Perceptually Valid Dynamics for Smiles and Blinks

Laura Trutoiu

TR-14-15

August 2014

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Jessica K. Hodgins (Chair), Carnegie Mellon University
Jeffrey Cohn, Carnegie Mellon University and University of Pittsburgh
Nancy Pollard, Carnegie Mellon University
Carol O'Sullivan, Disney Research and Trinity College Dublin

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

This research was sponsored by the National Science Foundation grant CCF-0811450 "Exploring the Uncanny Valley" and by Disney Research Pittsburgh.

Keywords: computer graphics, facial animation, data-driven models, perception, smiles, blinks, blendshape interpolation

“Whatever it is you’re seeking won’t come in the form you’re expecting.”
Haruki Murakami

Abstract

In many applications, such as conversational agents, virtual reality, movies, and games, animated facial expressions of computer-generated (CG) characters are used to communicate, teach, or entertain. With an increased demand for CG characters, it is important to animate accurate, realistic facial expressions because human facial expressions communicate a wealth of information. However, realistically animating faces is challenging and time-consuming for two reasons. First, human observers are adept at detecting anomalies in realistic CG facial animations. Second, traditional animation techniques based on keyframing sometimes approximate the dynamics of facial expressions or require extensive artistic input while high-resolution performance capture techniques are cost prohibitive.

In this thesis, we develop a framework to explore representations of two key facial expressions, blinks and smiles, and we show that data-driven models are needed to realistically animate these expressions. Our approach relies on utilizing high-resolution performance capture data to build models that can be used in traditional keyframing systems. First, we record large collections of high-resolution dynamic expressions through video and motion capture technology. Next, we build expression-specific models of the dynamic data properties of blinks and smiles. We explore variants of the model and assess whether viewers perceive the models as more natural than the simplified models present in the literature.

In the first part of the thesis, we build a generative model of the characteristic dynamics of blinks: fast closing of the eyelids followed by a slow opening. Blinks have a characteristic profile with relatively little variation across instances or people. Our results demonstrate the need for an accurate model of eye blink dynamics rather than simple approximations, as viewers perceive the difference.

In the second part of the thesis, we investigate how spatial and temporal linearities impact smile genuineness and build a model for genuine smiles. Our perceptual results indicate that a smile model needs to preserve temporal information. With this model, we synthesize perceptually genuine smiles that outperform traditional animation methods accompanied by plausible head motions.

In the last part of the thesis, we investigate how blinks synchronize with the start and end of spontaneous smiles. Our analysis shows that eye blinks correlate with the end of the smile and occur before the lip corners stop moving downwards. We argue that the timing of blinks relative to smiles is useful in creating compelling facial expressions.

Our work is directly applicable to current methods in animation. For example, we illustrate how our models can be used in the popular framework of blendshape animation to increase realism while keeping the system complexity low. Furthermore, our perceptual results can inform the design of realistic animation systems by highlighting common assumptions that over-simplify the dynamics of expressions.

Acknowledgments

This thesis came into being with the support and guidance of many people. First and foremost, my committee played a crucial role in the very existence and shape of this thesis. I am sincerely grateful to my advisor, Jessica Hodgins, for the opportunities and freedom she has granted me in the past six years. Under Jessica's supervision I have been able to explore many projects and find my path while also learning the high standards of academic research.

Jeff Cohn is a source of invaluable mentorship, guidance, and many rewarding conversations. Being optimistic about scientific research is difficult; I am grateful that Jeff frequently offered a ray of hope and a new way to look at data analysis paradigms and result graphs. Nancy Pollard taught me many practical ways to get out of what seemed like insurmountable difficulties. Nancy's vote of confidence in my thesis proposal ensured there is now a thesis. Carol O'Sullivan has shown me what a keen eye for research is. Carol's insight and comments have made this thesis significantly better both at the idea level and at the nitty-gritty level of analyzing psychological experiments.

Liz Carter is a great collaborator, friend, and above all a wonderful mentor. I'll cherish Liz's feedback and pragmatic advice on writing style and getting things done. A big thank you goes out to Moshe Mahler for modeling the characters used throughout these studies and for proving once again that artistic input is priceless. Justin Macey's help and expertise was priceless in collecting and cleaning the motion capture data. Thanks to Justin, even the many hours spent cleaning face marker glue were much more enjoyable than one would expect.

Jim McCann has been a sounding board for many half baked ideas and an inexhaustible source of creativity. I will always remember Jim's advice on how to read reviews positively — which may or may not involve funny voices. In addition to providing sound academic advice on research projects, Jim introduced me to the worthy pursuits of glass blowing, puzzle hunts, and mock-academic paper writing.

Reid Simmons stepped in at a delicate moment in the development of this thesis and played a significant part in mediating its course. I am deeply moved by the commitment and support the RI has shown its students in general and myself in particular.

At Disney Research, I've had the wonderful opportunity to interact with world-class researchers from whom I learned a lot about facial animation. Iain Matthews directed my first project and provided me with a great experience in recording a large dataset. Rafael Tena graciously shared his pipeline for cleaning up motion capture data and was a strong opponent at paint ball. Lavanya Sharan was a marvelous companion to spend that first SIGGRAPH winter with at CMU. Kiran Bhat at Industrial Light and Magic gave me the unique opportunity to experience a top R&D lab and was extremely patient with all my exercises in writing quality code.

If this thesis was a novel, the protagonist would be Ken, a talented and patient actor always ready to glue 300 markers on his face. Thanks to Ken and our other actors, professional or not, for taking part in our recording sessions. Similarly, over 3000 people viewed and rated a large number of animations clips in different stages of production.

Many people were incredibly generous with their time, feedback, and kindness. The graphics lab community in particular was always ready to give feedback, participate in long preliminary experiments, and provide much needed moral support.

Little did I know that the first person I met from CMU, Alyosha Efros, would influence my life in many ways. Alyosha was in charge of the PhD orientation in 2008 and he recruited Amar to help manage the crowd. That orientation opened up a world of beautiful friendships, peppered with many evenings of pondering the meaning of science while enjoying the more epicurean aspects of life. I've looked to Alyosha for advice in many situations and he was always ready with poignant, yet compassionate, words of wisdom.

Adrien Treuille showed me how important enthusiasm is. Thanks to Adrien's advice, my speaking skills talk changed for the better after only a two-hour feedback session. Kayvon Fatahalian showed me how to focus and highlight salient talk points. A research pitch I gave to Kayvon made it into a core idea of this thesis.

Ronit Slyper took me under her wing and showed me the graphics lab ropes in the early days. Ronit taught me that baking is always a good way to deal with stress and that cupcakes improve lab meetings. Eakta Jain taught me to take it one step at a time and to stop worrying about the big picture. Sophie Joerg was my commiseration buddy when perceptual results didn't turn out as expected. Q Youn Hong and Jenn Hyde, my academic siblings, have been great partners in this coming of age story. Similarly, my office mates Matt, Nico, and Tomas, have made Smith Hall 234 feel more like a home than an office.

Before focusing on facial animation, I collaborated on projects in robotics, biomechanics, gait analysis, and machine learning with Katsu Yamane, Akihiko Murai, Mark Redfern, and Fernando De La Torre. At University of Pittsburgh, I have received support with facial analysis and data acquisition from many folks including Jeff Girard, Nicki Siverling, and Shawn Zuratovic. Throughout the past six years, I have also had the pleasure to work with talented interns and undergraduate researchers: Chris Reynia, Katie Nestor, and Sharon Hoosein. They all deserve credit for what I have learned along the way.

CMU staff and resources helped me progress smoothly through grad school. Suzanne Lyons Muth, Deb Cavlovich, and Jess Butterbaugh frequently and adeptly tackled all sorts of administrative conundrums. Michele Keffer opened my eyes to a new world of productive introspection. Doug Phillips helped improve the writing in this thesis. Carol Frieze and Mary Widom provided me with many words of encouragement and the opportunity to be part of the tightly knit Women@SCS.

Bill Thompson, John Rieser, Claude Fennema, and Betty Mohler, my mentors before coming to CMU, have been my keen supporters in the big grad school adventure. Their influence and trust still carries through to this day.

My CMU batchmates and friends from here and there have made grad school the great experience I was hoping it would be. Thank you Or, Mark, Rui, Lin, Field, Iulian, Jenn, Erik, Matt, Brendan, Kate, Ricardo, Maria, Leon, Anca, Stacey, Ioana, Dani, Beatrice, Cristina, Natasha, Marynel. Whether it was potlucks, movie nights, board games get-together, craft sessions, pseudo-book clubs, outreach workshops, or road trips, we had a blast.

My parents are the constant friends that supported my academic endeavors and decisions every step of the way. They also instilled in me the key lesson of never giving up when things get hard. Lastly, I am amazingly grateful to have encountered the one person who is my endless source of energy, positive thinking, and inspiring altruism: my life partner, Amar.

Thank you to the entire community that made this journey possible!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

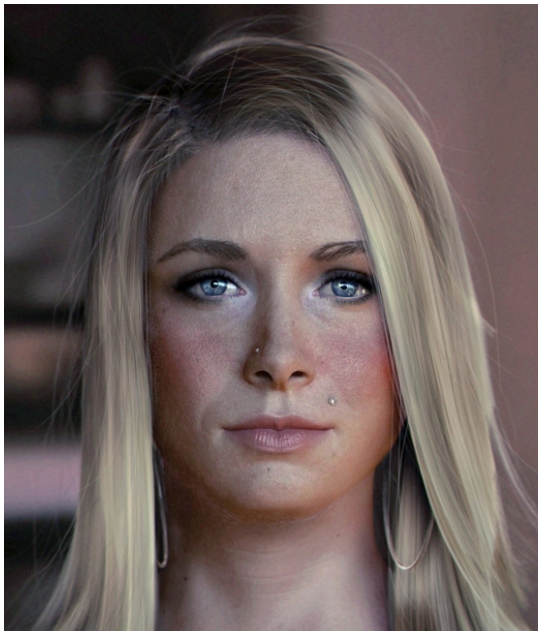
Realistic animated characters are essential for many computer graphics applications, such as movies, computer games, and embodied conversational agents. However, creating life-like human characters is challenging for developers and artists. Facial animation is particularly difficult because human observers are astoundingly good at perceiving and interpreting facial expressions: both physiological and psychological evidence supports the assertion that humans attend to the subtleties of facial expressions and emotion above most other signals. In fact, the human brain has a specific area, the fusiform face area, dedicated primarily to face processing [47]. Faces effectively communicate emotions [22], dominance [51], and approval [59].

Facial animation challenges are noticeable in the movie industry. Computer-generated (CG) movies have received negative film reviews because of anomalies in facial expressions, models and rendering. For example, critics stated that the CG faces in *The Polar Express* looked "lifeless" and "creepy" [13], that characters meant to be enraged in *Beowulf* "just look[ed] a little out of sorts" [38], and that animation in *A Christmas Carol* was "zombie-like" [60].

The negative reaction triggered in movie audiences and critics by anomalies in facial animations has been associated with the *Uncanny Valley* hypothesis. The hypothesis, proposed in 1970 by roboticist Masahiro Mori, explains emotional reactions to robots with increasing similarity to humans [69]. As the "human likeness" of a robot increases, viewers' emotional responses are initially positively correlated. However, beyond a certain similarity level, small details incongruent with human appearance and behavior may result in a strong negative correlation.

Mori defined his hypothesis for a robot's appearance as well as for dynamic motion. The hypothesis has since been extended to CG characters [37, 62, 81]. Despite many experiments the *Uncanny Valley* remains a hypothesis perhaps because the space is multi-dimensional and the axes are not well defined [39]. Current studies are beginning to address these issues [42, 43, 68, 94].

Title: Reg, The Normals
Artist: Liam Kemp
Software: 3DS Max, HairFX, Brazil
http://www.liamkemp.com/



Title: Project Blue
Artist: Dan Roarty
Software: Maya, Mudbox, Shave & Haircut, Mental Ray
http://www.danroarty.com/

Figure 1.1: Example of highly realistic CG faces produced with commercial software by professional artists. Animating these faces requires extensive time and effort. Images from the artists' websites.

Computer graphics researchers have significantly increased the realism of CG faces, particularly by acquiring and modeling detailed surface properties. For example, we now have sophisticated algorithms that model detailed skin characteristics such as light reflectance [27], wrinkles [8, 64], and color variations [19, 46]. Many of these research results are now available for use in commercial software packages (see Figure 1.1).

Actuating the face realistically, however, is more difficult because it involves creating spatial-temporal deformations, which we refer to as facial dynamics. For cinematic purposes, there are two main approaches to facial animation: key frame animation and performance-driven animation. Key frame animation considers facial expressions as a sequence of static deformations and relies largely on simple algorithms, such as blendshape interpolation. Performance-driven approaches use capture technology to reproduce an actor's unique performance, but the resulting animations are difficult to edit. These two methods lack models of the facial dynamics that can be used with current pipelines to create natural facial expressions that encapsulate the variability seen in human performances.

1.1 Approach

In this thesis, we investigate the dynamic characteristics of eye blinks, smiles, and the temporal connection between blinks and smiles. In particular, we investigate which data-derived models best capture these motion dynamics. Based on our perceptual experiments, we argue that facial animation algorithms should account for the temporal properties of naturally occurring facial expressions. We further provide evidence that animation textbook guidelines for key frame animation do not always generate natural facial expressions.

Our framework relies on three stages: (1) data collection, (2) models to produce new motions, and (3) perceptual evaluation of the dynamic properties of facial expressions. For model construction, our key insight is that because facial expressions are muscle-driven, they have consistent spatial deformations and characteristic dynamics defined as deformations over time. We exploit large data collections of ecologically valid facial expressions (smiles and blinks) and build models of the dynamics to create perceptually valid animations.

Eye blinks are relatively simple facial motions: the dynamics of the inter-eyelid distance can be described as fast closing followed by slower opening. Modeling and animating smiles is a more difficult task because of the expression complexity and the different meanings associated with different types of smiles. We further investigate the temporal connection between eye blinks and smile dynamics.

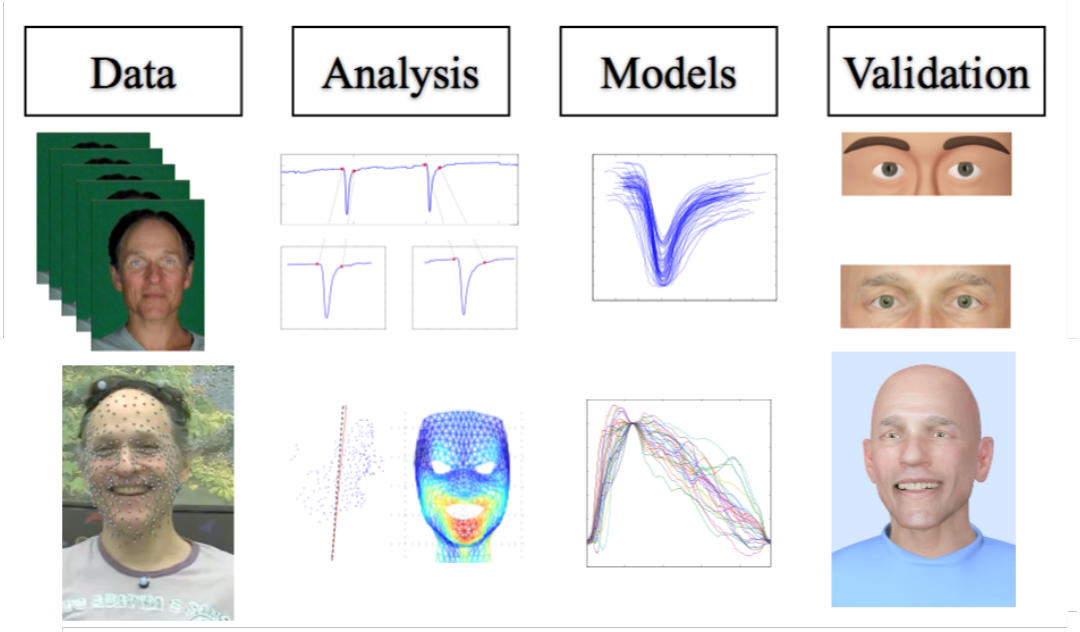


Figure 1.2: Overview of our approach: high-resolution data is processed and analyzed to obtain temporal models of facial deformations, which are then validated through perceptual experiments.

Creating the models for blinks and smiles follows the same overall framework: collect and analyze high-resolution data to build models, and conduct perceptual experiments to validate the models proposed (Figure 1.2). High-resolution data is crucial for obtaining accurate and generalizable models. For example, we used high-speed video to model eye blinks because eye blinks are very short in duration (less than 500 milliseconds). When using motion capture data, we opted for high spatial and temporal resolution. We recorded data from multiple participants to establish commonalities and differences across participants. Throughout the capture process, we elicited natural, spontaneous reactions and expressions by presenting participants with a wide range of stimuli. Working with high-density motion capture data of facial expressions is challenging because the 3D position of the face markers needs to be processed to consistently label the markers over time.

To build generative models of facial dynamics, we quantify and analyze relevant data parameters. For example, an eye blink is represented as the inter-eyelid distance over time. For smiles, we quantify the deformation of the mesh of connected vertices on the face. The primary data parameter we model is the relative deformation of the mesh at different instances in the smile. Similarly, we consider the correlations between head motion and mesh deformation in smiles.

We explain the problem of modeling the temporal properties of facial expressions in the context of blendshape-based systems. Blendshapes are static deformations of the face, usually extreme facial expressions, and arguably the most versatile and widely used parametrization of facial deformations currently in use (Figure 1.3). A basis set of N blendshapes B_i (ranging from forty to several hundred) can be combined linearly to create a facial pose P at time t such that

$$P_t = \sum_{i=1}^{N} (B_i \mathbf{f_i(t)}), \tag{1.1}$$

where $\mathbf{f_i(t)}$ is the interpolation function at time t (Figure 1.4). In practice, the interpolation function $\mathbf{f_i}$ is represented by a spline or polynomial expression that often approximates physical laws (acceleration and deceleration) as shown in Figure 1.4. To develop a realistic model, we aim to analyze high-resolution data to provide both the blendshapes $\mathbf{b_i}$ and the characteristic interpolation functions $\mathbf{f_i(w_i)}$ for a set of facial motions. We extend the blendshape model with data-driven interpolation functions to more accurately model the recorded data.

Our approach is to build expression-specific models allowing us to find a compact representation: data-driven interpolation functions for each expression. For blinks, an artist defines the simple blendshapes for eyes fully open and eyes fully closed. For smiles, the blendshapes are recovered from data as the extreme poses in the smile sequences. We then optimize blendshape weights to best reconstruct facial expressions. The models consist of new, data-driven, interpolation functions. To generate these new interpolation functions, we fit a Principal Component Analysis (PCA) generative model to the time series of computed weights or parameters.

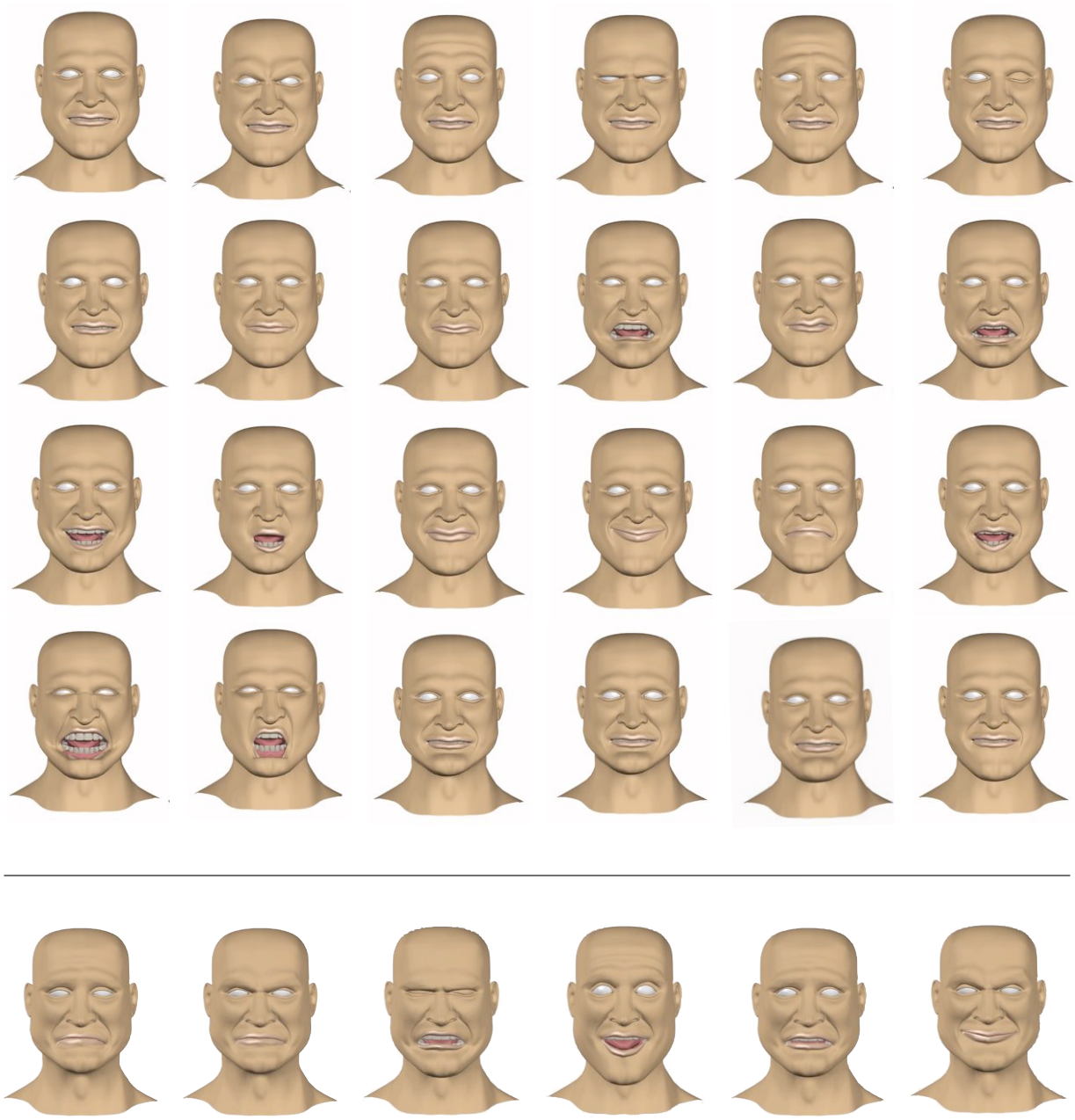


Figure 1.3: Example of a blendshape-based facial rig from a craft book on 3D modeling and animation: *Animating Facial Features & Expressions* [49]. Top: Twenty-five typical human expression blendshapes. Bottom: Sample facial expressions created by combining the blendshapes.

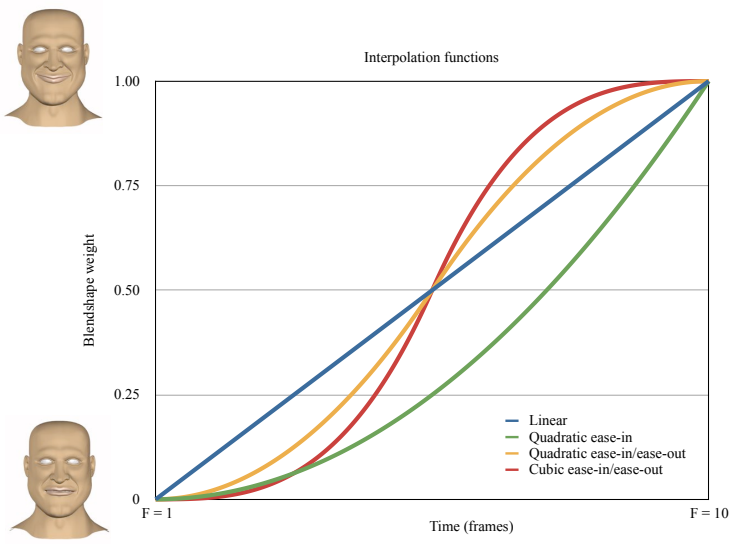


Figure 1.4: Example of interpolation functions used in blendshape systems [75]. Cubic interpolation splines (red) are often used in professional animation authoring tools such as Autodesk Maya.

Throughout the thesis, an important step is evaluating our models by creating animations based on the model and conducting perceptual experiments. Our hypothesis is that innate human knowledge about the characteristic dynamics of spontaneous facial expressions can be used to assess the realism of blendshape-based animation systems and identify the model parameters.

1.2 Research questions

The primary research question we address in this dissertation is *How should we design parsimonious models for realistic facial expressions such as blinks and smiles?* We hypothesize that these models need to include temporal information that characterizes the dynamics of the expression. Each chapter also addresses additional research questions as follows:

- How does temporal symmetry (same number of frames for the first half and the second half of the animation) impact how eye blinks are perceived?
- What eye blink durations are rated as most natural?
- How do spatial cues like lower eyelid motion and nonlinear vertex motion change how blinks and smiles respectively are perceived?
- What variability is present in eye blink motions?
- What are the quantitative differences between posed and spontaneous smiles?
- When do eye blinks occur relative to smile start and end?
- What guidelines can we provide animators who work with key frame-based systems to animate natural eye blinks and genuine smiles?

In the following sections, we summarize our answers to these questions. The overall contributions of this thesis are further discussed in Chapter 6.

1.3 Modeling and animating eye blinks

In Chapter 2, we focus on generating and evaluating dynamic parameters for one type of eye motion: blinking. Animation textbooks recommend using blinks as a way to "add life to a character" and to emphasize or influence personality and mood [63]. Unfortunately, oversimplifications and incorrect assumptions about the dynamics of eyelid motion can impact the quality of the final animation. These errors include the assumption that blinks are symmetric and thus the same number of frames should be used for opening and closing the eyelid [20, 63, 95], and that a linear or near-linear velocity profile of the eyelid motion is sufficient [20, 63]. This work was published in the journal *ACM Transactions on Applied Perception* 2011 [89].

Data capture. Blink motions are very short, ranging in duration from 100 to 500 milliseconds. In past research, eye blinks were quantified with invasive procedures such as magnetic coils [35]. Video technology has now improved sufficiently that we can measure eye blinks in high-speed video at 300 frames per second (fps). The video for three participants spanned different activities: performing emotional sequences, reading instructions on a teleprompter, and engaging in light conversation with the experimenters. Participants were aware that their facial expressions were recorded, but they did not know that eye blinks in particular were recorded and monitored.

Data analysis. We tracked the eyelid motion and surrounding eye area with an Active Appearance Model [67] (AAM) based computer vision algorithm. The eyelid position over time allowed us to quantify eye blink dynamics. We computed the inter-eyelid distance as a time series over the entire video. We then automatically determined the beginning and end of an eye blink.

Models. Using the computed data, we built a generative PCA model for the dynamics of eye blinks. Because eye blinks have different durations (eyelid closing time and eyelid opening time) and different closing amplitudes (minimum inter-eyelid distance during a blink), we parametrized these three variables as part of the model. The generated eye blink time series are equivalent to a set of key frame parameters that closely match naturally occurring blinks (one key frame per animation frame). The two blendshapes (eye open and eye closed) used to create animations were sculpted by an artist. Our model can generate a variety of blinks that follow a natural trajectory but show variability in closing amplitude and duration.

Perceptual experiments. We generated eye blink animations for two 3D characters: a photorealistic head and a cartoon head. We then asked participants to rate the naturalness of over 200 short animations. We statistically compared the perceived naturalness of blinks animated based on our model to those created using traditional animation techniques or generated using distorted position and velocity profiles. The results show that human observers are highly sensitive to recognizing natural eye blinks: blinks created with our model are perceived as significantly more natural than both symmetric and linear blinks.

1.4 Modeling and animating smiles

Though smiles are arguably the most common facial expression, textbook guidelines for animating smiles are minimal. Traditional blendshape interpolation functions such as cubic ease-in/ease-out may not reproduce the dynamics accurately. Common simplifications in blendshape systems, such as linearization, may alter the perceived genuineness of smiles. A model for smiles should be able to create realistic genuine smiles. Our goal is to (1) identify how linearizations in time or space impact smile genuineness and (2) model spontaneous smiles. These two goals are accomplished in Chapter 3, *Spatial and Temporal Linearities in Posed and Spontaneous Smiles*, and Chapter 4, *Modeling Smiles*, respectively.

Data capture. As in the other research in this thesis, we used a dataset of motion-captured expressions to construct models of smiles and assess whether those models are perceived differently from the simpler models in the literature and from ground-truth data. We collected high-resolution deformations by motion capturing smiles with dense marker placement. We elicited smiles with three types of stimuli: video, sharing jokes in conversation, and joke completion tasks similar to *The Cartoon Punchline Production Test* or *The Cartoon Caption Test* [36]. We annotated videos of the smiles with the start and end of each smile phase and with the triggering context: posed or spontaneous.

In Chapter 3, we investigate what properties of the smile expression need to be modeled to preserve genuineness. We explored how simplifications in space and time affect the perceived genuineness of smiles. We created realistic animations of spontaneous and posed smiles from high-resolution motion capture data for two computer-generated characters. The motion capture data was processed to linearize the spatial or temporal properties of the original animation to create stimuli that are approximations of the original, high-resolution, animation. The work in Chapter 3 was published in the journal *ACM Transactions on Applied Perception* 2014 [90].

Perceptual experiments to determine the impact of linearization. Through perceptual experiments, we evaluated the genuineness of smile animations, which is impacted by simplifications in both space and time. We also investigate the effect of head motion in genuineness perception and show that animations with and without head motion are impacted similarly by linearization. Additionally, the spontaneous smiles were more affected by temporal linearization than spatial linearization. Our results agree with and extend previous research on linearities in facial animation and allow us to conclude that a model of smiles must include a nonlinear model of velocities.

Data analysis. To account for the perceptual difference between linearized and nonlinearized animations in the perceptual experiments, we quantified the differences between posed and spontaneous smiles for the following characteristics: duration, spatial nonlinearity, and mouth corner vertex speed. This analysis provides a basis for future studies and highlights the variability between posed and spontaneous smiles: spontaneous smiles are more complex and nonlinear than posed smiles. Together with the perceptual results, the numerical analysis suggests that it is critical to use spontaneous rather than posed expressions in studies that quantify facial dynamics.

Models. In Chapter 4, we present a data-driven model for genuine smiles that preserves the temporal properties of smiles and is augmented by plausible head motions. Based on the previous perceptual experiment results, we model data-driven interpolation functions and plausible head motions. The smile model consists of interpolation functions generated from a PCA model and actor-specific blendshapes. We augment the model for facial deformations with correlated head motions as observed in the data.

Perceptual experiments to validate the model. The data-driven model produces more genuine animations than traditional blendshape-based approaches with ease-in/ease-out interpolation functions. We first used three sample animations each of (1) original motion-captured smiles, (2) model smiles with data-driven interpolation functions, and (3) smiles with ease-in/ease-out interpolation functions. The model smiles were rated to be as genuine as the original while the ease-in/ease-out smiles were significantly less genuine. Interestingly, this effect was observed only when animations were displayed with head motion. In a second experiment, we compared a larger number (twelve samples) of model smiles to their ease-in/ease-out counterparts for three characters (two photorealistic and one cartoon-like). The model animations, which were created from one participant's data, appeared more genuine than the ease-in/ease-out smiles for the two photorealistic characters.

1.5 Temporal connection between blinks and smiles

Facial expressions have been investigated primarily in terms of their spatial configuration. As a result, little is known about the relative timing of different facial motions. In the last part of the thesis (Chapter 5), we investigate where blinks are temporally located relative to smile start and end.

To determine the temporal relationship between blinks and smiles, we analyzed 43 videos of spontaneous smiles from the Cohn-Kanade Facial Expression database [50, 61]. The sequences were annotated with the start and end of the smile expression. To identify blinks in the video, we used Active Appearance Models [67], similar to the method used in *Modeling and Animating Eye Blinks* (Section 1.3). Using the temporal location of the smile and eye blinks, we computed the temporal distance between blinks and smile onsets and offsets.

Our data show that eye blinks are correlated with the end of a smile and occur close to its offset, but before the lip corners stop moving downwards. Furthermore, a marginally significant effect suggests that eye blinks are suppressed (less frequent) before smile onset. These results were published in the *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition* 2013 [91].

1.6 Organization

The approach described in this thesis aims to recover perceptually meaningful information from high-resolution data of blinks and smiles in order to generate realistic animations. The remainder of this thesis is organized as follows. In Chapter 2, we describe the details of modeling and animating eye blink dynamics. We present detailed results obtained using our framework. Then we consider animating genuine smiles in Chapters 3 and 4. In Chapter 5, we present the analysis of the connection between blinks and smiles. Finally, Chapter 6 discusses the contributions of this thesis and makes suggestions about future work.

Chapter 2

Modeling and Animating Eye Blinks

Facial animation requires laborious attention to detail because humans are attuned to subtle changes and anomalies in faces. In particular, eye motion, gaze, saccades, and blinks generally require significant artist input. Failure to properly animate eye motion may alter the intended emotional content of animated feature films. In this chapter, we focus on generating and evaluating parameters for eye blinks (Figure 2.1), one important component of facial animation.

Highly skilled animators can convey a wide range of emotions using subtle animation cues, including eye blink amplitudes and dynamics. Indeed, animation textbooks recommend using blinks as a way to "add life to a character" and to emphasize or influence personality and mood [63]. We hypothesize that the quality of an animation can suffer when incorrect assumptions are made about the dynamics of eye blinks. These simplifications include the directive that blinks are symmetric; therefore, the same number of frames should be used for opening and closing the eyelids [20, 63]. Additionally, it has been suggested that a linear or near-linear velocity profile for the eyelid motion is sufficient [20, 63, 95].

In practice, animation systems often make use of ease-in/ease-out animation profiles for motions, including eye blinks. However, as we present in the following sections, simple ease-in/ease-out motions do not accurately mimic human eyelid motion. Furthermore, 300 frames per second (fps) video makes it clear that there is non-negligible horizontal and vertical movement of the lower eyelid, as shown in Figure 2.2. This horizontal motion is not mentioned in the textbooks.

We challenge the common assumptions about animating eye blinks and show results indicating that observers distinguish and rate as more natural eye blink animations that are generated from actual human data. We propose using data-driven methods for inferring parameters in traditional facial animation techniques, such as blend shape animation. We use the Active Appearance Model (AAM) computer vision algorithm [67] to track unadorned eyes in high-speed video footage (Figure 2.1). The tracking information allows us to determine the accurate temporal and spatial dimensions of human blinks. We use a model based on principal component analysis (PCA) that can generate new blinks in the same space as the training data.

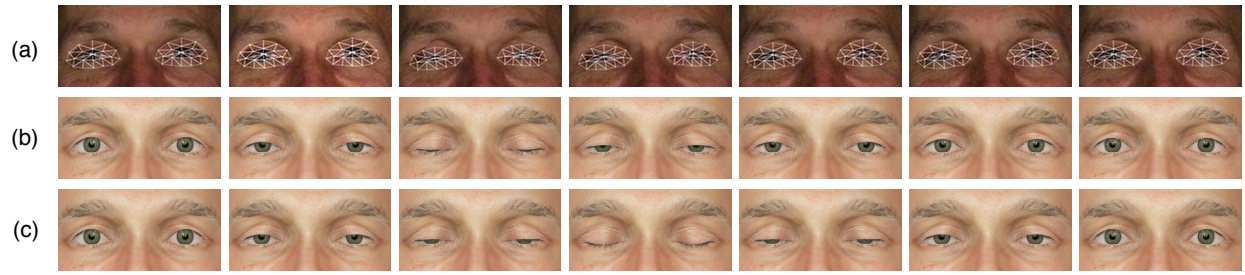


Figure 2.1: (a) High-speed video footage of human blinks was tracked with Active Appearance Models. (b) Based on the tracked data, a realistic model of human eye blinks was used to generate eye blink animations. (c) The symmetric blinks generated using common animation guidelines.

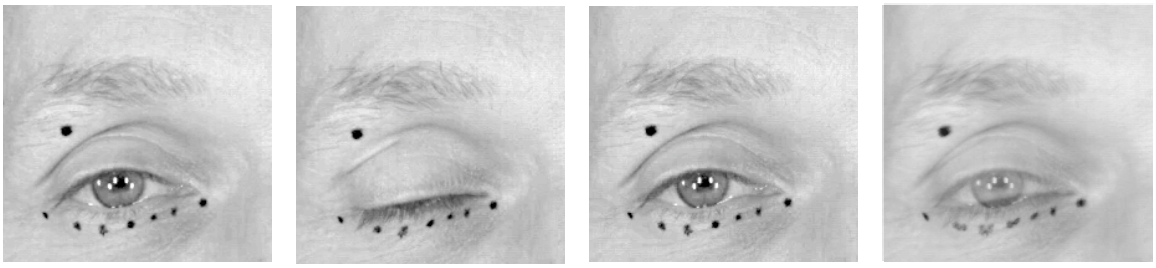


Figure 2.2: Images of the eye at (a) the beginning of the eye blink, (b) the maximum closed position, and (c) the end of the blink. (d) Overlaid images of the eye during blinking. The blurring of the markers on the lower eyelid demonstrates the displacement of the lower eyelid during the eye blink.

An extensive set of perceptual experiments shows the improvements in naturalness ratings that arise from the use of accurate eye blink motion. A set of representative eye blink profiles from video data are shown in Figure 2.3 while Figure 2.4 shows examples of eye blink profiles used in animation.

2.1 Related work

Eye movements are of interest to animators, computer graphics researchers, psychologists, and neuroscientists. In this section, we review three topics pertaining to eye motions and blinking: the physiology of human eye blinks, methods for measuring eyelid dynamics, and common animation methods.

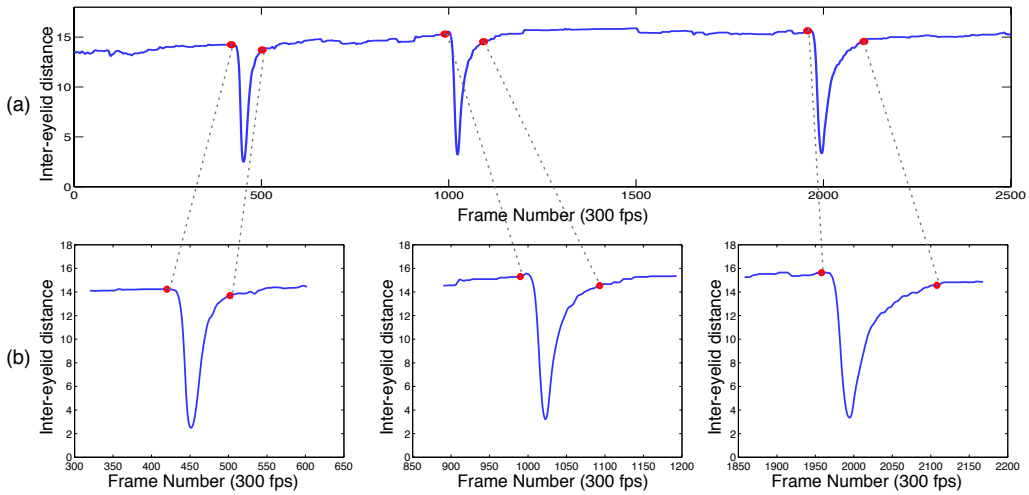


Figure 2.3: (a) A sequence (8.3 seconds) of inter-eyelid distance data. (b) Individual human eye blinks are characterized by a fast eyelid closing and a slower, asymptotically converging eyelid opening.

2.1.1 Physiology of eye blinks

Blinking is a natural eye motion defined as the rapid closing and opening of the eyelid [9]. Two antagonistic muscles are primarily responsible for generating a blink: the sphincter muscle, orbicularis oculi, closes the eyelids, and the levator palpebrae superiori muscle raises the upper lid [34]. Eye blinks can be put into three categories: spontaneous (unconsciously triggered), reflexive (elicited by a sudden impulse), and voluntary (intentionally triggered). These categories can be distinguished based on duration, amplitude, and context [5]. Throughout this study, we focus on naturally occurring, spontaneous blinks.

The dynamics of lid motion follow a highly asymmetrical motion pattern in time (Figure 2.3). The down phase, when the lid closes, is short in duration and achieves a high velocity with fast accelerations. The up phase lasts longer and decelerates more slowly. This pattern has been described by multiple research groups [35, 40, 84, 92].

The duration and variability of blinks have been recorded under various conditions, including voluntary and spontaneous blinks as well as those induced by air puffs and electrical stimuli. Note that accurately determining the end point of a blink is difficult because the eye opens slowly and the final inter-eyelid distance does not always return to the initial inter-eyelid distance in a brief amount of time. VanderWerf and colleagues [92] proposed defining the end of a blink as the instant when the inter-eyelid distance reaches 95% of the original value.

Changes in the speed, frequency, and strength of blinks provide information to the observer. For example, increased durations of eye closure and reopening are associated with drowsiness [11]. Blink rate is positively correlated with difficulty for some tasks, such as mental

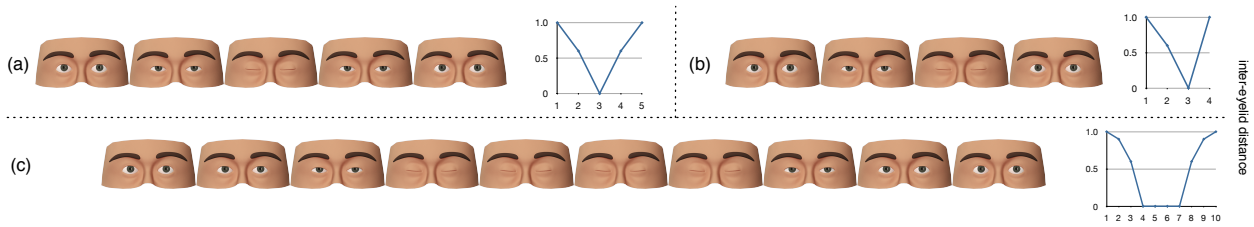


Figure 2.4: Illustrative figure of the animation timing suggested for a blink adapted from Maestri [1996]. (a) Temporally and spatially symmetric sequence that the author recommends for most situations. (b) A second, asymmetric, blink is used by the author to illustrate how to make the character look more alert. Note that the resulting dynamics are a reversal of human eye blink dynamics (as shown in Figure 2). (c) A long blink sequence that is fully symmetric and creates the appearance of sleepiness. None of these blinks accurately models spontaneous human eye blink motion.

arithmetic [86], and negatively correlated for others, including flight simulation tasks [93, 96]. Additionally, both blink frequency and blink duration have been associated with emotional states. For example, the amplitude of reflexive blinks was higher while viewing unpleasant pictures than during pleasant pictures [14].

Cues from blinking can also suggest whether or not a subject is telling the truth. Elevated blink rates are found in individuals who are masking their true emotions [78], and people show decreased blinking while lying, followed by increased blinking afterwards [55].

2.1.2 Quantifying eye blink dynamics

Evinger and colleagues [35] described two methods for measuring eyelid position: scleral search coils and electromyographic (EMG) recordings. The scleral search coils required insertion under anesthesia and a head restraint; therefore, they were used only in animals. For humans, EMG recordings were performed using electrodes pasted to the upper eyelid.

More recently, researchers have performed human eyelid recordings using the electromagnetic search coil technique, which involves positioning the participant in the center of a weak magnetic field, taping one or two coils to the eyelid, and recording changes in the current [34, 84, 92]. This method produces measures of blink amplitude, velocity, duration, rate, and time of occurrence. Moreover, it can measure the horizontal and vertical movements of both the upper and lower eyelid [92]. However, it is limited by the number of coils that can be used simultaneously, and participants might change their behavior because of the invasiveness of the procedure. Additionally, recordings must be done while the participant remains still inside the magnetic field. These limitations make it impractical for use in many settings, including those used for performance capture. Video is a rich source of information for measuring eye blink dynamics. For example, Bacivarov and colleagues [6] have shown, that active appearance models can be successfully used to detect eye blink events in video, and our approach is similar.

2.1.3 Eye blink animation

Few data-driven methods for eye blink animation exist. One of the seminal works on eye motion animation proposed a model for eye movements that used empirical models of saccades and statistical models of eye-tracking data, including when blinks are triggered [56]. However, their model provided no information regarding the dynamics of the eye blink.

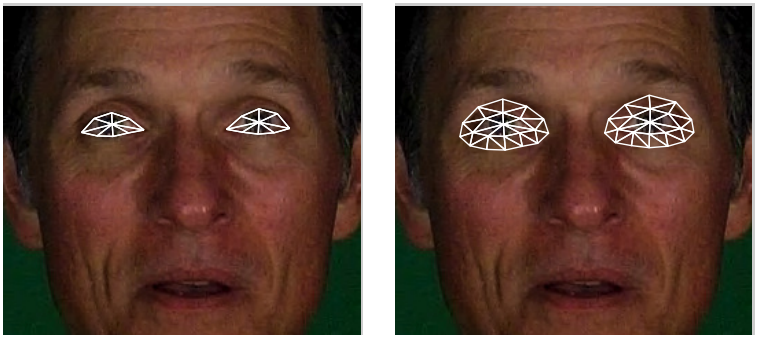
Deng and colleagues [23] proposed a texture synthesis-based technique to simultaneously generate realistic eye gaze and blink motion by modeling the correlation between eye gaze and blink motion. A video of one actor wearing face markers was tracked. Independent blink profiles generated with this motion appear similar to our data. However, their approach did not assess the benefits of accurate eye blink profiles. Steptoe and colleagues [82] investigated the kinematics of blinks and eyelid saccades. They used frames from video recordings (taken at 60 fps) of one individual during three blinks and three eyelid saccades to produce similar-looking animations with fully closed eyelids for the blinks and slightly closed eyelids for the saccades. These blink and saccade animations were then compared in a perceptual study to animations generated using the equations derived by Evinger and colleagues [34] and animations created using linear interpolation. Ten participants from their research group ranked the realism of exemplars of each type of animation using their memory of real blink dynamics and then ranked them based on their similarity to an eye movement from the source video. The animations generated from the video were ranked highest for realism and similarity to the source, followed by the clips created from the equations and then the sequences generated using linear interpolation. The authors assessed the joint effect of the blinks saccades separately. Only six stimuli (three blinks and three saccades) from each category were assessed, and they did not examine the full range of variability in blink duration and eyelid closing amplitudes seen in human eye blinks.

2.2 Approach

Unlike previous methods, the experimental framework we propose relies on high-resolution temporal and spatial measures of eye blink dynamics and we use perceptual experiments to validate the measured and modeled dynamic behaviors. Because we are quantifying both temporal and spatial characteristics, the results of this framework can easily replace or complement existing blink animation techniques. For example, traditional animation curves can be used to reproduce the eye blink dynamics that we have derived from data.

2.2.1 Active Appearance Model video analysis

Active Appearance Models are a non-rigid deformable tracking method that has been successfully used to track dynamic facial expressions [17, 67]. The model consists of two parts, a linear model of shape deformation and a linear model of shape-normalized appearance change, both of which are typically learned using PCA from labeled training data.



(a) (b)

Figure 2.5: (a) Basic eighteen point template used for video AAM labeling and tracking. (b) The basic template can be augmented to 42 points by adding border vertices.

Our AAM model tracks only the eyes of a given actor and is learned from a handful (20-25) of manually labeled images of each actor. Following the notation of Matthews and Baker [67], the shape, $\mathbf{s}$, of the AAM is the vector of vertices used to describe both eyes,

$$\mathbf{s} \;=\; (x_1, y_1, x_2, y_2, \ldots, x_v, y_v)^{\mathrm{T}}. \tag{2.1}$$

We label $v = 18$ 2D points in each training image as shown in Figure 2.5. The linear shape model is defined as,

$$\mathbf{s} \;=\; \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i, \tag{2.2}$$

where $\mathbf{s}_0$ is the mean shape, $\mathbf{s}_i$ are the shape PCA basis vectors and p_i are the shape parameters. The appearance model is similarly defined as,

$$A(\mathbf{x}) \;=\; A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) \qquad \forall\, \mathbf{x} \in \mathbf{s}_0, \tag{2.3}$$

where $A_0(\mathbf{x})$ is the mean appearance, $A_i(\mathbf{x})$ are the PCA appearance basis vectors and λ_i are the appearance parameters. We let $\mathbf{s}_0$ also denote the set of pixels $\mathbf{x} = (x, y)^{\mathrm{T}}$ that lie inside the base mesh $\mathbf{s}_0$. The appearance of an AAM is then an image $A(\mathbf{x})$ defined over the pixels $\mathbf{x} \in \mathbf{s}_0$. The original formulation of AAMs [17] included an additional PCA step to learn a single *coupled* parameterization of shape and appearance,

$$\mathbf{s} \;=\; \mathbf{s}_0 + \sum_{i=1}^{l} c_i w_p^{-1} \mathbf{s}_i \mathbf{j}_i^{\mathbf{s}} \;,\; A(\mathbf{x}) \;=\; A_0(\mathbf{x}) + \sum_{i=1}^{l} c_i A_i(\mathbf{x}) \mathbf{j}_i^{A(\mathbf{x})} \;. \tag{2.4}$$

The coupled parameters, $\mathbf{c} = (c_1, c_2, \ldots, c_l)^T$ are the parameter weights for the PCA basis of the concatenation of $\mathbf{p}$ and $\boldsymbol{\lambda}$,

$$\mathbf{c} = \begin{bmatrix} w_p \mathbf{p} \\ \boldsymbol{\lambda} \end{bmatrix} = USV^T = \mathbf{j} SV^T, \tag{2.5}$$

where $\mathbf{p} = (p_1, p_2, \ldots, p_n)^T$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T$, w_p is an energy normalizing weight, and $\mathbf{j} = (\mathbf{j}_1, \mathbf{j}_2, \ldots, \mathbf{j}_l)$ are the eigenvectors of the joint PCA basis. For applications where shape and appearance are highly correlated, such as eyelid tracking, the coupled parameterization can be significantly more compact, i.e., $l < m + n$.

For tracking faces, Matthews and Baker [67] do not use the coupled parameterization. The independent parameterization allowed them to introduce the fast, appearance *project out*, inverse-compositional, gradient descent fitting algorithm. However, for eyelid tracking, the coupled model has many advantages. For example, our model has $n = 8$ shape parameters (98% variance), $m = 10$ appearance parameters (90%), but only $l = 10$ coupled parameters (98%). Fewer parameters make the tracking faster and more reliable.

To accurately and efficiently fit a coupled-parameter AAM model to an image, we extend the *simultaneous* shape and appearance, inverse-compositional, gradient descent fitting algorithm described by Baker and colleagues [7],

$$\sum_{\mathbf{x}} \left[A_0(\mathbf{W}(\mathbf{x}; \Delta\mathbf{p})) + \sum_{i=1}^{m} (\lambda_i + \Delta\lambda_i) A_i(\mathbf{W}(\mathbf{x}; \Delta\mathbf{p})) - I(\mathbf{W}(\mathbf{x}; \mathbf{p})) \right]^2$$

where $\mathbf{W}(\mathbf{x}; \mathbf{p})$ denotes the piecewise affine warp over each triangle of the AAM mesh deformed by the shape parameters $\mathbf{p}$. We replace the independent shape and appearance parameterization with the coupled parameters and solve for the incremental warp update,

$$\Delta\mathbf{c} \;=\; -H^{-1} \sum_{\mathbf{x}} \mathbf{SD}^{\mathrm{T}}(\mathbf{x}) E(\mathbf{x}), \tag{2.6}$$

where the coupled-parameters *steepest descent images* are given by,

$$\begin{aligned} \mathbf{SD}(\mathbf{x}) \;=\; & [(\boldsymbol{\nabla} A_0 + \sum_{i=1}^{l} c_i \boldsymbol{\nabla} A_i \mathbf{j}_i^{\boldsymbol{\lambda}} \frac{\partial \mathbf{W}}{\partial c_1} \;,\; \ldots \;,\; (\boldsymbol{\nabla} A_0 + \sum_{i=1}^{l} c_i \boldsymbol{\nabla} A_i \mathbf{j}_i^{\boldsymbol{\lambda}}) \frac{\partial \mathbf{W}}{\partial c_n}] \\ + \; & [A_1(\mathbf{x}) \mathbf{j}_i^{A(\mathbf{x})},\; \ldots,\; A_m(\mathbf{x}) \mathbf{j}_i^{\boldsymbol{\lambda}}], \end{aligned}$$

where $\mathbf{j}_i^{\boldsymbol{\lambda}}$ denotes the subset of $\mathbf{j}_i$ that is modified by $\boldsymbol{\lambda}$ in (2.5),

$$H^{-1} \;=\; \sum_{\mathbf{x}} \mathbf{SD}^{\mathrm{T}}(\mathbf{x})\, \mathbf{SD}(\mathbf{x}), \tag{2.7}$$

is the Gauss-Newton approximation to the Hessian, and the coupled parameter AAM error function is

$$E(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{l} c_i A_i(\mathbf{x}) \mathbf{j}_i^{\boldsymbol{\lambda}} - I(\mathbf{W}(\mathbf{x}; \mathbf{c})). \tag{2.8}$$

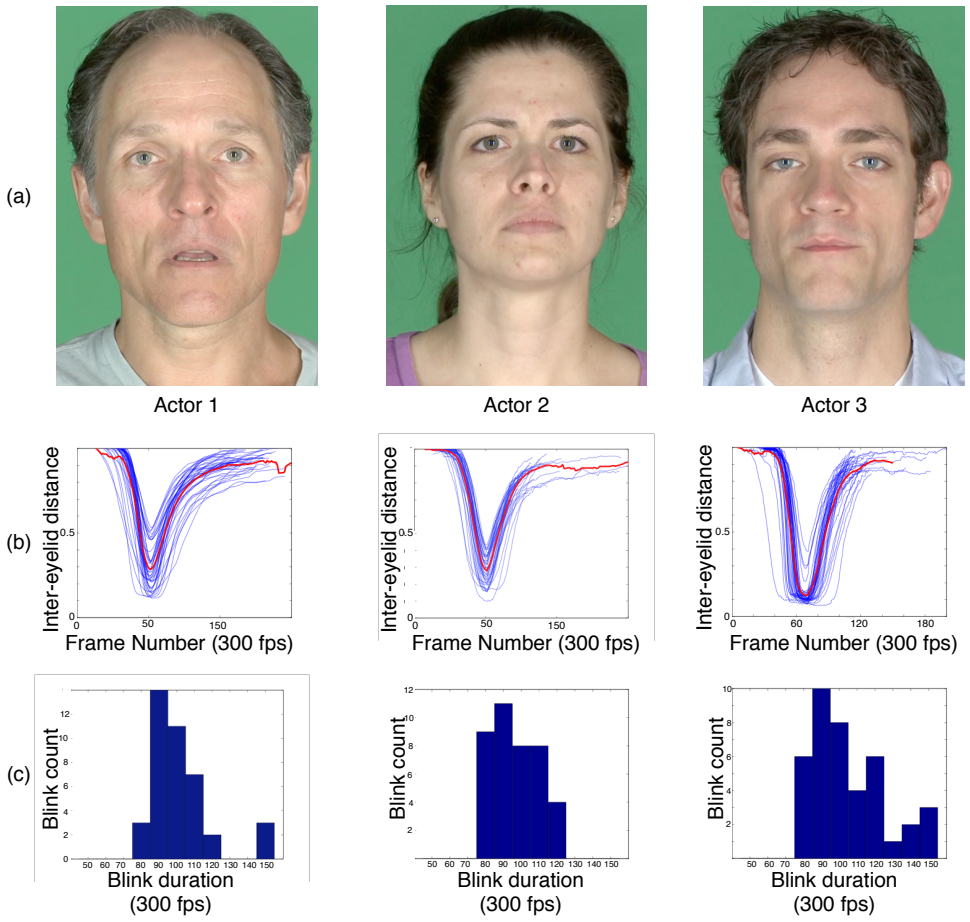


Figure 2.6: (a) Video data from three actors was used in this study. (b) For each actor, 40 eye blinks were aligned according to the minimum inter-eyelid distance to depict the variability in closing amplitude. The mean blink for each actor (red) was calculated by averaging the 40 blinks after alignment. (c) Histograms of blink durations.

2.2.2 Data processing

Three actors were recruited from the local community to record the video stimuli for this study (Figure 2.6). We recorded videos at 300 fps with a Casio Exilim FX1 camera. The actors were instructed to perform several two-minute vignettes, and video data was recorded during both the performances and the breaks in between.

Each video sequence at 300 fps was over 400,000 frames in duration. For each actor's sequence, a small number of frames were manually labeled with 18 points as shown in Figure 2.5 (a). The data was then tracked using the method described above. We automatically extend the eye model to include the additional border vertices shown in Figure 2.5 (b), resulting in a 42-point shape model. Because the data is 300 fps and does not change much between

frames, tracking is very reliable and fast. Our implementation runs at over 200 fps and is able to track an entire 400,000 frame sequence given an initial estimate for the first frame.

Blink frequency and inter-blink timing were quantified for the three actors shown in Figure 2.6. Blink frequency varied across the three actors: Actor 1 had an average blink rate of 8.2 blinks per minute, Actor 2 had an average blink rate of 27.0 blinks per minute, and Actor 3 had an average rate of 6.6 blinks per minute. Though all of the blink rates are within previously reported bounds [28], Actor 2 showed a relatively increased blink frequency, possibly due to wearing contact lenses. Further analysis can be conducted on the timing of eye blinks, however, the scope of this project is to investigate the dynamics of independent eye blinks.

2.2.3 Animated blink profiles

A basic model of an eye blink profile can be characterized as the inter-eyelid distance over time. For our perceptual experiments, we generate animations using six profile types. Two of the profiles (real and model) were created using tracked video data; the other four are based on traditional animation techniques (symmetric linear, asymmetric linear, ease-in/ease-out symmetric, and ease-in/ease-out asymmetric).

Real profiles

The inter-eyelid distance of Actor 1 was calculated from the AAM tracked video data (300fps) as the absolute distance between a horizontally centered upper eyelid marker and a horizontally centered lower eyelid marker. For the purpose of this study, we assumed that the motion of both eyes was identical. A zero-phase 10 frame filter was first applied to the signal. An automatic blink-labeling algorithm was used to identify the local minima in the inter-eyelid distance signal. The local minima correspond to periods when the eyes are closed, with a distance close to zero between the upper and lower eyelid. The algorithm then looked in the neighborhood of each minimum point (300 frames to the left and 300 frames to the right of the minima) for the beginning and end of the eye blink based on two criteria:

1. The gradient of the inter-eyelid distance is small. The gradient is measured as the change in consecutive values of the inter-eyelid distance and is averaged over a window of 10 frames. Small values ($< .1$) of the averaged gradient ensure that, over the given window, the inter-eyelid distance is constant, corresponding to little to no vertical motion in the eyelids.
2. The inter-eyelid distance value is above the baseline value computed for each blink. The baseline value is defined as the average of the inter-eyelid distance over 100 frames left and right of the minima. Though eye blinks are shorter than 200 frames, the baseline value

quantifies the average inter-eyelid distance over a window that contains the opening or closing of the eye. In this way, we ensure that the frames in the immediate vicinity of the eye-closed frame, which may have a small gradient but also have a value smaller than the baseline, are not labeled as the start or end of the blink.

To find the beginning of an eye blink we start at a local minima and search backward. The first point located before the minima that matches the two criteria listed above corresponds to the instant the eyes begin to close. Similarly, by starting at the minima and searching forward we can find the end of the eye blink (corresponding to the eye open position). Examples of the resulting start and end eye blink labels can be seen in Figure 2.3.

We supplemented the automatic blink annotation with visual inspection to remove false positives from the data (three segments) and verify the accuracy of the blink-labeling algorithm. Furthermore, observations of the video and audio content led us to recognize that some blinks were used as non-verbal communication. For several eye-closing points, the action was intended to represent a particular emotion, such as pride. These eye-closing intervals were removed from the data set to ensure that only spontaneous blinks were used.

The eye blink time series that were used in the subsequent experiments consist of 49 blinks from one male actor, collected over a six-minute period. To determine whether the blinks from this subject were idiosyncratic, we analyzed the blinks from two young adults, one male and one female. The blink profiles for all three actors are similar in both general shape, blink duration distribution, and closing amplitude. Figure 2.6 shows the histograms of blink duration and variability in closing amplitudes for the three actors.

Model profiles

The blink sequences from Actor 1 were used to build a model that can generate different blink durations. For each of the 49 blinks, two splines, one each for eye closing and eye opening, are fit to the inter-eyelid position data at 300 fps. The two splines allow the blink time series to be normalized for duration and aligned such that the minimum inter-eyelid distance always occurs at the same frame (Figure 2.7). To create the numerical entries that represent the blink time series, we sample each spline uniformly. Using the normalized blink time series we construct a matrix that contains on each row an independent eye blink. This matrix is augmented with the scaling factors that were used for the two blink parts, added as the last two columns entries for each blink. We then represent the data matrix (normalized time series and scaling coefficients) in PCA form as a weighted sum of eigenvectors.

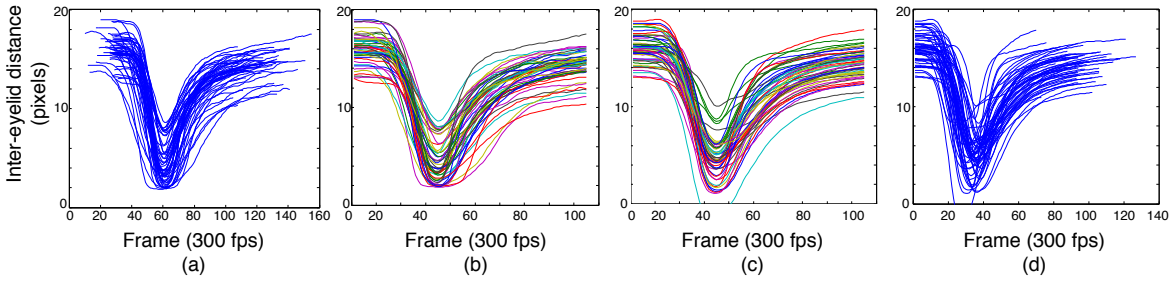


Figure 2.7: PCA-based eye blink model. (a) The blink model takes as input blink profiles derived from video. (b) The blinks are then normalized to the average blink duration independently for opening and closing. (c) The PCA-based model then generates duration-normalized blinks and scaling factors that are used to (d) generate blink profiles with the appropriate duration.

To generate new data, we use the first five dimensions (representing 98% variance) of the PCA model. For the five dimensions, we project new, random coefficients that are within one standard deviation of the original PCA coefficients. Animated blinks generated using this profile were used in all three experiments.

Linear profiles

The linear blink profiles are generated by linearly interpolating between the inter-eyelid position at the beginning of the blink to the maximally closed position and from there to the end of the eye blink (Figure 2.8 a). Two types of linear profiles can be generated: symmetric and asymmetric. The minimum inter-eyelid distance falls in the center frame for the symmetric linear profiles and in an off-center frame for the asymmetric linear profiles. These simple profiles of blink dynamics have been described in textbooks such as *Digital Character Animation* by George Maestri [63].

Ease-in/ease-out profiles

Ease-in/ease-out (Figure 2.8) curves are often used in animation as a way of portraying motion accelerations and decelerations. Motion with ease-in/ease-out is thought to convey a greater sense of realism because it obeys physical laws for inertia. The generated ease-in/ease-out profiles are Bezier curves with control points proportional to the duration of the two parts of the blink.

2.3 Perceptual studies

We conducted three perceptual studies to investigate the perceived naturalness of animated blinks. The animated clips were displayed on a 23-inch LCD display at actual size, where the inter-ocular distance equals 6 cm and the maximum inter-eyelid distance equals approximately 1.1 cm. For all experiments, only the eye region was visible (see Figure 2.1).

Trials were self-initiated: the participants pressed a key to play each clip. Before every experiment, there was a brief practice session to familiarize the participants with the procedures. The experimental procedure was approved by the Carnegie Mellon University Institutional Review Board, and participants were compensated for their time.

2.3.1 Experiment 1: Comparing eye blinks created with tracked data to our profile generation method

Experiment 1 confirmed the validity of the PCA-based model by comparing the perceived naturalness of model-generated eye blinks to the perceived naturalness of those that were animated based on the data obtained from the video sequences. In this experiment, we also investigated the effects of motion blur in rendering and the effect of closing amplitude. Eye blink naturalness ratings were collected for two 3D animated characters, one photorealistic, one cartoon-like.

We hypothesized that the animated blinks with motion blur would be rated as more natural than animated blinks without motion blur. Furthermore, we expected that the difference in naturalness ratings between fully closed and naturally closed eye blinks would be insignificant if rendered with the correct motion blur profile.

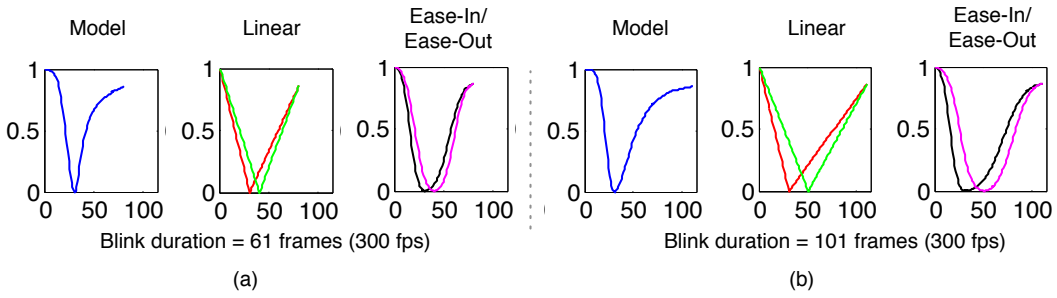


Figure 2.8: Eye blink profiles: model (blue), symmetric linear (green), asymmetric linear (red), symmetric ease-in/ease-out (purple) and asymmetric ease-in/ease-out (black). Two different blink durations (a) 61 frames and (b) 101 frames are shown at 300 fps.

Methods

Thirty-two adult participants completed this experiment. The participants were recruited from an online participant pool. Each participant watched 320 clips of blinks cropped to cover only the eye region. After each animation, participants were asked to rate the naturalness of the animation on a 7-point rating scale (1 = very unnatural, 7 = very natural). A very natural clip was defined as something you would expect to see in the real world. We elected to use a rating scale rather than a forced-choice preference paradigm in order to assess as many clips as possible in the allotted time. Each session lasted approximately 40 minutes.

Experiment 1 also explored the effect of rendering accurate motion-blur profiles and closing amplitudes on the perceived naturalness of eye blinks. In many blink animations, the eyes are fully closed in order to produce a maximum blink amplitude. However, the video data shows large variability in the amplitude of the blinks, with as many as 50% of the observed eye blinks not reaching the fully closed position (quantified as an inter-eyelid distance of zero). Twenty eye blinks were randomly selected from the set of real blinks. Similarly, twenty eye blinks were selected for duration and closing amplitude from a set of 100 eye blinks randomly generated with the PCA model. As a result of the linear model, a small number of generated profiles had negative values and were eliminated from the selection pool. For the fully closed condition (FC), the inter-eyelid distance was normalized to the maximum of the signal, while for the naturally closed (NC) condition, the inter-eyelid distance was normalized by the difference between the maximum and the minimum inter-eyelid distance. Both conditions were animated and then rendered at 300 fps.

To generate the motion blur (MB) effect, the final target frames were obtained by averaging over a window of ten frames centered on the current frame. For the no motion blur condition (NB), the 300 fps blink signal was down sampled to 30 fps by selecting every tenth frame.

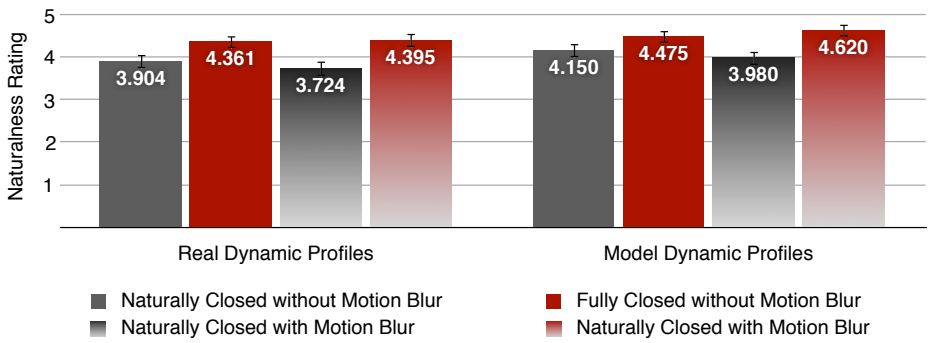


Figure 2.9: Experiment 1 results. Naturalness ratings for Real and Model dynamic profiles with or without motion blur.

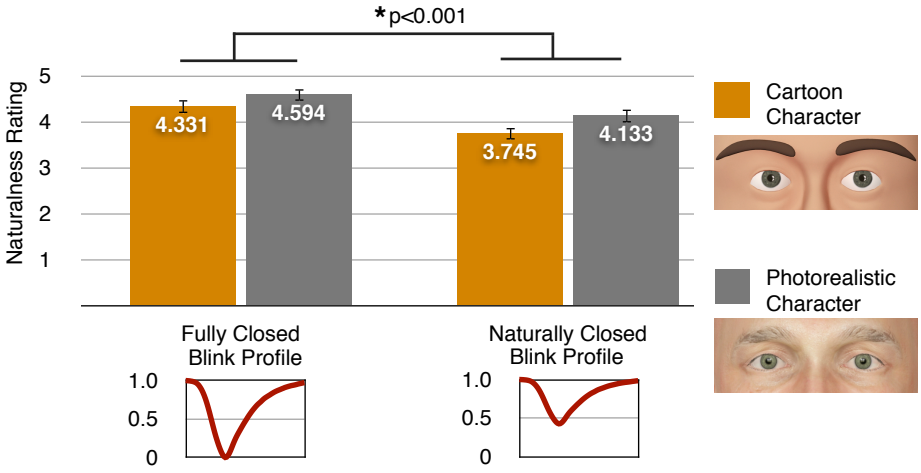


Figure 2.10: Experiment 1 results. The effect of closing amplitude on naturalness ratings for the cartoon and photorealistic characters.

In order to investigate the possibility that the complexity of the motion profile should correlate to that of the character as previously proposed [54], we animated two characters. A professional artist created two 3D computer-generated character heads in Maya (Autodesk). The first was a photorealistic character that was made using photographs and videos of Actor 1 for reference and texture information. The second character was created in a simple cartoon style. For the eye blinks based on tracked data, the Maya blend-shape weighting parameters were based directly on tracked data that was scaled to match the [0 1] interval. The stimuli for the two characters were presented in alternating blocks (photorealistic stimuli block and cartoon stimuli block) with the order for the two blocks randomized.

Results

We performed a 2 (character) $\times$ 2 (profile) $\times$ 2 (motion blur) $\times$ 2 (closing amplitude) repeated-measures ANOVA with Bonferroni corrections for multiple comparisons and a significance threshold of $p = .05$. There were higher naturalness ratings for the photorealistic character than for the cartoon character, $F(1, 31) = 5.91$, $p = .02$. There was a significant main effect of profile such that the model blinks were rated as more natural than the real blinks, $F(1, 31) = 39.07$, $p < .001$ (Figure 2.9). There was no main effect of motion blur, $F(1, 31) = .95$, $p = .34$. A significant main effect for amplitude was found, such that the fully closed blinks received higher naturalness ratings than the naturally closed blinks, $F(1, 31) = 15.55$, $p < .001$ (Figure 2.10).

All potential interactions were examined, and two were found to be significant. There was a significant interaction between blur and amplitude, $F(1, 31) = 20.35$, $p < .001$, such that the fully closed blinks received higher ratings with motion blur than without motion blur, but the reverse was true for the naturally closed blinks. There was also a significant interaction among character, blur, and amplitude, $F(1, 31) = 9.93$, $p = .004$, suggesting that the ratings for the naturally closed blinks on the cartoon character without motion blur received lower ratings than with motion blur while on the photorealistic character, naturally closed blinks without motion blur were rated higher than with motion blur. However, motion blur increased the ratings for both cartoon and photorealistic characters in the fully closed condition.

2.3.2 Experiment 2: Comparing model eye blinks to traditional methods

In this experiment, we contrast naturalness ratings for the animated eye blinks generated from our model with those created using traditional animation techniques, including ease-in/ease-out, asymmetric linear, and symmetric linear methods. Additionally, we examined the effect of blink duration.

Methods

Forty-three adult participants completed this experiment. For each of 400 trials, the participants watched a clip and rated the naturalness of the animation on a 7-point rating scale (1 = very unnatural, 7 = very natural). Each session lasted approximately half an hour.

Clips were created in a 5 (dynamic profile) $\times$ 4 (duration) $\times$ 2 (character) design. Five categories of blink profiles were included for comparison with our model: symmetric linear (SL), asymmetric linear (AL), an asymmetric ease-in/ease-out (AL Ease), and a symmetric ease-in/ease-out (SL Ease) as described in section 4. For each category of blinks, five exemplars each were generated for four durations: 7, 9, 11, and 13 frames (for videos shown at 30 fps). In all cases, the eyes fully closed and the motion blur parameters described in Experiment 1 were used. This process resulted in 100 unique blink animations that were then rendered with the Photorealistic and Cartoon characters. The animations for each character were kept in separate blocks. Each block of 100 clips was shown twice during the experiment, for a total of 400 trials over the course of four blocks.

Results

A 5 (profile) $\times$ 4 (duration) $\times$ 2 (character) repeated-measures ANOVA was performed with Bonferroni corrections for multiple comparisons and a significant level of $p = .05$. We found a significant main effect of profile, $F(4, 39) = 11.52$, $p < .001$, such that the highest ratings

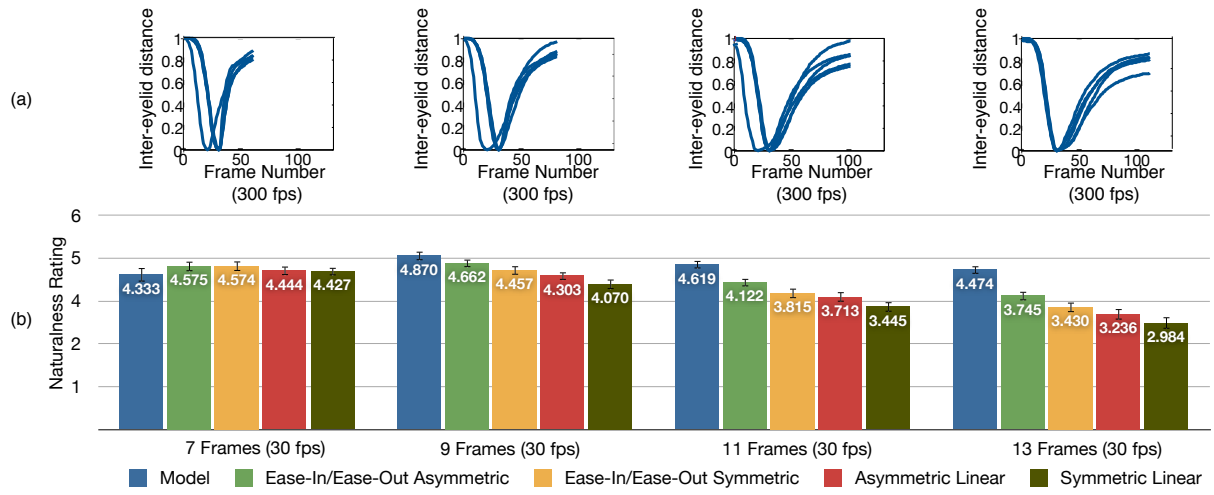


Figure 2.11: Experiment 2 conditions and results. (a) Model dynamic profiles for the four different durations used in Experiment 2. (b) Average naturalness ratings for the five dynamic profile conditions according to duration.

were given to blinks generated with our profile, followed by the AL Ease condition, the SL Ease condition, the AL condition, and finally the SL condition.

All pairwise comparisons between different dynamic profiles were significant at $p < .05$ (Bonferroni corrected). Additionally, there was a significant main effect for blink duration, $F(3, 40) = 23.40$, $p < .001$, such that naturalness ratings were highest for 9 followed by 7, 11, and 13 frames. Pairwise comparisons showed no significant difference between blinks of duration 7 and 9 frames. However, all other pairwise comparisons were significant with $p < .05$. The main effect of character was not significant, $F(1, 42) = .50$, $p = .48$.

The interaction among character, profile, and duration was not significant, $F(12, 31) = .51$, $p = .89$. Overall, the highest average naturalness rating was given for the photorealistic character with the model eye blink profiles and a 9-frame blink duration (mean 4.95). Interestingly, 9 frames was the dominant blink duration in the distribution of blink durations from our tracked data, as shown in Figure 2.6.

All pairwise interactions were significant, including between character and profile, $F(4, 39) = 3.95$, $p = .009$; character and duration, $F(3, 40) = 5.65$, $p = .003$; and profile and duration, $F(12, 31) = 4.21$, $p < .001$. Although there was a significant character by profile interaction, the naturalness ratings followed the same rank order for both characters such that the ratings for the model blinks were the highest, followed by asymmetric ease-in/ease-out, symmetric ease-in/ease-out, asymmetric linear, and symmetric linear.

For the cartoon character, the preferred duration was seven frames after collapsing across dynamic profiles (mean 4.50) while for the photorealistic character the highest rating is at a nine frame duration (mean 4.56). This result suggests that appearance plays an important role in determining what is perceived to be natural motion. In fact, according to the type of animation, in some cases it may be desirable for cartoon-like characters to preserve the less natural dynamics in order to emphasize their non-human characteristics.

The interaction between profile and duration is potentially due to an effect of profile on perceived blink duration. Participants gave low ratings ratings to the thirteen frame symmetric linear blinks. Some participants mentioned in informal conversation that those blinks seemed longest. Future work could examine whether perceived duration is affected by profile and, in turn, whether these combined factors affect naturalness ratings.

2.3.3 Experiment 3: Lower eyelid motion contribution

The third study examined the contribution of lower eyelid movements, including horizontal and vertical motion, to perceptions of naturalness.

Methods

Twenty adult participants completed this experiment. Participants rated the naturalness of 160 clips on a 7-point rating scale (1 = very unnatural, 7 = very natural). Sessions lasted 20 minutes.

Twenty samples for each of four types of clips were created. In the first type, the lower eyelid moved accurately horizontally and vertically (Both). The other conditions included vertical motion only (Vertical), horizontal motion only (Horizontal), and no lower lid motion (None). The horizontal and vertical motion parameters were those originally calculated from our model. All clips had motion blur, and the blink durations followed the natural distribution (Figure 2.6). All clips were rendered on the Photorealistic character because the Cartoon was not capable of horizontal lower lid motion. Participants saw two randomly ordered blocks, each containing all of the possible combinations of lower eyelid motion, resulting in two viewings of each clip.

Results

We performed a 4 (lower eyelid motion types) $\times$ 1 repeated-measures ANOVA with Bonferroni corrections for multiple comparisons. There was no significant effect of the various types of lower eyelid motion on perceived naturalness, $F(3, 17) = 1.28$, $p = .31$. The result suggests that viewers are not sensitive to this motion on our characters at the presented scale.

2.4 Discussion

Our perceptual studies underline the importance of using profiles of blinks that are based on a physiologically valid model rather than those created using simple animation curves. In Experiment 1, viewers rated the blinks animated using our model as more natural than those animated using measured blink profiles. We hypothesize that this is because the PCA model blinks were closer to the prototypical mean blink from the training set. Participants found fully closed blinks to be more natural than naturally closed blinks. This result may have arisen because viewers were unaware that the eyelids often fail to close fully during blinks and/or because this phenomenon is more obvious in animated characters. We also suspect that the effect may be a result of observing blinks independently rather than in a sequence. In future studies, we intend to investigate the naturalness of blink sequences. As expected, the participants found the blinks of the photorealistic character to be more natural than those of the cartoon character.

In our experiments, the blink models were derived from one actor's data (Figure 2.6: Actor 1). Our intention was to closely match the photorealistic character to the actor's features. Because the dynamic eye blink profiles across three different actors shown in Figure 2.6 appear to be similar, it seems reasonable to postulate that the results generalize. As an extra validation in future studies, we intend to compare eye blinks derived from different actors. In particular, we are interested in age correlates and the percentage of fully closed eye blinks in an actor's dataset. Another avenue that should be explored in future experiments is left/right eye blink symmetry and the correlation with naturalness ratings. Our system is capable of tracking both eyes simultaneously. In most situations, the left and right eyelid dynamics are highly correlated and we created our models and animations using data from only one eye in this study. However, we noticed occasional asymmetries in our video recordings that we would like to examine further.

The second experiment compared blinks generated from our model to those generated from simple profiles, including asymmetric linear profiles, symmetric linear profiles, asymmetric profiles with ease-in/ease-out curves, and symmetric profiles with ease-in/ease-out curves (Figure 2.8). The model blinks were rated as significantly more natural than the simple approximations. Viewers also compared these blinks at various durations. The highest naturalness ratings were assigned to blinks generated from our model that were nine frames in duration. Experiment 3 results suggest that there is no effect of the various possible techniques for lower eyelid motion on our 3D characters; this knowledge is particularly beneficial for animators using simple models that do not have the capability for this type of movement.

From a methodological perspective, this project introduces a new technique for detecting and describing blinks in video recordings. High-resolution spatial and temporal information were collected with a high-speed video camera. Then, AAMs were used to track the eyes and measure the blink profiles in long videos. The inter-eyelid distance during blinks was used to construct a PCA-based model that can generate eye blink dynamic profiles with properties (duration, closing amplitude) similar to the original distribution. Once the PCA-based blink model has been created, it can generate a variety of physiologically valid blink profiles for use in animation.

Our model is easy to include in an animation pipeline. For example, a script can be used to generate a variety of blinks for a character. Animation curves can then be driven from these dynamics. Technical Directors can create a blink button and seamlessly use this system without increasing the time required for animation. Additionally, the model can create variations in the blink sequences, an important component in making animations as true-to-life as possible. A limitation of our studies is that participants gave naturalness ratings of individual eye blinks in isolation: no facial context was given, and no inter-blink timing was considered. In the future, we intend to generate full facial animations that show sequences of eye blinks.

Future work will require improving our facial animation system to allow for small, subtle changes in the face (such as breathing) to match observations of faces in video. Additionally, we also intend to utilize the framework presented above to investigate blink frequency patterns and their correlations to head motion and facial expressions.

Chapter 3

Perception of Spatial and Temporal Nonlinearities in Genuine Smiles

Smiles vary dramatically in terms of geometric appearance and dynamics. People use this variation to convey subtle nuances of emotion and expression. Over eighteen labels, including polite, amused, embarrassed, and fearful, have been used to describe smiles and how viewers perceive and interpret them [3, 16, 30]. One important characteristic of smiles is genuineness. Genuine smiles are recognized as expressing positive emotions across cultures [33]. In contrast, non-genuine or posed smiles are intended to mask true emotions. In this chapter, we investigate how animation techniques affect the perceived genuineness of smiles. In particular, we explore the impact of motion linearization in time and space.

We investigate linearizations in time and space because cues in both of these domains affect how genuine a smile in perceived. For example, a genuine smile is often accompanied by a spatial cue: the slight eye wrinkling of the eye corner [31]. However, faking this cue can be more easily detected in video sequences than in static images, which implies that the timing of the cue plays a role in how the smile is perceived [53]. In Section 3.1, we discuss further the perceptual and quantitative characteristics of genuine smiles.

In animation, facial expressions, including smiles, can be created by specifying the deformations of the face, represented as a vertex mesh, over time. Thus, a smile has two components: (i) the spatial, or the geometric path of the vertices; and (ii) the temporal, or the rate of change in the vertex position (vertex speed). High-resolution motion capture data shows that these deformations are complex and nonlinear in both space and time. The geometric path of the vertices is nonlinear and the vertex speed is not constant (Figure 3.1).

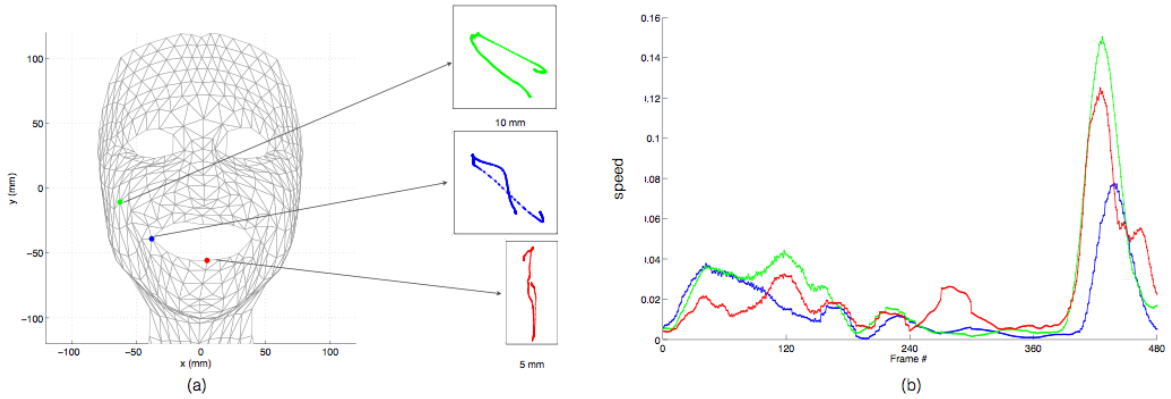


Figure 3.1: Spatial and temporal nonlinearities during a spontaneous smile. (a) Spatial nonlinearities represented by the nonlinear geometric paths of three vertices during the smile. (b) Temporal nonlinearities illustrated by the speed of the three vertices during a 4-second smile (480 frames). The smile is not symmetric: the speed profiles at the start (frames 1 to 180) and end (frames 300 to 480) of the smile are different.

Common animation techniques, which rely on keyframing, approximate the temporal or spatial properties of a smile. Craft books often describe smiles in terms of combinations of basic blendshapes [49]. The simplest model for a smile animation consists of two blendshapes, a neutral pose and a peak smile pose, and an interpolation function which operates globally on all vertices. The interpolation between two blendshapes results in a linear geometric path of the vertices, while the speed of vertices is determined by the interpolation function. The choice for the interpolation function has important consequences. If a linear interpolation function is used, the motion may look mechanical and unrealistic because of constant vertex speed. As in the case with eye blinks (Chapter 2 and [89]), the two-blendshape model with a linear interpolation function may be perceived differently than animations with data-driven, or nonlinear, interpolation functions.

We evaluate the perceptual benefit of preserving data-derived motion characteristics (geometric path, interpolation function) for realistic smile animation. Our perceptual results show how approximations in the temporal or spatial characteristics of the data affect the genuineness of a smile expression (Figure 3.2). We compare smiles animated with motion capture with smiles in which the geometric path of the vertices is linear or the interpolation function is linear. We find that linearizations lead to smiles being perceived as less genuine. We find similar results for animations with or without head motion. We contribute to previous results in the literature by disentangling the effects of spatial properties (geometric path) and temporal properties (interpolation function). Furthermore, animators will benefit from knowing how to avoid creating fake smiles, or, conversely, knowing exactly what parameters cause a smile to look posed.

Figure 3.2: (a) A spontaneous and (b) a posed smile animated from motion capture data. The posed smile is rated as significantly less genuine than the spontaneous smile. Linearizations in time and space reduce the perceived genuineness of the spontaneous smile to the level of the posed smile.

3.1 Related work

In this section, we discuss research related to smile perception and linearities in facial animation. We cover perceptual research on how genuine smiles are recognized and animation research linearizing expressions.

Smile genuineness is often associated with a slight wrinkling on the outer corner of the eyes known as the *Duchenne marker* [31]. In a more recent study, however, Krumhuber and Manstead [53] discovered that over eighty percent of participants could pose in photographs with smiles that included the Duchenne marker. Furthermore, when viewing static pictures of smiles, volunteers similarly rated both posed and spontaneous smiles as genuine. Conversely, participants recognized posed smiles more often in videos, which may indicate that the timing of different facial actions is relevant.

Multiple studies have analyzed the temporal properties of spontaneous and posed smiles. For example, spontaneous smiles have been found to have smaller amplitude and slower onset than posed smiles [16, 80]. When examining computer-generated smiles, Krumhuber and Kappas [52] found that perceived smile genuineness increased as a function of onset and offset durations and decreased as a function of apex duration.

Temporal and geometric cues also affect the perceived meaning of smiles in more subtle ways. Ambadar and colleagues [3] analyzed and annotated short movies of smile sequences

using the Facial Action Coding System [32]. The authors characterized them as amused, embarrassed, nervous, polite, or other. Smile categories were differentiated by temporal cues, including duration, onset velocity, offset velocity, asymmetry of velocity, and head movement. For example, amused smiles had larger maximum velocities and longer durations than polite smiles.

Several studies have begun to establish the importance of nonlinear motion for facial animation. For example, nonlinear temporal and geometric motion can affect the accuracy of emotion recognition [94]. Wallraven [94] created animations for seven posed expressions, including happiness, using either ground-truth data or linear interpolation between two blendshapes (peak and neutral). Linear interpolation in this case created animations with both a linear geometric path and linear timing (constant speed). Viewers were better able to recognize emotions conveyed in the animations created with the ground-truth data than in those with linear interpolation.

Cosker and colleagues [18] similarly showed that the originally recorded motion of short facial movements is preferred to linearly interpolated motions. The researchers captured posed expressions using dynamic 3D scanning. Then, participants viewed animations made from the non-geometric recorded data as well as animations created using linear geometric movement between data-recorded blendshapes. Viewers generally preferred the nonlinear geometric movement and rated it as more natural than the linear movement; however, this result did not hold for posed smiles [18].

Liu and colleagues [58] considered spatial nonlinearities for a on the face. They compared linear and nonlinear geometric paths for these points and determined that nonlinear geometric paths were rated as more realistic. The authors proposed an optimization scheme to explore the nonlinear relationship between spatial path and blendshape animation. With their proposed method they analyzed the motion of two points, center of the chin and eyelid. The spatial path of these points is in the sagittal plane. In our work we investigate spatial nonlinearities for vertices densely sampled on the entire face with deformations that occur in all three planes.

Our study contributes to the existing work in several ways. We consider the effect of linearization for both posed and spontaneous smiles, whereas previous research has focused on posed expressions. We expect that spontaneous expressions have greater nonlinear motion and will therefore be more heavily impacted by linearization. Furthermore, animations with a nonlinear geometric path while the timing information is linear have not been previously investigated. We investigate linearizations in both space and time and we hypothesize that these linearizations will decrease the perceived smile genuineness.

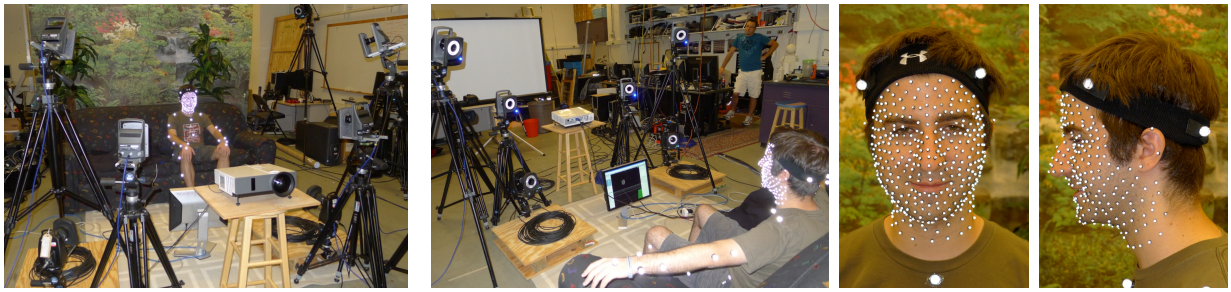


Figure 3.3: Setup for recording smiles: facial motion was recorded with a commercial motion capture system that tracks the position of 250 3 mm markers on the face.

3.2 Approach

In this section, we discuss our approach to capturing and processing our high-resolution dataset. We also describe the animation process.

3.2.1 Performance capture

We recorded over 100 smile sequences from two participants (one male, one female) during three-hour recording sessions. To elicit smiles, we asked participants to (a) view amusing videos, (b) rate one-panel comic captions, and (c) smile according to the experimenter's instructions.

Facial expressions were captured with an 18-camera Vicon system by recording the 3D positions of markers at 120 frames per second (fps); torso and head motion were also recorded. Participants wore 250 reflective markers spaced approximately 1 cm apart on the face, as shown in Figure 3.3. We applied the reflective markers in a similar pattern for both participants. However, due to differences in their facial geometry, the marker positions were not identical nor in direct correspondence. In the following section, we describe how the raw motion capture data was processed to obtain 3D meshes deforming over time, with the vertices in correspondence, for the two participants.

3.2.2 Data processing

The motion capture system records the 3D position of markers on a frame-by-frame basis. Because of marker density, the system does not generate consistent marker labels throughout a sequence. Furthermore, physical marker positions and distribution vary across participants. The goal of the data processing step is to obtain smile sequences with the markers for each participant in direct correspondence. To achieve this goal, we first cleaned all the motion capture data so that all markers were present in each frame. We then standardized the meshes for all sequences.

To clean the high-resolution motion capture data, we used the method and semiautomatic tool developed by Akhter and colleagues [2]. Their algorithm uses a bilinear spatiotemporal data representation and Expectation Maximization to simultaneously label, de-noise, and compute missing points in motion capture data.

The configuration of facial markers differed significantly between participants. Additionally, during the session, some markers were displaced from their positions at the start of recording. Following the approach of Tena and colleagues [88], we fit a dense 3D generic mesh template with more than 8000 vertices to our entire motion capture database. The mesh was subsampled to a limited number of vertices (approximately 400), which resulted in 3D meshes in full direct dense correspondence. Our goal was to analyze and animate facial expressions independently from head motion. For each sequence, rigid-body transformations, such as head motion, were removed by aligning each motion capture frame to the subsampled generic mesh template using ordinary procrustes analysis [29].

3.2.3 Original smile sequences

We used twelve basic smile expressions from two participants (six from each participant). To select the smile samples, we first ordered the smile videos for each participant according to duration. Each video contained at least one smile as determined by visual inspection. Expressions of fewer than ten seconds were selected for further annotation. From these short smile videos, we picked three spontaneous smile sequences and three posed smile sequences such that the smiles started and ended in a neutral expression. The start, end, and peak frames of the smile expression were identified based on the velocity of the vertices on the cheeks.

We evaluated the genuineness of the twelve selected smiles in a brief experiment on *Amazon Mechanical Turk*. Thirty participants rated each smile video twice on a scale from 0 (not genuine) to 100 (genuine). We described a genuine smile to the participants as a smile that someone shows when she/he is joyful, happy, or amused. The videos were recorded during the motion capture session and thus the actors KB and SD were wearing motion capture markers similar to the actor shown in Figure 3.3.

As expected, posed smiles (average rating = 27) were rated as less genuine than spontaneous smiles (average rating = 70). The ratings for the two characters are shown in Figure 3.4. We suspect that the ratings for spontaneous smiles do not reach an average closer to 100 because the video sequences selected are relatively simple. Based on preliminary testing of other video sequences, we posit that higher ratings of genuineness are generally associated with sequences that more closely resemble laughter than a simple simple.

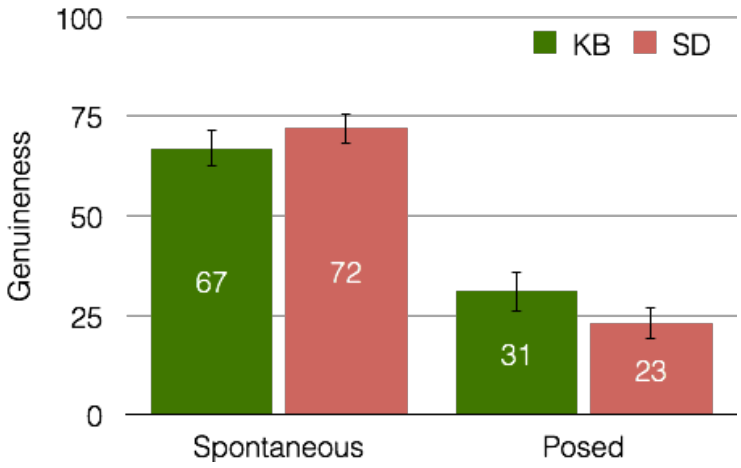


Figure 3.4: Average genuineness ratings for the smile videos selected for animation. Three smiles for each actor for each category (posed or spontaneous) were rated by thirty participants. KB is the male actor participant and SD the female actor participant. The values are plotted with standard error bars.

3.2.4 Animation

A professional artist created virtual replicas (CG) of the male (KB) and female (SD) actors from the motion capture sessions (Figure 3.5). Using a series of photographs for reference, the artist matched the geometric shape of the actors' faces to the CG characters. The photographs also provided a high-resolution texture for the characters.

In our framework, a smile sequence is parametrized as a matrix S of m (markers) $\times$ $3F$ (frames: for each motion axis x, y, and z). Using the 3D modeling and animation software *Autodesk Maya*, we created animations based on the motion of the markers, quantified by the matrix S. Spheres corresponding to the marker positions in the neutral expression are used as influence binds on the 3D mesh of the character to be animated. The spheres deform the skin surface by influencing position attributes (translation) of nearby mesh vertices. Influence objects deform smooth skin objects in the same manner that joints can influence smooth skin objects. Virtual markers, controlled by the matrix S, are parented to the spheres such that the position of the markers over time deforms the mesh accordingly. We linearized space and time to create the matrix S for the different animation conditions described in Section 3.3.2.

Figure 3.5: Actors whose smiles were recorded and their CG character counterparts used for the perceptual experiments: (a) KB, the male actor (b) KB's CG character (c) SD, the female actor (d) SD's CG character.

3.3 Experiment 1: Linearized animations with head motion

The goal of this first experiment was to determine how linearization impacts the perceived genuineness of smiles. This information can help us understand whether a smile model should preserve temporal and/or spatial information. We therefore examined the perception of four animation conditions resulting from linearizing time or space.

We chose genuineness as the dependent variable for two main reasons. First, genuineness is a useful metric for smile animation. Given the context of the animation, animators strive to create genuine or posed smiles. Controlling the genuineness of the smile is essential for the message of the animator to be conveyed. Second, "naturalness", the dependent variable we used for blinks, does not help distinguish between posed and spontaneous smiles as both types of expressions occur naturally. Additionally, in pilot experiments we found that participants are more comfortable rating one metric (genuineness) than multiple metrics such as the type of smile (i.e. amused, polite, nervous, posed). We suspect this is because genuineness is more easily understood.

We collected genuineness ratings of animated smiles through controlled experiments on *Amazon Mechanical Turk*. Fifty-seven viewers successfully rated 48 animations of the two CG characters. Viewers were at least 18 years old and located in the United States and other demographics such as age and gender were not recorded. The animations were displayed in a randomized order to control for possible order effects. After viewing each animation, participants were asked to rate the smile by moving a slider on a continuous scale from 0 (not genuine) to 100 (genuine). Similar to the method used by Krumhuber and Kappas [52], we described a genuine smile as a smile that someone shows when she/he is joyful, happy, or amused. Though

the characters in the clips are computer-generated and therefore not truly happy, we explained that the animations reflect certain aspects of smiles that also occur in real people. The experiment took no longer than 20 minutes.

All animations were displayed with the originally recorded head and torso motion and eye blinks. Trutoiu and colleagues ([91] and Chapter 5) showed that smiles have a characteristic blink placement relative to the smile start and end. We therefore considered it important to add eye blinks to the animation with the same timing as in the video sequence.

We used a within-subject repeated-measures experimental design. We focus on two independent variables that can be linearized: the geometric path of the vertices and the interpolation function. Interactions between these variables result in the animation conditions described below. The four conditions represent (1) a ground truth animation where both space and time are nonlinear and the original data is used without modification, (2) an animation where both space and time are linearized, (3) linearized time with the original spatial path, and (4) linearized spatial path with original timing. Additional independent variables are the CG character used (female or male), the type of smile (posed or spontaneous), and the smile sample. We used a total of twelve smile sequences: three genuine and posed smiles for each character.

3.3.1 Hypotheses

Our research question is how are are smile animations impacted by linearizing space or time. As discussed previously, spatial linearizations often occur when smile animations are created by interpolating between two blendshapes. Similarly, temporal linearizations occur if a linear interpolation function is used. We hypothesized that viewers will rate as less genuine smiles with linearized time because the constant speed of vertices may be perceived as mechanical. Based on how the original data was recorded, we expected spontaneous smiles to be rated as more genuine than posed smiles. Similarly, because spontaneous smiles are longer in duration and therefore viewers are exposed to them for longer, we expected perceptual differences across the animation techniques to be more pronounced for spontaneous smiles. We did not expect to observe a difference between CG characters.

3.3.2 Animation conditions

The animation conditions result from the combination of two independent variables:

- Spatial (S), determined by the geometric path of the vertices as data-derived and thus nonlinear (N) or linear (L)
- Temporal (T), determined by the interpolation function as data-derived and thus nonlinear (N) or linear (L).

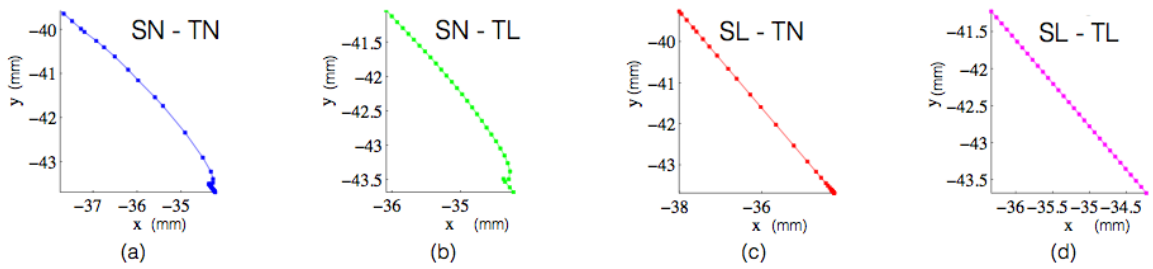


Figure 3.6: The geometric path of the right mouth corner vertex for all animation conditions in a short sequence of frames. The frequency of the dots reflects the velocity along the path. (a) Space Nonlinear-Time Nonlinear (SN-TN): Ground truth animation with both data-derived spatial path and data-derived interpolation function (b) Space Nonlinear-Time Linear SN-TL: data-derived spatial path with a linear interpolation function (c) Space Linear-Time Nonlinear SL-TN: linear spatial path with a data-derived interpolation function (d) Space Linear-Time Linear SL-TL: linear spatial path with linear interpolation function.

For a vertex i we define its position at time t as $V_i(t)$ where t ranges from 1, the first frame of the smile, to p, the peak displacement of the smile relative to frame 1, and n, the end of the smile. For a *data-derived geometric path (SN)*, V_i is directly taken from the motion capture data. For the *linear geometric path (SL)*, we define a piece-wise linear function composed of the linear path of vertex V_i between the position at frame 1 and the peak frame p and the linear path of vertex V_i between the position at frame p and the end frame n.

We next describe each of the four animation conditions, focusing on deriving the position of vertices for the first part of the smile, from frames 1 to p. Similar computations are defined for frames $p + 1$ to n. Visual representations of the vertex path and the interpolation functions used are shown in Figure 3.6 and Figure 3.7.

SL-TL: A linear geometric path for vertices $V_i t$ between 1 and p moving with constant speed based on *linear interpolation* is computed as

$$V_i(t) = V_i(1) + \frac{t-1}{p-1} * [V_i(p) - V_i(1)]. \tag{3.1}$$

This condition is equivalent to using two blendshapes (the neutral and the peak frame) with a linear interpolation between them.

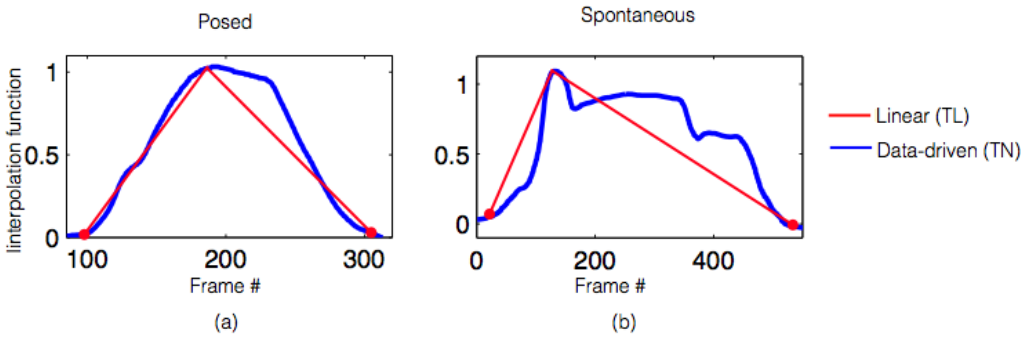


Figure 3.7: Data-driven (TN condition) and linear (TL condition) interpolation functions for (a) a posed smile and (b) a spontaneous smile at 120 fps.

SL-TN: For a data-derived speed and linear geometric path for the vertices, the position of vertices V_i at time t is computed as

$$V_i(t) = V_i(1) + rc(t) * [V_i(p) - V_i(1)], \tag{3.2}$$

where $rc(t)$ is the reconstruction coefficient computed as the least-squares solution to minimize

$$||V(t) - V(1) + rc(t) * [V(p) - V(1)]||. \tag{3.3}$$

Note that at every frame t, the positions of vertices are computed based on frames 1 and p and the sequence follows a linear path between 1 and p. This condition corresponds to using two blendshapes (the neutral and the peak frame) with a data-based interpolation function. The data-derived reconstruction coefficient may have values outside of the $[0 : 1]$ interval, resulting in the point moving forward and then backward along the same path. This effect can be seen in smiles when the smile is released and a small lip adjustment occurs: the lips are pressed together during the release and then relax in a natural position.

SN-TL: The vertices V_i move with constant speed across a data-derived geometric path. We first compute $PathLength$, the length of the path traversed by each vertex V_i from frame 1 to p. The position of each vertex V_i at time t is computed iteratively such that

$$V_i(t+1) - V_i(t) = \frac{PathLength}{p-1}. \tag{3.4}$$

SN-TN: The vertices have a data-derived geometric path and speed directly based on the recorded motion capture data. This condition is considered to be the ground-truth animation, closest in naturalness to the video.

Table 3.1: Significant results from Experiment 1: Linearized animations with head motion.

Effect	F-Test	Post-hoc
Main effects		
Space	F(1,840)=21.96, p<.0001	Nonlinear space (48.34) is rated as more genuine than linear space (44.21)
Time	F(1,840)=14.43, p=.0002	Nonlinear time (47.95) is rated as more more genuine than linear time (44.60)
Smile type	F(1,840)=297.58, p<.0001	Spontaneous smiles (53.87) are rated as more genuine than posed smiles (38.68)
Character	F(1,840)=99.06, p<.0001	The KB character (50.66) is rated as more genuine than the SD character (41.89)
Two-way Interactions		
Time*Smile type	F(1,840)=14.97, p<.0001	Linearizing time significantly impacts spontaneous but not posed smiles
Smile type * Character	F(1,840)=48.40, p<.0001	KB and SD are significantly different for posed but not for spontaneous smiles

3.3.3 Results

To explore the effects of linearization on perceived genuineness ratings we performed a 2 (Space) $\times$ 2 (Time) $\times$ 2 (Smile type) $\times$ 2 (Character) repeated-measures ANOVA followed by Tukey's HSD post-hoc significance tests. The main finding of this experiment is that linearizations in either time or space influence the perceived genuineness of smiles. The significant effects are shown in Table 3.1 and further detailed below.

Our hypothesis that linearizing time reduces the perceived genuineness of smiles was supported: animations with linearized timing (TL = 44.60), where the velocity of vertices is constant, are rated lower than animations in which the vertices follow the temporal profile of the original data (TN = 47.95), $F(1, 840) = 14.43$, $p = .0002$. Similarly, animations where the vertices move in a linear geometric path (SL = 44.21) are rated significantly lower than animations where the vertices follow the original path (SN = 48.34), $F(1, 840) = 21.96$, $p < .0001$.

As expected, we found that spontaneous smiles are rated as more genuine than posed smiles, with averages of 53.87 and 38.63 respectively, $F(1, 840) = 297.58$, $p < .0001$. The two CG characters were rated in significantly different ways: KB averaged ratings of 50.66, while SD averaged 41.89, $F(1, 840) = 99.06$, $p < .0001$. The ratings for the two CG characters do not differ for spontaneous smiles, while for posed smiles they are significantly different, with SD posed smiles being rated as the lowest, average of 31.23. This result is highlighted in the interaction between the type of smile and the CG character, $F(1, 840) = 48.40$, $p < .0001$.

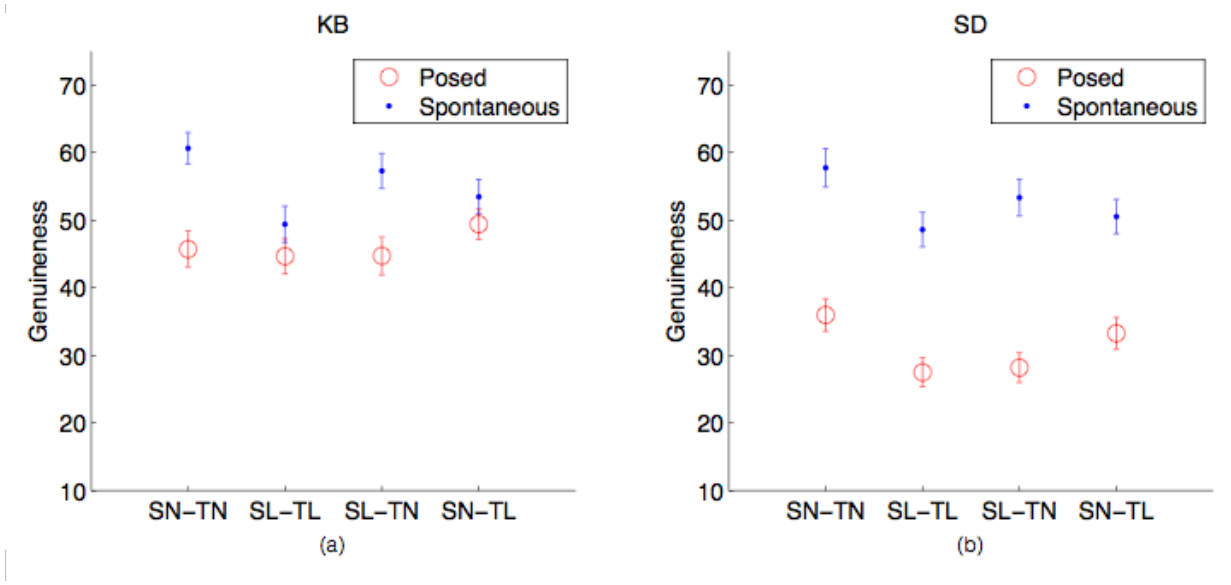


Figure 3.8: (a) Genuineness ratings for the KB character smiles. (b) Genuineness rating for the SD character smiles. The values are plotted with standard error bars. The four animation conditions in both graphs are (1) SN-TN: Space Nonlinear-Time Nonlinear (2) SL-TL: Space Linear-Time Linear (3) SL-TN: Space Linear-Time Nonlinear and (4) SN-TL: Space Nonlinear-Time Linear.

Another significant interaction occurs between the type of smile and the time variable. A constant velocity for the vertices (TL, linear time) impacts spontaneous ($p < .0001$) but not posed smiles.

We also explored how the space and time parameters interact to impact genuineness for spontaneous smiles in particular. We analyzed the differences in spontaneous smiles and found that the original animations (SN-TN = 59.19) are not significantly different than the animations in which space is linearized while temporal information is preserved (SL-TN =55.31). In contrast, the original animations (SN-TN = 59.19) are significantly more genuine than animations in which the geometric path of the vertices is preserved while the speed of the deformation is constant (SN-TL=52), $F(3, 210) = 16.68$,$p < .0001$. The overall interactions between space and time variables are shown in Figure 3.8.

It is interesting to contrast the ratings for animated smiles with the original data (condition SN-TN in Figure 3.8) with their corresponding video ratings (Figure 3.4). The animated spontaneous smiles had lower ratings than their video counterparts while the reverse was true for posed smiles. We suspect that presenting the smiles with CG characters makes it more difficult for participants to distinguish between posed and spontaneous smiles. Regardless, the animations preserved the ordering between posed and spontaneous smiles.

Table 3.2: Significant results from Experiment 2: Linearized animations without head motion.

Effect	F-Test	Post-hoc
Main effects		
Space	F(1,859.9)=19.95, p<.0001	Nonlinear space (47.65) is rated as more genuine than linear space (43.38)
Time	F(1,859.9)=32.44, p<.0001	Nonlinear time (47.65) is rated as more more genuine than linear time (43.38)
Smile type	F(1,859.9)=155.53, p<.0001	Spontaneous smiles (50.58) are rated as more genuine than posed smiles (39.92)
Character	F(1,859.9)=531.22, p<.0001	The KB character (54.93) is rated as more genuine than the SD character (35.58)
Two-way Interactions		
Smile type * Character	F(1,859.9)=16.76, p<.0001	KB posed smiles (51.32) were rated as more genuine than SD spontaneous smiles (42.63)

3.4 Experiment 2: Animations without head motion

One potential confound of Experiment 1 is that linearizing time may be desynchronizing the facial expressions from the head motions. To overcome this confound, we conducted the current experiment in which animations were displayed without head motion.

In this experiment, we explored the effect of linearization on genuineness in animations without head motion. The procedures, independent variables, and measures were in all other respects identical to those used in Experiment 1. In our video data, we observed posed smiles as exhibiting less head motion than spontaneous smiles. We therefore hypothesized that participants in this experiment rate both posed and spontaneous smiles as less genuine. Sixty-one participants successfully took part in this study.

3.4.1 Results

As in Experiment 1, we conducted a 2 (Space) $\times$ 2 (Time) $\times$ 2 (Smile type) $\times$ 2 (Character) repeated-measures ANOVA followed by Tukey's HSD post-hoc significance tests. The results of this experiment are in most respects similar to those from Experiment 1. The significant effects are shown in Table 3.2 and further detailed below.

Linearizing either the space or interpolation function resulted in significantly lower genuineness ratings. Animations with a linear interpolation function (TL = 42.86) were rated lower than animations with nonlinear interpolation functions (TN = 47.65), $F(1, 859.9) = 19.85, p < .0001$. Animations with linear geometric paths (SL = 43.38) were rated lower than animations with nonlinear geometric paths (SN = 47.13), $F(1, 859.9) = 32.44, p < .0001$.

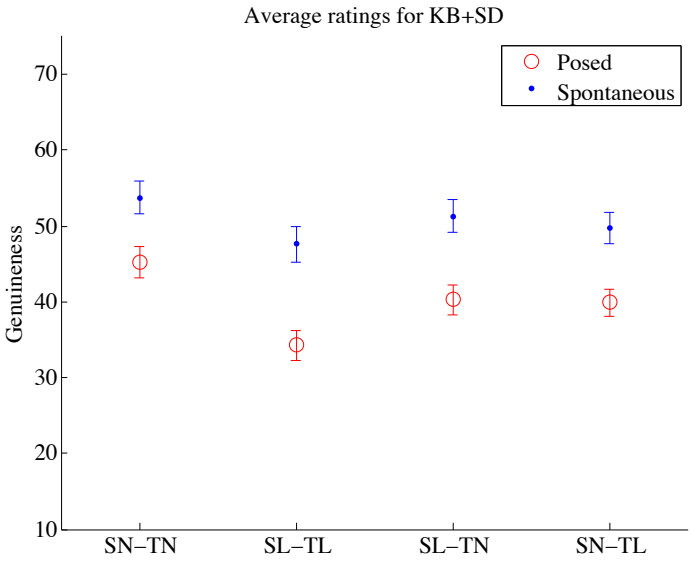


Figure 3.9: Average genuineness ratings for animations without head motion. The values are plotted with standard error bars.

In our data, we did not find a significant difference between spontaneous smiles using a data-driven interpolation function and a linear geometric path (SL-TN = 51.35) and original animations (SN-TN = 53.71), $F(1, 859.9) = 2.11$, $p = .146$. The interactions between the spatial and temporal independent variables are shown in Figure 3.9. Interestingly, all of the SD animations without head motion, including those of spontaneous smiles, were rated as less genuine than even the posed KB animations without head motion, which was not the case in Experiment 1.

3.4.2 Further analysis

Experiment 1 and 2 were conducted as independent within-subject experiments. Different participants, as observed by their unique *Amazon Mechanical Turk* ID, took part in Experiment 1 and Experiment 2. To compare animations with and without head motion statistically, we considered Experiment 1 and 2 as blocks in a mixed experiment design: head motion was the between-subject categorical variable while space, time, smile type, and character were within-subject independent variables. A repeated measures ANOVA with head motion as the between subject variable did not find a significant effect of head motion: animations without head motions (45.6) in Experiment 2 were rated only slightly lower than animation with head motions (mean = 46.2) in Experiment 1. In Chapter 4, we conducted a within-subjects experiment in which participants rated animations with and without head motion.

Table 3.3: Differences between posed and spontaneous smiles quantified in terms of duration, nonlinearity, and mouth corner vertex speed.

	Sample	Duration (seconds)	Spatial nonlinearity (mm)	Right mouth corner vertex speed (mm/second)	Left mouth corner vertex speed (mm/second)	Average mouth corner vertex speed (mm/second)	Difference in mouth corner speed (L-R) (mm/second)
KB posed	KB_p1	3.68	0.62	6.89	7.99	7.44	1.10
	KB_p2	3.45	0.60	7.38	9.33	8.36	1.96
	KB_p3	3.96	0.62	6.98	8.13	7.55	1.15
	average	3.70	0.61	7.08	8.49	7.78	1.40
KB spont	KB_s1	4.71	0.85	4.61	6.72	5.66	2.10
	KB_s2	4.09	0.65	3.68	5.26	4.47	1.59
	KB_s3	5.70	0.64	2.44	3.83	3.14	1.40
	average	4.83	0.71	3.57	5.27	4.42	1.70
SD posed	SD_p1	2.10	0.46	4.03	3.73	3.88	-0.30
	SD_p2	1.70	0.08	0.78	1.06	0.92	0.28
	SD_p3	2.10	0.44	3.85	5.23	4.54	1.38
	average	1.97	0.33	2.89	3.34	3.11	0.46
SD spont	SD_s1	5.00	0.49	3.89	4.41	4.15	0.52
	SD_s2	5.25	0.58	3.22	4.50	3.86	1.28
	SD_s3	5.17	0.32	3.13	3.94	3.53	0.81
	average	5.14	0.46	3.41	4.28	3.85	0.87

3.5 Differences between posed and spontaneous smiles

As shown in the previous experiments, linearizing time in particular decreases genuineness more for spontaneous smiles than for posed smiles. In this section, we quantify the differences between posed and spontaneous smiles for the following characteristics: duration, spatial nonlinearity, and mouth corner vertex speed. We chose these variables because they reflect spatial and temporal properties of the vertex motion. Furthermore, quantifying these differences helps us relate the smile samples used in our experiments with previous research on the differences between posed and spontaneous smiles. The quantified variables are shown in Table 3.3.

For our data, the six spontaneous smiles were longer (average duration of 4.9 seconds) than posed smiles (average duration of 2.9 seconds). These differences in duration are consistent with previous research. Ambadar and colleagues [3] found that on average, perceived amused

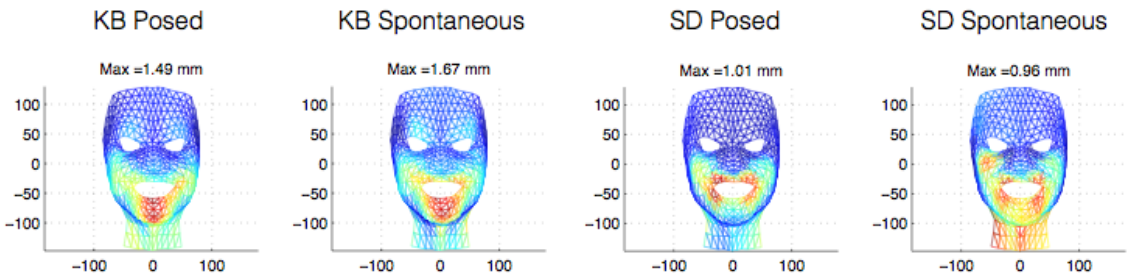


Figure 3.10: Visual representation of the difference between the original nonlinear animation and its spatially linearized counterpart. The differences were averaged across the three smile samples. In each figure, the maximum Euclidean distance for a vertex is noted in the title.

smiles lasted about 4 seconds, whereas perceived polite or embarrassed/nervous smiles lasted for 2 seconds and 2.9 seconds, respectively. In our case, perceived amused smiles equate with spontaneous smiles (which were rated as highly genuine); posed smiles, which lack genuineness, likely include both polite and embarrassed smiles. The durations for the twelve smiles used in the perceptual experiments are reported in the first column of Table 3.3.

Linearizing space impacted spontaneous smiles more than posed smiles. We hypothesized that spontaneous expressions are more complex and therefore more nonlinear. To quantify spatial nonlinearity, we computed the differences between the original and the linearized geometric path. We used a Euclidean distance-based measure as proposed by Cosker and colleagues [18]. For each frame, we computed the Euclidean distance between the vertices in the original animation and their linear counterparts. The second column of Table 3.3 shows the average nonlinearity measure per vertex normalized by smile duration. As expected, spontaneous smiles had more nonlinear motion: the Euclidean distance is larger for spontaneous (average of 0.59 mm) than for posed (average of 0.47 mm) smiles. Note that though these values are small, they are averaged over 430 vertices. In Figure 3.10, the nonlinear motion of spontaneous smiles appears to be more diffuse on the face. Different patterns are visible for the two actors. For example, KB, the male actor, shows more nonlinear motion in the eyebrow and lower jaw region. In contrast, SD, the female actor, shows more nonlinear motion in the mouth corner region.

The measure for spatial nonlinearity described above is the difference between ground truth animations (SN-TN) and animations with linearized time (SL-TN). We computed the remaining two differences between ground truth animations and animations with both time and space linearized (SL-TL) and animations with nonlinear space and linear time (SN-TL) respectively. Figure 3.11 shows the error of the three linearized conditions relative to the ground truth animations. Not surprisingly, the SL-TL condition has the highest error and was also rated as the least genuine.

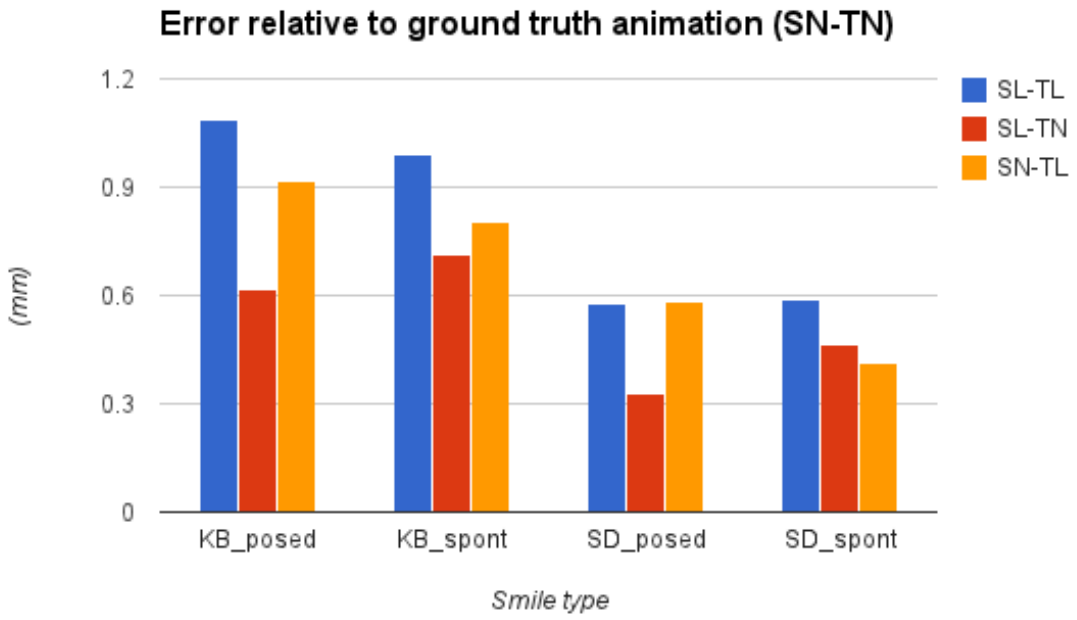


Figure 3.11: Error for the linearized conditions computed as the average per vertex Euclidean distance between ground truth animations (SN-TN) and their respective linearized conditions.

We considered the differences in the vertex speed for the two mouth corner vertices. There is evidence that posed and spontaneous smiles differ in vertex speed and symmetry. For posed expressions, Schmidt and colleagues [80] showed that movement asymmetry (measured by change in pixel values over time) was significant for expressions of joy, including smiles, with more movement on the left side of the face. Our results similarly show that the left mouth corner speed was consistently higher than the right right mouth corner speed (Table 3.3, last column).

In our analysis, for the KB actor, posed smiles had larger mouth corner vertex speeds than spontaneous smiles. However, the opposite was true for the SD actor. Ekman first posited that spontaneous smiles are more symmetric than posed smiles [33]. However, their study did not quantitatively assess the movement asymmetry. The left mouth corner for both KB and SD showed more vertex speed than the right mouth corner. In future work, we intend to analyze a larger dataset of posed and spontaneous expressions across a broader pool of subjects to better quantify motion asymmetry.

A representative example of vertex speed over time for posed and spontaneous smiles is shown in Figure 3.12. In previous research, viewers associated irregularities in the offset of smile, as measured by the frequency of phasic change, with posed smiles [41]. However, in our examples, spontaneous smiles showed more changes in velocity.

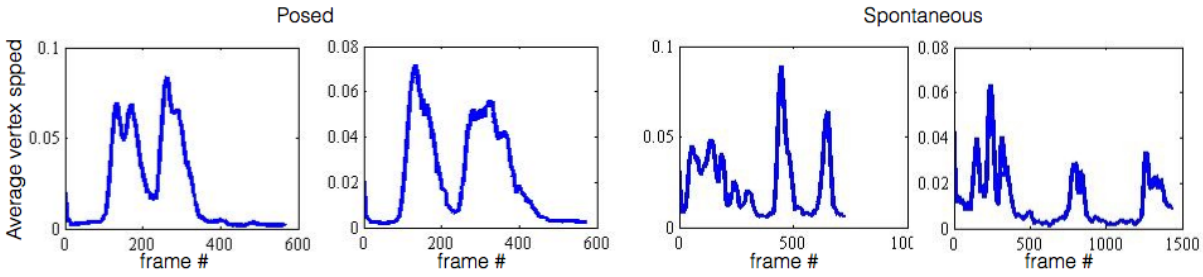


Figure 3.12: Average vertex speed for four smiles computed at 120 fps. Spontaneous smiles have more peaks compared to posed smiles. The first and last peaks correspond to the onset and offset of the smile.

3.6 Discussion

Our experiments reveal that linearizing either space or time decreases the perceived genuineness of smiles. However, our data indicates that spontaneous smiles with a linearized geometric path (space) and a data-driven interpolation function (time) are rated as being as genuine as the original high-resolution animations. Based on these results, in Chapter 4 we investigate a parsimonious model for smiles consisting of data-driven interpolation functions that capture the dynamics of the facial expressions but linearize the spatial path.

There are several limitations to our study. We recorded and animated smiles for two CG characters. Genuineness ratings for the two characters differed in their respective animations, with KB's smiles rated more genuine than SD's. Several explanations may account for these differences: intrinsic differences in the smile expressions between the two participants, differences in the quality or rendering style of the CG character, and perceptual differences related to age and gender. For example, the male participant, KB, is a professional actor while the female participant, SD, had no acting experience. This difference in acting experience could potentially explain why KB's posed smiles were rated more genuine than SD's posed smiles even though their spontaneous smiles had similar values.

The main limiting factor in conducting this kind of study is determined by the availability of high-resolution CG characters. We aimed for our CG characters to be of similar quality. However, previous research has shown that small differences in rendering styles can influence perceptual judgments of CG characters [44, 68]. SD, the female character, had fewer wrinkles, leading to a smoother face appearance. In future work, we will consider conducting a larger study using more smiles from more actors and counterbalancing for age and gender.

We used two types of interpolation functions (data-driven and linear) and found that, in this particular case, data-driven interpolation functions are needed for animating genuine smiles. However, many animation techniques use ease-in/ease-out interpolation functions that mimic the effects of acceleration and deceleration seen in physical systems. It may be the case that ease-in/ease-out interpolation functions are sufficient to create genuine smiles. In Chapter 4, we further explore how ease-in/ease-out interpolation functions compare to data-driven interpolation functions. Though more complex than posed smiles, the spontaneous smiles chosen started and ended in neutral expressions and were relatively short. It is possible that with more complex spontaneous smiles, preserving temporal information would be insufficient.

In Experiments 1 and 2, participants rated animations with and without head motion, respectively. Our results show that the effect of linearization on smile genuineness is similar in both experiments. However, previous research has shown a correlation between head motion and the dynamics of smiles [16]. Furthermore, for animations without head motion, SD's animations of spontaneous smiles were rated lower than KB's posed smile animations. This result indicates that the contribution of head motion to the perception of smile genuineness may vary across individuals. Further analysis is required to determine the relationship between head motion and smiles.

Our experiments indicate that some simplifications that occur in traditional blendshape animation may lead to smiles being perceived as posed and less genuine. These results suggest that if animators want to create genuine smiles, they should use nonlinear, preferably data-driven, interpolation functions. The data-driven interpolation functions we observed in our data show multiple peaks which add motion complexity and somewhat resemble a laughter-like pattern. We suspect that both the multiple peaks and the laughter-like pattern are cues that our viewers used to rate the smiles as more genuine. Furthermore, our study underlines the importance of using spontaneous rather than posed expressions in studies that quantify facial dynamics.

Chapter 4

Modeling Smiles

In this chapter, we use high-resolution motion capture data to build a parsimonious model of spontaneous smiles. Our smile model consists of two parts: (1) smile expressions and (2) plausible head motions. In Chapter 3, perceptual experiments showed that spontaneous smiles generated with two blendshapes (neutral and peak) and data-driven nonlinear interpolation functions are rated as genuine as high-resolution animations. We therefore build, from recorded smiles, a generative model that produces interpolation functions nonlinear in time. For each data-driven interpolation function we provide a plausible head motion. The complete model (interpolation function and plausible head motion) can be used to create genuine smiles in the traditional framework of blendshape animation. We demonstrate that smile animations from this model are comparable to high-resolution animations and are more genuine than animations with ease-in/ease-out interpolation functions. with identical head motion.

For the expression model, we start with high-resolution motion capture data of smiles. We reconstruct each smile sequence as linear combination of two blendshapes to obtain a data-driven interpolation function. These interpolation functions are nonlinear and and capture the plausible velocity as well as the multiple peaks that occur in natural smiles. Finally, we build a generative model of the reconstructed data-driven interpolation functions that allows us to create new interpolation functions.

We complement the newly generated interpolation functions with plausible head motions to create a model for genuine smiles. Spontaneous smiles are strong nonverbal signals that are often accompanied by moderately correlated head motions [15]. Furthermore, we suspect that lack of head motion in animation may make characters look rigid and less life-like. Based on finding moderate correlations in our own data between head motion and interpolation functions, we create plausible head motions which are proportional to the smile amplitude.

Through perceptual studies, we demonstrate that our model outperforms the commonly used ease-in/ease-out interpolation functions. We evaluate our model based on how the smiles are rated for genuineness. In a first perceptual experiment, we compare model smiles with recorded high-resolution spontaneous smiles, and also smiles generated with ease-in/ease-out interpolation functions. Our data showed no significant difference between the high-resolution spontaneous smiles and the model smiles. In a second experiment, we find that our model-based interpolation functions coupled with appropriate head motions are not character specific. That is, we show that the model derived from high-resolution data from one actor can be used for two different CG characters, both resulting in similarly high genuineness ratings.

4.1 Related work

In this section, we discuss methods for generating smiles. For a discussion on the perceptual differences between different types of smiles please refer to Section 3.1.

Though many graphics research articles use smiles as example expressions [48, 57, 65, 73, 74, 75] few explicitly consider generative models for smiles and laughter. For example, DiLorenzo and colleagues modeled and animated laughter but their goal was to synthesize anatomically inspired torso movements and deformations rather than facial expressions [25].

A first attempt at generating different types of smiles was proposed by Krumhuber and colleagues [52]. The authors did not provide an explicit smile model; however, they used a data-based heuristic to generate genuine smiles characterized by a long onset and offset duration with a shorter peak. The smiles were temporally symmetric and varied little in the expression. The perceptual contributions of their study are discussed in Chapter 3.

A discrete model for smiles with a limited number of parameters was proposed by Ochs and colleagues [72]. Their work introduced an algorithm to create three categories of smiles: polite, embarrassed, and amused. The authors asked participants to animate a 2D character for each smile category in a Flash-based web application. To animate a desired smile, participants chose among two or three discrete values for seven parameters: (1) amplitude, (2) mouth opening, (3) symmetry of lip corners, (4) lip press, (5) cheek raising, (6) duration, and (7) velocity of onset and offset. Smiles of different types were then generated with a decision tree trained with the user-selected smiles. Perceptual studies validated that naive users were able to recognize the categories for the generated smiles.

As discussed in Chapter 3, animators often use rule of thumb heuristics that rely on generic interpolation functions. A smile can thus be animated as the interpolation between two blend-

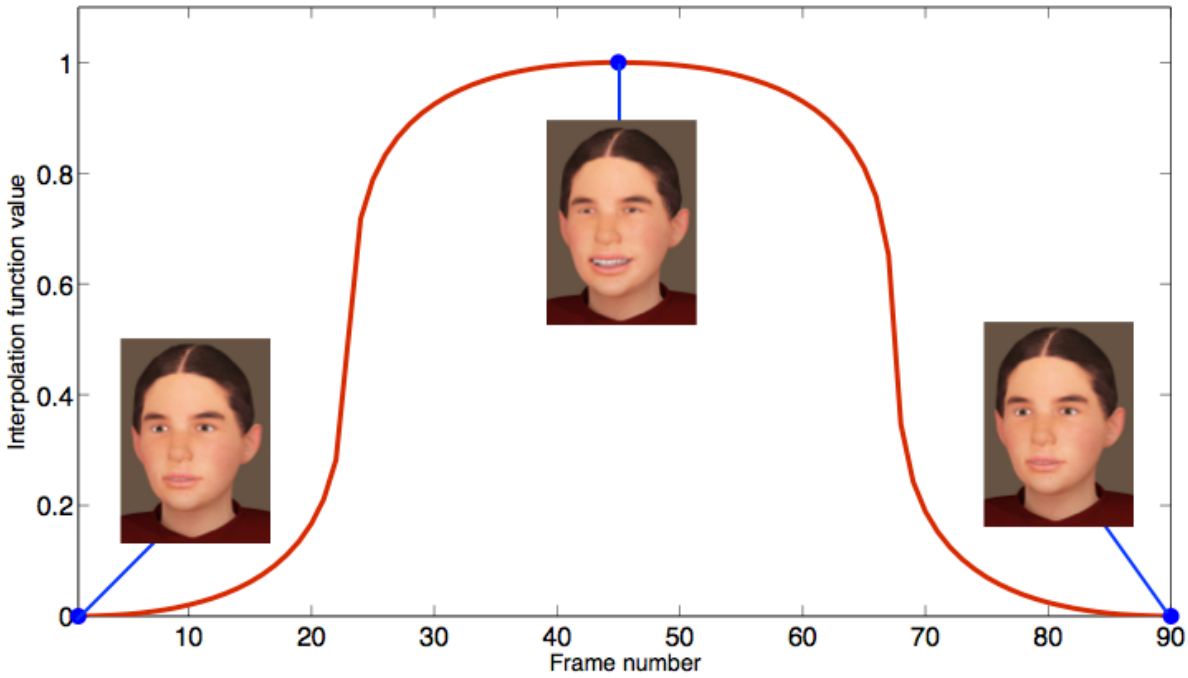


Figure 4.1: Ease-in/ease-out interpolation functions used to create a three-second smile animation with two blendshapes: a neutral face expression blendshape and a peak smile expression blendshape. This blendshape interpolation approach is recommended by traditional animation textbooks (see for example [63]).

shapes: (1) a neutral expression and (2) a peak smile, as shown in Figure 4.1. Our model has several advantages over previous models. First, we model the temporal progression of the smile which takes into account the durations of accelerations and decelerations that occur in natural smiles. Second, our model preserves motion complexity (multiple smile peaks). Our data indicate that most spontaneous expressions do not follow a smooth transition from neutral to peak and back to neutral. Rather, the spontaneous expressions we observed consist of multiple peaks in which the smile amplitude increases and decreases.

4.2 Smile model

In this section, we present a smile model that consists of two parts: (1) a generative model for smile expressions, represented as interpolation functions, and (2) plausible head motions. For the smile expression model, we chose to represent smiles as data-driven interpolation functions based on the perceptual results in Chapter 3. In our perceptual experiments, we found that temporal nonlinearities are needed to preserve smile genuineness while spatial nonlinearities are not. We capture temporal nonlinearities in data-driven interpolation functions with a generative Prin-

cipal Component Analysis (PCA) model. A second part of our smile model consists of plausible head motions because animations presented without head motion appeared artificial and rigid.

For both facial expressions and correlated head motions we used 25 spontaneous smiles from one female participant (SD). The dataset of smile profiles for the facial expression model is shown in Figure 4.2. The data recording process is described in the previous chapter. Each part of the smile model is described in more detail in the following subsections.

4.2.1 Generative model for smile expressions

Our approach to building a smile model relies on modeling the temporal properties of the expression represented as interpolation functions. Following our approach to creating an eye blink model (Chapter 2), we used PCA to construct a generative model of data-driven interpolation functions.

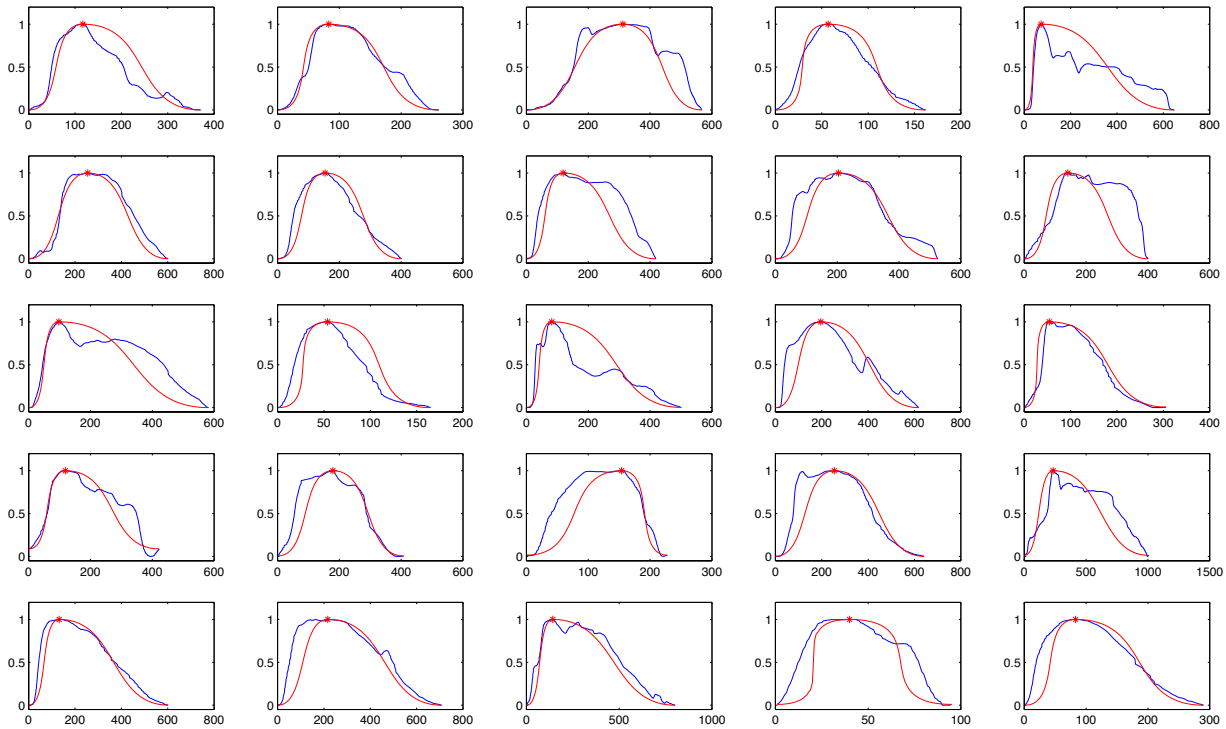


Figure 4.2: Dataset of smile profiles used to build a genuine smile model. The time series represent data-driven interpolation functions (blue). For contrast, we also show the ease-in/ease-out interpolation functions that are computed given the peak (red). For each sample, the y-axis shows the interpolation function value while the x-axis shows the smile duration as a frame number at 120 fps.

Twenty five spontaneous smiles from the SD character were used for the PCA model. We chose smiles from SD because of the availability of data for her spontaneous smiles, even though, in Chapter 3, SD's smile animations were rated as overall less genuine than the male participant's (KB). We suspect that a larger dataset based on KB's smiles would help strengthen the results and potentially create a more robust model.

We reconstructed high-resolution expressions of motion capture data with two blendshapes to build a time series dataset for smile dynamics. The time series are represented by the coefficients rc, which are the least square reconstruction of each smile as a linear combination of start frame ($V(1)$) and peak frame ($V(p)$) such that the following expression is minimized:

$$||V(t) - V(1) + rc(t) * [V(p) - V(1)]||. \tag{4.1}$$

For each smile sequence, $V(p)$ is the mesh at the peak frame p, the time instance with the maximum deformation (sum of per vertex euclidean distance) relative to a neutral expression (frame 1). This representation is equivalent to representing the smile as a linear combination of the blendshapes $V(1)$ and $V(p)$. The time series for each smile are shown in Figure 4.2.

PCA models the variability in the data-driven interpolation functions. The input to the PCA is a matrix with time series data for the 25 interpolation functions computed from the motion capture data (the reconstructed smile dynamics described above). In order to use PCA, each time series must be the same duration. As with blinks, we normalized the rc time series such that the duration from 1 to p and p to n is the same for all sequences. These fixed durations were determined as the median duration in the original data. The two scaling coefficients were appended at the end of each time series. The scaling coefficients are important because they model the variability in the duration of the generated smiles.

We represented the original dataset using only the first ten principal components accounting for 98% variance. Next, we projected new, random coefficients within one standard deviation of the original coefficients onto these ten PCA dimensions. Figure 4.3 shows the input and output of the PCA model: the original time series correspond to the input and the generated time series to the output. Using the last two terms of the newly generated time series, we scaled back each part of the smile to create interpolation functions of different durations. Animated smiles generated using these interpolation functions were used in both experiments.

Note that the generative PCA model of interpolation functions does not take into account the spatial information of smile expressions. We modeled smiles as data-driven interpolation functions where the motion occurs between two predetermined blendshapes. The smile sequence is thus a linear combination of two static poses. In this representation, the nonlinear spatial

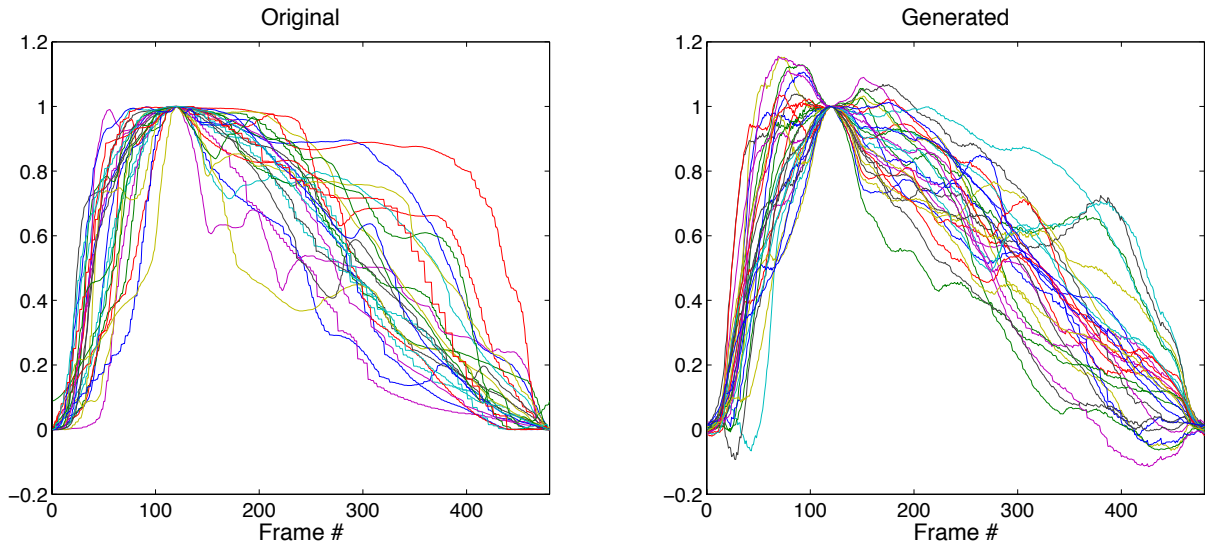


Figure 4.3: Principal Component Analaysis model for genuine smiles. Original smile profiles used in the model (left) and profiles generated (right).

movement of vertices on the mesh is lost. For the smile samples used in Chapter 3, temporal information was found to be more important than spatial information in the perceptual experiments. We therefore found it reasonable to build a model that considers the nonlinearity of temporal information.

4.2.2 Plausible head motions

Head motion is an important non-verbal cue that communicates or emphasizes the meaning of facial expressions. For example, animators often start by creating the important static head motion poses (”blocking out”), and then add facial expressions [63]. Furthermore, as we showed in Chapter 3, for participant SD, in particular, head motion plays a role in whether smiles are rated as genuine.

Our goal was to find a plausible head motion that will augment the genuineness of the facial expression. We hypothesized that a plausible head motion is proportional to the smile amplitude, similar to laughing, where sound correlates to torso movements. Existing studies have reported correlations between head motion and spontaneous smiles. Cohn and colleagues investigated the relationship between smile dynamics, head motion, and eye motion [15]. Their data showed moderate correlation between head pitch and the lip corner displacement: the smile intensity increased as the head moved downwards. The authors attribute this correlation to the types of smiles used, which likely signaled embarrassment. The authors thus hypothesize that smiles associated with the experience of joy and especially surprise would show smile intensity increasing and decreasing together with the head tilting backward.

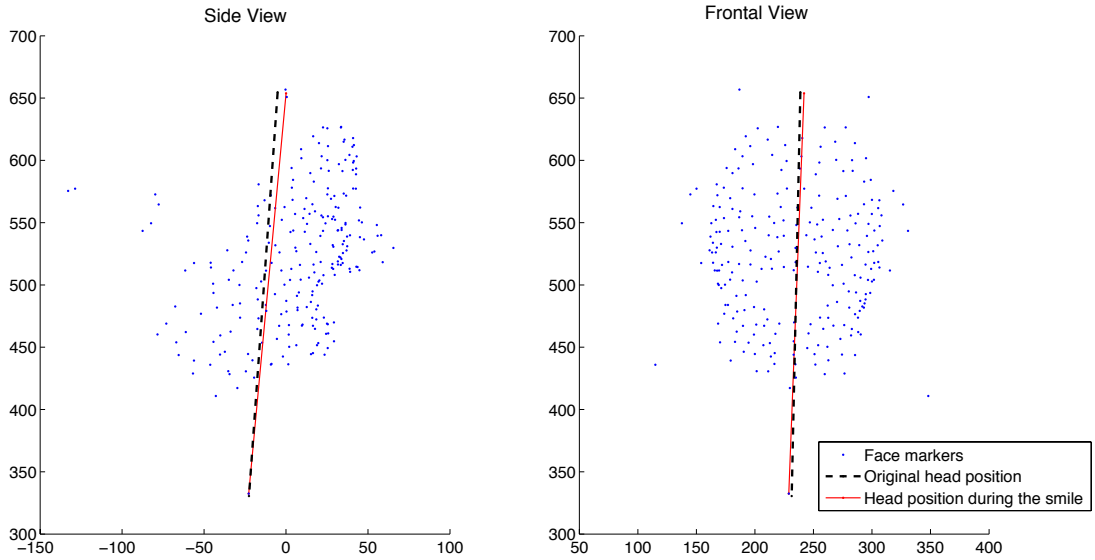


Figure 4.4: Head motion angle computation relative to a joint on the sternum for a sample smile sequence. The head pose at the beginning of the smile is represented in black. The head pose at the peak of the smile is represented in red.

We evaluated the correlation between head motion and smiles in our dataset of 25 smiles. We first computed 3D rotations for the head relative to a joint at the base of the neck (sternum) as shown in Figure 4.4. We then calculated the correlation coefficient between head pitch and their respective data-driven interpolation functions. The average correlation for all smile sequences is moderate, -0.38 on a scale from -1 to 1 where 0 means no correlation is occurring. However, some smile samples show stronger correlations with a maximum of -0.8. Figure 4.5 shows a smile profile and head rotation relative to a joint on the sternum. The smile profile is aligned and proportional with one of the head angle motions (head pitch).

Previous research and computed correlations provided evidence that plausible head motion is proportional to smile amplitude. We therefore considered generating head motions derived from the interpolation functions. Fully synchronized motions appeared unnatural and we opted for adding a small amount of noise to the motion. The noise level was determined by trial and error. Similarly, we determined the maximum amplitude of the motion, 12 degrees backward pitch, by inspecting existing smile samples.

We first generated a proportional lower neck rotation (joint located at the sternum) by adding small amplitude white noise (signal-to-noise ratio of 5) to the smile profile multiplied by 12, the value we expect for the head pitch amplitude. Because we were animating only one joint, the

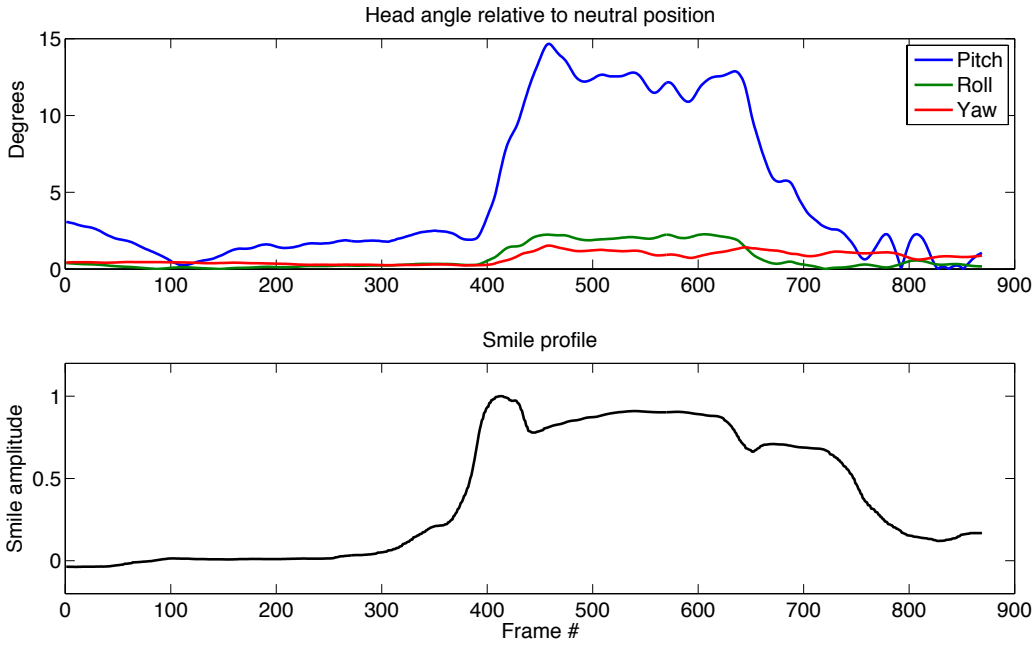


Figure 4.5: Head motion correlations for a smile video. Note that this is a strong correlation (0.75) between head pitch and smile profile.

motion looked mechanical. We therefore added a similar motion, with a smaller amplitude (3 degrees) to the upper neck joint. Examples of the head motion generated for the pitch of each neck joint are shown in Figure 4.6.

It is important to note that moderate correlations between facial expressions and head motions found in data analysis imply that there are likely many plausible head motions that could accompany a smile. Our intention was to generate one such motion to enhance the perceived genuineness of the facial expressions. Therefore, we refer to the head motions that are part of our smile model as *plausible motions* rather than as *a data-driven model for head motion*.

4.3 Perceptual experiments

We conducted a set of perceptual experiments to validate our model. In Experiment 1, our goal was to evaluate a small number of samples from the model relative to ground truth animations and ease-in/ease-out animations. In this experiment, we wanted to evaluate the model relative to original high-resolution animations so we used only SD's animations. In Experiment 2, our goal was to test a large sample of model smiles and apply the model to multiple CG characters.

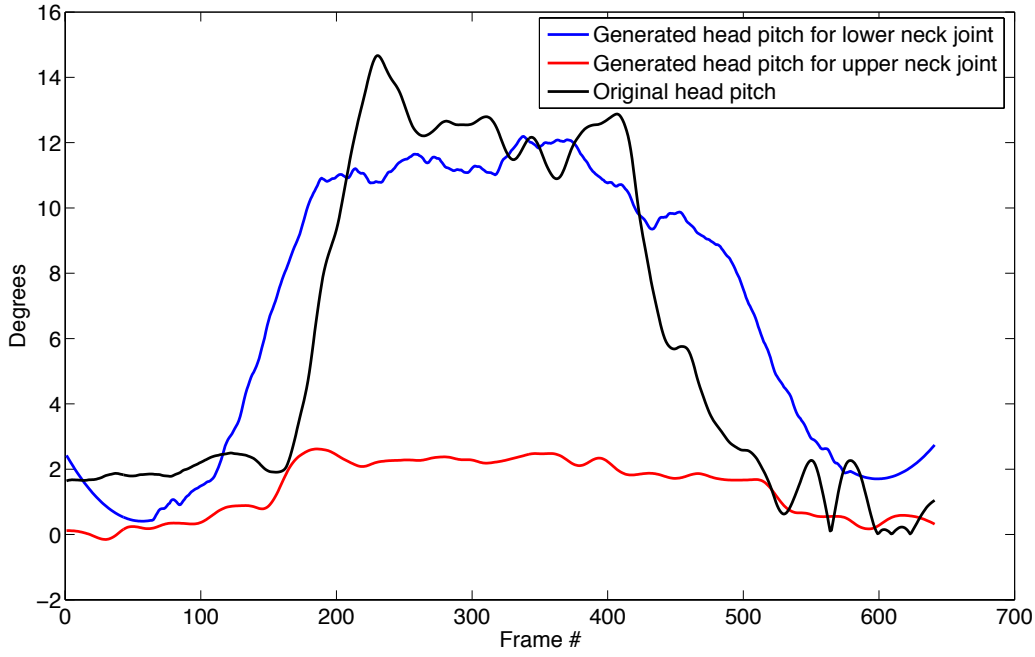


Figure 4.6: Comparison between recorded and generated head pitch.

4.3.1 Hypotheses

The experiments in Chapter 3 suggested that temporal information is required to maintain the genuineness of smile expressions when linearizing spatial motion, as in the case of blendshape interpolation. In that experiment, animations with linear spatial motion and data-driven interpolation functions had similarly high genuineness ratings to the high-resolution animations of spontaneous smiles. The timing information (interpolation function) for smiles with linearized spatial motion was derived directly from existing high-resolution smile data. In this experiment, we hypothesized that interpolation functions generated from our PCA model would result in smile animations rated as genuine as recorded high-resolution animations. We further expected that adding plausible head motions, proportional to the smile amplitude, to the smile expressions would increase the perceived genuineness of the animation. We also expected that model data-driven interpolation functions could be used for multiple CG characters.

4.3.2 Experiment 1: Spontaneous, posed, model, and ease-in/ease-out smiles

To evaluate our model, we conducted a within-subjects experiment with the following independent variables: smile type (spontaneous, posed, model, and ease-in/ease-out smiles) and head motion (with and without). As in Chapter 3, the dependent variable was perceived genuineness

on a scale from 1 to 100. In this experiment, the head motion for the model and ease-in/ease-out condition were identical and proportional to the model smile profile as described above. Sixty-one viewers rated 24 smile animations on *Amazon's Mechanical Turk*.

Smile types:

- **Spontaneous**: The three high-resolution animations of spontaneous smiles for the SD character used in Chapter 3.
- **Posed**: The three high-resolution animations of posed smiles for the SD character used in Chapter 3.
- **Model**: Three data-driven interpolation functions were randomly chosen (samples 1 to 3) from the model. Two static blendshapes (neutral and peak) were obtained from SD's collection of spontaneous smiles. The neutral blendshape was chosen as a neutral expression from an existing smile. The neutral blendshape appeared similar to the start of several smile expressions. The peak blendshape was defined as the static frame, from all smile samples, with the highest deformation in the cheek and mouth region relative to the neutral blendshape.
- **Ease-in/ease-out**: For each model interpolation function we created counterpart ease-in/ease-out curves with the same durations from neutral to peak and from peak to the end the smile. The ease-in/ease-out curves are, as in the case of blinks, two cubic Bezier curves: one curve from neutral to peak and one from peak to the end of the smile. The acceleration and deceleration for the Bezier curves were proportional to the duration of each smile phase. The Bezier curves were equivalent to using the default option *flat tangents* with a length of 0.5 of the smile phase duration in the 3D animation software *Maya* (Autodesk).

Results

We conducted a repeated-measures 4 (smile type) $\times$ 2 (head motion) ANOVA to investigate possible effects of the independent variables on smile genuineness. Both independent variables and their interaction significantly impacted genuineness ratings.

We found a significant main effect of smile type on smile genuineness $F(3, 420) = 19.13$, $p < .0001$. Posed smiles (rating of 35.62) were rated as being significantly less genuine than all other conditions, which were not significantly different amongst each other. Similarly, head motion had a significant effect on smile genuineness: animations with head motion had an average rating of 52.56 while animations without head motion averaged 37.51, $F(1, 420) = 120.80$, $p < .0001$.

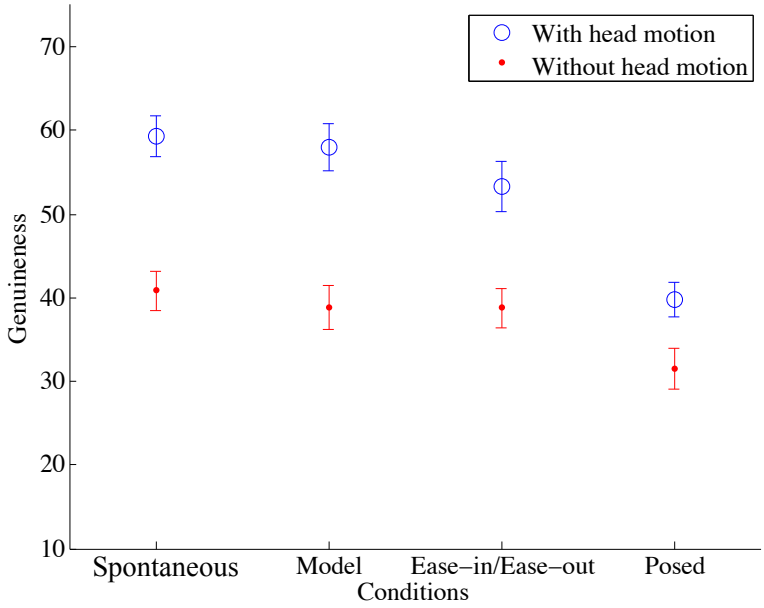


Figure 4.7: Interaction between smile type and head motion. Animations with head motion from the model are rated similarly to spontaneous animations. In contrast, ease-in/ease-out animations are significantly different than spontaneous smiles with head motion.

We further investigated the significant interaction between smile type and head motion $F(3, 420) = 2.83$, $p = .0379$. For animations with head motion, spontaneous smiles (59.31) were significantly different than ease-in/ease-out smiles (53.28), $F(1, 420) = 4.12$, $p = .043$, but not significantly different than model smiles (57.93), $F(1, 420) = .215$, $p = .643$. Without head motion, only posed smiles were significantly less genuine. This interaction is shown in Figure 4.7.

In summary, the complete smile model (PCA-generated interpolation functions accompanied by plausible head motion) generates animations that, based on our data, are not statistically different in genuineness ratings than the original high-resolution data. On the other hand, ease-in/ease-out animations, even with head motion, result in animations that are rated as significantly less genuine than the high-resolution animations. We found significantly lower ratings for animations without head motion. Furthermore, without head motion there was no difference between model (only PCA-generated interpolation functions), ease-in/ease-out, or spontaneous smiles.

4.3.3 Experiment 2: Multiple characters

The results in Experiment 1 confirm that our model outperforms ease-in/ease-out animations and that head motion plays a key role in generating genuine smiles. However, only three samples were used for each animation type. We therefore conducted a second experiment with a larger sample size.

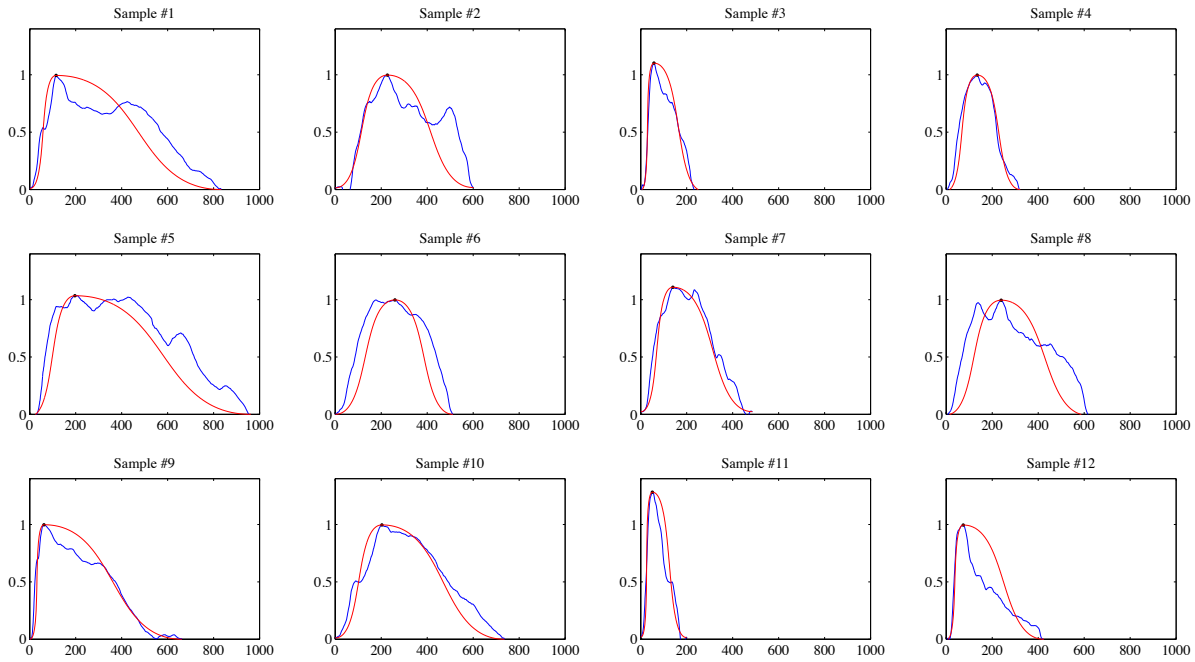


Figure 4.8: Model interpolation functions (blue) and ease-in/ease-out interpolation functions (red) samples used in Experiment 2. For each sample, the y-axis shows the interpolation function value while the x-axis shows the smile duration as a frame number at 120 fps.

The independent variables used in this experiment are the smile type (model or ease-in/ease-out), the CG character (female SD, male KB, or cartoon-like CP), and the smile sample (1 to 12). The dependent variable is smile genuineness. The twelve data-driven interpolation functions (Figures 4.8) and corresponding head motions are from the SD model. The blendshapes, neutral and peak are shown in Figure 4.9. All of the animations in this experiment were shown with head motion which was generated based on the model smiles.

We used a mixed experiment design, 3 (character) $\times$ 2 (smile type) $\times$ 12 (smile sample), with character as a between-subjects variable while smile type and smile sample were within-subjects variables. Each participant saw each animation type (n = 24) for only one character. Fifty-eight participants viewed and rated SD animations, 57 participants rated KB animations, and 63 participants rated CP animations.

Results

To examine the effects of smile type, sample, and character on genuineness ratings, we conducted a repeated-measures ANOVA with smile type and sample as a within-subjects variables and character as a between-subjects variable.

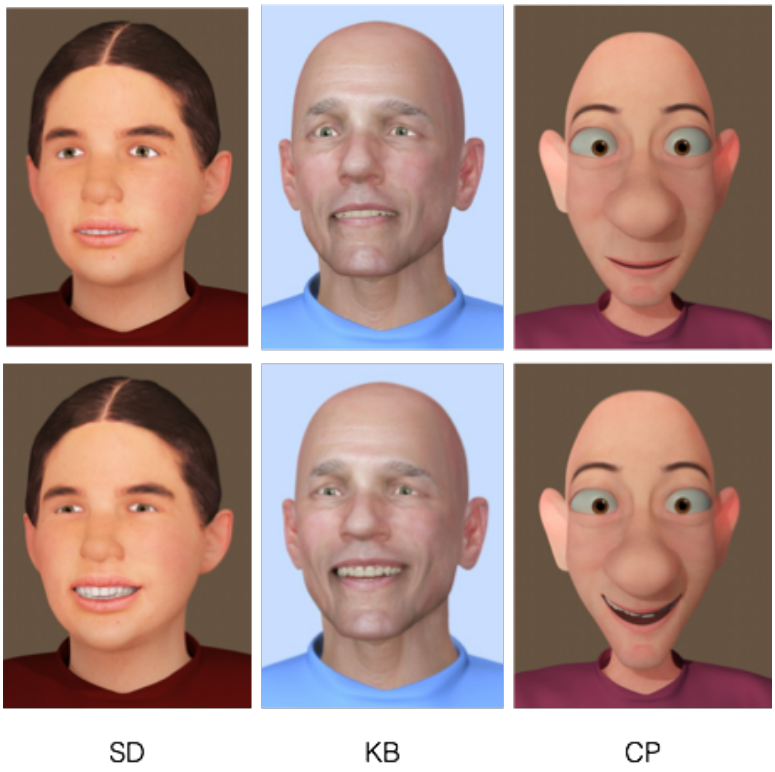


Figure 4.9: The three CG characters used in Experiment 2. Neutral (top) and peak (bottom) blendshapes for each character were either derived from data (KB and SD) or sculpted by an artist (CP). CP's blendshapes were sculpted by an artist because the character is cartoon-like and no real data was available.

Significant main effects were observed for smile type and sample but not for CG character. We found significant interactions between smile type and sample as well as between smile type and character. The significant effects are shown in Table 4.1 and further detailed below.

For the smile type, model smiles (56.13) were significantly more genuine than ease-in/ease-out smiles (52.22), $F(1, 4031) = 35.12$, $p < .0001$. Smile samples were also a significant main effect, $F(11, 4030) = 28$, $p < .0001$, with samples 5 (64.79), 7 (61.82), and 6 (59.55) were rated the most genuine while samples 3 (45.82), 11 (46.22), and 12 (47.21) were rated the least genuine. The interaction between smile type and sample, $F(11, 4030) = 7.15$, $p < .0001$, showed that both the two highest rated and the two lowest rated animations (collapsed across the three characters) are from the model category. This result implies that the model may need additional data to better model variability or that a stronger model is required.

The interaction between smile type and character was significant, $F(2, 4031) = 17.30$, $p < .0001$. Post-hoc comparisons indicated that for SD and KB, the photorealistic charac-

Table 4.1: Significant results from Experiment 2: Multiple characters with model and ease-in/ease-out interpolation functions.

Effect	F-Test	Post-hoc
Main effects		
Smile type	F(1,4031)=35.13, p<.0001	Model smiles (56.13) are rated as more genuine than ease-in/ease-out smiles (52.22)
Sample	F(11,4030)=28, p<.0001	The ratings for samples varied from 64.79 (sample 5) to 45.82 (sample 3)
Two-way Interactions		
Smile type*Sample	F(11,4030)=7.15, p<.0001	Model smile sample 5 is rated highest (70.42) while model smile sample 12 is rated lowest (43.51)
Smile type*Character	F(2,4031)=17.30, p<.0001	For SD and KB, model smiles are higher than ease-in/ease out

ters, the model was rated significantly more genuine than ease-in/ease out, $F(1, 4030) = 17.87$, $p < .0001$, and $F(1, 4030) = 48.18$, $p < .0001$ respectively. There was no difference between the model and ease-in/ease-out for the cartoon, CP, character: $F(11, 4035) = 1.18$, $p = .277$. The interactions between smile type and character are shown in Figure 4.10.

The results in this experiment indicate that our model is appropriate for use with photorealistic characters such as KB and SD. More research is required to investigate how this type of model can be used with cartoon-like characters. The three-way interaction between character, smile type, and sample is shown in Figure 4.11.

4.4 Discussion and future work

In this chapter, we showed the advantages of a smile model consisting of a data-driven, generative model of interpolation functions with plausible correlated head motions. New smile expressions are generated from a PCA model of interpolation functions. Given a smile profile, we generate a proportional head motion for two joints: lower neck and upper neck.

The difference between model and ease-in/ease-out animations was most pronounced for the SD character (Figure 4.10). This result is expected since the model was generated using only SD's data. However, we have also shown that the model, which was trained with a limited data set, can be used with a different CG photorealistic character, KB. More research is needed to determine how gender affects expressions of spontaneous smiles and whether the same model can be used for male and female CG characters.

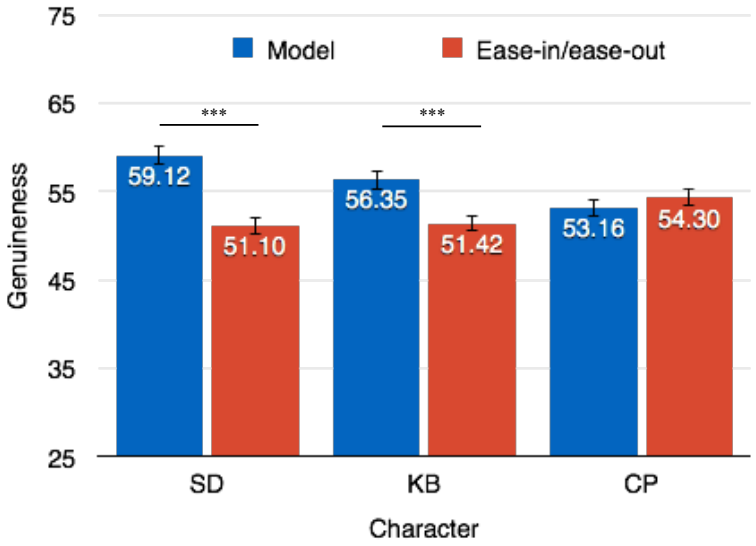


Figure 4.10: Genuineness ratings for three characters and two smile types (model and ease-in/ease-out). For photorealistic characters SD and KB the model is rated as significantly more genuine than ease-in/ease-out.

Our model can be used to generate spontaneous smiles of different durations. Interestingly, in Experiment 2 not all model samples were rated high on the genuineness rating nor were they all significantly more genuine than ease-in/ease-out. Consistent with previous research, the short smiles samples were rated lowest. In contrast, the longest smile sample had the highest rating across characters. We suspect that we need a larger training dataset to accurately model smile duration, in particular the duration from neutral to peak and peak to neutral. With a larger dataset, it may be possible to discretize the smile dynamics into multiple regions based each local peak rather and one global peak. Furthermore, more perceptual labels for the spontaneous smiles would help increase the consistency of the original data.

Our smile model includes a plausible head motion for each generated expression. The plausible head motion is head pitch proportional to the expression: the head tilts back as the smile increases in magnitude. We proposed this particular motion because we observed examples of this pattern in the data and, similar to previous research, moderate correlation between head motion and smile dynamics. More work is needed to explore variations in head motion and create a more comprehensive model. As we mentioned above, we adjusted the head pitch amplitude and amount of noise to a level that appeared reasonable. One potential avenue for future work is to change these two values to be data-driven.

We conducted several pilot experiments before considering adding plausible head motions. These experiments compared model facial expressions with ease-in/ease-out expressions and high-resolution animations of spontaneous smiles. In those limited experiments, we did not ob-

serve a difference between model and ease-in/ease-out animations and both of them had low genuineness ratings compared to the high-resolution spontaneous smiles. In Experiment 1, the head motion for both model and ease-in/ease-out smiles was based on the model interpolation functions.

An additional lesson learned from pilot experiments is related to the peak expression. In initial variants of the model, we considered smiles of various amplitudes that do not always reach the peak blendshape. Variability in the data modeled ensured some smiles had low amplitudes and not surprising were consistently rated as less genuine. There is evidence that posed smiles in fact show increased amplitude [79]. We suspect that participants in our experiments tried to identify salient cues for rating smiles as genuine and high amplitude smiles could be more similar to laughter. This effect may also have been visible in the experiments we ran within-subjects for animations with and without head motions. Animations without head motion were easily distinguishable from animations with head motion and it is possible that our viewers used that as a dominant cue.

The results for the cartoon character in Experiment 2 point to another avenue for future investigation. The photorealistic characters in the same experiment had blendshapes derived from data. The cartoon character, CP, had blendshapes defined manually according to common heuristics: the mouth corners move up, the eyes squint slightly, the nostrils dilate, the mouth opens. It may be the case that the peak blendshape we picked was not genuine enough. Following the completion of the experiment we also noted that the initial head pose for CP was with the head and chin slightly pointing downward, unlike the head pose for the photorealistic characters (Figure 4.9). At the end of the smile, the head returns close to its initial pose. In our opinion, the motion at the end of the smile gives the character a particular puppet-like appearance with the head dropping into an inert state. A potential continuation of this work is to consider a set of static blendshapes for both peak and neutral and how they interact with data-driven interpolation functions.

In summary, the primary contribution of this chapter is to demonstrate that data-driven interpolation functions accompanied by correlated head motions are appropriate for modeling smiles. Our smile model of interpolation functions and plausible head motions is rated as more genuine than animations based on the commonly used ease-in/ease-out interpolation functions. The model preserves naturally occurring smile accelerations, decelerations, and multiple smile peaks. In contrast, animations with ease-in/ease-out interpolation functions are smooth with a single peak and therefore may not accurately represent spontaneous smiles. Using our model, animators should be able to generate more genuine smiles expressions with plausible head motions.

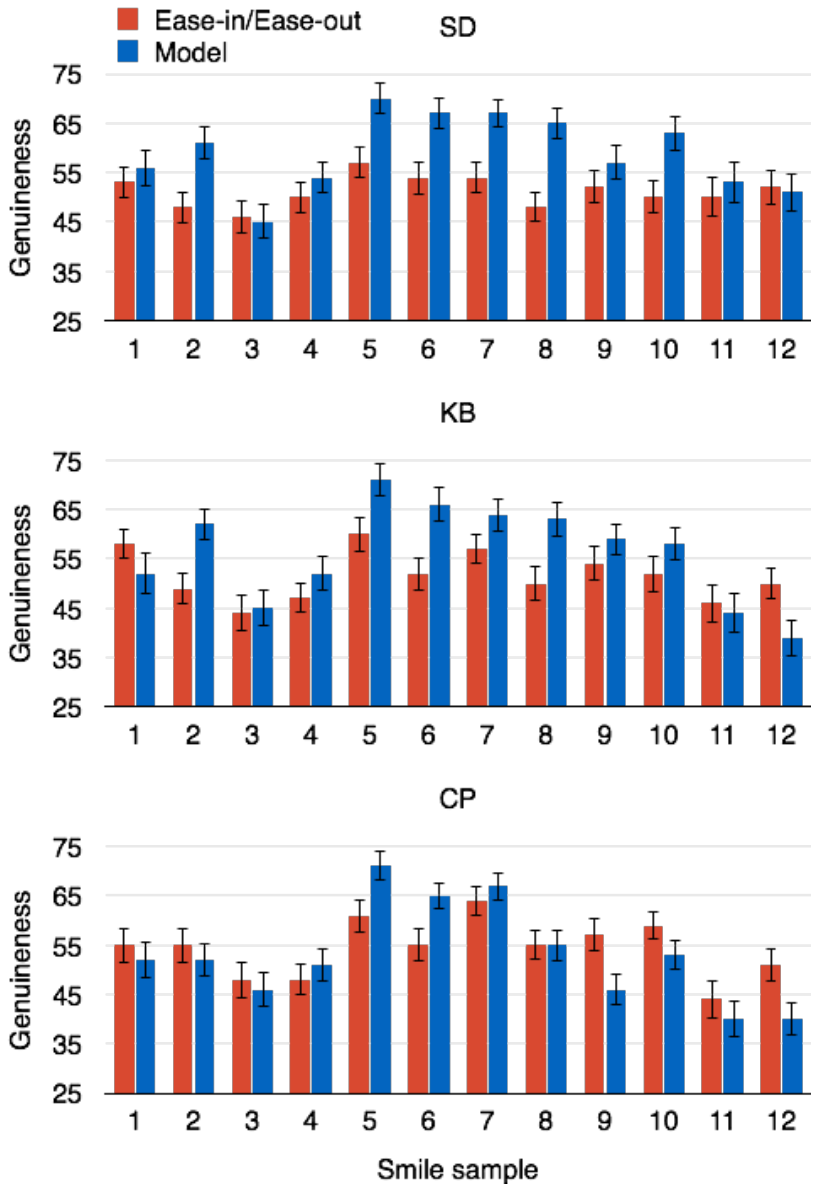


Figure 4.11: Genuineness ratings for three characters, twelve smile samples, and two smile types. For the SD character, more model smiles are rated as genuine than their ease-in/ease-out counterpart. In our data, the difference between model and ease-in/ease-out animations for the CP character is not statistically significant.

Chapter 5

The Temporal Connection between Blinks and Smiles

Facial expressions are driven by muscles. Arguably, independent anatomic actions activated by muscles form a basis for facial expressions. The most detailed visual description of these actions, known as action units (AU), is provided by Ekman's Facial Action Coding System (FACS) [32]. In the past decade, significant progress has been made toward automatic detection of action units in spontaneous facial behavior; however, relatively little is known about the timing of individual AUs and their temporal coordination, especially in natural behavior. To extract meaning from facial expression and inform realistic computer animation, it is necessary to both detect and synthesize facial actions and their temporal coordination.

In this chapter, we investigate the relative timing of two distinct facial motions: *blinks* and *smiles*. We are interested in the relative timing of blinks and smiles for two reasons. First, understanding where blinks occur relative to smiles can improve animation quality by better approximating the natural relationship between the motions. Second, the timing of blinks relative to smiles may have communicative value and is thus relevant to creating expressive animations and facilitating communication with avatars.

In the next section, we provide background information regarding the purpose and communicative value of both smiles and blinks. In Section 5.2, we analyze 43 videos of spontaneous smiles and demonstrate how eye blink sequence relates to smile start (onset) and smile end (offset). Our data show that eye blinks correlate with the end of the smile and occur close to the offset, before the lip corners stop moving downwards (Section 5.3). An illustrative example of the pattern observed in our data is presented in Figure 5.1. Finally, in Section 5.4, we discuss the limitations of the current approach and future research.

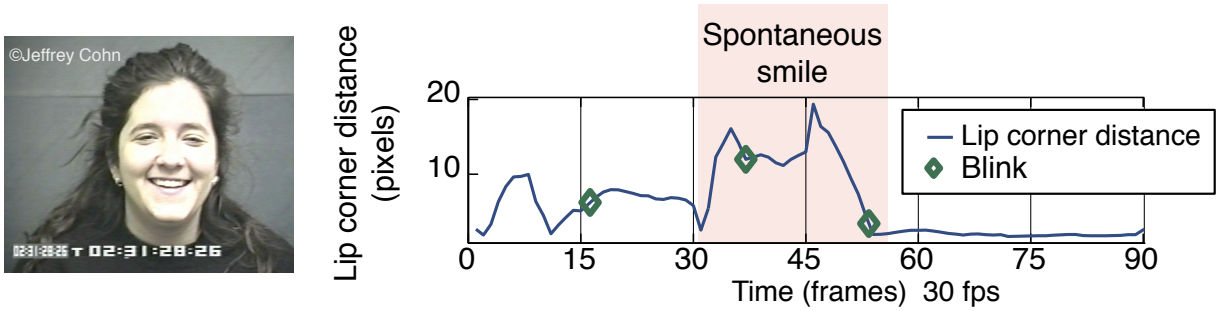


Figure 5.1: Blink occurrences for a participant during a short spontaneous smile. The expressions were recorded for the Cohn-Kanade facial expression database [50, 61].

5.1 Related work

Prior research on blinks and smiles was discussed in more detail Chapters 2 and 3. In this section, we describe prior research on the function of spontaneous blinks in relation to visual information processing.

5.1.1 Blinks

Blinks appear to serve both physiological and information processing functions. A likely physiological function of blinks is to lubricate the cornea and clear debris particles [34]. However, experimental evidence and hypotheses to support this position are mixed. Ponder and Kennedy [77] found no significant differences between the blink rates of participants exposed to variation in ambient humidity, which would affect eye lubrication. Although one would anticipate more rapid drying of the eye and thus more frequent or longer blinks to provide lubrication when humidity is low, their results suggest that moderate changes in environmental factors do not directly impact blink frequency. In light of these results, Ponder and Kennedy further hypothesized that blinking is likely a relief mechanism associated with "mental tension" [77].

Additional evidence for blink regulation comes from information processing research where blink patterns have shown task-dependent variability. Researchers have determined that blinking is modulated according to conditions such as increased mental activity [86], negative stimuli [14], or attempts to mask deceit [55, 78]. In light of the results described above, Ponder and Kennedy further hypothesized that blinking is likely a form of relief mechanism associated with a participant's "mental tension" [77]. Evinger and colleagues speculated that blink signals regulate eye blinks to occur at times when they are least disruptive to visual processing [35].

Nakano and Kitazawa found that eye blinks synchronized between subjects who simultaneously viewed the same short video story [71]. Because participants' behavior was presumably uninfluenced by others' behavior, they concluded that this synchronization must have arisen from

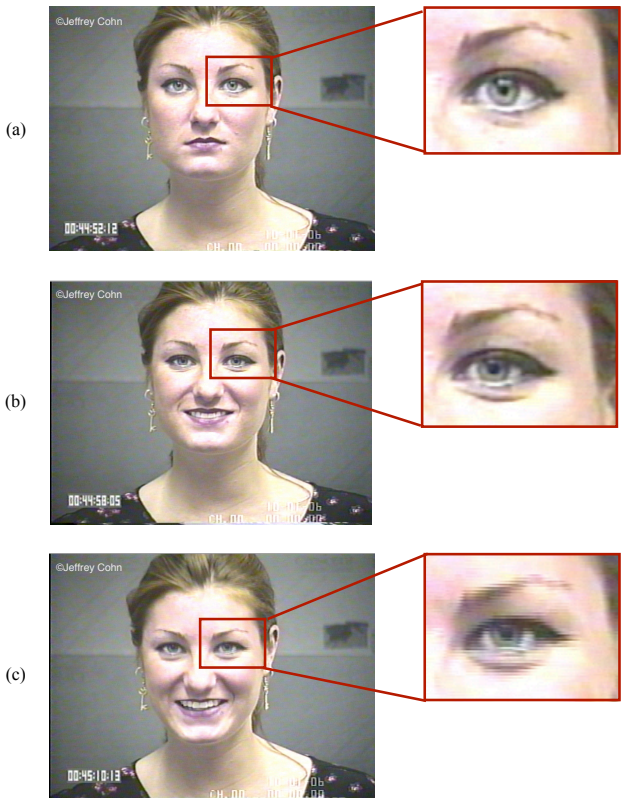


Figure 5.2: (a) Neutral face pose for participant in the Cohn-Kanade facial expression database [50, 61]. (b) The participant demonstrating a posed smile. (c) A spontaneous smile with the eyes narrowed as a result of the orbicularis oculi activation.

either the video imagery or the sound track, and was therefore related to information flow. Based on their results, and following Evinger's results., Nakano and Kitazawa hypothesized that humans share a mechanism for regulating blinks which naturally times the motion to minimize the risk of losing critical information.

In another study, Cummins investigated the interpersonal regulation of gaze and blinking in dyadic conversation [21] and found that individuals display markedly different patterns of covariation in blink and gaze patterns. Though participants showed systematic modulation of blink and gaze behavior as a function conversation coordination (speaking and gaze state), individual subject behaviors varied greatly. They concluded that a participant's blinking and gaze pattern during dyadic conversations exhibits an "idiosyncratic individual style" that characterizes their communicative style. Their results indicate that blinks could serve a communicative function at the message level and as an individual difference.

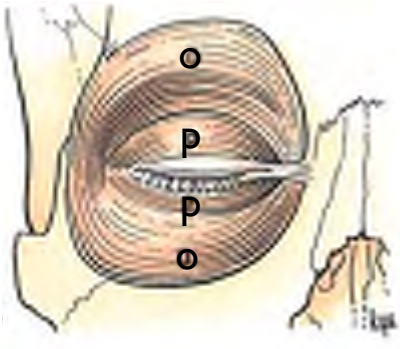


Figure 5.3: The orbicularis oculi muscle is associated with both blinking and smiling. Blinking is a result of the palpebral part (P) contraction and the Duchenne marker observed in spontaneous smiles is a result of the orbital part (O) contraction. Figure adapted from *Grant's Atlas of Anatomy* [1].

5.1.2 Connections between blinks and smiles

Though blinks and smiles are apparently unrelated facial movements, under certain circumstances they are activated by a common muscle: the orbicularis oculi. A basic smile is represented by the raising of the mouth corners in a U shape with the zygomaticus major muscle; enjoyment smiles are further distinguished by a slight wrinkling around the exterior corner of the eye produced by the activation of the orbicularis oculi muscle. This wrinkling, known as the Duchenne marker, is similar in appearance to slight squinting (Figure 5.2) [33]. The orbicularis oculi is also the principal muscle that affects voluntary or involuntary eye closing during blinking [35]. However, different sub-parts of the orbicularis oculi (Figure 5.3) are generally associated with blinking (orbicularis oculi palpebral part) and the Duchenne marker (orbicularis oculi orbital part).

Dibeklioglu and colleagues provide indirect evidence that blinks increase during spontaneous smiles [24]. Their study showed that distance-based eyelid features can classify smiles as spontaneous or posed. The eyelid features (average and standard deviation of opening distance) were computed for short videos of posed and spontaneous smiles. Though their study does not explicitly search for a connection between blinks and smiles, the authors noted that spontaneous smile sequences exhibited a larger number of blinks than posed smiles sequences [24].

We hypothesize that during spontaneous smiles, blinks are more likely to occur at either the onset or the offset of the smile. The release of a spontaneous smile can be associated with a shift in attention that can be consistent with the low disruption hypothesis described earlier.

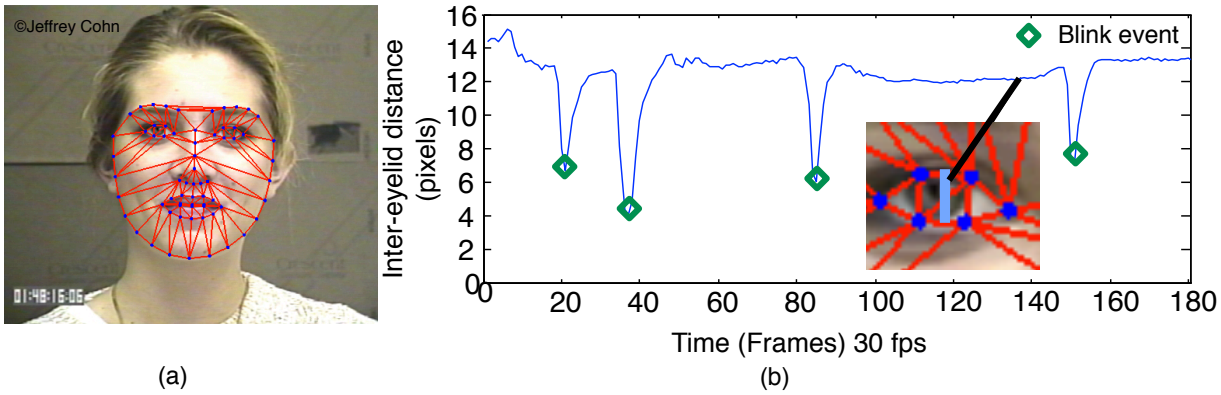


Figure 5.4: (a) We used the Active Appearance Models [67] algorithm to track the face and identify eye blinks in the videos. (b) Sample sequence of inter-eyelid distance and detected blinks.

5.2 Approach

To understand how eye blinks synchronize with spontaneous smiles, we analyzed 43 videos of smiles from female participants in the Cohn-Kanade Facial Expression database [50, 61]. In the study, participants were recorded performing facial expressions of basic emotions and completing directed facial action tasks (Action Units). Participants conversed with the experimenter and acknowledged her presence by completing the requested task. Though they were aware that they were being recorded, the participants were unaware that their blinking or smiles were the focus of the study.

We selected videos of spontaneous smiles that occurred during the recording process. We defined spontaneous smiles as facial expressions that include skin deformations characteristic of zygomaticus major muscle activation, such as up-turning lip corners and rising cheeks, and which were not directly prompted by the experimenter.

Each video was recorded at 30 frames per second. Criteria for further inclusion were (a) at least 30 seconds of video data before and after the smile (approximately 1 minute of data), (b) absence of facial occlusion, and (c) absence of image artifacts (e.g., camera motion). Twenty-seven videos from 27 individuals met these criteria. Similar to Nakano and Kitazawa [71], we excluded from the results participants whose mean blink rate was greater than one standard deviation from the mean. The final data set consisted of 22 videos of spontaneous smiles.

5.2.1 Video annotation

To detect eye blinks in the videos, we used a computer vision algorithm, Active Appearance Models (AAMs) [67]. Each smile was tracked with AAM, using a 66-point face model individually trained for each subject, as shown in Figure 5.4.a. We then used the AAM data to detect when blinks occurred during each video. For each frame in the video, we computed the absolute distance between the midpoint of the upper and lower eyelids, similar to the method used in Chapter 2. Local minima in the inter-eyelid distance time series corresponds to the eyelids being at their closest point. This event was labeled as a blink (Figure 5.4.b).

Next, we manually annotated each video to mark the start and end of the smile. The start of the smile was determined to be the frame in which the lip corners start moving upward. For the end of the smile, we annotated the last frame of lip corner movement. Though onset and offset are generally time intervals, throughout the rest of the chapter we will refer to the instant in time when a smile starts as onset and the end of the smile as offset. The annotators were qualified in recognizing AUs as proposed by Ekman [32]. Average smile duration was 81 frames (2.7 seconds). Figure 5.1 shows a representative example of smile dynamics dynamics by plotting the lip corner absolute distance..

5.2.2 Data analysis

We measured whether eye blinks are correlated to smile onset and offset. To examine the temporal relationship between eye blinks and smiles, we computed temporal distances (in frames) between smile onset or smile offset and blinks immediately before or after those points. Our hypothesis is that eye blinks will occur closer than expected to smile onset and offset. To test our hypothesis, we compare the values measured in the original data (measured value) with values computed for a sequence of blink time series not correlated with the smile event (expected value).

To compute the expected value for the blink-smile event distance, we used the approach of Nakano and Kitazawa [71]. Their approach decorrelates the blink signal from the smile event yet preserves the inter-eye blink interval (IBI) distribution for each participant. We computed expected values for each participant independently because blink distribution patterns are idiosyncratic (Figure 5.5). IBIs are measured by computing the distance in frames between consecutive eye blinks in the video data, as discussed in Section 5.2.1. Figure 5.6 illustrates how the expected value for blinks occurring immediately before smile onset is computed using surrogate data.

Nakano and Kitazawa [71] measured the correlation between two time series of eye blink data by comparing measured values with surrogate, uncorrelated time series. We adapted their method and generated surrogate blink time series decorrelated from smile data. For each participant, we

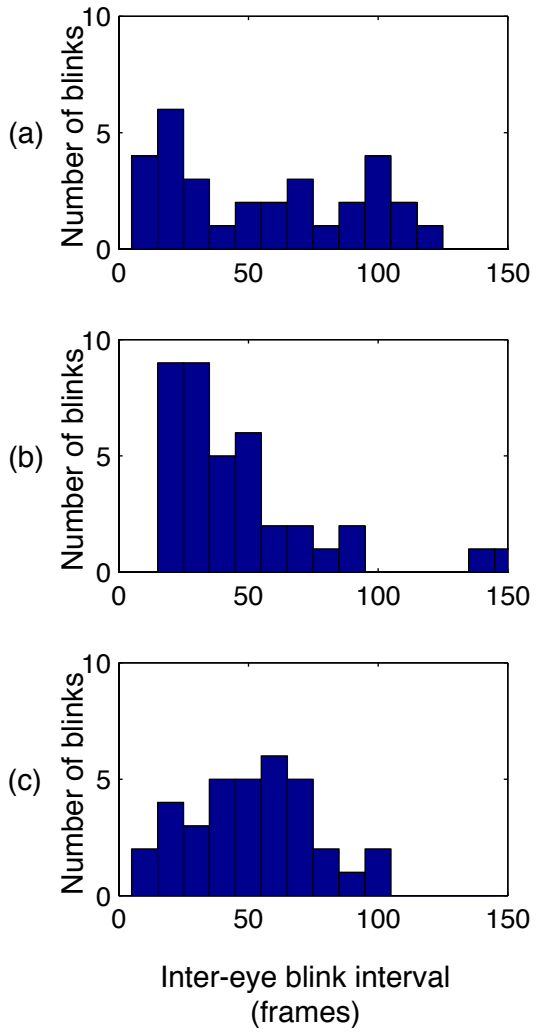


Figure 5.5: Distributions of inter-eye blink intervals for three participants: (a) bi-modal distribution, (b) J-shaped distribution, and (c) Gaussian distribution.

first generated 1000 surrogate time series by randomly reordering the original IBIs for that individual (Figure 5.6). The randomized time series preserve the distribution of the original IBIs but remove the causal relationships with participant facial expressions, particularly smiles. We then compute blink-smile distances for the randomized time series. The average blink-smile distance in the 1000 randomized time series is the expected value for the null hypothesis.

Nakano and Kitazawa [71] analyzed longer blink time sequences, and we needed to determine whether their analysis could be used for shorter sequences. Their IBI time series were recovered from three-minute long recordings of participants viewing videos on a computer monitor. In the current study, the available video data was one minute surrounding the smile.

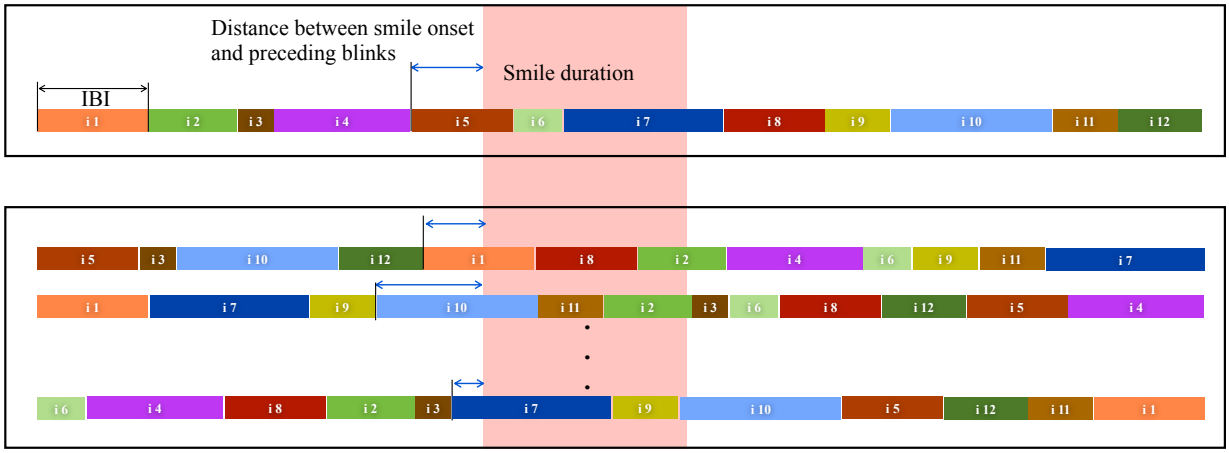


Figure 5.6: Top: Original inter-eye blink time series. Bottom: In order to calculate the expected value for the blink-smile event distance, we create surrogate data by randomizing the original inter-eye blink time series similar to the method proposed by Nakano and Kitazawa [71].

The blink frequency in the current study was slightly higher (33 blinks per minute, standard deviation of 13) than the blink rate reported in the Nakano study (25 blinks per minute, standard deviation of 16) [71]. This difference, though not statistically significant (independent samples t-test p value of .0807), is consistent with the difference in activity for the participants (light conversation in our study vs. viewing video in the Nakano study) [28].

5.3 Results

For both smile onset and offset, four two-tailed paired t-tests (confidence interval of 95%) were computed between the measured and expected value for blink-smile distance for both smile onset and offset. Consistent with our hypothesis, eye blinks occurring before smile offset were significantly closer to the offset than expected values (M1 $p = .016$, Cohen's d:.768) than expected values. Eye blinks occurred in close proximity to the smile offset, specifically at an average frame distance of -21 frames. However, for the de-correlated blink sequence the expected value is -38 frames. The Cohen's d coefficient was computed with pooled standard deviations for the sample population, which indicates that the population size was sufficient. The results for blink-smile offset distances are shown in Figure 5.7.

Blinks occurred before smile onsets at a distance of -51 frames. This effect is equivalent to a suppression effect because the expected value was -38. This result is marginally significant ($p = .098$). For a non-correlated blink-smile sequence we can expect a blink to happen 38 frames before the smile onset (Figure 5.7).

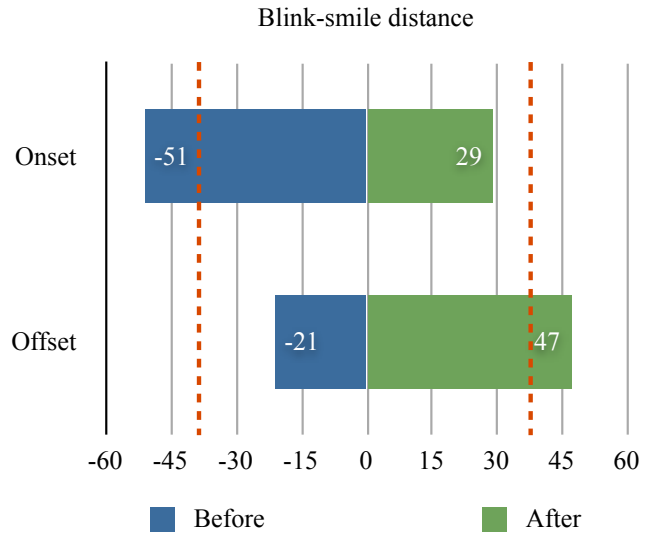


Figure 5.7: Average distance in frames between a smile onset, offset, and blinks. The average expected value for the distance between an uncorrelated smile event and a blink is shown by the dotted line (38 frames).

The distance between a smile onset and an immediately following blink was measured to be 29 frames compared to the 38 frames for the expected distance, but the difference was not significant ($p = .127$). Similarly, the measured distance between the smile offset and the following blink was 46 frames and not significantly different ($p = .244$) from the expected value of 38 frames.

Our results suggest a correlation between smile events (onset and offset) and eye blinks. Consistent with our hypothesis, eye blinks occurred closer to the smile offset and immediately preceding it. In addition, as a marginally significant effect, blinks were suppressed before smile onset. These combined results suggest that the eye blink sequence is modulated such that blinks that would have occurred shortly before a smile are postponed until the end of the smile. This effect is local, and the average IBI in the 30 seconds before the smile (67 frames) is not significantly different ($p = .158$) than the average IBI after the smile offset (51 frames).

5.4 Discussion and future work

Our analysis demonstrates a connection between two discrete spontaneous facial movements: smiles and blinks. We suspect that eye blinks may in fact punctuate the end of facial expressions. An extension of this work should include correlation analyses for other facial expressions.

There may be other intermediate correlations with either smile offset or blinks. Head motion may also play a role in the blink-smile relationship. The correlation between head motion and smiles was investigated by Cohn and colleagues with moderate results (correlation coefficients of +/- 0.36 to 0.50) [15]. Similarly, blinks have been correlated with head motions of certain amplitudes [76].

A limitation of this study is that only female participants took part in the recording. Gender differences have been previously established in the communicative gaze patterns of males and females [85]. It is therefore possible that blinking, which can interact with the gaze pattern shows similar gender differences during smiles. A future study should investigate the effect of gender by analyzing a set of smiles recorded from a gender-balanced participant pool.

Another continuation of this work is to further segment smile dynamics and examine the relationship with the peak of the smile. This segmentation is difficult because in many spontaneous smiles suppression actions occur. A continuous quantification of smile dynamics, similar to the one proposed in Figure 5.1, could better support the correlation analysis.

Similarly, further analysis is needed to analyze gaze-blink-smile correlations. Several studies have shown that gaze and eye blinks are tightly correlated [23]. Gaze is also a very informative non-verbal cue [70] that may show a particular pattern during smiling. Subjective observations indicated that there was no clear pattern of gaze during smiling in our dataset. Stifter and Moyer found that infants avert their gaze more frequently following high-intensity smiles than low-intensity smiles, and they hypothesize that gaze aversion is the infants' way of lowering their positive arousal when it gets too high [83]. Blinking may be a similar form of positive arousal control, with the implication that blinks occur towards the end of high-intensity/high-arousal smiles.

Chapter 6

Conclusion

The overarching goal of this thesis is to guide and inform animation practice for artists and developers. To achieve this goal, we explored expression-specific models for two key facial expressions: blinks and smiles. For developers, we provide generative statistical models for blinks and smiles as well as an approach to derive new models. For traditional animators, the models provide useful alternatives to commonly used ease-in/ease-out interpolation functions. These models are primarily aimed to create realistic motion for photorealistic characters.

We conclude with a summary of our contributions, limitations, and avenues for future research.

6.1 Contributions

The primary contributions of this thesis can be summarized as follows:

1. Perceptual results: The numerous studies presented leverage viewers' perception in determining the parameters for realistic animations. The experiments explore parameters such as temporal symmetry, duration, motion amplitude, or spatial and temporal nonlinearity. Based on viewers' ratings, we incorporate the most perceptually significant information into our models. For example, from our data we find it is not necessary to preserve original high-resolution geometry data as long as the model accurately captures temporal information. Furthermore, our work challenges common assumptions that over-simplify expression dynamics; we find that some guidelines in animation textbooks insufficiently or inaccurately describe the complex motion in both blinks and smiles.

2. Models: We introduced a method to generate parsimonious data-driven models for blinks and smiles, and showed that these data-derived models are perceptually better than current heuristics. Our data analysis and models confirm that spontaneous blinks and smiles

are highly complex, coordinated activities that exhibit characteristic motion profiles. The models, data-driven interpolation functions, aim to bridge the gap between performance capture and traditional animation methods. In particular, we use the temporal data to accurately model, and thus preserve, realistic acceleration and deceleration of facial expressions, as well as motion complexity. Additionally, the generative quality of our models can be used to add variability when creating several motions at a time, such as when animating secondary characters.

3. Through data analysis, we found a pattern for the temporal relationship between blinks and smiles: blinks occur close to the end of the smile, before the lips stop moving downward.

4. The framework: The broader impact of this thesis is to initiate research in expression specific data-driven models of facial expressions. We argue that facial animation needs a generative basis of actions that includes temporal information. This position is supported by research in other fields. Speech animation research, for example, has shown that phonetic units are better represented as short dynamic lip motions rather than static mouth poses [12, 87, 97]. We envision a similar development in facial animation research with dynamic facial units replacing static blendshape interpolation.

For traditional animators, the most important results of our studies are summarized in the following guidelines:

- When animating eye blinks, the eyes should close in a fast, linear, motion and open in a slower motion with very slow motion in the last frames. The motion for the opening of the eyes resembles a logarithmic curve. At 30 frames per second, nine-frames blinks are most common though blink duration ranges from seven to 13 frames. For the nine-frame blink, three frames can be used for closing and six for the slow opening. It is important to avoid ease-in/ease-out interpolation, particularly at the three-frame mark where the eyes close, because the eyes appear closed for longer than necessary.

- Spontaneous genuine smiles are generally longer (3-5) seconds than posed smiles with dynamics that resemble laughter: the smile should exhibit some motion complexity rather than smoothly going from neutral to peak to neutral. Based on our data, we speculate that animators should add multiple peaks to a smile expression to create a genuine smile. Head motion is important for spontaneous smiles and without any head motion smiles frequently appear posed. Because head motion is to some degree correlated to the smile dynamics, we propose a plausible head motion proportional to the smile amplitude.

- Eye blinks should occur well in advance of the start of the smile (2 seconds) and very close to the smile end (less than half a second), before the face reaches a fully neutral position.

6.2 Future directions

Chapters 2-5 conclude with a discussion of specific limitations and future work for each study. In this section, we discuss broad avenues of future work at the perceptual methodology and model levels.

6.2.1 Perceptual methodology

Perceptual research manipulates only a small number of variables at a time and, as such, many factors remain constant. Our studies tested several variables that can influence the perception of blinks and smiles (e.g., symmetry, duration, or spatial and temporal nonlinearity). However, there are many other potential variables. For example, we limited the number of computer-generated characters in our animations to minimize the difference in static appearance between animations and videos. Further, we were limited by the availability of high-quality photorealistic characters and focused on animations for two CG characters: a male and a female. More CG characters, particularly photorealistic characters matched to our actors, will generate more robust results.

Throughout this dissertation, we were primarily interested in whether blinks and smiles appear realistic. In this case, realism was measured through subjective ratings of naturalness and genuineness. Additional dependent variables could have included rating an animation's accuracy relative to video. This measure is suitable for determining how close ground truth animations match videos but is less applicable to newly generated animations without a video reference. Two-alternative forced choice experiments, in which viewers choose the more realistic of two animations, may provide a richer and potentially stronger differentiation between animation conditions. However, forced choice experiments require numerous trials to compare all possible combinations, which can be time prohibitive. A complimentary approach to evaluating perceptual research in facial animation should include objective measures. Facial expressions can exhibit a social mimicry effect [26]. An objective measure of the realism of facial expression could be based on how viewers' expressions change while observing various animations of smiles. More complex tasks would involve evaluating the degree to which animated characters with genuine smiles convey information, for example.

Future research should investigate how to best apply data derived from human actors to non-photorealistic CG characters, and the degree to which this is appropriate and beneficial. Our

study achieved mixed results for blinks and smiles. With blinks, we obtained similar results for both the photorealistic and non-photorealistic character. However, for smiles, model animations were rated as highly genuine only for the photorealistic characters.

More data is needed to calibrate the physical appearance and appeal of the CG characters used. As we discussed in Chapter 3, the female character may have appeared less photorealistic than the male character. This difference may have affected the results by biasing users to rate her smile animations as less genuine. More research is needed to determine how realism in appearance and dynamics influence a viewer's perception. This interaction is important because various levels of realism, either in appearance or dynamics, may be required for different applications using CG characters.

Future work should also consider gender and cultural differences in raters, CG characters, and actors. A study focused on those variables would require significantly more data. We analyzed the blinks and smiles of a relatively small number of actors. However, individuals showed similar dynamics for their expressions. For example, the blink profiles were consistent across three participants and similar to results in previous literature.

As we reach a high level or realism with facial expression, more research that relies on nonverbal behavior theory and practices will help create CG characters that effectively communicate and convey information. For example, the timing of certain expressions could influence how viewers perceive the personality of a character. Isbister and Nass [45] found that viewers can identify the personality of a character from nonverbal cues such as posture. Furthermore, consistent nonverbal cues for the CG character more strongly influenced peoples' behavior.

Context also significantly influences viewers' interpretation of facial expression. Because facial expressions are part of a rich set of nonverbal cues, such as eye motion, head motion, and body pose, presenting animations in a richer context with appropriately modeled motions for each of these cues should advance our understanding of realistic motion. Hodgins and colleagues [43] took a step towards better understanding the relative importance of different anomalies in full body character animation. The authors started with high-resolution motion capture data and iteratively introduced anomalies for either the eyes, facial expressions, head pose, or hand motions. Their study underlines the difficulty in considering the perceptual effects of different parameters when dealing with a complex scene. Both their work and our observations emphasize that eye gaze is a particularly powerful cue that needs further study. Eye-tracking data is facilitating the development of eye gaze models for CG characters [4, 23, 56] and ideally accurately animated eye motion should be part of all perceptual experiments on facial expressions.

6.2.2 Improving the models

Our model of smile dynamics is based on spontaneous smiles. However, many different types of smiles have been identified. These smiles vary in terms of communicative intent, context, shape and motion characteristics. A limitation of the current work is that it distinguishes only between spontaneous and posed. It is conceivable that, within the spontaneous smile category, different classes of smiles could exhibit different dynamics and could therefore be better modeled independently. For example, differences in duration, velocity, symmetry, and amplitude have been found across smiles perceived as amused, polite, and embarrassed [3]. It is particularly interesting whether the multiple peaks we observed in our spontaneous smiles can be better characterized in multiple categories of smiles. More research is needed to explore and build models for these different types of smiles.

Similarly, for eye blinks it would be beneficial to consider whether blink dynamics vary across tasks and emotional state. It may be possible that eye blink duration is influenced by the experienced emotional state. For example, shorter eye blinks could be more common when an individual is fearful.

A data-driven model for the head motion, rather than our method for creating a plausible head motion correlated with the smile, could provide more insight into the relationship between head motions and smiles. A promising avenue is to generate joint models for head motion and facial expressions. Candidates for how to generate these joint models include Hidden Markov Models and dynamic Bayesian networks which have been used to generate joint models for head motions and prosodic features. For example, Busso and colleagues [10] modeled the temporal dynamics for four categories of emotional head motion (sadness, happiness, anger and neutral state) driven by prosodic features. The authors generated head motion sequences by interpolating between discrete poses that corresponded to states in a Hidden Markov Model. More recently, Mariooryad and Busso [66] used dynamic Bayesian networks to generate head and eyebrow motions based on speech input.

Different head motion models may be required for different classes of smiles. Ambadar and colleagues [3] found that perceived embarrassed/nervous smiles were accompanied by greater downward head movements that perceived amused and polite while motions in the other axes, yaw and roll, did not vary across smile types. This result emphasizes the need to consider models for more discrete categories of smiles.

A continuation of this dissertation work would consider how to integrate the current smile model with other expressions or speech occurring at the same time. The expressions we analyzed

and modeled were considered in isolation without additional facial expression. High-resolution smiles were selected to start and end in relatively neutral expressions and to not contain verbal utterances. We made this choice in order to produce a consistent dataset because other utterances and expressions add variability. The next question that arises is how to integrate our smile model with speech or other facial expressions.

One approach to integrating the smile model with speech is to make use of localized blendshapes, which are common in facial animations systems. These localized deformations can correspond to muscle groups or, in the case of blendshapes for speech, to phonemes. Our model would have to be adapted and decomposed into localized blendshapes to work with existing blendshapes. We believe this is feasible because the motion in smiles is dominated by zygomaticus muscle activation pulling the cheeks diagonally and the masseter muscle lowering and closing the jaw, which are all localized deformations. Alternatively, given a set of localized blendshapes and a collection of smiles it would be interesting to determine the data-driven interpolation functions that best represent the smiles given the entire blendshape basis. This process could be conducted by optimizing the blendshapes to fit given expressions and analyzing the resulting interpolation functions for each blendshape.

An even more pressing need for understanding facial expression is for rich datasets with high-resolution deformations. The success of expression-specific models is based on how much high-resolution data is available to train the models. We were able to collect substantial data because we limited our work to only two expressions. As the technology improves, markerless facial motion capture may be the answer in acquiring sufficient high-resolution data. For a basis for facial expressions beyond the two we studied, a much larger collection of spontaneous expressions is needed.

The primary research question we addressed in this thesis is how to build parsimonious models for two frequently occurring expressions, blinks and smiles. Based on data-analysis followed by perceptual experiments we found that models need to preserve temporal information. Our models for data-driven interpolation functions can be used by animators as a better alternative to generic ease-in/ease-out interpolation functions. Furthermore, we studied other variables of the models (temporal symmetry, spatial and temporal nonlinearities) that provide insights into perceptually valid facial animation.

Bibliography

[1] Agur, A. M. and Dalley, A. F. (2009). *Grant's atlas of anatomy*. Lippincott Williams & Wilkins.

[2] Akhter, I., Simon, T., Khan, S., Matthews, I., and Sheikh, Y. (2012). Bilinear spatiotemporal basis models. *ACM Transactions on Graphics*, 31(2):17:1–17:12.

[3] Ambadar, Z., Cohn, J. F., and Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, 33(1):17–34.

[4] Andrist, S., Pejsa, T., Mutlu, B., and Gleicher, M. (2012). Designing effective gaze mechanisms for virtual agents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 705–714.

[5] Bacher, L. F. and Smotherman, W. P. (2004). Spontaneous eye blinking in human infants: A review. *Developmental Psychobiology*, (44):95–102.

[6] Bacivarov, I., Ionita, M., and Corcoran, P. (2008). Statistical models of appearance for eye tracking and eye-blink detection and measurement. *IEEE Transactions on Consumer Electronics*, pages 1312–1320.

[7] Baker, S., Gross, R., and Matthews, I. (2003). Lucas-Kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

[8] Bickel, B., Botsch, M., Angst, R., Matusik, W., Otaduy, M., Pfister, H., and Gross, M. (2007). Multi-scale capture of facial geometry and motion. In *ACM Transactions on Graphics*, volume 26, page 33.

[9] Blount, W. P. (1928). Studies of the movements of the eyelids of animals: Blinking. *Quarterly Journal of Experimental Physiology*, (18):111–125.

[10] Busso, C., Deng, Z., Grimm, M., Neumann, U., and Narayanan, S. (2007). Rigid head motion in expressive speech animation: Analysis and synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1075–1086.

[11] Caffier, P. P., Erdmann, U., and Ullsperger, P. (2003). Experimental evaluation of eye-blink

parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89(3–4):319–325.

[12] Caldognetto, E. M., Zmarich, C., Cosi, P., and Ferrero, F. (1997). Italian consonantal visemes: Relationships between spatial/temporal articulatory characteristics and coproduced acoustic signal. In *Audio-Visual Speech Processing: Computational & Cognitive Science Approaches*, pages 5–8.

[13] Clinton, P. (2010). Review: 'Polar Express' a creepy ride. `http://articles.cnn.com/2004-11-10/entertainment/review.polar.express_1_polar-express-film-series-sensors?_s=PM:SHOWBIZ`. This is an electronic document. Date of publication: November 10, 2004. Date retrieved: September 9, 2012.

[14] Codispoti, M., Bradley, M. M., and Lang, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, (38):474–478.

[15] Cohn, J. F., Reed, L. I., Moriyama, T., Xiao, J., Schmidt, K., and Ambadar, Z. (2004). Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. *IEEE Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 129–135.

[16] Cohn, J. F. and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *Journal of Wavelets, Multi-resolution and Information Processing*, 2:1–12.

[17] Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685.

[18] Cosker, D., Krumhuber, E., and Hilton, A. (2010). Perception of linear and nonlinear motion properties using a FACS validated 3D facial model. In *Proceedings of the 7th ACM Symposium on Applied Perception in Graphics and Visualization*, pages 101–108.

[19] Cula, O. G., Dana, K. J., Murphy, F. P., and Rao, B. K. (2005). Skin texture modeling. *International Journal of Computer Vision*, 62(1-2):97–119.

[20] Culhane, S. (1988). *Animation from script to screen*. St. Martin's Press.

[21] Cummins, F. (2011). Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. *Language and Cognitive Processes*.

[22] Darwin, C. (1872/2009). *The Expression of the Emotions in Man and Animals*. Harper Perennial, anniversary edition.

[23] Deng, Z., Lewis, J. P., and Neumann, U. (2005). Automated Eye Motion Using Texture Synthesis. *IEEE Computer Graphics and Applications*, 25(April):24–30.

[24] Dibeklioglu, H., Valenti, R., Salah, A. A., and Gevers, T. (2010). Eyes do not lie: spontaneous versus posed smiles. *Proceedings of the International Conference on Multimedia*, pages 703–706.

[25] DiLorenzo, P. C., Zordan, V. B., and Sanders, B. L. (2008). Laughing out loud: Control for modeling anatomically inspired laughter using audio. In *ACM Transactions on Graphics*, volume 27, page 125.

[26] Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, 11(1):86–89.

[27] Donner, C., Weyrich, T., d'Eon, E., Ramamoorthi, R., and Rusinkiewicz, S. (2008). A layered, heterogeneous reflectance model for acquiring and rendering human skin. *ACM Transactions on Graphics*, 27(5):140:1–140:12.

[28] Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: During reading and video display terminal use, in primary gaze, and while in conversation. *Optometry & Vision Science*, 78(10).

[29] Dryden, I. L. and Mardia, K. V. (2002). *Statistical Shape Analysis*. John Wiley and Sons.

[30] Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised and updated edition)*. W. W. Norton & Company, 2 rev sub edition.

[31] Ekman, P., Davidson, R. J., and Friesen, W. (1990). The Duchenne smile: emotional expression and brain physiology. *Journal of Personality and Social Psychology*, 58(2):342–53.

[32] Ekman, P. and Friesen, W. (1978). *Facial action coding system: A technique for the measurement of facial movement.* Consulting Psychologists Press, Palo Alto.

[33] Ekman, P. and Friesen, W. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6(4):238–252.

[34] Evinger, C., Manning, K. A., and Sibony, P. A. (1991). Eyelid movements mechanisms and normal data. *Investigative Opthalmology & Visual Science*, 32(2):387–400.

[35] Evinger, C., Shaw, M. D., Peck, C. K., Manning, K. A., and Baker, R. (1984). Blinking and associated eye movements in humans, guinea pigs, and rabbits. *Journal of Neurophysiology*, 52(2):323–339.

[36] Feingold, A. and Mazzella, R. (1993). Preliminary validation of a multidimensional model of wittiness. *Journal of Personality*, 61(3):439–456.

[37] Flach, L. M., de Moura, R. H., Musse, S. R., Dill, V., Pinho, M. S., and Lykawka, C. (2012). Evaluation of the uncanny valley in CG characters. *Proceedings of SBGames*.

[38] Gallagher, D. F. (2007). Digital actors in 'Beowulf' are just uncanny. `http://bits.blogs.nytimes.com/2007/11/14/digital-actors-in-beowulf-are-just-uncanny/`. This is an electronic document. Date of publication: November 14, 2007. Date retrieved: September 9, 2012.

[39] Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applica-*

tions, 28(4):11–17.

[40] Guitton, D., Simard, R., and Codere, F. (1991). Upper eyelid movements measured with a search coil during blinks and vertical saccades. *Investigative Ophthalmology & Visual Science*, 32(13):3298–3305.

[41] Hess, U. and Kleck, R. E. (1994). The cues decoders use in attempting to differentiate emotion-elicited and posed facial expressions. *European Journal of Social Psychology*, 24(3):367–381.

[42] Ho, C.-C. and MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518.

[43] Hodgins, J., Jörg, S., O'Sullivan, C., Park, S. I., and Mahler, M. (2010). The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception*, 7(4):1–14.

[44] Hyde, J., Carter, E., Kiesler, S., and Hodgins, J. (2013). Perceptual effects of damped and exaggerated facial motion in animated characters. In *Proceedings of the* 10^{th} *IEEE International Conference on Automatic Face and Gesture Recognition*.

[45] Isbister, K. and Nass, C. (2000). Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, 53(2):251–267.

[46] Jimenez, J., Scully, T., Barbosa, N., Donner, C., Alvarez, X., Vieira, T., Matts, P., Orvalho, V., Gutierrez, D., and Weyrich, T. (2010). A practical appearance model for dynamic facial color. *ACM Transactions on Graphics*, 29(6):141:1–141:10.

[47] Josh, N. K., Mcdermott, J., and Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17:4302–4311.

[48] Kalra, P., Mangili, A., Magnenat-Thalmann, N., and Thalmann, D. (1991). Smile: A multilayered facial animation system. *Modeling in Computer Graphics*, 199(1).

[49] Kalwick, D. (2006). *Animating facial features & expressions, (Graphics Series)*. Charles River Media, Inc.

[50] Kanade, T., Cohn, J. F., and Tian, Y. (2000). Comprehensive database for facial expression analysis. *Robotics*, 4:46–53.

[51] Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20:165–182.

[52] Krumhuber, E. and Kappas, A. (2005). Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior*, 29(1):3–24.

[53] Krumhuber, E. G. and Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion*, 9(6):807–820.

[54] Lasseter, J. (1987). Principles of traditional animation applied to 3D computer animation. *Computer Graphics*, pages 35–44.

[55] Leal, S. and Vrij, A. (2008). Blinking during and after lying. *Journal of Nonverbal Behavior*, 32(4):187–194.

[56] Lee, S. P., Badler, J. B., and Badler, N. I. (2002). Eyes alive. *ACM Transactions on Graphics*, 21(3):637–644.

[57] Li, H., Weise, T., and Pauly, M. (2010). Example-based facial rigging. *ACM Transactions on Graphics*, 29(4):32.

[58] Liu, X., Xia, S., Fan, Y., and Wang, Z. (2011). Exploring non-linear relationship of blendshape facial animation. In *Computer Graphics Forum*, volume 30, pages 1655–1666.

[59] Lochman, J. E. and Allen, G. (1981). Nonverbal communication of couples in conflict. *Journal of Research in Personality*, 15(2):253 – 269.

[60] Long, T. (2009). Review: Creepy 3-D 'A Christmas Carol' deserves a lump of coal. `http://www.detroitnews.com/article/20091106/OPINION03/911060320/1034/ent02/Review--Creepy-3-D--A-Christmas-Carol--deserves-a-lump-of-coal`. This is an electronic document. Date of publication: November 6, 2009. Date retrieved: September 9, 2012.

[61] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis*, pages 94–101.

[62] MacDorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, 25(3):695–710.

[63] Maestri, G. (1996). *Digital character animation*. New Riders Publishing.

[64] Magnenat-Thalmann, N., Kalra, P., Luc Leveque, J., Bazin, R., Batisse, D., and Querleux, B. (2002). A computational skin model: Fold and wrinkle formation. *IEEE Transactions on Information Technology in Biomedicine*, 6(4):317 –323.

[65] Magnenat-Thalmann, N., Primeau, E., and Thalmann, D. (1988). Abstract muscle action procedures for human face animation. *The Visual Computer*, 3(5):290–297.

[66] Mariooryad, S. and Busso, C. (2012). Generating human-like behaviors using joint, speech-driven models for conversational agents. *IEEE Transactions on Audio, Speech, and Language*

Processing, 20(8):2329–2340.

[67] Matthews, I. and Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision*, 60:135–164.

[68] McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render me real?: Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, 31(4):91:1–91:11.

[69] Mori, M. (1970). The Uncanny Valley. *Energy*, 7(4):33–35.

[70] Mutlu, B., Shiwa, T., Kanda, T., and Ishiguro, H. (2009). Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Human-Robot Interaction*, volume 2, pages 61–68.

[71] Nakano, T. and Kitazawa, S. (2010). Eyeblink entrainment at breakpoints of speech. *Experimental Brain Research*, 205(4):577–81.

[72] Ochs, M., Niewiadomski, R., Brunet, P., and Pelachaud, C. (2011). Smiling virtual agent in social context. *Cognitive Processing*, pages 1–14.

[73] Parke, F. I. (1972). Computer generated animation of faces. In *Proceedings of the ACM Annual Conference-Volume 1*, pages 451–457.

[74] Pasquariello, S. and Pelachaud, C. (2001). Greta: A simple facial animation engine. In *6th Online World Conference on Soft Computing in Industrial Appications, Session on Soft Computing for Intelligent 3D Agents*.

[75] Patel, N. M. and Zaveri, M. (2010). Parametric facial expression synthesis and animation. *International Journal of Computer Applications IJCA*, 3(4):34–40.

[76] Peters, C. and O'Sullivan, C. (2003). Attention-driven eye gaze and blinking for virtual humans. In *ACM SIGGRAPH Sketches & Applications*, pages 1–1, New York, NY, USA.

[77] Ponder, E. and Kennedy, W. (1927). On the act of blinking. *Experimental Physiology*, 18(2):89.

[78] Porter, S. and ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 19(5)(5):508–514.

[79] Schmidt, K. L., Ambadar, Z., Cohn, J. F., and Reed, L. I. (2006a). Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *Journal of Nonverbal Behavior*, 30(1):37–52.

[80] Schmidt, K. L., Liu, Y., and Cohn, J. F. (2006b). The role of structural facial asymmetry in asymmetry of peak facial expressions. *Laterality*, 11(6):540–61.

[81] Seyama, J. and Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*,

16(4):337–351.

[82] Steptoe, W., Oyekoya, O., and Steed, A. (2010). Eyelid kinematics for virtual characters. *Computer Animation and Virtual Worlds*, 21(3-4):161–171.

[83] Stifter, C. A. and Moyer, D. (1991). The regulation of positive affect: Gaze aversion activity during mother-infant interaction. *Infant Behavior and Development*, 14(1):111 – 123.

[84] Sun, W. S., Baker, R. S., Chuke, J. C., Rouholiman, B. R., Hasan, S. A., Gaza, W., Stava, M. W., and Porter, J. D. (1997). Age-related changes in human blinks. *Investigative Opthalmology & Visual Science*, 38(1):92–99.

[85] Swaab, R. I. and Swaab, D. F. (2009). Sex differences in the effects of visual contact and eye contact in negotiations. *Journal of Experimental Social Psychology*, 45(1):129–136.

[86] Tanaka, Y. and Yamaoka, K. (1993). Blink activity and task difficulty. *Perceptual and Motor Skills*, 77(1):55–66.

[87] Taylor, S. L., Mahler, M., Theobald, B.-J., and Matthews, I. (2012). Dynamic units of visual speech. In *ACM/Eurographcs Symposium on Computer Animation*.

[88] Tena, J. R., De la Torre, F., and Matthews, I. (2011). Interactive region-based linear 3D face models. In *ACM Transactions on Graphics*, volume 30, page 76.

[89] Trutoiu, L. C., Carter, E. J., Matthews, I., and Hodgins, J. K. (2011). Modeling and animating eye blinks. *ACM Transactions on Applied Perception*, 8(3):1–17.

[90] Trutoiu, L. C., Carter, E. J., Pollard, N., Cohn, J. F., and Hodgins, J. K. (2014). Spatial and temporal linearities in posed and spontaneous smiles. *ACM Transactions on Applied Perception*, 8(3):1–17.

[91] Trutoiu, L. C., Hodgins, J. K., and Cohn, J. F. (2013). The temporal connection between smiles and blinks. In *Proceedings of the* 10^{th} *IEEE International Conference on Automatic Face and Gesture Recognition*.

[92] VanderWerf, F., Brassinga, P., Reits, D., Aramideh, M., and de Visser, B. (2003). Eyelid movements: Behavioral studies of blinking in humans under different stimulus conditions. *Journal of Neurophysiology*, 89(5):2784–2796.

[93] Veltman, J. A. and Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5):656–669.

[94] Wallraven, C., Breidt, M., Cunningham, D. W., and Bülthoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. *ACM Transactions on Applied Perception*, 4(4):1–20.

[95] Williams, R. (2001). *The animator's survival kit*. Faber and Faber.

[96] Wilson, G. F., Purvis, B., Skelly, J., Fullenkamp, P., and Davis, I. (1987). Physiological

data used to measure pilot workload in actual flight and simulator conditions. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 31(7):779–783(5).

[97] Ypsilos, I. A., Hilton, A., Turkmani, A., and Jackson, P. J. B. (2004). Speech-driven face synthesis from 3D video. In *Proceedings of the 2nd IEEE International Symposium on 3D Data Processing, Visualization and Transmission*, pages 58–65.