Lip Sync - Making Characters Speak

By Michael B. Comet - This article, all images and character designs are Copyright 1998 Michael B. Comet All Rights Reserved.

What is Lip Sync?

Chances are if you've ever animated a character, you've needed or wanted to make it talk, sing, or otherwise communicate via dialogue of some sort. Lip sync is the art of taking a pre-recorded track of dialogue, and making a character appear to speak it. This involves figuring out the timings of the speech (breakdown) as well as the actual animating of the lips/mouth. In addition making the actual setup or mouth phonemes needed can also be considered a part of the entire Lip Sync process for 3D.

This article will give an overview of the process of animating dialogue. While I may focus on specific software in some cases, I'll try to stay general enough that you can use this information with any software. You should have a basic grasp of standard animation techniques and terminology. If you don't you might want to check out the Frequently Asked Questions (FAQ) document on the CG-Char web pages.

Part I - Phoneme Setup

As with any 3D animation, you can't actually animate until you have your object, and it has to been setup for animation. For characters bodies, this usually consists of segmented joints parented to each other, or for smooth characters, a setup with bones.

Lip sync in 3D has similar requirements. In this case though we're only focusing on the head of the character. (Or at least the mouth if you have some freaky creature that doesn't have a 'head' per se.)

When thinking about 3D facial setup, it helps to explain how 2D cartoons achieve lipsync. The appearance of speech in cartoons is created by drawing different mouth shapes for a character on different frames. By animating the mouth shapes timed to dialogue, the character appears to speak. Each of these mouth shapes is usually referred to as a **phoneme**.

A phoneme is the unique shape a mouth takes to make a specific sound. For example when you make the sound "ooo" as in "gr**oo**vy" your mouth tends to pucker into a small circle. When you make the sound "Aaa" as in "**A**pple" you mouth tends to open up wider and fuller. The word "Animation" could be described phonetically as "a-nee-may-shun" or more like a breakdown as "a-n-ee-m-ay-sh-u-n".

For basic breakdown purposes, the lowest number of phonemes most people use is 9. These usually are broken down as:

1. A, I
2. O
3. E (as in sweet)
4. U
5. C, K, G, J, R, S, TH, Y, Z
6. D, L, N, T
7. W, Q
8. M, B, P
9. F, V
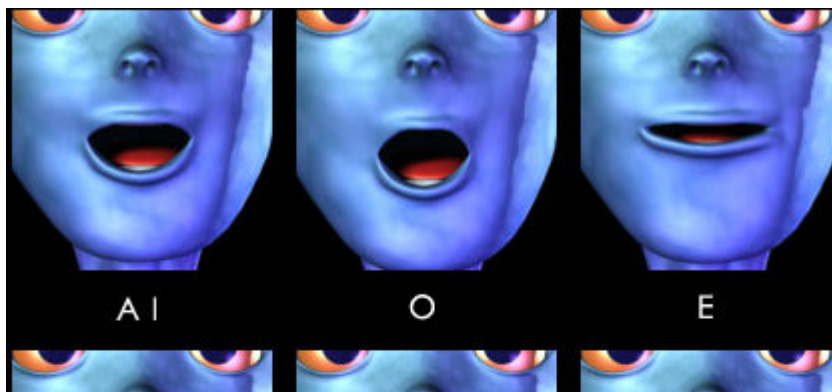
Sometimes these shapes overlap. For example, you may need to use the F phoneme for TH. So this should be looked at as a basis that you can work from and not something that must be ardently followed. In fact for more realistic animation I tend to divide these phonemes even more getting as specific as I can.

The phoneme set I now use which is derived from a larger set the Ventriloquist tool usually requires (see below) is as follows:

1. **M B P** - The lips are pressed firmly closed and can actually intersect a little to show the difference between the base default pose. I generally rotate the lower lip up and in, and the bottom lip down and in and then move them together to make this pose. Sample: Map, Bang, toP
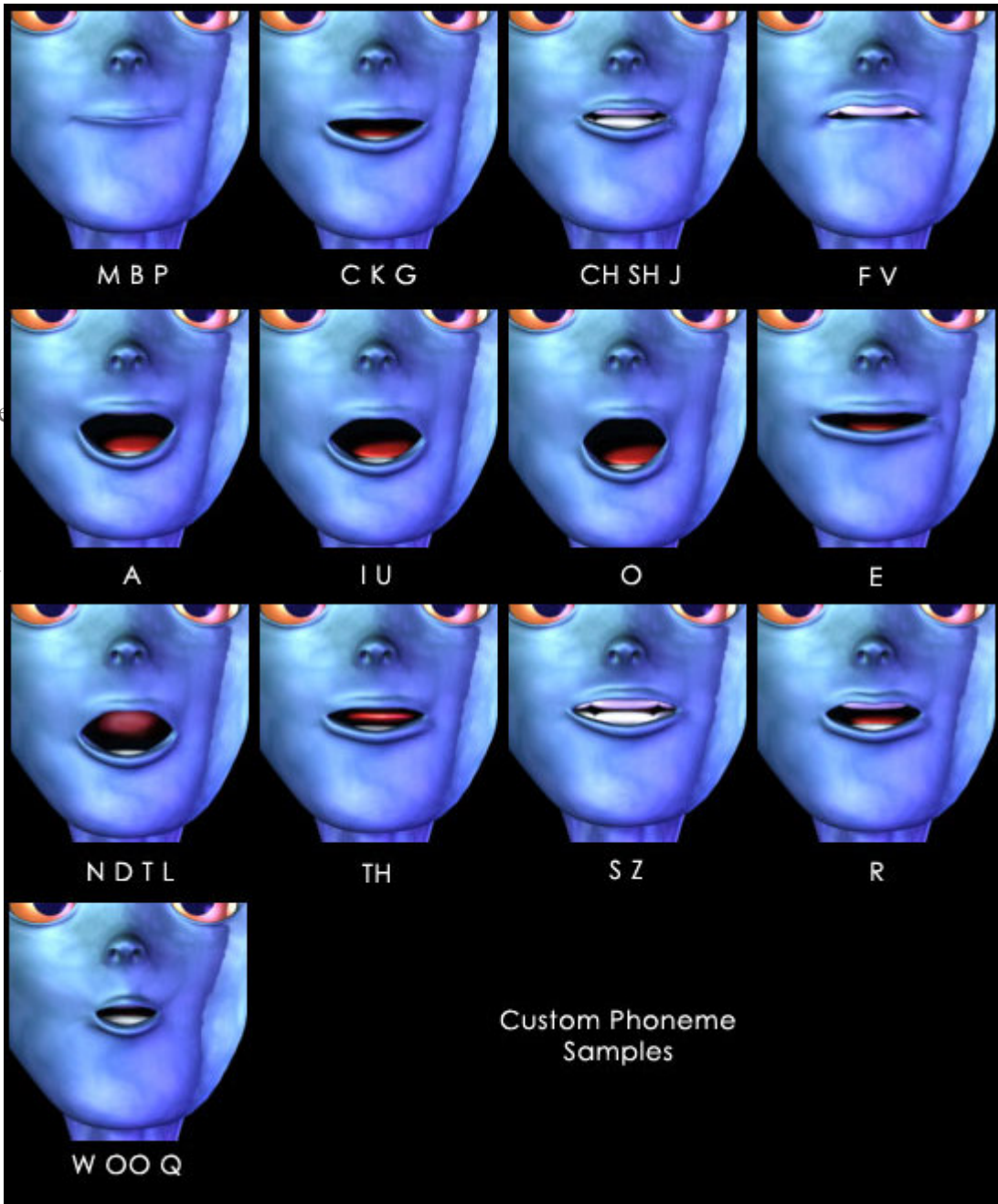

2. **C K G** - This is the generic, a sound if coming out of the mouth pose. It is not opened as far as the vowel phonemes. The jaw rotates down, and the teeth are separated about the height of a tooth. Sample: Carry, looK
3. **CH SH J** - The jaw does not actually open or rotate, but the lips are puckered outwards and narrow the mouth a bit. The teeth are together. Sample: CHerry, SHout, Jump
4. **F V** - In this pose the jaw can rotate down and actually back slightly, then the lower lip is curled/rotated up and underneath the upper teeth, so that it is pressed between the upper and lower teeth. I also tend to bring up the upper lip a bit so the upper teeth are exposed. Sample: Fine, LoVe

5. **A** - This phoneme is created by rotating the jaw down so the mouth is open. I place the corner of the lips above the vertical center point of the open mouth. Sample: Apple, blAde, Ape

6. **I U** - This target is also an open rotated jaw. I tend to open the jaw slightly more than in the A pose, but less than an O pose. The corner of the lips are placed at or below the vertical mid point of the open mouth. Sample: If, Under, wOnder, pIck

7. **O** - This is the most extreme jaw open vowel target. The corner of the lips are placed at the mid point of the vertically opened mouth, and the mouth is narrowed/puckered naturally since it needs to narrow as it opens. Sample: Open, OAt, Over

8. **E** - This is the smalled jaw opened vowel pose, the mouth opens about the same as a C K G, but is widened out left and right. Sample: swEEt, EAt, feet

9. **N D T L** - For poses where the tongue touches the teeth, rotate the jaw about the same as an A or I U (I usually vary it slightly) and then rotate the tongue up and behind the front teeth. Sample: Name, Dove, abouT, faLL

10. **TH** - In this pose, the jaw is rotated open just enough that the tongue can protrude out and be pressed between the upper and lower teeth. Sample: teeTH, forTH, THat

11. **S Z** - This pose has the jaw not opened, or only slightly rotated, with the teeth together. The lips expand up and down and widen so the teeth are visible. Sample: Snow, Zoo, Sneer

12. **R** - The R phoneme is one of those things that really vary from person to person. You can probably get away with a regular C K G pose, but I tend to open the jaw a little and then pucker the mouth so it narrows, and bring up the upper lip into kind of a sneer. Sample: Roll, dooR, wondER

13. **W OO Q** - This can be one of the most difficult poses to model, the mouth contracts/narrows into a small opening with the lips also puckered outwards. The jaw can rotate slightly down. Sample: WOOd, drEW, QUIet, fOOd

Having a more detailed breakdown like this can help when doing more realistic facial animation. But it may or may not be needed for your character. I prefer it simply because it gives me more variation to work from.

There is one other item you need to be aware of for setup. That is emotion or expression. Unless you want your character to remain perfectly flat, you'll need to make it look happy, sad or a wide variety of other expressions. Typically there are 6 base emotions. These are:

1. Sorrow
2. Anger
3. Joy
4. Fear
5. Disgust
6. Surprise

Six Basic Expressions
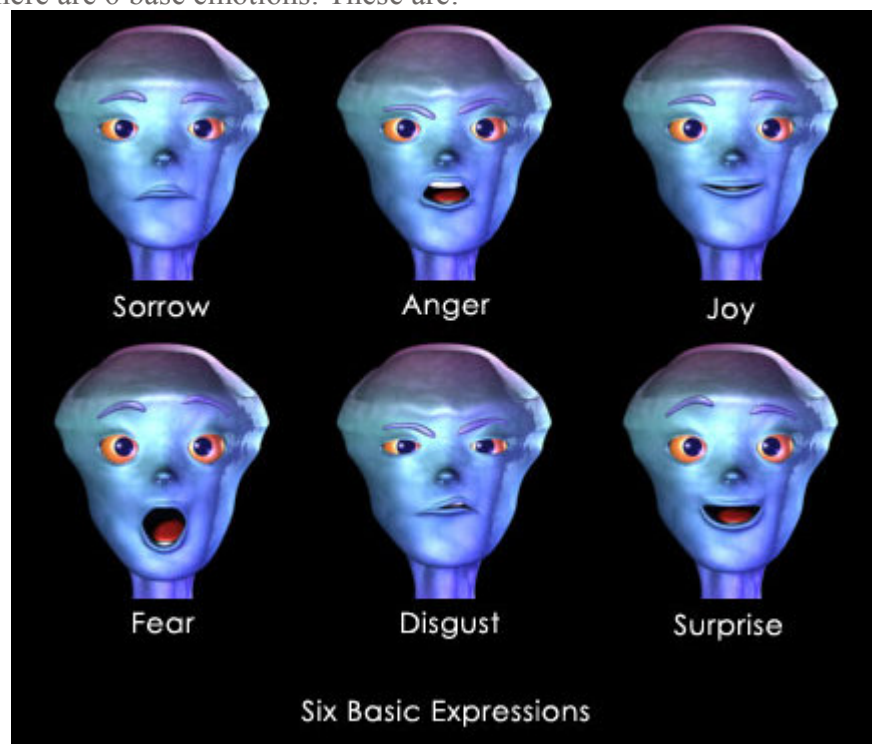
In most cases you will want the ability to create phonemes not only in a flat style but also in each of

these expressions as well. So you might have the phoneme for "oh" as "**oh** my!" seven different ways. Imagine how someone's mouth may change between saying "oh my" really loud and scared, versus very coy or sly with a grin. You will see that this is one of the reasons the "weighted morphing" method discussed below is preferred.

Getting back to our cartoons...the animator would draw one of these pictures of the mouths at the right point to create the lipsync. Naturally this carries over to 3D animation as well. You must keyframe your characters mouth to match the right phoneme for the sound at that frame. This means you must setup your character's head to deform into each of these shapes.

The method you plan to use for animating will directly influence how you want to approach setup. For example, if you are going to use bones to modify your head and mouth, you don't need to model different phonemes, you'd just setup the bones. On the other hand for a weighted morphing system like Morpher, Blend Shapes, Morph Magic, Smirk or the Morph Gizmo, you'd need to create different models. The following is a list of some ways to setup and animate a head:

1. Animated Image maps
Animation is achieved by replacing the color map. Much like 2D. You draw each of the mouths and apply the maps at different frames. You could make a matching bump and specularity map to go along with the color map...though this method still looks pretty cheesy for real 3D characters. However if you are going for a cel shaded look this may actually be all you need.

2. Object Replacement
Animation is done by swapping out head models of the object. Much like the puppet animation in "Nightmare Before Christmas", There can be a different head used for almost every frame. This method has the downfall of requiring tons of head models and therefore RAM usage (not to mention modeling time). The animation will look very snappy or quick just like true puppet animation, so it may not fit the 3D style in all cases.

3. Standard Morphing
With this method you model many different poses of phonemes and expressions. Then the software can interpolate smoothly between each head. This results is a much smoother look than object replacement.

4. Bones
Setup bones for the characters head. By animating the bones on different frames the face is deformed. You can create an archive of bone positions for all the phonemes and copy the keyframes, or animate every keyframe by hand.

5. FFD
Same as bones but uses a Free Form Deformation tool instead.

6. Weighted Morphing
Different head targets are modeled for each facial muscle group, or phoneme and expression. Then these shapes can be mixed and matched in different percentages yielding a wide variety of poses. This is usually the preferred method.

Rather than detailing each of these, I am going to focus simply on the last 3 methods with an emphasis on weighted morphing. If you are interested in learning more, I'd recommend reading some of the materials listed in the bibliography section at the end of this document.

Bones and FFD setup consist of modeling and texturing one head. Then you setup bones or FFD controls around the mouth. Set them up such that the lips can bend into the required shapes and also so the lower jaw can drop. Once the skinning is done, you can continue and manually position the bones or FFD controls on each frame to animat them. Another alternative is to create an "archive" of keyframes for the phonemes. For example, 3D Studio MAX allows you to set keyframes for objects before frame 0. So you might set up your head with bones and make the A phoneme on -1, the E phoneme on -2 and so on. Later you could very quickly copy the bones keys from a negative frame

over to a real part of the animation. This saves a lot of time, and in fact there are plugins to allow you to automate this. If you use 3D Studio MAX you can download the 47k ZIP file of the Magpie Importer by Andrew Reid

One problem with Bones and FFDs is it can be difficult to get the original model to deform nicely for each pose. Getting the proper creases, folds and lips to move correctly can be a pain to setup. I've personally done both Bones and FFD's for facial setup. Compared to actually modeling in a change, it can be a lot of work.

Another downside of Bones and FFDs is they can be slow to animate, and difficult to add emotion or variation to. For example if you're animating bones by hand without using an archive, then you will have to manually pose each bone for each keyframe. If you are using the archive, your animation may look robotic since every same phoneme will look identical. ie: all the A phonemes will look alike unless you manually change them. Plus unless you create a large archive for phonemes with different expressions (such as happy, sad,...) you'll need to go back by hand and tweak the keys.

Enter weighted morphing. Weighted morphing is just what it sounds like, different morph targets mixed together at different strengths. What you do is model a different head for each phoneme. Typically these heads must have the same point/vertex count and ordering as the original base. For that reason the other target models are usually created by simply pulling, stretching, using bones or FFDs on the original model and saving out the copies. However, as it's just a model you can use any non-destructive modeling tool. I usually work with a spline model, or low poly (smoothed after morphing), and simply move the points around to make the new targets.

When animating, instead of always hitting an A pose perfectly you could use the A phoneme target at say, 35% and get something between an A phoneme and a closed mouth. Or you could mix the A phoneme with others...to get a new variation. But it gets even better. Remember those expressions. Now you could mix say an A phoneme at 80% with say the Anger phoneme at 90% to get a very mad looking face still saying ahhh. Now let's go one step further. Rather than model phonemes, model targets for each muscle of the face. By modeling facial muscles, or rather the effect they have on the face, any phoneme or expression can be made from only a few targets.

As an example of muscle modeling, by making a target where the jaw drops and then targets for the zygomatic major, risorius and platysma muscles, you can make a smile and E phoneme. Simply put, using muscle targets mimicks the way a real head works. You simply animate the muscle groups to shape the head for each keyframe into the desired phoneme and expression.

One catch with the muscle approach however is sometimes you will have to mix greater than three shapes, and this can mess up how the targets look. For this reason you can still model regular phoneme shapes, and then also model additional muscle or expression shapes for tweaking.

Weighted morphing can be used to create very realistic and fluid lipsync. You may be able to obtain or write a plugin or expression to import a breakdown and directly alter the morph keyframes for the lipsync. Then you could tweak it and add in keyframes for expressions. Or you can animate by hand. While slower, it is still quicker to adjust a slider than manually moving or rotating tons of bones or FFD controls. For example, with bones, you might have 16 controls for the lips, and one for the jaw. To open the mouth, you would have to manually position and rotate each of these bones. With weighted morphing, the same task could be accomplished by dragging one slider.

The images to the right show the basic

expression
targets used for weighted morphing that work in conjunction with the phoneme targets listed earlier. Many books or papers on creating facial targets show a diagram of the face with the muscles used. Rather than do that here I've decided to actually show sample expression targets. This basically encompasses much of the same information when mixed with the phonemes, includes a bit more detail and should be more useful. However it is still helpful to look at and understand the facial muscles. I'd highly recommend taking a look at Gary Faigin's "Facial Expression" book listed in the bibliography at the bottom of this article. It is a fantastic reference for anyone doing facial animation, whether or not you will use the muscle targets or not. I'd also highly recommend using a mirror to study your own face.

The following is a quick description of each of these targets and how they are used:

1. **Grin L & Grin R** - A basic half smile. The mouth remains closed. This allows it to blend with an open vowel phoneme, such as an A to get an opened mouth smile. In some cases this blend won't be perfect and for long lasting smile shots you may want to make an open mouth smile as well. The area underneath the eyes can also be made to buldge out and flatten the bottom of the eye as well.
2. **Sneer L & Sneer R** - This is pretty much a one muscle target, usually its good to try to get a crease line from the nose to the side of the mouth. I tend to expose a bit of the upper gums as well. These two targets can be used to expose the upper teeth more when needed during talking, as well as for anger, disgust, and other expressions.
3. **Frown** - A basic frown pose. Technically there are two separate muscles pulling the corners of the lips down, and in some cases you may want to make 2 different targets. I also make the lower lip pout outwards in this target sometimes.
4. **Eyebrow Up L & Eyebrow Up R** - Your basic eyebrow up target. For cartoony characters you may want more targets than the simple up and down. For realistic, the outer edge of the eyebrow tends not to move very much, while the center area moves more.
5. **Eyebrow Down L & Eyebrow Down R** - The eyebrow moves down and in and the center of the eyebrow area starts to wrinkle.
6. **Squint** - In this target I have the eyelids close to meet at the exact middle, and tend to also buldge up the area beneath the eye.
7. **Blink L & Blink R** - Here each eye closes individually so you can offset blinks. By having this separate from the squint, you can use the squint to get an overall acting of the eye shape and then easily go back and animate in blinks later. I tend to have the lids meet at an area 3/4 of the way down from the top vs. in the center like a squint...so the upper lids rotate more.

For some of these targets I mention creating the left and right sides as separate poses. The reason for this is that you will have more control. You can make your character grin to one side, or blink one eye and so on. If you had both sides as one target this wouldn't be possible. Another big reason is to easily remove the "twins" syndrome of too much symmetry in the pose. If you are trying to or need to save space and create some of these as one target, you should make the target offset left and right a bit...such as the smile on one side slightly higher than the other. This will help make the face assymetrical and look more natural. For more information on basic animation principles and "twins" see the CG-Char Web Pages and the FAQ document there.

One thing to note with weighted morphing is that mixing the various targets won't always result in exactly what you want. This is usually a result of three or more targets having some odd interaction when mixed together at high percentages. Therefore it is typical to start with phonemes or targets for each muscle, but then to add new targets with specific snapshots of the head as needed. This is especially true for hyper-exaggerated poses.

Finally I'd like to mention a few points about the actual mouth model itself. This applies to any of

the techniques mentioned above. Since the mouth of your character will be open it is generally a good idea to have some amount of detail in there. Teeth, gums, tongue and some inside mouth cavity. You may not think you need gums but for poses like the sneer, you'll see them, so they should be modeled. For a tongue I tend to use a simple deformed sphere. In allit really depends on your character. I imagine some creatures might not have teeth or could be more cartoony, but in many cases you'll want this detail. Once your character is setup, you are ready for the next step in lipsync, Breakdown.

Part II - Breakdown / Track Analysis

Track Analysis, or breakdown, is the art if listening to pre-recorded dialogue and sound, and figuring out the timing. Traditionally the audio would be played on a device with a counter. The person analyzing it would listen over and over and write down the best estimate of when each phoneme occured.

For cartoons this information was (and is) recorded on what is called an Exposure sheet, or X-Sheet for short. An X-Sheet is really nothing more than a glorified table. It has numbers representing frames down one column, and then other areas where you can pencil in dialogue notes, camera instructions and so on.

A section from a traditional paper X-Sheet.
The one on the right has been filled out as a sample.

Of course, computer animation once again mimicks it's traditional history. The breakdown is still recorded...many times on a digital X-Sheet. However, the actual work of figuring out the timing has gotten a bit easier. The first method one can use is to load the audio file onto the computer with software that shows the waveform and timecode. Then repeatedly play the whole or parts of the audio figuring out the timing. This is essentially the digital counterpart of the traditional method.

Yet, track analysis can get even easier than this. There is software available that not only shows the waveform and timecode, but actually has a digital exposure sheet and allows bitmaps of phonemes to be played in real time. So you can see and check how your breakdown is working as you create it. Another variation of this is to simply work right in the 3D animation program and if it allows, scrub the time slider back and forth to hear the audio and then key the pose at the right place. Assuming fast enough playback, you may be able to view results in near realtime.

At the top of the line is automatic voice recoginition. These packages automatically look at a WAV file and figure out the timing and phonemes for lip sync. All you need to do isgo back and tweak to correct any errors. My personal favorite is Ventriloquist (a plugin 3D Studio MAX, the generic standalone is called "Echo") by Lips Inc.. This utility can take an audio file, a text line of what is in the file, and automatically generate nice weighted morph fcurves for phoneme based weighted morphing. It can also create a digital X-sheet for you if you still want to enter the data by hand, or want to write your own importer. It still requires tweaking, but is pretty good for a first step, or as a final step in cases where time doesn't permit. Since the point of this article is to teach how to manually do analysis, I won't go further into this topic.

The shareware software and my personal choice for manually breaking down audio is Magpie. Magpie is available at: http://thirdwish.simplenet.com/magpie.html and runs on Windows 95 and NT machines. It allows you to drag and drop custom phonemes into a digital exposure sheet as well as

view 2d images for each phoneme in realtime for playback.

Sample screenshot from magpie. Note the phonemes on the left, the Waveform, sample image and main window housing a digital exposure sheet.

Analyzing voice is simply a matter of listening to small chunks of the audio and marking the proper phoneme for the proper frames. As an example, I have used Treason.wav. This is a woman saying "You will pay for your treason pilot!".

In figuring out what phoneme comes where you can almost "see" the breakdown visually. Notice the pattern of the waveform from frames 2-10. These are the words "You will". 3-5 is the "You" and 6-10 is the "Will". You can tell this because vowels tend to make large balloon shapes in the waveform, while the M/B/P, F/V and to a lesser extent W, Y and S tend to flatten out the waveform. So

Treason - Magpie Exposure Sheet

Frame	Waveform	Timecode	Key	Mouth name	Comments
0		00.00.00.00	X	Closed	
1		00.00.00.01		Closed	
2		00.00.00.02	X	Y	You
3		00.00.00.03		Y	
4		00.00.00.04	X	OO	
5		00.00.00.05		OO	
6		00.00.00.06		OO	Will
7		00.00.00.07	X	L	
8		00.00.00.08		L	
9		00.00.00.09	X	P	Pay
10		00.00.00.10	X	P	
11		00.00.00.11		P	
12		00.00.00.12	X	A	
13		00.00.00.13		A	
14		00.00.00.14		A	
15		00.00.00.15	X	E	
16		00.00.00.16		E	
17		00.00.00.17	X	F	For
18		00.00.00.18	X	F	
19		00.00.00.19		F	
20		00.00.00.20	X	I	
21		00.00.00.21		I	
22		00.00.00.22		I	
23		00.00.00.23		I	
24		00.00.00.24	X	OO	Your
25		00.00.00.25		OO	
26		00.00.00.26	X	R	
27		00.00.00.27		R	
28		00.00.00.28	X	S	Treason
29		00.00.00.29		S	
30		00.00.01.00		S	
31		00.00.01.01	X	I	(pucker here a bit)
32		00.00.01.02	X	I	
33		00.00.01.03		I	
34		00.00.01.04		I	
35		00.00.01.05	X	R	
36		00.00.01.06		R	
37		00.00.01.07	X	E	
38		00.00.01.08		E	
39		00.00.01.09		E	
40		00.00.01.10	X	C	
41		00.00.01.11		C	
42		00.00.01.12	X	I	
43		00.00.01.13		I	
44		00.00.01.14		I	
45		00.00.01.15	X	N	
46		00.00.01.16		N	
47		00.00.01.17		N	
48		00.00.01.18	X	P	Pilot
49		00.00.01.19	X	P	
50		00.00.01.20		P	
51		00.00.01.21		P	
52		00.00.01.22	X	O	
53		00.00.01.23		O	
54		00.00.01.24	X	I	
55		00.00.01.25		I	

frames 6 and 7
are likely the "W" phoneme of "will".

Look at frames 10-13. This is a very flat area. The next word "pay" has the M/B/P phoneme which is
what causes this. Frames 14-15 tend to show an increase in the waveform as the mouth opens a bit
and some of the air escapes. 16-20 is that "ay" sound of the word "pay".

You may be noticing that my breakdown doesn't follow this...that is I have actually placed the
phoneme F on frames 17-19 and the A and E actually occurs much earlier. There are 2 reasons for
this. First, it is very common to have the mouth shapes lead before the sound. That is, you'll want to
shift your mouth poses to actually appear 2-4 frames before when it actually occurs in the WAV file.
This method simply tends to make the lipsync look more correct. Tony White in "The Animator's
Workbook" mentions that at Disney, some actions were even anticipated 12-16 frames ahead of
when they occured.

In addition, it is very common to need to lead the M/B/P and sometimes F/V sounds even more. For
example, the word "Pilot". Frames 49-51 are the flat section of the P phoneme. At 52 there is a
breath of air as the mouth starts to open. 55 starts the vowels. However, I have placed the breakdown
for this word about 3-4 frames ahead simply because it looks more correct when played back. Even
so, I still use the visual cues to help me when I'm breaking down the track.

There are two ways to handle this leading of the poses. First, you can breakdown your track 100%
accurate to the waveform. Look at the pattern, listen to the sound and place the right phoneme there.
Then later on, simply slip your audio to start a little later. The other option is to actually compensate
as I did in the breakdown itself. This is the method I prefer since I feel I may want to adjust the keys
more or less for different phrases. It allows me to tweak the timing of the poses within magpie and
then the final version in MAX without any extra mucking about later.

The other thing to keep in mind when analyzing the voice is the sound of each phoneme. Remember
that a phoneme or sound of the voice doesn't need to actually match the spelling of the word. One
famous example is that "ghoti" spells the word "fish". Granted, that looks like it would sound like
"goat-tea" but phoentically it can just as easily sound like the aquatic animal. Here's how: Take the
"gh" from the word "enou**gh**". The "o" from the word "w**o**men". Finally "ti" from the word "na**ti**on".
As you can see, all of these words have sections that are spelled one way but sound another. Keep
this in mind and really listen to the sound of each phrase as you do the breakdown.

Finally, it's OK to drop phonemes. For example, look at the final word "Pilot". The "ihhh" sound of
"lot" isn't there. It simply is dropped. One of the key tricks in getting lipsync to work
right...especially fast paced speech, is to learn what to drop and what to keep. Think of the mouth as
flowing and figure out what the best shape is at that point. Do this by listening and looking at what
phonemes are around the current frame. In general, if you hit the M/B/P phoneme and the large
vowels that follow, your animation will usually look correct. Everything else is really just an
inbetween of the extreme mouth closed, mouth open and mouth tighten poses. Think about puppets.
Typically they have a mouth open, and a mouth closed pose. Yet in many cases, people accept a
puppet as actually talking. If your lip sync looks off, check the timing of these poses first. Chances
are correcting them will correct your animation.

When I first start to do breakdowns, I tend to put a keyframe on every frame. For example, take a
look at the breakdown I have for this WAV file. If you can't see the image, you can also use the text
version saved by Magpie. Note how there are X's on certain frames. This represents where the
keyframe for that phoneme would go. Frame 4 has a W/OO sound. However, as shown in the text
version of the breakdown, I start by placing a keyframe on all frames. This does NOT mean I have a
different phoneme on each individual frame. Rather, there are no inbetweens.

What this does is make the animation very snappy. Each phoneme literally pops in. This makes it easier to see mistakes in the timing. You can view an AVI of this initial rough breakdown (301K). If this animation looks a bit harsh or wrong, you're quite right.

Essentially this version suffers from a lack of tweening, making it look too rough. Also the fact that I'm using only 14 basic phonemes...none of which really look like a yelling type expression makes it seem out of place. Plus the phonemes don't really smoothly correlate to each other. For example the "f" in "for" should probably be a bit wider because it comes right after "pay" where the mouth gets a bit wider. However even with some inbetweens this starts to look a little better. This is shown in the tweened default AVI (334k).

This second AVI is identical to the first, except there are key frames only where marked as on this text copy of the breakdown. This makes the phrases flow a little better...though it starts to look a bit too floaty.

Once you have finished the track analysis you should have your finished X-sheet With that information you are ready for the final phase of actually animating your character.

Part III - Animation

We have finally arrived at the animation stage! This is where some of the real fun happens. As you saw, doing breakdowns is really pretty simple and in some ways tedious work. At this point, you can start to actually make your character have emotion and come alive.

Probably the simplest way to get from a breakdown to an actual animated character is to simply import the breakdown directly. Magpie can export various formats, including some methods specific to Animation Master and Lightwave. If you use 3D Studio MAX you can download the 47k ZIP file of the Magpie Importer by Andrew Reid. Also the newer versions of Magpie may allow direct importing into MAX. All you really need is to get a keyframe of the pose on the right frame as listed on your X-Sheet.

There are two problems with this method though. First of all it's probable you'll still need to animate the upper face by hand. Second, the animation will probably look "canned" or generic. The result of simply taking a breakdown and importing it is the same as the animation of the simple tweened AVI mentioned before. what happens is every "A" phoneme looks the same, every "T" phoneme looks the same, and so on. Even if you have more phonemes, chances are things will just look kind of blah. (Note: some automated packages, like Ventriloquist/Echo mentioned above are smart enough to make the phoneme a % based on volume and speed of the phrase so the generic look isn't quite as prominent and it needs less tweaking.)

The best way to animate the face is to start with the object, your X-sheet, and then manually pose each keyframe yourself. If there is one thing I'd highly recommend when doing facial animation it's buying a mirror. You can pick up a little bathroom mirror at a drugstore for relatively cheap, and it's invaluable to have it sitting on your desk as you animate.

All you do is look at your X-Sheet and copy the pose. Take the above "Treason" breakdown, the first key after frame 0 is the Y pose on frame 2. Go to that frame, then say the word and look at yourself in the mirror. Then match what you see to your character in your 3D package. Then go to the next keyframe and repeat this process.

The benefits of this method are that you are not limited to any preset phonemes. The mouth will look very natural since you will be posing every frame as it would look. You can go extreme when needed, smaller when not. Pay attention to how letters change especially consonants when they're

around vowels. Compare my custom hand animated AVI (360k) to the generic tweened (334k) version from before.

Hand animating allows you to pay close attention to snap, offsetting the left and right sides of the face to make things look more natural, and generally results in a better looking result. One thing to keep an eye on in with any method though is the interpolation.

By default, most spline interpolation has errors when two keyframes are close together, or nearby other keys that are radically different. For example, look at the image on the right. The top version has a default smooth spline type. Note how the area the red arrow is pointing to dips downward even though the next key is higher. Actually, almost every section of this curve is incorrectly interpolated.
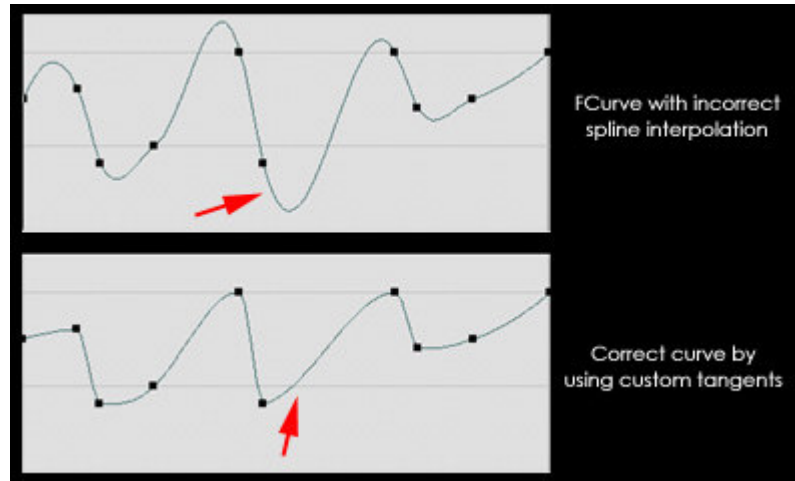
Now examine the lower image. This spline has been adjusted with custom bezier tangents. This allowed me to direct the path of the curve to more closely follow the keyed points. Imagine if the point at the red arrow was equal to 0 percent. The top version would have tweened that target or bone to a negative value. If your morph program didn't cap at a 0 value you might end up with something very odd looking indeed. However, in some cases the default interpolation is fine and results in smoother motion or even adding a bit of 'anticipation' to a movement. The catch is to pay attention to the curves and adjust them where necessary. One quick way to fix fcurves is to take all the lower values, such as those at 0 and make them "linear", then take the tops of the curves and set them to be "ease-in" and "ease-out" types. This generally fixes any overshoot of the curves and looks pretty nice.

Back to our lipsync animation. At this point, you should have a pretty well adjusted version of the character speaking. Up until now, chances are you haven't animated the upper face. Simply adding in some eyebrow and eye movement can really add a lot to the character and expression. In fact, the eyes are what most people focus on when talking so in that sense it's even more critical than the mouth. What I tend to do is write down ideas on the X-Sheet as I animate. I just make notes at different frames with things like, expand eyes, raise eyebrows, etc... Most importantly again, look at the mirror! I repeatedly act out the dialogue looking in a mirror experimenting with different versions. Then, I just match that to the animation.

This can actually be kind of easy, since you already have the breakdown and know when each word occurs. You might notice that when you say a certain word, or part of a word, your eyebrows go up. All you need to do is look at your X-Sheet, then find what frame the word starts on and set your eyebrow keys. Actually, it's a good idea to offset the upper face and even parts of the lower face so things don't always hit keys on the same frame. This is shown in the previous custom AVI (360k). When the whole face works together the results are even better.

At this point, all that is left is to animate the actual body and head motion. Once again, I use a mirror, act it out and animate. You can view the final version of the AVI (434k) with custom mouth animation that has been tweaked more, eyes and head motion. I'd recommend studying how people move when they speak. There can be some very subtle head motion...and I think there is a propensity for new animators to either overdo or underdo the head motion. Think about what words you want to accent and which parts you want to play down. It also helps to get inside your character's head to

figure out what it's thinking. That helps a lot especially with animating the eyes. As in all cases, study from life, and one more time, use that mirror!

Tips, Tricks and Suggestions

Hopefully, this article provided you with enough information to try out your own facial animation. Doing lipsync breakdown is a great way to work on building your timing skills. It really forces you to pay attention to every frame. There's nothing like moving a phoneme 1 frame back and having everything suddenly work. Plus, it can be darn fun to do. Below is a list of key points to remember:

- If you are using Magpie or a similar tool, or especially if you are going to simply import the basic breakdown to get the animation, more phonemes are better. You might want want to consider making 2 extra sets for each base sound. One where the mouth moves only slightly as if the character is murmuring, another where it is open and exaggerated more as in a yell. With these extra sets, it should be possible to get finer control during the breakdown and better results in a direct import animation.

  Another idea is to also make phonemes for consonants paired with vowels. Compare how your mouth is shaped for the "L" in "Lease" vs. "Loop". In the first case it's wider, in the second much more puckered and closed. So for each consanant, you might want to make extra phonemes where it's paired with a vowel, such as "la", "li", "lo", "lu", "le" and "loo".

- If you are animating someone with an accent, pay close attention to how they pronounce certain words. Actually breaking down an accented speech is a great way to force yourself to really listen to the track, instead of just dropping in what you think are correct phonemes.

- Remember, one of the key points, especially for fast speech, is knowing when NOT to put phonemes in. If it looks like your characters mouth is opening and closing too fast, it means you need to drop frames. Think about how the mouth would blend between the phonemes, and then stick that pose in place of several others.

- Generally, I allow for at least one inbetween for every phoneme. That is, I will never have 2 different mouth positions keyed on consecutive frames. If the speech is fast, I will drop one. The general exception to this rule is the M/B/P and F/V poses, followed by a vowel. In some cases, you may want to go from a closed mouth straight to a large open mouth. This gives more snap and may look better. Ususally I tend to use at least one inbetween though.

- Make sure you hit the M/B/P phonemes and large vowels at the right point. As long as you get those extremes at the right point, your animation should look 90% correct. In addition I usually hold the M/B/P and F/V poses for at least 2 frames. So that the mouth is closed at least fairly tight for 2 frames, then has one inbetween, and finally is followed by a vowel. Note this is a general rule and sometimes has to be disregarded.

- Consonants tend to take on the shapes of the vowels around them. The word "kind" and the word "look" would have different shaped 'K' phonemes. Remember to take this into account when hand animating. The W/OO vowels tend to have a very big effect on consanants they are close to.

- Buy a mirror and use it when animating!

Special Thanks

I'd like to thank **Doug Kelly** for his ongoing support and tips, and for helping me get my first real job in the industry. **Kim and Steve Oravecz** for their continuing advice, comments, and general zany

creativity. **Jeremy Bernal** for being my first true animation mentor, and for helping me "learn the ropes". **Mitri Vanichtheeranont** for his critical eye, help with proofreading this article, and putting up with my spontaneous singing at work. My **parents and grandparents** who helped make my dream of animating become reality through their loving support (and funding of expensive computer equipment). **Rick May** and the entire CG-Char Mailing List for all the advice and help it has given me.

About the Author

Michael Comet is currently a Rigger/T.D. at Blue Sky Studios in New York. Previosuly he was Video Team Lead Rigger and Cg-Supervisor and a 3D/Animator/Artist at Big Idea in Lombard, IL where he is worked on 3-2-1 Penguins and Veggie Tales. Prior to that he was lead animator at the video game company Volition, Inc., where he animated most of the cinematic sequences for Descent: Freespace, and headed up much of the realtime character animation and cinematics for their RPG title, "Summoner". He can be reached via email at comet@comet-cartoons.com, and has a personal homepage at: http://www.comet-cartoons.com/ which has more information and samples of his work

About the Sample Images

The head used was modeled with the Surface Tools plugin and 3D Studio MAX r1.2. The character is part of a personal project I am working on.

Bibliography

"Cartoon Animation", Preston Blair
> Walter Foster Publishing, Laguna Hills, CA 1994. ISBN: 1-56010-084-2

"Digital Character Animation", George Maestri,
> New Riders Publishing, Indianapolis, IN 1996. ISBN: 1-56205-559-3

"Lightwave 3D 5 Character Animation f/x", Doug Kelly
> Ventana Communications Group, Inc., Research Triangle Park, NC 1997. ISBN: 1-56604-532-0

"The Animator's Workbook", Tony White,
> Watson Guptill Publications, New York, NY 1990. ISBN: 0-8230-0229-2

"The Artists Complete Guide to Facial Expression", Gary Faigin,
> Watson Guptill Publications, New York, NY 1990. ISBN: 0-8230-1628-5