An Empirical Rig for Jaw Animation

GASPARD ZOSS, Disney Research
DEREK BRADLEY, Disney Research
PASCAL BÉRARD, Disney Research, ETH Zurich
THABO BEELER, Disney Research

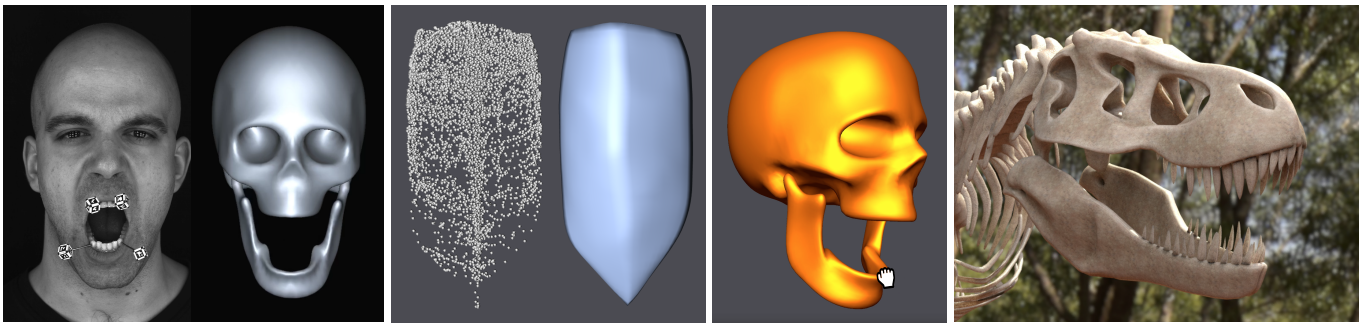

Fig. 1. We present an empirical rig for jaw animation, built from accurate capture data. Our rig is based on Posselt's Envelope of Motion, allowing intuitive 3-DOF control yet highly expressive and realistic jaw motions, and it can be retargeted to new characters and fantasy creatures.

In computer graphics the motion of the jaw is commonly modelled by up-down and left-right rotation around a fixed pivot plus a forward-backward translation, yielding a three dimensional rig that is highly suited for intuitive artistic control. The anatomical motion of the jaw is, however, much more complex since the joints that connect the jaw to the skull exhibit both rotational and translational components. In reality the jaw does not move in a three dimensional subspace but on a constrained manifold in six dimensions. We analyze this manifold in the context of computer animation and show how the manifold can be parameterized with three degrees of freedom, providing a novel jaw rig that preserves the intuitive control while providing more accurate jaw positioning. The chosen parameterization furthermore places anatomically correct limits on the motion, preventing the rig from entering physiologically infeasible poses. Our new jaw rig is empirically designed from accurate capture data, and we provide a simple method to retarget the rig to new characters, both human and fantasy.

CCS Concepts: • **Computing methodologies** → **Motion processing**; *Motion capture*;

Additional Key Words and Phrases: Jaw Rig, Data Driven Animation, Jaw Animation, Facial Animation, Motion Capture, Acquisition

Authors' addresses: Gaspard Zoss, Disney Research, gaspard.zoss@disneyresearch.com; Derek Bradley, Disney Research, derek.bradley@disneyresearch.com; Pascal Bérard, Disney Research, ETH Zurich, pascal.berard@disneyresearch.com; Thabo Beeler, Disney Research, thabo.beeler@disneyresearch.com.

0730-0301/2018/8-ART59 $15.00
https://doi.org/10.1145/3197517.3201382

ACM Reference Format:
Gaspard Zoss, Derek Bradley, Pascal Bérard, and Thabo Beeler. 2018. An Empirical Rig for Jaw Animation. *ACM Trans. Graph.* 37, 4, Article 59 (August 2018), 12 pages. https://doi.org/10.1145/3197517.3201382

1 INTRODUCTION

When looking at the human face, the mandible (or *jaw-bone*) plays a central role in defining the facial structure and appearance. Its position fundamentally determines the shape of the skin as well as the placement of lower teeth, both of which are extremely important and salient visual features, and misplacement by even a few millimiters can be perceived. As a consequence, one of the most common orthognathic procedures is to extend or shorten the mandible by a few millimeters to correct for malocclusions, such as under- or overbites. In computer graphics, the jaw plays a particularly important role during animation. Many facial rigs employ skinning to deform the skin as a function of the underlying bone motion, and hence it is important that this motion is correct. Such rigs employ abstract bones that are connected to each other via joints which limit their relative motion and offer an intuitive control structure.

The mandible is attached to the skull via the temporomandibular joint (TMJ), which is one of the most complex joints in the human body. Unlike a simple hinge joint (such as, for example, the elbow joint), the mandible slides over the surface of the skull bone while rotating, which means that the jaw does not have a fixed center of rotation (see Fig. 2). Furthermore, the final motion of the mandible is governed by the interplay of two such joints, one on each side of the head, with the consequence that the articulation takes place on a complex manifold in $\mathbb{R}^6$.

In computer animation this complexity is usually overlooked. Most commonly, animation rigs model the jaw joint by two rotations

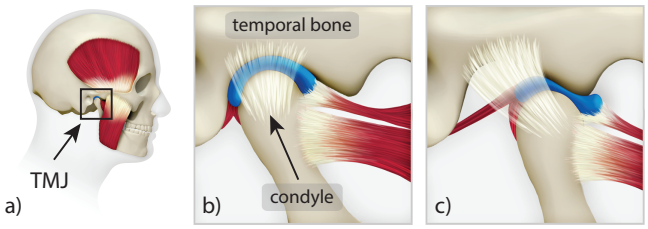

Fig. 2. **Anatomy.** The mandible is attached to the skull via the temporomandibular joint (TMJ) and held in place by ligaments and muscles (a). For small openings, the TMJ acts mostly rotational (b), but when the jaw is opened further, the posterior condyle leaves its socket and slides over the temporal bone of the skull (c), causing the rotational pivot to translate along a curve. A cartilage disc (blue) serves as cushion and prevents abrasion of the bone.

and one translation, simplifying the motion to three basic parameters - jaw-open, left/right, and forward/backward. While this simplification allows for intuitive control as it is only three degrees of freedom, it fails to reproduce the correct jaw articulation in $\mathbb{R}^6$, and can also allow anatomically impossible poses. When manual control is not required, such as for simulation, oftentimes jaw rigs with more degrees of freedom are employed to better resemble the correct jaw articulation [Ichim et al. 2017]. But such rigs are even more susceptible to producing physiologically infeasible articulation as they do not explicitly model the complex behaviour of the TMJ.

The goal of this paper is to provide an empirical jaw rig that models and exploits the manifold structure of the jaw articulation in order to provide more realistic jaw motions. Additionally, to manipulate our new rig we expose a compact and intuitive set of controls that allow for easier manual animation than current rigs. While this paper focuses purely on the articulated motion of the jaw bone, our work has the potential to significantly impact facial animation, since traditional face rigs deform the skin surface as a function of the jaw pose. Our empirical jaw rig is built from a corpus of highly accurate motion capture data that explores the entire manifold of jaw motion. The rig is parameterized by Posselt's Envelope of Motion [Posselt 1952; Posselt and D. 1958], a common anatomical representation of jaw movement, which maps the motion of a single point on the front of the mandible to a 3D volume that resembles the shape of a shield (see Fig. 3). In this work we will show that our compact rig representation can be controlled intuitively to create realistic jaw animations using several different user interfaces. Furthermore, we will demonstrate that the rig can be retargeted to new subjects (both human and fantasy creatures) given only a small number of input poses of the jaw for calibration, making our empirical jaw rig practical for immediate use in computer animation and visual effects.

2 RELATED WORK

In the following we outline related work in the areas of jaw motion analysis, face and jaw rigging, facial capture, and the use of facial anatomy in computer graphics.

2.1 Jaw Motion Analysis

In the medical field, a host of research studies have analyzed the motion of the mandible, in particular for dentistry [Ahlers et al. 2015; Bando et al. 2009; Ferrario et al. 2005; Knap et al. 1970; Mapelli et al. 2009; Okeson 2013; Posselt 1952; Posselt and D. 1958; Villamil and Nedel 2005; Villamil et al. 2012] and facial muscle control [Laboissière et al. 1996]. Since the mandible slides over the surface of the skull in a complex way while rotating, jaw articulation occurs on a complex manifold in $\mathbb{R}^6$. Early studies by Posselt et al. [1952; 1958] indicate that the range of motion of the mandible can be parameterized by tracing the trajectories of a single point at the anterior of the jaw. These trajectories form an intuitive constraint manifold with the shape of a shield, known as Posselt's Envelope of Motion (refer to Fig. 3). Our empirical jaw rig is based on this intuitive parameterization.

Another field that relies on jaw motion models and movement prediction is forensics. Here, several studies have analyzed the position and motion of the mandible with the goal of identifying humans from their skeletal remains [Bermejo et al. 2017; Kähler et al. 2003], or predicting what would have been in-vivo mandibular motion given only the geometry of a jaw-bone [Lemoine et al. 2007].

Aside from dentistry and forensics, jaw motion has been studied in the context of speech analysis [Ostry et al. 1997; Vatikiotis-Bateson and Ostry 1999, 1995] and chewing motion for food science [Daumas et al. 2005]. It is interesting to note that Ostry et al. [1997] criticize the parameterization of jaw motion based on Posselt's Envelope, since, in theory, an infinite combination of jaw orientations and positions can yield the same position of a single point at the front of the jaw. In their research, they suggest that a full 6-DOF parameterization is thus required. Theoretically this is correct, even though a real human jaw cannot undergo every possible combination of positions and orientations, there can in fact be ambiguities when mapping from Posselt's Envelope to the 6-DOF jaw pose. For some applications (such as medical or dental) this ambiguity may be critical, however for computer animation we believe the benefit of an intuitive mapping with less degrees of freedom outweighs the potential for ambiguity in the parameterization. In fact, as an experiment we performed this analysis and verified empirically that Posselt's Envelope in $\mathbb{R}^3$ has a nearly-unique mapping to the full position and orientation of the jaw in $\mathbb{R}^6$. As we will detail in Section 6.1, we found that ambiguities only occur in certain corner cases with a negligible difference in jaw positions. Therefore, we base our jaw rig on the parameterization of Posselt's Envelope in order to provide intuitive 3-DOF control while still providing highly accurate jaw motion.

2.2 Face and Jaw Rigging

In computer graphics, faces are often animated through the use of a facial rig. The most common facial rig is based on blendshapes, where the facial motion is created by blending linear combinations of individual expressions. We refer to the state-of-the-art reports of Orvalho et al. [2012] and Lewis et al. [2014] for a review of facial rigging and blendshape face models, respectively. Facial animation can also be achieved using data-driven statistical face models, like the morphable model [Blanz and Vetter 1999], or other multi-linear

models [Chen et al. 2014; Vlasic et al. 2005]. Recently, facial rigs are also starting to include volumetric information in order to represent tissue deformation below the surface [Ichim et al. 2017, 2016; Kozlov et al. 2017], and can even be built automatically from monocular video [Garrido et al. 2016a].

Some face rigs contain an underlying jaw rig as one of the components, often simplifying to three degrees of freedom (jaw open rotation, left/right rotation, and forward/backward translation). The jaw rig is typically used to skin the facial surface geometry, using methods such as linear blend skinning. For the application of physical simulation of faces, some more advanced jaw rigs do exist. In their pioneering work on muscle-based facial modeling, Sifakis et al. [2005] propose a different 3-DOF jaw rig specifically designed for easy linearization of the jaw constraints in an application of fitting to mo-cap data. Their rig allows a single rotation around a horizontal axis whose endpoints are located at the sides of the cranium and are allowed to slide forward and backward asymmetrically. The rotation models mouth opening, while the endpoint sliding models left/right rotation (when sliding is asymmetric) and forward/backward translation (when sliding is symmetric). Ichim et al. [2017] propose a physics-based facial animation system with a 5-DOF jaw rig, modeling rotation about both the horizontal and vertical axes and a full 3-DOF positional offset. While jaw rigs for simulation can be more complex and provide more degrees of freedom, they are not necessarily more accurate as they do not explicitly model the complex behaviour of real human jaw motion.

Our work is the first to explicitly focus on jaw rigging, as we build an empirical jaw rig with intuitive control that yields more accurate jaw motion than traditional jaw rigs.

2.3 Facial Capture

In this work we build a dataset of real jaw motion data using a capture setup similar to traditional facial performance capture. The field of facial capture has seen tremendous progress in recent years, both in the area of multi-view studio production capture [Beeler et al. 2010, 2011; Bradley et al. 2010; Fyffe et al. 2011, 2017] [Fyffe et al. 2014] and more lightweight consumer capture [Garrido et al. 2013; Shi et al. 2014; Suwajanakorn et al. 2014; Tewari et al. 2017; Valgaerts et al. 2012; Wu et al. 2016b] [Laine et al. 2017], even in real-time [Bouaziz et al. 2013; Cao et al. 2015, 2014; Hsieh et al. 2015; Li et al. 2013; Thies et al. 2016; Weise et al. 2011] [Thies et al. 2015]. Specific efforts have focused on complex components of the face including the eyes [Bérard et al. 2016, 2014], eyelids [Bermano et al. 2015], lips [Garrido et al. 2016b], teeth [Wu et al. 2016a], tongue [Luo et al. 2017], facial hair [Beeler et al. 2012], and even audio-driven animation [Karras et al. 2017]. To the best of our knowledge, our work is the first to go beyond traditional face capture and reconstruct detailed jaw movement for the purpose of rigging jaw animation.

2.4 Facial Anatomy in Computer Graphics

Since we study movement of the jaw bone for facial animation, our work is akin to other methods in computer graphics that consider facial anatomy during animation. In recent years we have seen an increasing trend in incorporating underlying anatomy (e.g. bones, muscles and tissue) in facial animation and tracking, as anatomy can provide very realistic constraints on motion and skin deformation. Historically, one of the first to build a complete muscle, tissue and bone model for simulating facial animation was Sifakis et al. [2005], mentioned earlier. More recently, new methods for physical simulation of faces also constructed at least partial models of bone or muscle [Cong et al. 2015, 2016] [Ichim et al. 2017; Kozlov et al. 2017]. In a different application, Beeler and Bradley [2014] fit a skull to facial scans using anatomically-motivated skin tissue thickness for the purpose of rigid stabilization of facial expressions. Wu et al [2016b] go even further and use the skull and jaw bones together with an expression-dependent skin thickness subspace and local deformation model to perform anatomically-constrained monocular face capture. We believe that having an anatomically accurate jaw rig can only help such techniques and promote further incorporation of anatomy in the field of data-driven facial animation.

3 EMPIRICAL JAW RIG

As illustrated in Fig. 2, the jaw bone or mandible is attached to the skull (more precisely to the temporal bone of the skull) via the temporomandibular joint (TMJ). Unlike a simple rotational joint, the TMJ contains both a rotational and translational component. This comes from the fact that for large openings of the mouth the posterior condyle of the mandible leaves its socket and slides over the surface of the temporal bone, effectively translating the rotational pivot along a curve in 3D. The two bones are held together by ligaments and, to prevent abrasion of the bones, are separated by a small disc of cartilage. To complicate things even more, two such joints operate in harmony to produce the motion of the jaw. For example, when rotating the mandible to the right, the right condyle remains within its socket and acts as a pure rotational joint, while the left one leaves its socket and translates forward. As a consequence, the motion of the jaw is constrained to a highly complex manifold. While the manifold is embedded in $\mathbb{R}^6$ it is itself lower dimensional. Medical literature reports the dimensionality to be $\mathbb{R}^4$ [Ostry et al. 1997], but for the purpose of computer animation, we show that it can be approximated in $\mathbb{R}^3$, which allows for convenient and intuitive parameterization.

3.1 Jaw Coordinate Frame

Given a mesh of the mandible in neutral pose (Section 4.2), we setup a convenient and intuitive coordinate frame for the jaw compatible with existing jaw rigs. We initialize the origin $\bar{\mathbf{o}}_{init}$ to be halfway between the left and right condyles, and choose the vector running from the right to the left condyle as x-axis. The z-axis is orthogonal to the x-axis and points from the origin towards the reference point $\mathbf{p}$ on the tip of the mandible, and the y-axis is chosen to form a right hand coordinate frame, roughly pointing upwards. For convenience, we define $\mathbf{C}$ as the transformation matrix of the coordinate frame in world space. See Fig. 3 (a) for a schematic depiction of the coordinate frame.

3.2 Traditional Jaw Rig

A generic jaw rig $\mathbf{J} = \mathcal{J}(\boldsymbol{\Theta}, \mathbf{C})$ computes a rigid transformation matrix $\mathbf{J} \in \mathbb{R}^6$ from the input parameterization domain $\boldsymbol{\Theta}$ relative to the neutral pose of the jaw in coordinate frame $\mathbf{C}$.

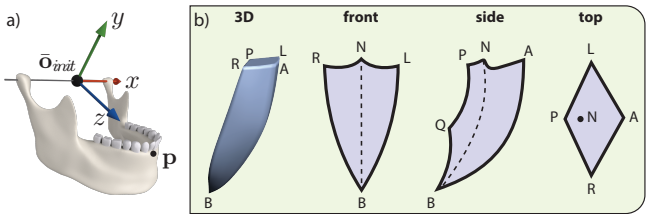

Fig. 3. **Jaw Model.** (a) The right hand coordinate frame C is setup such that its origin $\bar{o}_{init}$ is halfway between the condyles, with the x-axis pointing to the left, and the z-axis pointing towards the reference point $\mathbf{p}$ on the anterior part of the jaw. (b) The motion of this reference point lies in a subspace that resembles the shape of a shield, known as Posselt's Envelope of Motion. On the top the surface of the shield is determined by the teeth and the other surfaces are due to the limits of the TMJ: left (L) to right (R), anterior (A) to posterior (P), as well as fully open (B). N denotes the neutral jaw position. From P to Q the jaw operates purely rotational during opening, but from Q downwards to B the rotation axis translates as the condyle slides over the temporal bone (Fig. 2).

Skeletal rigs typically parameterize the bone motion via rotation $\phi_{\{xyz\}}$ and translation $t_{\{xyz\}}$ and traditional jaw rigs follow this strategy. By setting $\boldsymbol{\Theta} = [\phi_x, \phi_y, \phi_z, t_x, t_y, t_z]$ the jaw rig can be formulated as

$$\mathbf{J} = \mathbf{C} \cdot \mathbf{R}_x(\phi_x) \cdot \mathbf{R}_y(\phi_y) \cdot \mathbf{R}_z(\phi_z) \cdot \mathbf{T}_x(t_x) \cdot \mathbf{T}_y(t_y) \cdot \mathbf{T}_z(t_z) \cdot \mathbf{C}^{-1}, \quad (1)$$

where $\mathbf{R}_i(\phi_i)$ constructs the rotation around the i-axis by ϕ_i and $\mathbf{T}_i(t_i)$ the translation matrix along the i-axis by t_i. To allow for artistic control, the dimensionality of the parameterization is often reduced by constraining some components to 0. Probably the most commonly used parameterization is $\boldsymbol{\Theta} = [\phi_x, \phi_y, 0, 0, 0, t_z]$ as it offers intuitive control with three degrees of freedom (jaw opening, jaw rotation to the sides and forward/backward translation). However, as we will show in Section 6.1, this parameterization is too simplistic and fails to explain the physiologically correct jaw motion. Other parameter vectors of higher dimensionality have been proposed as well, and we analyze several of them in Section 6.1, with the conclusion that in order to explain real world observations of jaw motion a full rigid transformation in $\mathbb{R}^6$ is required.

3.3 Jaw Manifold

Instead of constraining the jaw motion to a subspace in $\mathbb{R}^n$, with $n < 6$ to allow for artistic control, we propose to constrain it to a manifold in $\mathbb{R}^6$, where the manifold itself has lower dimensionality. As discussed at the beginning of this section, this design choice is motivated by the anatomical function of the temporomandibular joints. The shape of the manifold is learned from captured data (Section 4.4) by a non-parametric regression (Section 5.1). The regression will output the full rigid jaw transformation $\mathbf{J} \in \mathbb{R}^6$ from a lower dimensional parameter vector. A good bijective parameterization domain is key for this approach to be successful.

3.4 Jaw Parameterization

A good parameterization should be as compact as possible and the individual dimensions should be semantically meaningful to allow for intuitive control. It should further be flexible enough such that all desired jaw poses can be reached while at the same time ensuring that anatomically infeasible poses cannot be generated. Finally, a jaw parameterization should ideally allow for different modes of control, for example, ranging from direct manipulation where a user directly grabs and moves the jaw to indirect control via a set of sliders. We base our parameterization on Posselt's Envelope of Motion, and show that such a parameterization can fulfill all these requirements.

Posselt's Envelope of Motion. In 1952 Dr. Ulf Posselt made the observation that a reference point on the anterior part of the mandible traces the shape of a shield in 3D during jaw articulation, nowadays referred to as Posselt's Envelope of Motion (Fig. 3 (b)). The envelope is bounded on the sides by the limits of the TMJ and on the top by the teeth occlusion when the jaw is fully closed. Any point within the envelope can be reached by the jaw, and as such it concisely describes the feasible subspace of motion for that point in $\mathbb{R}^3$. As we show in Section 6.1, we found that the mapping between a point in this envelope in $\mathbb{R}^3$ and the jaw pose in $\mathbb{R}^6$ is sufficiently bijective for the purpose of computer animation, and hence we suggest to use Posselt's Envelope of Motion as the parameterization domain and to learn a mapping $\boldsymbol{\Theta} = \Phi_{3D\to 6D}(\mathbf{p})$ that predicts jaw rotation and translation for any given point $\mathbf{p}$ within Posselt's Envelope $\mathcal{P}$, and from these the jaw pose $\mathbf{J}$ can be computed using the traditional jaw rig formulation (1)

$$\mathbf{J} = \mathcal{J}(\Phi_{3D\to 6D}(\mathbf{p}), \mathbf{C}). \quad (2)$$

Manifold Mapping. We represent the mapping $\Phi_{3D\to 6D}(\mathbf{p})$ using radial basis functions (RBFs), which provides a compact representation that lends itself well to interpolation within the shield. Each RBF kernel has a standard deviation σ_i, and is defined by its weight vector $\boldsymbol{\psi}_i \in \mathbb{R}^6$ and the RBF centers $\boldsymbol{\mu}_i \in \mathbb{R}^3$, which are uniformly distributed within the shield.

$$\Phi_{3D\to 6D}(\mathbf{p}) := \frac{\sum_{i=0}^{N-1} \boldsymbol{\psi}_i \exp\left(-\frac{1}{2}\frac{\|\mathbf{p}-\boldsymbol{\mu}_i\|^2}{\sigma_i^2}\right)}{\sum_{i=0}^{N-1} \exp\left(-\frac{1}{2}\frac{\|\mathbf{p}-\boldsymbol{\mu}_i\|^2}{\sigma_i^2}\right)}. \quad (3)$$

Please see Section 5.1 for details on how the mapping weights $\boldsymbol{\psi}_i$ are learned from captured data. The envelope naturally imposes limits on the parameters such that any generated jaw pose is anatomically feasible and all possible jaw poses may be created.

Unit Cube Parameterization. The envelope has a non-trivial shape and also varies between subjects. Because of this, it can be difficult to control the parameterization point $\mathbf{p}$ during animation, or semantically map jaw poses from one subject to another. For this reason, we define a mapping of the envelope to the unit cube. Except for the bottom point of the shield, this mapping is fully bijective such that

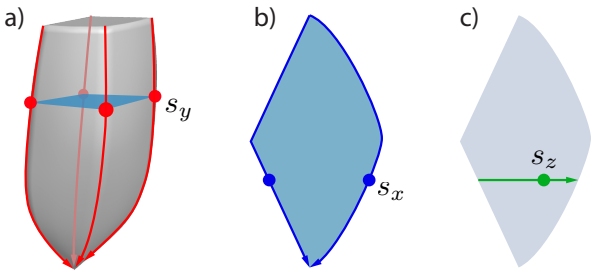

Fig. 4. **Curves.** (a) Four top-down curves at the extremal corners of the shield are sampled at s_y defining a horizontal slice through the envelope. (b) Within that slice the anterior and posterior curves running from left to right are sampled at s_x. (c) Lastly, the vector running from the posterior to the anterior s_x is sampled at s_z, providing the location $\mathbf{p}$ for the parameter vector $[s_x, s_y, s_z]$.

every point $\mathbf{p} \in \mathcal{P}$ has a unique correspondence $\mathbf{s} = [s_x, s_y, s_z]$ in the unit cube, and every point in the unit cube maps to a single valid point in the envelope, semantically equal between envelopes. The bottom point maps to the bottom face of the cube, which, however, does not pose a problem for the purpose of this paper. We choose the axes of the unit cube to be semantically meaningful with respect to jaw motion by setting the x-axis to encode left (s_x=0) to right (s_x=1), the y-axis top (s_y=0) to bottom (s_y=1), and the z-axis to represent back (s_z=0) to front (s_z=1). The surface of the unit cube will correspond to the surface of the envelope. Given a point $\mathbf{s} = [s_x, s_y, s_z]$ in the unit cube we compute the corresponding position of the reference point $\mathbf{p}$ within the envelope as follows. At the four extremal corners, where the jaw is all the way to the left/right and back/front we trace four curves from top to bottom (Fig. 4 (a)). Sampling each curve at the given value s_y produces a horizontal slice through the envelope (Fig. 4 (b)). Within this slice, two curves are defined, one on the anterior and one on the posterior surface of the envelope, running from left to right and sampled at the given value s_x. Finally, these two sample points describe the back-front vector, which is sampled at s_z yielding the final position of the reference point $\mathbf{p}$ within the envelope (Fig. 4 (c)). As indicated, this parameterization will come in handy for indirect control (Section 3.5) as well as rig adaptation to new subjects (Section 5.2).

3.5 Jaw Control

Two different paradigms exist for character rigging. One paradigm is based on directly manipulating the rig, typically leading to inverse kinematics, and the other paradigm is centered around indirect control or forward kinematics, where the rig is typically controlled via a set of sliders. Both have their strengths and we show how both paradigms can be implemented within the proposed jaw rig.

Direct Manipulation. Direct manipulation is straightforward in the presented context. A user may grab the jaw and translate it by moving the cursor, which will translate the reference point $\mathbf{p}$ within the envelope. Using the proposed mapping $\Phi_{3D\to 6D}(\mathbf{p})$ the six dimensional pose of the jaw is computed and applied. If the user moves $\mathbf{p}$ outside of the shield, it is projected back onto it ensuring the animation conforms with the anatomical limits of the character.

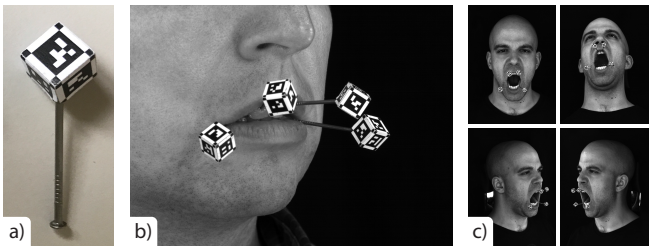

Fig. 5. **Data Capture**. (a) We designed 3D printed cubes with a fiducial marker on each face. (b) The marker cubes are mounted on steel pins and attached to the subjects teeth using dental glue, two on the lower and two on the upper teeth. (c) Eight cameras capture the jaw movement from four distinct viewpoints.

Indirect Control. To allow for indirect control we define three sliders by which a point within the envelope can be moved up-down, left-right and backwards-forwards. Mapping these sliders to the unit cube parameterization introduced above solves this task. The range of each slider is defined within 0 and 1, where the point will lie on the surface of the unit cube and consequently also on the surface of the shield for these two extremes. The point will trace a curve on the surface of the shield and follow interpolated curves within the envelope. Any point $\mathbf{s} = [s_x, s_y, s_z]$ within the unit cube will map to a point $\mathbf{p}$ in the envelope, which can then be mapped to the six dimensional pose of the jaw using the presented manifold mapping $\Phi_{3D\to 6D}(\cdot)$ analogous to the direct manipulation use case.

4 DATA ACQUISITION AND PREPARATION

Our jaw rig is empirically designed based on a corpus of highly accurate real jaw motion data, which we collect specifically for this purpose. The motion data is represented as a sequence of precisely tracked jaw poses $\hat{\mathbf{J}}_f \in \mathbb{R}^6$ for a number of frames f. The hat on the variable indicates that the quantity has been reconstructed from data without a rig prior. Reconstructing jaw motion from real subjects is extremely difficult, since the jaw is never directly visible. The subject's teeth are rigidly connected to the jaw, but even those are at least partially occluded almost all the time. To alleviate these problems, we attach marker tags to both the upper and lower teeth, providing consistently visible proxies for tracking the invisible skull and jaw bones (see Fig. 5).

4.1 Marker Design

We designed four 1cm^3 3D printed cubes (Fig. 5 (a)) on which we glued binary tags generated from a dictionary of 4x4 markers with a minimum hamming distance of 5 using the AruCo Library [Garrido-Jurado et al. 2014, 2016]. The markers are mounted on steel pins, which we glue to the teeth (two on the top, two on the bottom) using uv-hardened dental composite, ensuring sturdy attachment (See Fig. 5 (b)). This design provides a total of 32 markers, 16 per bone. In theory a single marker would be sufficient to recover the six dimensional pose of a bone, but by combining the information of several tags we can achieve much higher precision and robustness.

4.2 Data Capture

Once glued to the teeth we record the subject undergoing various jaw movements, including basic jaw articulation as well as more complex motion patterns such as chewing and speech. Recording was done using eight *Ximea CB120MG* monochrome machine vision cameras, which captured 4K imagery at 24 frames per second. The cameras were positioned in pairs of two, one pair on each side of the face, one in front and one slightly from below, and were geometrically calibrated using a checkerboard of fiducial markers [Garrido-Jurado et al. 2014].

We additionally acquired a single 3D face scan using a multi-view photogrammetry system [Beeler et al. 2010]. This allows us to relate the 3D position of the individual markers with the facial geometry, and the underlying bones. To determine the shape and relative positioning of the bones, we follow the approach presented in Beeler and Bradley [2014] and Ichim et al. [2016] and fit a generic skull to the face scan using forensic measurements [De Greef et al. 2006]. Once the skull is in place, we repeat the process for the jaw bone, additionally constraining the posterior condyle of the mandible relative to the temporal bone of the skull (Fig. 2 (a)). From the input imagery corresponding to the face scan, we reconstruct the pose of the individual marker cubes (as discussed below), establishing the relationship between the bones and the marker cubes. For convenience, we employ the same hardware setup to acquire both 3D scan and jaw movement, and so all data is inherently registered in the same world coordinate frame. The world coordinate frame is chosen as a right handed coordinate system with the origin inside the subjects head, the y-axis pointing up, and the z-axis through the nose.

4.3 Marker Cube Pose Estimation

The marker tags are detected in the captured images (Fig. 5 (c)) using the AruCo library with additional subpixel corner refinement (OpenCV). AruCo provides a unique ID for each marker tag, as well as an estimate of its pose. As we know which marker tags belong to which marker cube we can combine these independent estimations into a single pose prediction $\hat{\mathbf{T}}_{cube}$ per marker cube per frame, which is more precise than the individual estimates. The pose is computed by projecting the 3D marker tag corners into each visible camera view and minimizing the distance to their corresponding 2D locations, posing a least squares problem which we solve using Ceres [Agarwal et al. 2016].

Still, the pose is not perfect since the detected corners have slight inaccuracies (for example, due to foreshortening) and hence we refine $\hat{\mathbf{T}}_{cube}$ following the approach of [Wu et al. 2017]. We densely sample the marker cubes generating 3D positions $\mathbf{x}_i$ and associated colors $\mathbf{c}_i$ from the tags. With these we formulate a photometric loss by projecting the 3D points $\mathbf{x}_i$ into each visible camera view v, sampling the camera images $\mathbf{I}_v$ at those locations and computing the difference to the expected colors $\mathbf{c}_i$

$$E_{photo}(\,\hat{\mathbf{T}}_{cube}\,) = \sum_i \left\| \mathbf{c}_i - \mathbf{I}_v(\,\Gamma_v(\hat{\mathbf{T}}_{cube}\cdot\mathbf{x}_i)\,)\,\right\|_2^2, \tag{4}$$

where $\Gamma_v(\cdot)$ denotes the camera projection. We solve for the optimal transformation $\hat{\mathbf{T}}_{cube}$ starting from the previous guess using the Ceres solver. This yields extremely stable transformations per cube, removing any visible temporal jitter.

4.4 Jaw Pose Estimation

Given the individual poses of the two marker cubes attached to a bone, our goal is to infer the pose of that bone ($\hat{\mathbf{S}}_{world}$ and $\hat{\mathbf{J}}_{world}$ respectively). Since both bone and marker cubes are within the same coordinate frame, we can set the transformation of the bone to correspond to the average transformations of its marker cubes. As we are not interested in absolute jaw motion $\hat{\mathbf{J}}_{world}$ but rather its motion relative to the skull, we apply a change of coordinate frames by multiplying with the inverse of the skull bone transform $\hat{\mathbf{S}}_{world}$, yielding

$$\hat{\mathbf{J}} = \hat{\mathbf{S}}_{world}^{-1}\cdot\hat{\mathbf{J}}_{world}. \tag{5}$$

These poses are estimated independently per frame and serve as input data for fitting the rig in the next section.

5 RIG FITTING

Given a jaw rig $\mathcal{J}(\boldsymbol{\Theta}_f, \mathbf{C}(\bar{\mathbf{o}}))$, with $\mathbf{C}(\bar{\mathbf{o}})$ being the transformation matrix of the coordinate frame where $\bar{\mathbf{o}}$ is the transformed jaw origin, the goal is to find the optimal origin $\bar{\mathbf{o}}$ along with per frame rig actuation parameters $\boldsymbol{\Theta}_f$ that match the tracked jaw poses $\hat{\mathbf{J}}_f$ computed in the previous section, for all frames f. To this end, we formulate an energy that minimizes the difference between $\hat{\mathbf{J}}_f$ and the jaw transformation predicted by the rig $\mathcal{J}(\boldsymbol{\Theta}_f, \mathbf{C}(\bar{\mathbf{o}}))$ for all frames $f \in \mathcal{F}$

$$E_{data}(\boldsymbol{\Theta}, \bar{\mathbf{o}}) = \sum_{f\in\mathcal{F}} \left\| \hat{\mathbf{J}}_f - \mathcal{J}(\boldsymbol{\Theta}_f, \mathbf{C}(\bar{\mathbf{o}}))\right\|_{\mathbf{F}}, \tag{6}$$

where $\|\cdot\|_{\mathbf{F}}$ denotes the Frobenius norm. Depending on the degrees of freedom of the jaw rig, this formulation yields an underdetermined problem since the translational components of $\boldsymbol{\Theta}$ and the origin $\bar{\mathbf{o}}$ are ambiguous. Hence we add a weak regularization term that adds an additional constraint on the origin

$$E_{reg}(\bar{\mathbf{o}}) = \|\bar{\mathbf{o}}_{init} - \bar{\mathbf{o}}\|_{1^+}, \tag{7}$$

biasing the optimized origin to stay close to the initialization. Since we employ a soft L1 norm this presents only a weak bias even for larger deviations. The origin is initialized to a reasonable location as described in Section 3.1 and we further downweight the regularization term by $\lambda = 0.1$ relative to the data term yielding the following non-linear optimization problem

$$\min_{\boldsymbol{\Theta},\bar{\mathbf{o}}}\; E_{data}(\boldsymbol{\Theta}, \bar{\mathbf{o}}) + \lambda\cdot E_{reg}(\bar{\mathbf{o}}), \tag{8}$$

which we solve using the Ceres solver [Agarwal et al. 2016].

5.1 Manifold Regression

Fitting full rigid transformations to each frame f as described above will provide a set of jaw poses $\mathbf{J}_f$ that are essentially equivalent to the measured jaw poses $\hat{\mathbf{J}}_f$. Applying the transformation $\mathbf{J}_f$ to the reference point $\mathbf{p}$ on the mandible gives per frame positions $\mathbf{p}_f$, tracing the envelope of motion. From this dataset we regress the

manifold map $\Phi_{3D\to 6D}(\mathbf{p}_f) \to \mathbf{J}_f$ relative to coordinate frame $\mathbf{C}$ presented in Section 3.4. To perform the regression we minimize

$$E_{RBF}(\boldsymbol{\psi}_i, \sigma_i) = \left\| \mathcal{J}(\Phi_{3D\to 6D}(\mathbf{p}_f), \mathbf{C}) - \mathbf{J}_f \right\|_{\mathbf{F}} \tag{9}$$

to find the weight vectors $\boldsymbol{\psi}_i$ and supports σ_i for each of the RBF kernels. This regression closely predicts the fit jaw poses $\mathbf{J}_f$, while interpolating jaw poses throughout the parameterized shield. Please refer to Section 6.1 for a validation of the regression accuracy.

5.2 Rig Adaptation

Acquiring the data we leverage to construct our empirical jaw rig is an involved process and it would be desirable not to require such dense measurements for every new subject. Therefore, we propose to adapt the fit rig to novel subjects using just a few measurements, without the need for marker cubes. Specifically, we capture measurements of extremal jaw poses (e.g. all the way open, left, right, etc.), which map to the surface of the shield for the new subject. Since we only require a sparse set of poses we propose to manually annotate the teeth, alleviating the need for glueing on marker cubes. For a full adaptation we require at least three 3D landmarks on the teeth per pose, but we also introduce a reduced adaptation where a single 3D landmark per pose is sufficient. We require that one of the landmarks corresponds to the front of the mandible (i.e. the bottom of the lower teeth) in each pose, such that we are sure to measure $\hat{\mathbf{p}}_k$ for the sparse poses k, and we can additionally compute the full jaw pose $\hat{\mathbf{J}}_k$ in the case of three landmarks per pose.

Now, given the reference rig $\mathcal{J}^*(\boldsymbol{\Theta}, \mathbf{C}^*)$ and the sparse measurements, the goal is to compute an optimal rig $\mathcal{J}(\boldsymbol{\Theta}, \mathbf{C})$ which matches the target subject. We denote a variable with $*$ to indicate it refers to the reference rig. The first step in retargeting is to deform Posselt's Envelope of Motion using a thin shell deformation energy [Botsch and Sorkine 2008], where the data constraints are given by the correspondences $\{\mathbf{p}_k^*, \hat{\mathbf{p}}_k\}$ and a regularization term is given by the surface Laplacian. Since the Laplacian is not rotation invariant, we first compute a transformation $\mathbf{T}_\mathbf{C}$ between coordinate frames, which best aligns $\{\mathbf{p}_k^*\}$ to $\{\hat{\mathbf{p}}_k\}$ using the Procrustes algorithm [Gower 1975], and pre-transform the Laplacians. Once the shield surface is deformed, we can establish a bijective mapping within the entire volume using our unit cube parameterization (Section 3.4), which maps both shields to the unit cube. From this mapping, we can identify the reference point in the new shield $\mathbf{p}$ by mapping $\mathbf{p}^*$ from the reference shield to the new shield. Using the computed coordinate frame transformation, we also estimate an initial origin for the new rig as $\bar{\mathbf{o}}_{init} = \mathbf{T}_\mathbf{C} \cdot \bar{\mathbf{o}}^*$.

Our strategy for computing the new rig $\mathcal{J}(\boldsymbol{\Theta}, \mathbf{C})$ will be to retarget a dense set of jaw poses from $\mathcal{J}^*$, simulating a corpus of capture data, and then fit the rig parameters as described in Eq. 8 and recompute the RBF regression as described in Section 5.1.

Reduced Adaptation. Using the unit cube parameterization we sample the two envelopes jointly, producing a set of corresponding sample points $\{(\mathbf{p}_i^*, \mathbf{p}_i)\}$, and subsequently evaluate $\Phi^*_{3D\to 6D}(\mathbf{p}_i^*)$ to obtain a dense set of reference jaw poses $\mathbf{J}_i^*$. The problem now is to find a set of transformations $\mathbf{T}_i$ along with an optimal origin

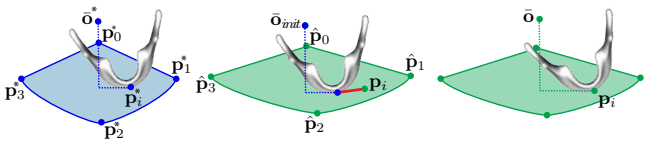

Fig. 6. **Rig Adaptation** (Top-down View). Our fit rig (blue) can be adapted to a new subject (green) by supplying a number of measured extremal positions $\{\hat{\mathbf{p}}_i\}$, which are used to deform the shield and initialize the origin $\bar{\mathbf{o}}_{init}$. Naïvely retargeting the jaw poses will violate the rig assumption that the front of the mandible lies at point $\mathbf{p}_i$ for pose $\mathbf{J}_i$ (red line). We solve for the optimal transformation $\mathbf{T}_i$ and origin $\bar{\mathbf{o}}$, which satisfies the rig assumption while also remaining close to the reference jaw pose.

$\bar{\mathbf{o}}$, such that the retargeted jaw poses $\mathbf{J}_i = \mathbf{C}(\bar{\mathbf{o}}) \cdot \mathbf{T}_i \cdot \mathbf{C}^{-1*}(\bar{\mathbf{o}}^*) \cdot \mathbf{J}_i^*$ align the front of the mandible to the points $\mathbf{p}_i$. This is illustrated in Fig. 6, where naïvely retargeting the jaw pose for $\mathbf{p}_i^*$ without a transformation and updated origin will violate a fundamental property of our parameterization in the new rig, i.e. that the reference point $\mathbf{p}$ on the anterior of the mandible must lie at position $\mathbf{p}_i$ for pose i. We seek to remove this discrepancy (shown as a red line in Fig. 6). Since this set of transformations is under-constrained, we add additional regularization to keep the jaw poses similar to the reference poses, aiming to maintain natural jaw motion where possible, and we prefer the retargeting transformations to be smooth. We also constrain the origin to remain close to the initial guess. With that in mind, we represent the transformations as $\mathbf{T}_i = \mathbf{T}(\mathbf{q}_i, \mathbf{t}_i)$, which converts to the quaternion $\mathbf{q}_i$ plus translation vector $\mathbf{t}_i$, and then formulate an energy residual for retargeting as

$$E_{retarget}(\mathbf{q}_i, \mathbf{t}_i, \bar{\mathbf{o}}) = \left\| \mathbf{C}(\bar{\mathbf{o}}) \cdot \mathbf{T}_i \cdot \mathbf{C}^{-1*}(\bar{\mathbf{o}}^*) \cdot \mathbf{J}_i^* \cdot \mathbf{p} - \mathbf{p}_i \right\|_2^2 . \tag{10}$$

In order to keep the resulting jaw poses similar to the reference poses, we add a term to penalize large transformations

$$E_{similar}(\mathbf{q}_i, \mathbf{t}_i) = \| \mathbf{T}_i - \mathbf{I} \|_2^2 . \tag{11}$$

To further constrain the solve we regularize the adaptation translations and quaternions to vary smoothly within the volume

$$E_{smooth}(\mathbf{q}_i, \mathbf{t}_i) = \sum_{j \in \mathcal{N}_i} \left\| \mathrm{conj}(\mathbf{q}_i) \cdot \mathbf{q}_j \right\|_2^2 + \left\| \mathbf{t}_i - \mathbf{t}_j \right\|_2^2 , \tag{12}$$

where $\mathcal{N}_i$ denotes the adjacent neighbours of i and $\mathrm{conj}(\cdot)$ computes the conjugate. Finally we constrain the origin the same way as during rig fitting by computing $E_{reg}(\bar{\mathbf{o}})$ using Eq. 7. We solve for optimal transformations $\{\mathbf{T}_i\}$ and origin $\bar{\mathbf{o}}$ by minimizing

$$\begin{aligned} \min_{\{\mathbf{q}_i, \mathbf{t}_i\}, \bar{\mathbf{o}}} \; & \lambda_0 \cdot E_{retarget}(\mathbf{q}_i, \mathbf{t}_i, \bar{\mathbf{o}}) + \lambda_1 \cdot E_{similar}(\mathbf{q}_i, \mathbf{t}_i) \\ & + \lambda_2 \cdot E_{smooth}(\mathbf{q}_i, \mathbf{t}_i) + \lambda_3 \cdot E_{reg}(\bar{\mathbf{o}}) . \end{aligned} \tag{13}$$

Once solved, we can compute the jaw poses $\mathbf{J}_i = \mathbf{C}(\bar{\mathbf{o}}) \cdot \mathbf{T}_i \cdot \mathbf{C}^{-1*}(\bar{\mathbf{o}}^*) \cdot \mathbf{J}_i^*$ for every point $\mathbf{p}_i$, compute the rig parameters $\boldsymbol{\Theta}_i$, and then recompute the RBF regression as described in Section 5.1.

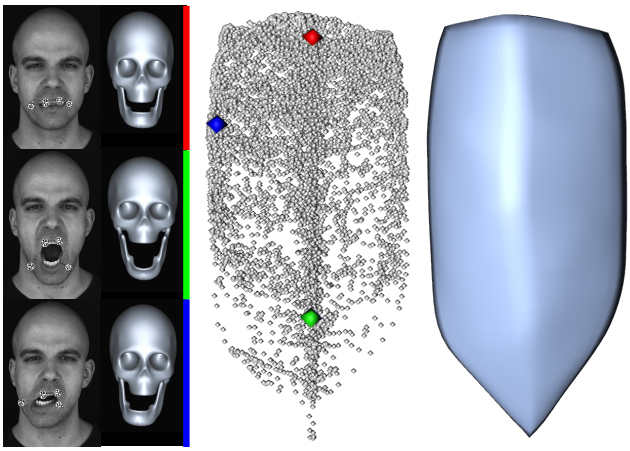

Fig. 7. **Jaw Motion Data**. We capture a large corpus of jaw motion data using a highly accurate marker-tracking approach. Here, three poses are shown, along with their corresponding location in Posselt's Envelope. The 9000 captured points span the entire envelope, allowing us to build a surface representation of the shield.

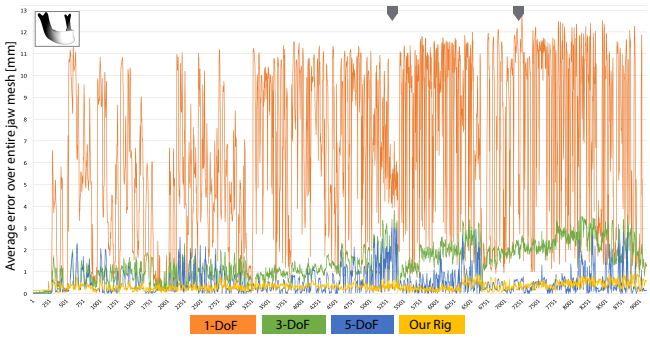

Fig. 8. **Rig Fitting**. We fit a naïve 1-DOF rig (orange line) and a traditional 3-DOF rig (green line) to our captured jaw motion and show the weighted error in pose per frame. Since errors at the front of the jaw would be most perceivable, this region is penalized higher than the back, as indicated by the weight map (inset top left). Even a higher-dimensional 5-DOF rig (blue line) cannot fit the data without occasionally large residual. Our intuitive 3-DOF rig (yellow line) fits the data much better, proving to be both accurate and easy to control. A visualization of the errors for two frames (indicated by the gray arrows) is shown in Fig. 9.

Full Adaptation. The reduced adaptation introduced above will transfer the pose space of the jaw to a novel user, but will not take into account person specific variations in the poses, for example if the target subject has stronger rotation around the z-axis when moving the jaw laterally. This requires a full adaptation of the rig, based on a sparse set of measured tuples $\{(\hat{\mathbf{p}}_k, \hat{\mathbf{J}}_k)\}$ which we get, for example, from the manually annotated frames used to deform the envelope. The full adaptation takes the same approach as the reduced adaptation, with one additional energy term in the optimization

$$E_{pose}(\mathbf{q}_j, \mathbf{t}_i, \bar{\mathbf{o}}) = \left\| \mathbf{C}(\bar{\mathbf{o}}) \cdot \mathbf{T}_k \cdot \mathbf{C}^{-1*}(\bar{\mathbf{o}}^*) \cdot \mathbf{J}_k^* - \hat{\mathbf{J}}_k \right\|_{\mathbf{F}}, \qquad (14)$$

for all sample poses k. The energy residual is combined with (13) and solved to retarget the pose space. We refer to Section 6.2 for applications of retargeting to both human and fantasy characters.

6 RESULTS

We now evaluate the different components and strengths of our empirical jaw rig, and present applications of jaw animation using our rig.

6.1 Evaluation

Our evaluation is based on a corpus of jaw motion data, captured in high quality as described in Section 4. We illustrate the dataset in Fig. 7, which shows three of the 9000 measured jaw poses, and the entire corpus as front mandible points which form a unique envelope for this subject.

Many traditional jaw rigs in computer animation model the motion with 3 degrees of freedom, a rotation to open the jaw, another rotation for lateral motion and a forward/backward translation such as the one used by Wu et al. [2016b]. While intuitive to control, we show that such a rig does not accurately model real jaw motion. To this end, we compute the optimal 3-DOF rig parameters and pivot point which best matches the jaw motion capture data of Fig. 7. Per frame errors are computed as the average Euclidean distance over all vertices of the mandible, spatially weighted to account for the fact that errors at the front of the jaw are more perceivable than at the back. The 3-DOF rig errors are plotted in Fig. 8 (green line). Higher dimensional rigs, such as the 5-DOF rig used by Ichim et al. [2017] are able to match the data better, but still contain significant errors (blue line). For completeness we also show a how a naïve 1-DOF rig fits the data, modeling only a rotational jaw open parameter (orange line). Only a full 6-DOF rig can model the data without residual. Our proposed rig is based on a 3D manifold in $\mathbb{R}^6$, parameterized by a mapping function $\Phi_{3D\to 6D}$ which is learned from the captured data. Fig. 8 (yellow line) shows that our 3-DOF rig has consistently low residual when fit to the data compared to other 3-DOF and even 5-DOF rigs. This suggests that our rig can remain faithful to real human motion while lending itself to easy manipulation thanks to only three control parameters. Fig. 9 visualizes the spatial distribution of the per vertex error for two frames of the captured sequence from Fig. 8 for each of the rigs (please see the supplemental material for videos of the entire sequence). It is worth noting that even though our rig fits the captured data well, there is nearly always *some* residual error. This is due to the nature of the RBF mapping framework described in Section 3.4, since the optimization tends to spread a little error evenly over all the RBF centers. For this reason, a few poses (such as the neutral jaw at the beginning of the sequence in Fig. 8) can actually be fit more accurately by the simple rigs, however our rig performs better on the full sequence with a consistently low error.

Our rig parameterization is based on Posselt's Envelope of Motion, which does not guarantee a unique mapping $\Phi_{3D\to 6D}$. That is to say, in theory, an infinite combination of jaw poses in $\mathbb{R}^6$ could map to the same envelope point in $\mathbb{R}^3$, making our rig ambiguous. However, we show that jaw motion is sufficiently constrained such that this does not occur in practice, except for negligible corner cases that

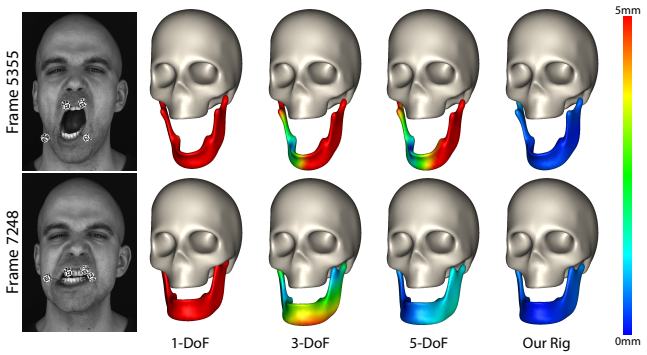

Fig. 9. **Rig Fit Error Visualization**. We visualize the Euclidean fitting error for each rig on two frames of the captured sequence shown in Fig. 8.

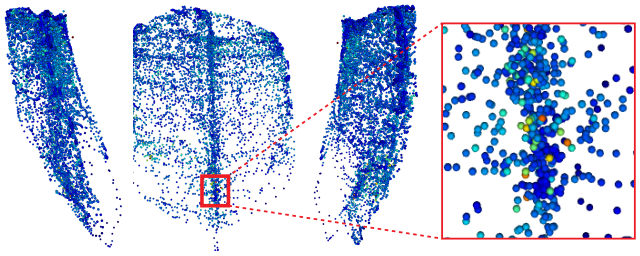

Fig. 10. **Bijective Mapping**. We plot d_i (Eq. 15) over the whole shield (front and two side views) to indicate how bijective the Posselt's Envelope is for our data. Jaw transformations mapping to the shield are generally bijective (blue) except for a few poses (red) due to hysteresis of the TMJ, as best viewed in the zoom region.

have little impact in computer animation. In order to validate this, for each point $\mathbf{p}_i$ in the shield we determine the set $\mathcal{N}_i$ of k nearest neighbors, and then compute

$$d_i = \max_{j \in \mathcal{N}_i} \mathcal{G}_\sigma \left(\|J_j - J_i\|_F \right) , \tag{15}$$

where $\mathcal{G}$ applies a Gaussian falloff with $\sigma = 1$mm, and we set $k = 10$. This measure aims to determine if there are close neighbors in the shield who's corresponding jaw poses differ substantially. Fig. 10 illustrates d_i for all measured points in our dataset. It is clear that the mapping is quite bijective almost everywhere (dark blue indicates $d_i = 0$ while red is $d_i = 35$). There are a few outlying points that exhibit some ambiguity in the mapping, for example as shown in the zoom region. These points represent jaw poses right before/after opening the jaw fully, and can be explained by hysteresis of the jaw motion, e.g. the jaw takes a different path when opening versus closing since the condyle is being pulled out of its socket during opening and pushed back in during closing. As the anterior of the mandible is at the same point for each pose, most of the effect happens at the back of the jaw and is imperceptible at the front.

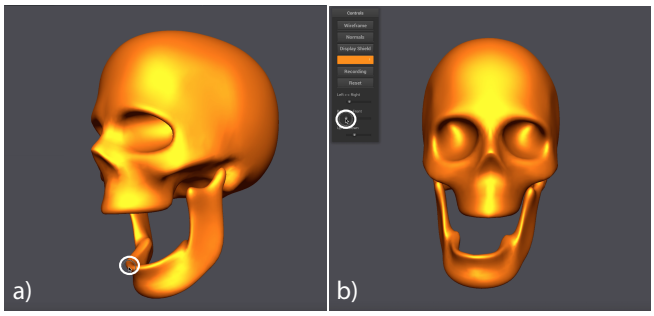

Fig. 11. **Rig Control**. We provide two methods for intuitive control of our rig. (a) Direct manipulation allows to directly grab and move the jaw. (b) Indirect manipulation allows jaw control through a set of sliders. Both methods are intuitive and provide accurate jaw motions. The mouse position is indicated by a white circle.

6.2 Applications

We demonstrate our jaw rig in action, with several applications of facial animation. For all applications, please refer to the supplemental video to see the full animations. As described in Section 3.5, we provide two different intuitive methods for controlling the rig, as shown in Fig. 11. The first is direct manipulation, where the user can directly grab the jaw at the front of the mandible and manipulate it with motion of the cursor (Fig. 11 (a)). The second is indirect manipulation, where the user can control the jaw pose through a set of sliders, which dictate the motion up-down, left-right, and backwards-forwards within the envelope (Fig. 11 (b)). At all times, the resulting jaw position is an anatomically feasible one, and the full space of motion can be reached with our simple rig interface.

Another application of our rig is retargeting, where the rig from one character can be adapted to another. We start by retargeting from one human to another. As described in Section 5.2, rig adaptation can be applied in two different ways - *full adaptation* where the input is a small number of jaw poses for the target person, or *reduced adaptation* where only the position of the anterior point on the jaw is provided for the target. In order to evaluate the effectiveness of both methods, we demonstrate rig retargeting from one captured subject to another subject where we also capture ground truth jaw motion data, as shown in Fig. 12. Specifically, the envelope from the source subject is adapted to the target using five measured extremal poses (front, back, left, right, and down). In this case the poses are recovered from the marker positions, but other approaches such as hand-labelling the teeth is also a viable option for such a small number of poses. The resulting envelope is shown in Fig. 12 (top row). Since we have a captured sequence of jaw positions for the target actor, we can use the corresponding anterior point to drive the motion of the adapted rig and measure the error with respect to the ground truth poses. This is illustrated in Fig. 12 (middle row) for both the full and reduced adaptation. As expected, the full adaptation (yellow line) performs better than the reduced one (blue line), in particular around the input poses. Fig. 12 (bottom row) illustrates the error distribution for two frames from the sequence (marked by gray arrows). It is worth noting that the target subject was unable to open his jaw as wide as the source subject, resulting in two very

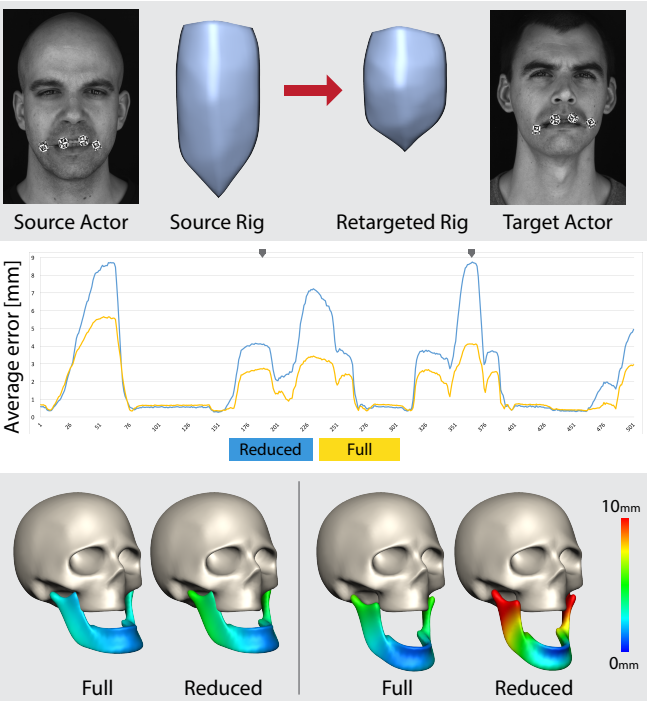

Fig. 12. **Retargeting to Human**. We demonstrate both full and reduced rig adaptation from one human subject to another. In this case, the source subject can open his jaw much wider, resulting in very different envelopes of motion (top row). Using a captured ground truth jaw motion sequence, we can analyze the performance of both adaptations (middle row). The error for individual frames shows how the full adaptation naturally performs better than the reduced one, although both are very plausible (bottom row).

different envelope shapes where the target one is much shorter. For this reason, the reduced adaptation struggles to adapt the jaw poses when the mouth is open wide, as the only constraint is at the anterior of the jaw ($E_{retarget}$ from Eq. 10). The full adaptation performs much better, since the actual pose of the open jaw is constrained as well (E_{pose} from Eq. 14). Nevertheless, both adaptions produce plausible jaw motion for the target subject. The optimization weights we use for retargeting from human to human are $\lambda_0 = 10$, $\lambda_1 = 0.1$, $\lambda_2 = 10$, $\lambda_3 = 0.1$ and $\lambda_4 = 100$.

Finally, we show an application of retargeting our jaw rig onto a fantasy creature in Fig. 13, where we adapt the rig from a human to a dinosaur using the full adaptation approach given five extremal jaw poses. We then drive the jaw motion using a captured sequence from an actor. Even though the envelope of motion for the two rigs are very different, our automatic retargeting provides natural animation transfer from the human actor to the creature. The optimization weights we use for retargeting from human to fantasy creature are $\lambda_0 = 10$, $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 0.1$ and $\lambda_4 = 1000$.

7 DISCUSSION

We present a novel jaw rig that models the physiological jaw motion more faithfully than existing rigs employed in computer animation while still offering intuitive artistic control. Furthermore, the rig

Fig. 13. **Retargeting to Fantasy Creature**. We retarget our rig from human to a fantasy dinosaur creature and drive the rig using motion capture data. Even though the envelope of jaw motion is very different, the transferred animation looks natural.

imposes realistic limits to the animator preventing anatomically infeasible jaw poses. Unlike prior art we do not constrain the jaw motion to lie in a subspace but explore the fact that the jaw motion lies on a constrained manifold embedded in $\mathbb{R}^6$. We show that for computer animation applications the manifold can be parameterized

by three degrees of freedom only, which can be mapped to intuitive dimensions. The design of the rig is motivated by anatomical considerations and derived from precise measurements of the jaw motion. We further show how the model can be retargeted to other actors and even fantasy characters using just a few data points. Once the rig is adopted to a new person, it can be evaluated efficiently, rendering itself very well for interactive and real-time applications. We demonstrate both direct and indirect manipulation controls.

7.1 Limitations and Future Work

The largest remaining residual between the captured jaw poses and the poses predicted by the presented rig can be attributed to hysteresis of the TMJ. This means that the condyle is at different positions relative to the reference point when opening and closing the jaw beyond the point of pure rotation, since it is being pulled out of its socket during opening and pushed back into it during closing. This is, amongst other things, due to the cartilage disc that serves as a cushion between the condyle and the temporal bone. As this effect is only really visible at the condyle itself and imperceivable towards the anterior part of the jaw, we did not address it in this work. Another interesting extension would be to investigate and model higher order motion patterns. Different activities such as speaking and chewing activate different muscle groups in the face and can cause unique (and often repetitive) motion of the mandible. It could be beneficial to animators if these complex motion patterns were modeled into higher order controls within our rig. These motion patterns are also very person specific and it would be intersting to retarget them to novel characters. Finally, it would be valuable to extend the model to include a concept of the overlying tissues. During jaw motion, these tissues slide over the bones, which is typically not accounted for with standard skinning techniques. On the other hand this could allow to track the jaw motion underneath the skin and allow to predict its pose even when the teeth are invisible.

ACKNOWLEDGEMENTS

We wish to thank our 3D artists Maurizio Nitti and Alessia Marra for their help in digital modeling and rendering, as well as Roman Cattaneo for posing as a capture subject. Jan Wezel was also instrumental in helping to design and 3D print the fiducial marker cubes. We further would like to thank Sabrina Wehrli for organising the dental resin "Telio CS Link" required to glue the marker cubes to the teeth, as well as Laurent Schenck and Michael Dieter from Ivoclar Vivadent for lending us the polymerisation lamp to cure it. The 3D dinosaur was modeled by Alvaro Luna Bautista and Joel Andersdon, and was obtained from the public domain[1].

REFERENCES

Sameer Agarwal, Keir Mierle, and Others. 2016. Ceres Solver. http://ceres-solver.org. (2016).

M Oliver Ahlers, Olaf Bernhardt, Holger A Jakstat, Bernd Kordass, Jens C Turp, Hans-Jurgen Schindler, and Alfons Hugger. 2015. Motion analysis of the mandible: guidelines for standardized analysis of computer-assisted recording of condylar movements. *International journal of computerized dentistry* 18, 3 (2015), 201–223.

Eiichi Bando, Keisuke Nishigawa, Masanori Nakano, Hisahiro Takeuchi, Shuji Shigemoto, Kazuo Okura, Toyoko Satsuma, and Takeshi Yamamoto. 2009. Current status of researches on jaw movement and occlusion for clinical application. *Japanese Dental Science Review* 45, 2 (2009), 83–97.

Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. 2010. High-quality single-shot capture of facial geometry. *ACM Transactions on Graphics* 29, 4 (2010), 1.

Thabo Beeler, Bernd Bickel, Gioacchino Noris, Paul Beardsley, Steve Marschner, Robert W. Sumner, and Markus Gross. 2012. Coupled 3D Reconstruction of Sparse Facial Hair and Skin. *ACM Trans. Graph.* 31, 4, Article 117 (2012), 117:1–117:10 pages.

Thabo Beeler and Derek Bradley. 2014. Rigid stabilization of facial expressions. *ACM Transactions on Graphics* 33, 4 (2014), 1–9.

Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. 2011. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics* (2011), 1.

Pascal Bérard, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. Lightweight eye capture using a parametric model. *ACM Transactions on Graphics* 35, 4 (2016), 1–12.

Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus Gross. 2014. High-quality Capture of Eyes. *ACM Trans. Graph.* 33, 6 (2014), 223:1–223:12.

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

Enrique Bermejo, Carmen Campomanes-Álvarez, Andrea Valsecchi, Oscar Ibáñez, Sergio Damas, and Oscar Cordón. 2017. Genetic algorithms for skull-face overlay including mandible articulation. *Information Sciences* 420 (2017), 200–217.

Volker Blanz and Thomas Vetter. 1999. A morphable model for the synthesis of 3D faces. *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99* (1999), 187–194.

Mario Botsch and Olga Sorkine. 2008. On linear variational surface deformation methods. *IEEE transactions on visualization and computer graphics* 14, 1 (2008), 213–230.

Sofien Bouaziz, Yangang Wang, and Mark Pauly. 2013. Online modeling for realtime facial animation. *ACM Trans. Graphics (Proc. SIGGRAPH)* 32, 4, Article 40 (2013), 40:1–40:10 pages.

D. Bradley, W. Heidrich, T. Popa, and A. Sheffer. 2010. High Resolution Passive Facial Performance Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 29, Article 41 (2010), 41:1–41:10 pages. Issue 4.

Chen Cao, Derek Bradley, Kun Zhou, and Thabo Beeler. 2015. Real-time high-fidelity facial performance capture. *ACM Transactions on Graphics* 34, 4 (2015), 46:1–46:9.

Chen Cao, Qiming Hou, and Kun Zhou. 2014. Displaced Dynamic Expression Regression for Real-time Facial Tracking and Animation. *ACM Trans. Graph.* 33, 4, Article 43 (2014), 43:1–43:10 pages.

Cao Chen, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. 2014. FaceWarehouse: a 3D Facial Expression Database for Visual Computing. *IEEE TVCG* 20, 3 (2014), 413–425.

Matthew Cong, Michael Bao, Jane L. E, Kiran S. Bhat, and Ronald Fedkiw. 2015. Fully Automatic Generation of Anatomical Face Simulation Models. In *Proc. SCA*. 175–183.

Matthew Cong, Kiran S. Bhat, and Ronald Fedkiw. 2016. Art-directed Muscle Simulation for High-end Facial Animation. In *Proc. SCA*. 119–127.

B. Daumas, W. L. Xu, and J. Bronlund. 2005. Jaw mechanism modeling and simulation. *Mechanism and Machine Theory* 40, 7 (2005), 821–833.

S. De Greef, P. Claes, D. Vandermeulen, W. Mollemans, P. Suetens, and G. Willems. 2006. Large-scale in-vivo Caucasian facial soft tissue thickness database for craniofacial reconstruction. *Forensic Science International* 159, 1 (2006).

Virgilio F. Ferrario, Chiarella Sforza, Nicola Lovecchio, and Fabrizio Mian. 2005. Quantification of translational and gliding components in human temporomandibular joint during mouth opening. *Archives of Oral Biology* 50, 5 (2005), 507–515.

Graham Fyffe, Tim Hawkins, Chris Watts, Wan-Chun Ma, and Paul Debevec. 2011. Comprehensive Facial Performance Capture. In *Eurographics*.

Graham Fyffe, Andrew Jones, Oleg Alexander, Ryosuke Ichikari, and Paul Debevec. 2014. Driving High-Resolution Facial Scans with Video Performance Capture. *ACM Trans. Graph.* 34, 1, Article 8 (2014), 8:1–8:14 pages.

G. Fyffe, K. Nagano, L. Huynh, S. Saito, J. Busch, A. Jones, H. Li, and P. Debevec. 2017. Multi-View Stereo on Consistent Face Topology. *Comput. Graph. Forum* 36, 2 (2017), 295–309.

Pablo Garrido, Levi Valgaerts, Chenglei Wu, and Christian Theobalt. 2013. Reconstructing Detailed Dynamic Face Geometry from Monocular Video. In *ACM Trans. Graph. (Proceedings of SIGGRAPH Asia 2013)*, Vol. 32. 158:1–158:10.

Pablo Garrido, Michael Zollhoefer, Dan Casas, Levi Valgaerts, Kiran Varanasi, Patrick Perez, and Christian Theobalt. 2016a. Reconstruction of Personalized 3D Face Rigs from Monocular Video. 35, 3 (2016), 28:1–28:15.

P. Garrido, M. Zollhöfer, C. Wu, D. Bradley, P. Perez, T. Beeler, and C. Theobalt. 2016b. Corrective 3D Reconstruction of Lips from Monocular Video. *ACM Transactions on Graphics (TOG)* 35, 6 (2016).

S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292.

[1] http://www.3drender.com/challenges/

S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer. 2016. Generation of fiducial marker dictionaries using Mixed Integer Linear Programming. *Pattern Recognition* 51, October (2016), 481–491.

S. Garrido-Jurado, R. Mu noz Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280 – 2292. https://doi.org/10.1016/j.patcog.2014.01.005

John C Gower. 1975. Generalized procrustes analysis. *Psychometrika* 40, 1 (1975), 33–51.

Pei-Lun Hsieh, Chongyang Ma, Jihun Yu, and Hao Li. 2015. Unconstrained Realtime Facial Performance Capture. In *Computer Vision and Pattern Recognition (CVPR)*.

Alexandru-Eugen Ichim, Petr Kadleček, Ladislav Kavan, and Mark Pauly. 2017. Phace: Physics-based Face Modeling and Animation. *ACM Transactions on Graphics* 36, 4 (2017), 1–14.

Alexandru-Eugen Ichim, Ladislav Kavan, Merlin Nimier-David, and Mark Pauly. 2016. Building and Animating User-Specific Volumetric Face Rigs. *Ladislav Kavan and Chris Wojtan* (2016).

K Kähler, J Haber, and H Seidel. 2003. Reanimating the Dead : Reconstruction of Expressive Faces from Skull Data. *ACM/SIGGRAPH Computer Graphics Proceedings* 22, July (2003), 554–567.

Tero Karras, Timo Aila, Samuli Laine, Antti Herva, and Jaakko Lehtinen. 2017. Audio-driven Facial Animation by Joint End-to-end Learning of Pose and Emotion. *ACM Trans. Graph.* 36, 4, Article 94 (2017), 94:1–94:12 pages.

F. J. Knap, B. L. Richardson, and J. Bogstad. 1970. Study of Mandibular Motion in Six Degrees of Freedom. *Journal of Dental Research* 49, 2 (1970), 289–292. https://doi.org/10.1177/00220345700490021501

Yeara Kozlov, Derek Bradley, Moritz Bächer, Bernhard Thomaszewski, Thabo Beeler, and Markus Gross. 2017. Enriching Facial Blendshape Rigs with Physical Simulation. *Comput. Graph. Forum* 36, 2 (2017), 75–84.

Rafael Laboissière, David J. Ostry, and Anatol G. Feldman. 1996. The control of multi-muscle systems: Human jaw and hyoid movements. *Biological Cybernetics* 74, 4 (1996), 373–384. https://doi.org/10.1007/BF00194930

Samuli Laine, Tero Karras, Timo Aila, Antti Herva, Shunsuke Saito, Ronald Yu, Hao Li, and Jaakko Lehtinen. 2017. Production-level Facial Performance Capture Using Deep Convolutional Neural Networks. In *Proc. SCA*. 10:1–10:10.

Jeremy J. Lemoine, James J. Xia, Clark R. Andersen, Jaime Gateno, William Buford Jr., and Michael A.K. Liebschner. 2007. Geometry-Based Algorithm for the Prediction of Nonpathologic Mandibular Movement. *Journal of Oral and Maxillofacial Surgery* 65, 12 (2007), 2411–2417.

J.P. Lewis, Ken Anjyo, Taehyun Rhee, Mengjie Zhang, Fred Pighin, and Zhigang Deng. 2014. Practice and Theory of Blendshape Facial Models. In *Eurographics State of The Art Reports*.

Hao Li, Jihun Yu, Yuting Ye, and Chris Bregler. 2013. Realtime facial animation with on-the-fly correctives. *ACM Trans. Graphics (Proc. SIGGRAPH)* 32, 4, Article 42 (2013), 42:1–42:10 pages.

Ran Luo, Qiang Fang, Jianguo Wei, Weiwei Xu, and Yin Yang. 2017. Acoustic VR of Human Tongue: Real-time Speech-driven Visual Tongue System. In *IEEE VR*.

Andrea Mapelli, Domenico Galante, Nicola Lovecchio, Chiarella Sforza, and Virgilio F. Ferrario. 2009. Translation and rotation movements of the mandible during mouth opening and closing. *Clinical Anatomy* 22, 3 (2009), 311–318. https://doi.org/10.1002/ca.20756

Jeffrey P. Okeson. 2013. Mechanics Of Mandibular Movement. In *Management of Temporomandibular Disorders And Occlusion*.

Verónica Orvalho, Pedro Bastos, Frederic Parke, Bruno Oliveira, and Xenxo Alvarez. 2012. A Facial Rigging Survey. In *Eurographics State of The Art Reports*.

David J. Ostry, Eric Vatikiotis-Bateson, and Paul L. Gribble. 1997. An Examination of the Degrees of Freedom of Human Jaw Motion in Speech and Mastication. *Journal of Speech, Language, and Hearing Research* Volume 40 (1997), 1341–1351.

U. Posselt. 1952. *Studies in the Mobility of the Human Mandible*. Acta Odontologica Scandinavica. https://books.google.ch/books?id=1MBpAAAAMAAJ

Ulf Posselt and Odont. D. 1958. Range of movement of the mandible. *The Journal of the American Dental Association* 56, 1 (1958), 10–13. https://doi.org/10.14219/jada.archive.1958.0017

Fuhao Shi, Hsiang-Tao Wu, Xin Tong, and Jinxiang Chai. 2014. Automatic Acquisition of High-fidelity Facial Performances Using Monocular Videos. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2014)* 33 (2014). Issue 6.

Eftychios Sifakis, Igor Neverov, and Ronald Fedkiw. 2005. Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM Transactions on Graphics* 24, 3 (2005), 417. https://doi.org/10.1145/1073204.1073208

Supasorn Suwajanakorn, Ira Kemelmacher-Shlizerman, and Steven M. Seitz. 2014. Total Moving Face Reconstruction. In *ECCV*.

Ayush Tewari, Michael Zollöfer, Hyeongwoo Kim, Pablo Garrido, Florian Bernard, Patrick Perez, and Christian Theobalt. 2017. MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction. In *Proc. of IEEE ICCV*.

Justus Thies, Michael Zollhöfer, Matthias Nießner, Levi Valgaerts, Marc Stamminger, and Christian Theobalt. 2015. Real-time Expression Transfer for Facial Reenactment. *ACM Trans. Graph.* 34, 6, Article 183 (2015), 183:1–183:14 pages.

J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner. 2016. Face2Face: Real-time Face Capture and Reenactment of RGB Videos. In *Proc. of IEEE CVPR*.

Levi Valgaerts, Chenglei Wu, Andrés Bruhn, Hans-Peter Seidel, and Christian Theobalt. 2012. Lightweight Binocular Facial Performance Capture under Uncontrolled Lighting. In *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2012)*, Vol. 31. 187:1–187:11.

E. Vatikiotis-Bateson and D.J. Ostry. 1999. Analysis and modeling of 3D jaw motion in speech and mastication. *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028)* 2 (1999), 442–447. https://doi.org/10.1109/ICSMC.1999.825301

Eric Vatikiotis-Bateson and David J. Ostry. 1995. An analysis of the dimensionality of jaw motion in speech. *Journal of Phonetics* 23, 1-2 (1995), 101–117. https://doi.org/10.1016/S0095-4470(95)80035-2

Marta B. Villamil and Luciana P. Nedel. 2005. A model to simulate the mastication motion at the temporomandibular joint. *Proceedings of SPIE* 5746, February 2014 (2005), 303–313. https://doi.org/10.1117/12.595742

Marta B. Villamil, Luciana P. Nedel, Carla M D S Freitas, and Benoit Macq. 2012. Simulation of the human TMJ behavior based on interdependent joints topology. *Computer Methods and Programs in Biomedicine* 105, 3 (2012), 217–232. https://doi.org/10.1016/j.cmpb.2011.09.010

Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. 2005. Face Transfer with Multilinear Models. *ACM Transactions on Graphics* 24, 3 (2005), 426–433.

Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. 2011. Realtime Performance-Based Facial Animation. *ACM Trans. Graphics (Proc. SIGGRAPH)* 30, 4 (2011), 77:1–77:10.

Chenglei Wu, Derek Bradley, Pablo Garrido, Michael Zollhöfer, Christian Theobalt, Markus Gross, and Thabo Beeler. 2016a. Model-based teeth reconstruction. *ACM Transactions on Graphics* 35, 6 (2016), 1–13.

Chenglei Wu, Derek Bradley, Markus Gross, and Thabo Beeler. 2016b. An anatomically-constrained local deformation model for monocular face capture. *ACM Transactions on Graphics* 35, 4 (2016), 1–12.

Po-Chen Wu, Robert Wang, Kenrick Kin, Christopher Twigg, Shangchen Han, Ming-Hsuan Yang, and Shao-Yi Chien. 2017. DodecaPen: Accurate 6DoF Tracking of a Passive Stylus. *ACM Symposium on User Interface Software and Technology* (2017), 365–374. https://doi.org/10.1145/3126594.3126664