STOP STARING

Facial Modeling
and Animation
Done Right

Third Edition

JASON
OSIPA

Stop Staring

Facial Modeling and Animation Done Right

THIRD EDITION

Stop Staring

Facial Modeling and Animation Done Right

THIRD EDITION

JASON OSIPA

WILEY

Wiley Publishing, Inc.

Dear Reader,

Thank you for choosing *Stop Staring: Facial Modeling and Animation Done Right, Third Edition.* This book is part of a family of premium-quality Sybex books, all of which are written by outstanding authors who combine practical experience with a gift for teaching.

Sybex was founded in 1976. More than 30 years later, we're still committed to producing consistently exceptional books. With each of our titles, we're working hard to set a new standard for the industry. From the paper we print on, to the authors we work with, our goal is to bring you the best books available.

I hope you see all that reflected in these pages. I'd be very interested to hear your comments and get your feedback on how we're doing. Feel free to let me know what you think about this or any other Sybex book by sending me an email at `nedde@wiley.com`. If you think you've found a technical error in this book, please visit `http://sybex.custhelp.com`. Customer feedback is critical to our efforts at Sybex.

Best regards,

Neil Edde
Vice President and Publisher
Sybex, an Imprint of Wiley

For my girls.

Acknowledgments

First and foremost, thank you to everyone at Wiley, who did most if not all of the work on this book.

Third edition: Mariann Barsolo, acquisitions editor; Kathryn Duggan, development editor; Christine O'Connor, Liz Britten, and Angela Smith, production editors; Paul Thuriot, technical editor; Judy Flynn, copyeditor; Jen Larsen, proofreader; Ted Laux, indexer.

Second edition: Willem Knibbe, acquisition editor; Jim Compton, development editor; Keith Reicher, technical editor; Rachel Gunn, production editor; Judy Flynn, copyeditor; Chris Gillespie, compositor; Jen Larsen, proofreader.

First edition: Pete Gaughan, development editor; Dan Brodnitz, associate publisher; Mariann Barsolo, acquisitions editor; Liz Burke, production editor; Keith Reicher, technical editor; Suzanne Goraj, copyeditor; Maureen Forys, compositor; Margaret Rowlands, cover coordinator; the CD team of Kevin Ly and Dan Mummert.

For helping with the book and bringing to it so much more than I could alone, I thank Juan Carlos Larrea and Jason Hopkins, animation; Chris Robinson, character design; Kathryn Luster, contact and casting; Chris Buckley, Craig Adams, Joel Goodsell, and Robin Parks for voice work; Jeremy Hall for Joel's recording.

Professionally, for supporting me and putting up with me, I thank Phil Mitchell, Owen Hurley, Jennifer Twiner-McCarron, Michael Ferraro, Ian Pearson, Chris Welman, Gavin Blair, Stephen Schick, Tim Belsher, Derek Waters, Sonja Struben, Glenn Griffiths, Chuck Johnson, Casey Kwan, Herrick Chiu, Chris Roff, and James E. Taylor. Thanks to all the good people at Surreal Software and everyone at Maxis/EA; the Sims EP team, the Sims 2 team, the Sims "next gen" team. Thanks to Glenn, Brian W., Paul L, Kevin, Clint, Ryo, Toru, Hakan, Frank, and Rudy; to Jesse, Lisha, and of course, the lovely miss Tee; to "fight club," my robots; to Andy, Sergey, Lucky, Yasushi, Daisuke, Paddy, and Brian Lee! To the best what-if team you could ever imagine: Paul, Brian, Jim, Matt A., Charles, Kelvin, Sean, Damon, Ian, Dale, Matthew, and Howard.

Mom, Dad, Veronica, Tom, Jorge, and all my great family in Winnipeg and Acapulco: I can never quite wait until the next time I get to see you; I'm always thinking of you. Thanks to my California family: you guys have enriched my life more than I tell you; Nick,

About the Author

Jason Osipa has been a working professional in 3D since 1997, touching television, games, direct-to-video, and film in both Canada and the United States. Carrying titles from modeler and animator to TD and director, he has seen and experienced the world of 3D content creation and instruction from all sides. Jason currently owns and

CONTENTS AT A GLANCE

operates Osipa Entertainment, LLC, offering contracting and consulting services for any kind of 3D production, including pipeline and tools design and sales as well as efficiency and workflow training in animation, modeling, and rigging.

Contents

Introduction

Animation has got to be the greatest job in the world. When you get started, you just want to do everything, all at once, but can't decide on one thing to start with. You animate a walk, you animate a run, maybe even a skip or jump, and it's all gratifying in a way people outside of animation may never be lucky enough to understand. After a while, though, when the novelty aspects of animation start to wear off, you turn deeper into the characters and find yourself wanting to learn not only how to move, but how to act. When you get to that place, you need more tools and ideas to fuel your explorations.

Animation is clearly a full-body medium, and pantomime can take years to master. The face, and subtleties in acting such as the timing of a blink or where to point the eyes, can take even longer and be more difficult than conquering pantomime. Complex character, acting, and emotion are almost exclusively focused in the face and specifically in the eyes. When you look at another person, you look at their eyes; when you look at an animated character, you look at their eyes too. That's almost always where the focus of your attention is whether you mean for it to be or not. We may remember the shots of the character singing and dancing or juggling while walking as amazing moments, but the characters we fall in love with on the screen, we fall in love with in close-ups.

Stop Staring is different than what you may be used to in a computer animation book. This is not a glorified manual for software; this is about making decisions, really learning how to evaluate contextual emotional situations, and choosing the best acting approach. You're not simply told to do A, B, and C; you're told *why* you're doing them, *when* you should do them, and then, *how* to make it all possible.

Why This Book

There is nothing else like *Stop Staring* available to real animators with hard questions and big visions for great characters. Most references have more to do with drawing and musculature and understanding the realities of what is going on in a face than with the application of those ideas. While that information is invaluable, it is not nearly tangible and direct enough for people under a deadline who need to produce results fast. Elsewhere, you can learn about all of the visual cues that make up an expression, but then you have

to take that and dissect a set of key shapes you want to build and joints you have to rig. You'll likely run into conflicting shapes, resulting in ugly faces, even though each of those shapes alone is fantastic.

Stop Staring breaks down, step-by-step, how to get any expressions you want or need for 99 percent of production-level work quickly and easily—and with minimum shape conflict and quick, easy control. You'll learn much of what you *could* learn elsewhere while also picking up information more pertinent to your immediate tasks that you *might not* learn elsewhere. Studying a brush doesn't make you a painter, using one does, and that is what this book is all about—the doing and the learning all at once.

Who Should Read This Book

If you've picked it up and you're reading this right now, then you have curiosity about facial modeling, animation, or rigging, whether you have a short personal project in mind, plan to open your own studio, or already work for a big studio and just want to know more about the process from construction all the way through setup to good acting. If you're a student trying to break into the industry, this book will show you how to add that extra something special—how to be the one that stands out in a pile of demo reels—by having characters that your audience can really connect with.

If you have curiosity in regard to creating facial setups, or just animating them, you're holding the answer to your questions. I'll show you how to get this stuff done efficiently, easily, and with style.

Maya and Other 3D Apps

There are obviously some technical specifics in getting a head set up and ready for character-rich animation, so to speak to the broadest audience possible, the instruction centers primarily around Autodesk's Maya. The concepts, however, are completely program-agnostic, and readers have applied the concepts to almost every 3D program there is.

How *Stop Staring* Is Organized

While *Stop Staring* will get you from a blank screen to a talking character, it is also organized to be a reference-style book. Anything you might want to know about the underlying concepts of the how and the why of facial animation is in Part I. Everything to do with the mouth—all animation, modeling, and shape-building—is in Part II. Part III takes you

through everything related to the brows and eyes. Part IV brings all of the pieces together, both literally and conceptually.

Part I, "Getting to Know the Face," teaches you the basic approach used throughout the book. Each chapter in this part is expanded into detailed explanation in a later part of the book: Chapter 1 in Part II, Chapter 2 in Part III, and Chapter 3 in Part IV.

Chapter 1, "Learning the Basics of Lip Sync," introduces speech cycles and visemes.

Chapter 2, "What the Eyes and Brows Tell Us," defines and outlines the effect of the top of the face on your character.

Chapter 3, "Facial Landmarking," brings in broader effects such as tilts, wrinkles, and even the back of the head!

Part II, "Animating and Modeling the Mouth," refines the viseme list and sync technique, then shows how to build key shapes and set them up with an interface.

Chapter 4, "Visemes and Lip Sync Technique," delves deeply into how to model for effective sync and shows that building good sync is less work than you thought but harder than it seems.

Chapter 5, "Constructing a Mouth and Nose," attacks the detailed modeling you'll need for a full range of speech shapes.

Chapter 6, "Mouth Keys," shows you a real-world system for building key sets— one that invests time in the right shapes early so you can later focus on artistry undistracted.

Part III, "Animating and Modeling the Eyes and Brows," guides you through creating a tool to put the book's concepts in practice beyond the mouth. From there you'll learn how to create focus and thought through the eyes.

Chapter 7, "Building Emotion: The Basics of the Eyes," shows you which eye movements do and don't have an emotional impact—and how years of watching cartoons have programmed us to expect certain impossible brow moves!

Chapter 8, "Constructing Eyes and Brows," guides you through building the eyeballs first, then the lids/sockets, and connecting all of that to a layout for the forehead and eventually shows you how to make a simple skull to attach everything else to.

Chapter 9, "Eye and Brow Keys," applies the key set system from Chapter 6 to the top of the face, bringing in bump maps for texture and realism.

Part IV, "Bringing It Together," takes all the pieces you've built in Parts II and III and brings them together into one head and then shows you how to weight and rig them for use.

Chapter 10, "Connecting the Features," teaches you to take each piece of the head—eyes, brows, and mouth, plus new features such as the side of the face and the ears—pull all of it into a scene together, and attach them to each other cleanly.

Chapter 11, "Skeletal Setup, Weighting, and Rigging," focuses on rigging your head, including creating the necessary skeleton and weighting each of your shapes for the most flexibility in production. In this chapter, you'll learn to use a system to control any eye and lid setup and how to create sticky lips.

Chapter 12, "Interfaces for Your Faces," demonstrates the benefit of arranging and automating your setup to make all your tools accessible and easy to use. There are ways to share interfaces as well as get very intricate shape relationships with very little work.

Chapter 13, "Squash, Stretch, and Secondaries," takes all the concepts taught up to this point and turns them a little sideways. This chapter introduces a few key ideas and integrates them into the rig in a way that you'll start to see your characters *really* start to bend, and you'll create a layer of control that can sit on top of any other rig.

Chapter 14, "A Shot in Production," presents five different scenes through the complete facial animation process, taking you inside the mind of three animators to see how and why every pose and move was made.

What's on the Website

The *Stop Staring* website, `www.sybex.com/go/stopstaring3`, provides all of the tools and scene files you need to work through the techniques taught in this book—source images and audio, and even Maya interface controls that you can use as-is or practice with to learn to build your own. Click the Resources & Downloads link to access chapter files, resources, and extras.

Use the chapter-by-chapter files as you walk through the step-by-step instructions on how to model parts of the face, rig them all to simplify your work, and then animate them quickly and naturally.

Resources include the head models, interface setups, and other elements of the scenes and shapes taught in the book. Here you'll find a new Maya shelf and scripts (MEL and Python) to speed up your work.

You will also find bonus movies that continue the demonstration of effective animation. And you get several extra sound files to practice animating your own work!

Getting to Know the Face

Before *we start animating, building, or rigging anything, let's be sure we're speaking the same language. In Chapter 1, I talk about talking, pointing out the things that are important in speech visually and isolating the things that are not. Narrowing our focus to lip sync gives a good base from which to build the more complicated aspects of the work later. In Chapter 2, I define and outline, in the same focused way, the top half of the face. In Chapter 3, we zoom back to the entire face—the tilt of the head, wrinkles being a good thing, and even parts of the face you didn't know were important.*

Each chapter in this part is expanded into a detailed explanation in a later part of the book: Chapter 1 in Part II, Chapter 2 in Part III, and Chapter 3 in Part IV.

Learning the Basics of Lip Sync

In modeling for facial animation, mix and match is the name of the game. Instead of building individual specialized shapes for every phoneme and expression, like for an F or a T, we'll build shapes that are broader in their application, like wide or narrow, and use combinations of them to create all those other specialized shapes. On the animation front, it's all about efficiency. You want to spend your time being creative and animating, not fighting with the complexities that often emerge from having a face with great range. It doesn't sound like there's much to these concepts for modeling and animating, and, yeah, they really are small and simple—but they're huge in their details, so let's get into them.

Before we can jump into re-creating the things we see and understand on faces, we need to first identify those things we see and understand. Starting on the ground floor, this chapter breaks down the essentials of lip sync. Next, we'll go into how basic speech can be broken into two basic cycles of movement, which is what makes the sync portion of this book so simple. Finally, at the end of this chapter, we'll take those two things— what's essential and the two cycles—and build them into a technique for animating.

- **The bare-bones essentials of lip sync**
- **The two speech cycles**
- **Starting with what's most important: visemes**
- **Building the simplest sync**

The Essentials of Lip Sync

People overcomplicate things. It's easy to assume that anything that looks good must also be complex. In the world of 3D animation, where programs are packed with mile after mile of options, tools, and dialog boxes, overcomplication can be an especially easy trap to fall into. Not using every feature available to you is a good start in refining any technique in 3D, and not always using the recommended tools is when you're really advancing and thinking outside the box. Many programs have controls and systems geared for facial animation, but you can usually find better tools for the job in their arsenals.

If you're fairly new to 3D, and have dabbled with lip sync, it has probably been frustrating, complicated, difficult, and unrewarding. In the end, most people are just glad to be done with it and regret deciding to involve sync in their project. We're starting to see some amazing results come from facial motion capture techniques, but at least for now, that's probably beyond the cost range for readers of this book. Automated techniques are always improving too, but so far, they aren't keeping up with what a good animator or capture technique can deliver.

Don't despair. I will get you set up for the sync part of things quickly and painlessly so you can spend your time on performance (the fun stuff!). If your bag is automation, there's still a lot of information in here you can use to bump the quality of that up too.

When teased apart properly, the lip sync portion of facial animation is the easiest to understand because it's the simplest. You see, people's mouths don't do *that* much during speech. Things like smiles and frowns and all sorts of neat gooey faces are cool, and we'll get to them later, but for now we're just talking sync. Plain old speech. Deadpan and emotionless and, well, *boring*, is where our base will be. Now, you're probably thinking, "Hey! My face can do all sorts of stuff! I don't want to create boring animation!" Well, you're right on both counts: Your face *can* do all sorts of things, and who really *wants* to do boring animation? Nobody! For the basics, however, this is a case of learning to walk before you can run. For now, we're not going to complicate it. If we jumped right into a world with hundreds or even thousands of verbal and emotional poses (which is how they do it in the movies), we'd never get anywhere. So, to make sure you're ready for the advanced hands-on work later, we're focusing on the most basic concept now: bare-bones lip sync. When dealing with the essentials of lip sync and studying people, there are just two basic motions. The mouth goes Open/Closed, and it goes Wide/Narrow, as illustrated in Figure 1.1.

Figure 1.1

A human mouth in the four basic poses
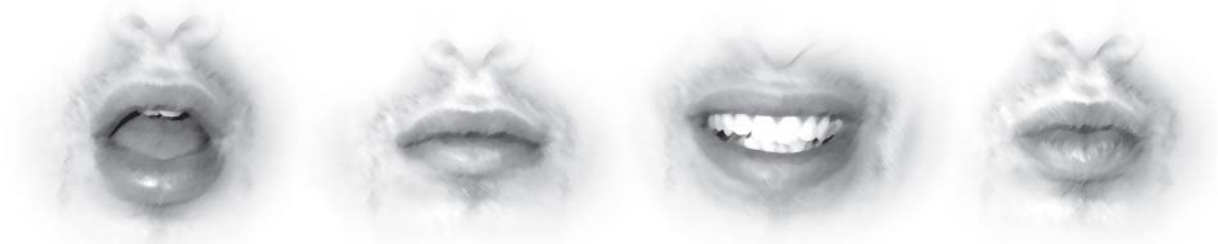
At its core, that's really all that speech entails. When lip-syncing a character with a plain circle for a mouth (which we'll do in just a minute), the shapes in Figure 1.2 are all that's needed to create the *illusion* of speech.

Your reaction to this very short list of two motions might be, "What about poses like F where I bite my lip, or L where I roll up my tongue?" *Ignoring* that kind of specificity is precisely the point right now. We're ignoring those highly specialized shapes and stripping the building blocks down to what is *absolutely necessary* to be understood visually. If these two ranges—from Open to Closed and Wide to Narrow—are all you have to draw on, you become creative with how to utilize them. Things like F get pared back to "sort-of

Figure 1.2

A circular spline mouth in the same four basic poses

closed." When you animate this way and stop the animation on the frame where the "sort of closed" is standing in for an F, it is easy to say, "That's not an F!" But in motion, you hardly notice the lack of the specific shape—and motion is what I'm really talking about here. You should be less concerned with the individual frames and more concerned with the motion and the impression that it creates. For most animators, there is a strong instinct to add more and more complexity too early in the lip-sync process, but too much detail in the sync can actually detract from the acting.

Animating lip sync is all illusion. What *would really be happening* isn't nearly as relevant as the *impression* of what *is* happening. How about M? You may be thinking, "I need to roll my lips in together to say M, and I can't do that with a wide-narrow-mouth-thingamajig." Sure you can, or at least you can give the impression in motion that the lips are rolled in—just close the mouth all the way—and that's usually going to be good enough. When you get the lip sync *good enough* to create an impression of speech and then focus your energies on the acting, others will also focus on the acting, which is precisely what you want them to do.

Analyzing the Right Things

Let me take you on a small real-world tutorial of what is and what is not important in speech.

Animators have a tendency to slow things down to a super-slow-mo or frame-by-frame level and analyze in excruciating detail what happens so as to re-create it. This is not necessarily a bad thing, but here's an example of how that can break down as a method: Look in the mirror, and then slowly and deliberately overenunciate the word *pebble*: PEH-BULL. You're trying to see exactly what happens with your face. Watch all the details of what your lips are doing: the little puff in your cheeks after the B; the way
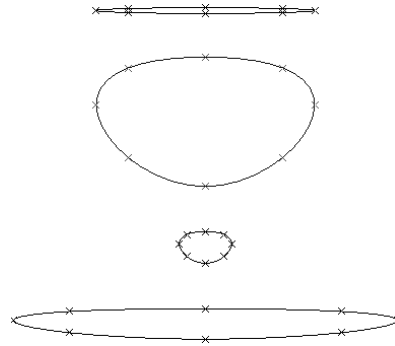
the pursing of your lips for P is different than for B; how your tongue starts its way to the roof of your mouth early in the B sound and stays there until just a split second after the end of the word. You'd think that all these details give you a better idea of how to re-create the word *pebble* in animation, right? Wrong! Most often, that would be exactly the wrong way to do it. It would be the right way to animate the word *pebble* if, and only if, a character was speaking slowly and deliberately, and overenunciating. This hopefully illustrates how a mirror can be misleading if used incorrectly. It can very easily lead to overanalysis, and then to animation that looks poppy and disjointed. This time, at regular, comfortable, conversational speed, say, "How far do you think this pebble would go if I threw it?" How did the word *pebble* look that time? Check it out again, resisting the urge to do it slowly or deliberately. As far as the word *pebble* is concerned in this context, the overall visual impression is merely closed, a little open, closed, a little open. That's it. In a regular delivery of that line, the word *pebble* will generally look the same as the word *mama* or *papa*. Say the sentence twice more, using the word *mama* and then *papa* in place of *pebble* and compare them. Try not to change what your mouth does, but instead notice that opening and closing the mouth are the most significant things happening during *pebble*, *mama*, and *papa*. The mouth doesn't even open wide enough to see a tongue, so there's no need to worry about it. Animating things you think should be there, but in context are not, would be like animating a character's innards. You can't see them, so animating them would be a silly waste of the time you could otherwise spend on—you guessed it—the *acting*.

Not just for our *pebble*, but in the vast majority of situations, the Opens and the Closeds are the most important things a mouth does. That's why puppets work. Does it *really* look to anyone like a puppet is *actually* saying anything? Of course it doesn't, but when a skilled puppeteer times the opening and closing of the mouth to the vocals, your brain wants to make that connection. You *want* to believe that the character is talking, and that's why the single most important action in the word *pebble* and this entire system is simply Open/Closed.

This is how you properly focus on the right things in basic sync: Search for the overall impressions, and fight the urge to bury yourself in the details too quickly.

Speech Cycles

This approach of identifying the two major cycles and *visemes* (a term you'll learn more about in just a moment) is likely very different than what you know now if you come from an animation background. If you're looking for phonemes and a letter-to-picture chart, you're going to be disappointed. In this approach, there is no truly *absolute* shape for every letter, and in a system like this, to point you in such a direction would do far more harm than good, despite what you might *think* you want to see. Each *sound's* shape is going to be unique to its context, and you'll learn to think of it not as a destination

shape, but as the sum of its critical components. To start, let's talk about the two major speech cycles.

In its simplest form, there are two distinct and separate cycles in basic sync: *open and closed*, as in jaw movement, and *narrow and wide*, as in lip movement.

> When I use the word *cycle*, I'm merely referring to how the mouth will go from one shape to the other and then back again. There are no other shapes along the way. The mouth will go open, closed, open, closed; and the lips will go wide, narrow, wide, narrow.

These two cycles don't necessarily occur at the same time, nor do they go all the way back and forth from one extreme to the other all the time. The open-and-closed motions generally line up with the puppet motion of the jaw, or flow of air—with *almost any* sound being created—whereas the wide-and-narrow motions have more to do with the *kind* of sound being created. For example, the following chart shows the Wide/Narrow sequence you get with the sentence "Why are we watching you?"

WORD	WIDE/NARROW SEQUENCE
Why	Narrow, wide
Are	No change in shape
We	Narrow, wide
Watching	Narrow, slightly wide
You	Narrow

Simple, right? Now take a look at the jaw, or the Open/Closed cycle described in the next chart. In this case, *Closed* refers to a position not completely closed, but closer to closed than to open.

WORD	OPEN/CLOSED SEQUENCE
Why	Closed, open, closed
Are	Closed, open, closed
We	Closed, slightly open
Watching	Closed, open, closed, slightly open, closed
You	Closed; no change

That's it for the essentials. The backbone of this book's lip-sync technique has to do with this simple analysis of the Wide/Narrow and Open/Closed cycles. You will be adding more and more layers to create complex, believable performances, but that is all going to be based upon this foundation. Taking the lead from the human mouth, I've based this approach on the "simpler is better" mindset. Your mouth is lazy. If it can say something with less effort, it will. In contrast, you've probably had textbooks, teachers, and/or tutorials tell you that for good sync, you need shape keys that include things like G. My question is, why would you build a shape for or pay any special attention to the letter G? Whether it's a hard G or a soft G, you can say it with your mouth in *any* of the shapes shown in Figure 1.3.

Figure 1.3

All varieties of G

What this tells us is that G has few visual *requirements*, so it won't be something we build a specific shape for. Further, we just proved that any single pose we picked would already be wrong two-thirds of the time, even in our small test. Given that, even if we *did* want to build a G, how would we ever pick a single shape?

Both G sounds are created *invisibly*—solely using mechanisms inside the mouth, not by the lips or even noticeable open/closed cues. This G example is here to begin to illustrate what is and, more importantly, what is *not* a viseme.

Starting with What's Most Important: Visemes

For this noninclusive approach, where you're trying to exclude extraneous mouth-to-sound pairings, something you'll need to know is what *must* be included. There are certain sounds that we make that absolutely need to be represented visually, no matter what. These are called *visemes*. Examples of visemes are Narrow for OO, as in *food*, and Closed for M, as in *mom*. You just can't make those sounds without those contortions. Looking back, do you think G is a viseme? It isn't. It couldn't possibly be any *less* of a viseme. It requires no contortion, and it did not suffer from any other contortions. It is visually meaningless. There are going to be more visemes to address than the Open, Closed, Wide, and Narrow variety I've touched on, but even this greater list of must-see shapes can be "cheated" to fit into the simple circle-mouth setup you've seen and are about to build.

Why Phonemes Aren't Best for CGI

Phonemes work fantastically in classical animation, where nothing comes for free and every frame has to be drawn. Used merely as a guide, with an animator drawing a new picture for each frame, phonemes are great. In CGI, when you're working with phonemes as *actual shapes*, each a discreet pose in the rig, sync animation tends to end up overly choppy, and counteranimation becomes too large a portion of the work. In other words, when phonemes are an *idea*, they can and do work very well. When phonemes are unique physical manifestations built deep into the core of a character rig, they can *and often do* just get in the way of good sync.