND WATERS, K. 1995. Realistic modeling for facial animation. *Comput. Graph. (SIGGRAPH Proc.)*, 55–62.
- LEE, S. P., BADLER, J. B., AND BADLER, N. I. 2002. Eyes alive. In *Proc. of ACM SIGGRAPH*, ACM Press, 637–644.
- MAGNETAT-THALMANN, N., PRIMEAU, E., AND THALMANN, D. 1988. Abstract muscle action procedures for human face animation. *The Visual Computer* 3, 5, 290–297.
- MORISHIMA, S., ISHIKAWA, T., AND TERZOPOULOS, D. 1998. Facial muscle parameter decision from 2D frontal image. In *Proc. of the Int. Conf. on Pattern Recognition*, vol. 1, 160–162.
- MOVING PICTURE EXPERTS GROUP, 1998. Information technology - coding of audio-visual objects part 2: Visual. Final draft of international standard ISO/IEC JTC1/SC29/WG11 N2501 14496-2.
- NA, K., AND JUNG, M. 2004. Hierarchical retargeting of fine facial motions. In *Proc. of Eurographics*, vol. 23.
- NOH, J., AND NEUMANN, U. 2001. Expression cloning. In *Proc. of ACM SIGGRAPH*, ACM Press, E. Fiume, Ed., 277–288.
- PARKE, F. I., AND WATERS, K. 1996. *Computer Facial Animation*. AK Peters, Ltd.
- PARKE, F. I. 1972. Computer generated animation of faces. In *Proc. of ACM Conference*, ACM Press, 451–457.
- PIEPER, S., ROSEN, J., AND ZELTZER, D. 1992. Interactive graphics for plastic surgery: A task-level analysis and implementation. In *Proc. of Symp. on Interactive 3D graphics*, ACM Press, 127–134.
- PIGHIN, F., HECKER, J., LISCHINSKI, D., SZELISKI, R., AND SALESIN, D. H. 1998. Synthesizing realistic facial expressions from photographs. In *Proc. of ACM SIGGRAPH*, ACM Press, 75–84.
- PIGHIN, F., SZELISKI, R., AND SALESIN, D. 1999. Resynthesizing facial animation through 3D model-based tracking. In *Proc. of Int. Conf. on Comput. Vision*, 143–150.
- PLATT, S. M., AND BADLER, N. I. 1981. Animating facial expressions. *Comput. Graph. (SIGGRAPH Proc.)*, 245–252.
- PYUN, H., KIM, Y., CHAE, W., KANG, H. W., AND SHIN, S. Y. 2003. An example-based approach for facial expression cloning. In *Proc. of ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 167–176.
- ROTH, S. H., GROSS, M., TURELLO, M. H., AND CARLS, S. 1998. A Bernstein-Bézier based approach to soft tissue simulation. In *Proc. of Eurographics*, vol. 17, 285–294.
- SUMNER, R., AND POPOVIĆ, J. 2004. Deformation transfer for triangle meshes. In *Proc. of ACM SIGGRAPH*, vol. 23, 32–39.
- TERAN, J., BLEMKER, S., NG, V., AND FEDKIW, R. 2003. Finite volume methods for the simulation of skeletal muscle. In *Proc. of the 2003 ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, 68–74.
- TERAN, J., SIFAKIS, E., IRVING, G., AND FEDKIW, R. 2005. Robust quasistatic finite elements and flesh simulation. *ACM Trans. on Graphics (to appear)*.
- TERAN, J., SIFAKIS, E., SALINAS-BLEMKER, S., NG-THOW-HING, V., LAU, C., AND FEDKIW, R. 2005. Creating and simulating skeletal muscle from the visible human data set. *IEEE Trans. on Vis. and Comput. Graph.* 11, 3, 317–328.
- TERZOPOULOS, D., AND WATERS, K. 1993. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 15, 6.
- TESCHNER, M., GIROD, S., AND GIROD, B. 2000. Direct computation of nonlinear soft-tissue deformation. In *Proc. of Vision, Modeling, and Visualization*, 383–390.
- TESCHNER, M., HEIDELBERGER, B., MÜLLER, M., POMERANETS, D., AND GROSS, M. 2003. Optimized spatial hashing for collision detection of deformable objects. In *Proc. of Vision, Modeling, Visualization (VMV)*, 47–54.
- U.S. NATIONAL LIBRARY OF MEDICINE, 1994. The visible human project. <http://www.nlm.nih.gov/research/visible/>.
- WANG, Y., HUANG, X., LEE, C. S., ZHANG, S., LI, Z., SAMARAS, D., METAXAS, D., ELGAMMAL, A., AND HUANG, P. 2004. High resolution acquisition, learning and transfer of dynamic 3-D facial expressions. In *Proc. of Eurographics*, 677–686.
- WATERS, K., AND FRISBIE, J. 1995. A coordinated muscle model for speech animation. In *Proc. of Graphics Interface*, 163–170.
- WATERS, K. 1987. A muscle model for animating three-dimensional facial expressions. *Comput. Graph. (SIGGRAPH Proc.)*, 17–24.
- WILLIAMS, L. 1990. Performance-driven facial animation. In *Computer Graphics (Proc. of Int. Conf. on Computer Graphics and Interactive Techniques)*, ACM Press, 235–242.
- WU, Y., KALRA, P., MAGNETAT-THALMANN, N., AND THALMANN, D. 1999. Simulating wrinkles and skin aging. *The Visual Computer* 15, 4, 183–198.
- ZAJAC, F. 1989. Muscle and tendon: Properties, models, scaling, and application to biomechanics and motor control. *Critical Reviews in Biomed. Eng.* 17, 4, 359–411.
- ZHANG, Q., LIU, Z., GUO, B., AND SHUM, H. 2003. Geometry-driven photorealistic facial expression synthesis. In *Proc. of ACM SIGGRAPH/Eurographics Symp. on Comput. Anim.*, ACM Press, 16–22.
- ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. 2004. Spacetime faces: High resolution capture for modeling and animation. In *Proc. of ACM SIGGRAPH*, ACM Press, vol. 23, 548–558.