-INNOVATIONS REPORT-

Animation Lip Sync

Sebastian Huart d1159930
BACVA 3

Contents:

Abstract:

The overall goal of this project is to create a short lip sync animation from a chosen sound clip, then follow the animation pipeline of designing and modelling a character from scratch that would best fit that voice clip and then search for a facial setup method that would allow me to quickly animate the character to produce the final product.

Introduction:

My initial innovations idea was to animate a realistic head as a lip sync animation with all the facial details [11], but after reading an article in 3D World magazine[4] about Masahiro Mori's "Uncanny Valley" [5], I then searched for more information about this topic. It describes the emotional response of humans towards robots and how the closer the robot's likeness becomes to humans, the more empathy it may receive until an extreme where it becomes strongly repulsive. However when the appearance and motion become more indistinguishable, the emotional response is positive again.

In the article previously mentioned, they discuss how this phenomenon is now being applied to games and films. I learned that "the closer the animators get to photo real characters, the higher the [viewer's] expectations are for a character's behaviour".

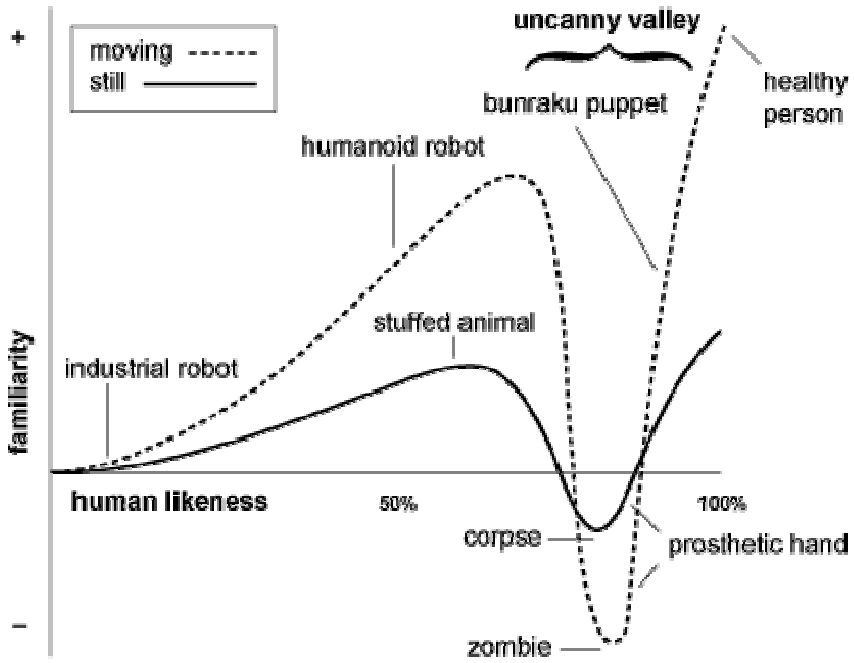

Fig 1: Mori's "Uncanny Valley" describes how our response to robots or CG characters alters according to how realistic their appearance is.

This is especially true when talking about modern day films such as the "Polar Express" [15] or "Final Fantasy – The Spirits Within" [16].

Fig 2 & 3: Realistic modelled and rendered humans, but lacking in natural and convincing animation.

The human models look convincingly realistic but the animation fails to complement it. This made me think back to what I really wanted to do: to create a lip sync animation. I therefore decided to do it as a cartoon animation. This would allow me to use extreme poses and allow for more creative freedom on the project. The innovation in this project would be the fact that this all new to me and seemed like fun. I also wanted to approach this project with a trial and error philosophy, as I have found this to be the best way to learn.

The Audio file:

I started my project by looking for a suitable audio file to work with. While a famous quote from a popular movie would be an obvious choice, an interview in the style of Creature Comforts [14] interested me more for the sheer anonymity of the character. I obtained some sample interview files from a fellow journalism student but found them to be fairly long in length and generally uninspiring. The need for a short and characterful interview was essential but hard to find, even with the vast resourcefulness of the internet. Finally the search led me to the popular video site www.youtube.com. I looked into stop motion dialogue studies made by students. I found a small interview clip that had an interesting voice sample, strong in tone and character that would be both challenging and fun to animate to. Here is a transcript of it:

"My greatest fear from childhood... think would be having the closet door open... I... well no closed! I like it open."

Before designing the character that would go with my sound clip, I spent time sketching the key poses that I thought were in keeping with the audio.

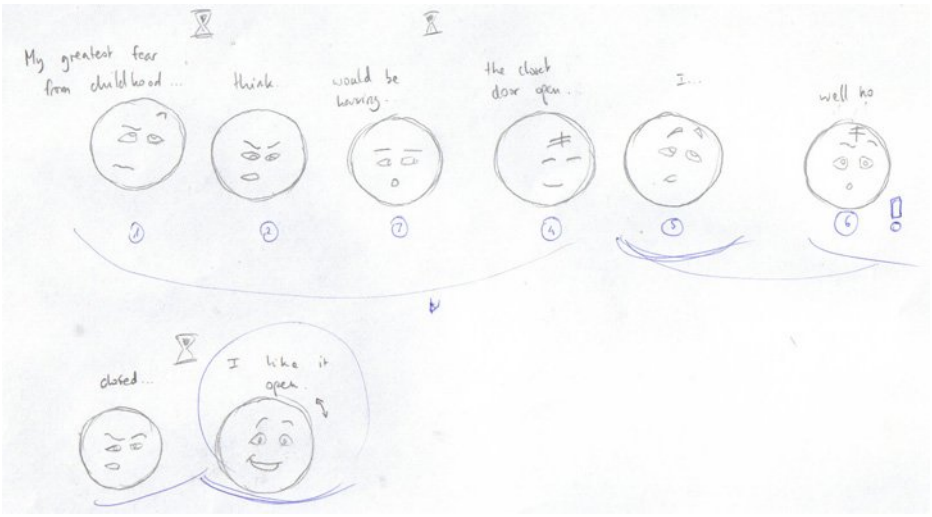

Fig 4: Key poses from the audio clip

Designs:

Finding a character to fit the voice was the next challenge. I wanted there to be a stronger link between the voice and final character as this is something that lacks in peoples animation. Designing a human character proved to be quite hard and I was not happy with any of the drawings I had done. I therefore also looked into anthropomorphism [8], using an animal to voice my animation. Through careful study of the voice, I chose to have a stereotypically big imposing animal as my character, playing on the fear part of the clip thus giving the piece a small comical twist. My initial idea was a polar bear as the voice had a small tone of lethargy and suited the animal well. In my sketches, I went for quite an exaggerated look with a big nose and mouth with small ears and eyes, to change it from the ordinary.

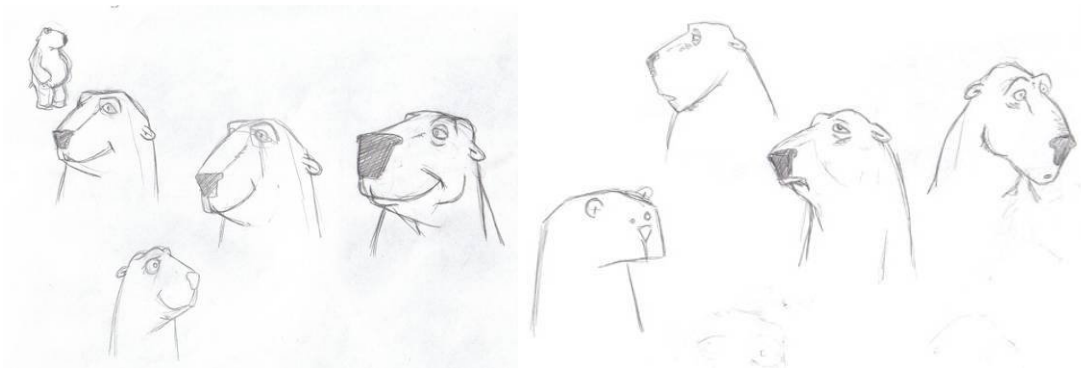

Fig 5: Early concept drawings.

Fig 6: A few concept design sketches of the polar bear.

Modelling:

This stage was pretty straight forward. With a front and side image, I quickly modelled the bear close to my design sketches. The model went through various stages based on feedback from fellow students. The feedback allows you to judge whether the personality of the character comes through enough, as in the end you want the viewer to make the logical link too. I kept the polygon count low to keep it simple for the rigging and skinning stages. I could then smooth the model afterwards. One important fact to remember when modelling for animation is to have clean topology and continuous edge loops, especially around the mouth when doing lip sync, as this will allow for nice deformation later on.

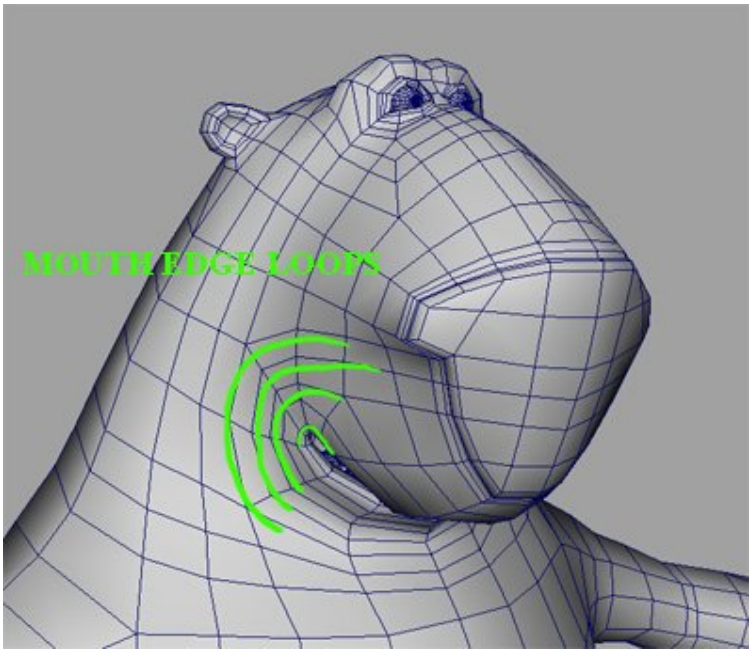

Fig 7: Clean topology will help deform properly in the animation stage.

Bad topology will distract the viewer and the animation will consequently suffer, so getting it right at the beginning is critical.

Rigging:

Rigging was one of the most important parts of the project and can determine if the animation looks good or bad. The main aim is a rig that will make animation easy and has quick feedback for positioning and playback. I explored several methods and the results were diverse.

Deformer methods:

Curve based:

As I searched the internet for a facial rigging tutorial to help me create a bone rig, I found an intuitive method from a tutorial on highend3d.com [9] that used clusters to control curves that in turn deformed the mesh. The curves follow the muscle structure to allow for natural deformation.

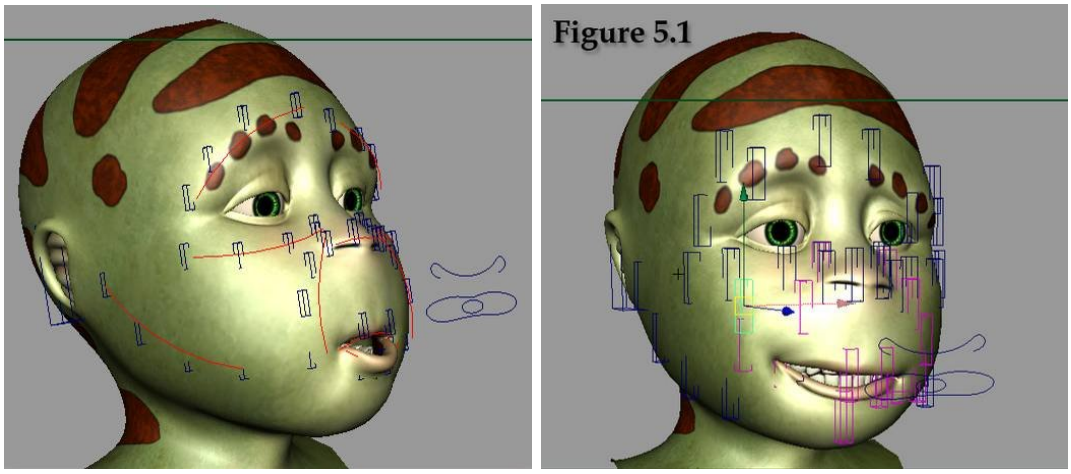

Fig 8: Curves controlled by clusters, deforming the underlying mesh to create natural looking facial expression.

As I went through the tutorial, I had to use their custom scripts to create the curves and bind them to the mesh. But as I tried to get the curves to deform the mesh, it did not work unless the clusters were in relative mode. A big problem arose as this made the mesh completely black. Even after countless efforts to resolve the problem, the polar bear was completely black and no texture material would be able to change it.

In hindsight, spending more time on figuring out the problem would have maybe resulted in a usable rig, but from testing the darkened model, the curves would not deform the bear's face properly and gave bad results. This may have been because the model was low poly and maybe required more detail and extensive weight painting, something to keep in mind for future reference.

As a whole I was disappointed with this method as I was expecting a much better result given to me by the images from the tutorial.

Joint based:

I then decided to move onto a joint driven facial rig. This is a method widely used in games and was the original method I was going to use to animate my character with. To help me with placing the joint on the bears' face, I looked at the www.cane-toad.com [12] animation website. The tutorial page gives us an overview of the facial rig they used in their production, and from the animation looked promising.

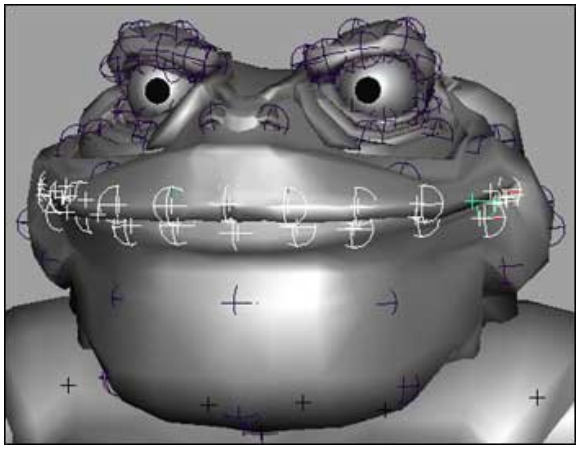

Fig 9: Joint based facial rig from the cane-toad animation

The facial rig involves having joints on many of the vertices to allow control over the whole face. The joints rotate around the locator it is parented to. This method is sometimes used for eyelids on rigs to have them rotate around the eyes. This concept was just extended onto the whole face. The joints can then be animated as if they were sliding over the underlying skeleton and give convincing deformation, especially around the mouth and eyebrows. With the joints in place, their movement would be controlled through attributes and would create the different poses needed for the animation. In turn these poses can be saved into attributes where an animator can just slide the values to get the desired facial expressions.

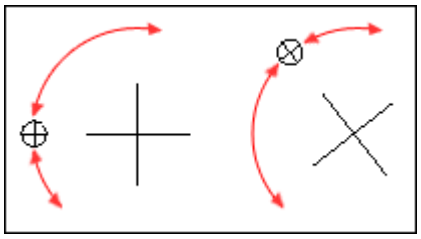

Fig 10: Rotation of the joint around the locator

d_mouth_Expression	
Squeeze In Out	-3.3
Smile	0
Pout	10
Frown	10
Grimace	10
Think	0
Oooo	0
Left Right	0
Rotate Mouth	2.7
Top Lip Up D	-6.8
Bot Lip Up D	-7.7
Arghh R	0
Arghh L	0
Grrr R	0
Grrr L	10
Bot Lip DR	0
Bot Lip DL	0
Smile Up DR	0
Smile Up DL	2.5
Sharp Up DR	0
Sharp Up DL	0
Grimace R	10
Grimace L	0

Fig 11: Facial expression library from cane-toad animation

I also read through “Maya Character Creation” from Chris Maraffi [3] as he introduces a joint based facial rig at the end of his book. The setup looked impressive although some of the shapes created looked unnatural.

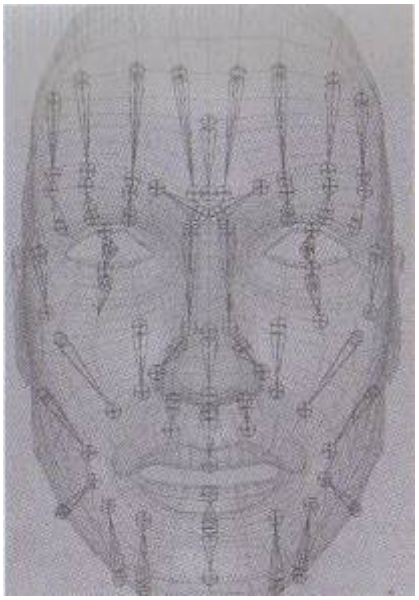

Fig 12: Joint based rig based on the anatomy allowing correct anatomical movement of mesh

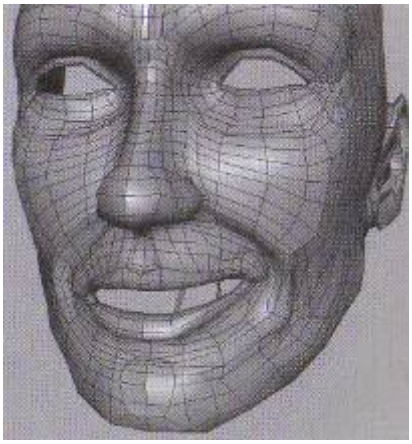

Fig 13: Small problems with deformation. Target expression will be based on good position of joints and extensive weight painting.

This concept is extremely powerful and gives the animator the ability to save individual facial poses into a library of expressions. The amount of joints that must be placed to get enough control over all the head is the only concern I had when looking at this method. The time taken to place the bare minimum of joints made me wonder if it was the right method to choose for an animation that would only last a few seconds. This method, in its basic form, is close to that of blendshape. The only differences are that the animators create the target blendshapes, not the modellers, and they have the control to modify the different shapes.

As a result, I have learned that the use of joints in a facial rig is suited for a more team based production, where one person would rig the character and another would animate. From a pragmatic perspective, I felt that this did not suit me as the amount of time and work involved in setting up this joint rig for a short lip sync animation was far too great. I was therefore determined to use blendshapes for my animation.

Blendshape methods:

Phonemes:

A good lip sync tutorial website I had come across in the past was Michael Comet's website www.comet-cartoons.com [10] where he gives us an overview of the process of animating dialogue using phonemes. These are single shapes a mouth can take to make a particular sound. He gives the example of "ooo" in groovy and how the mouth puckers to create a small circle. Each word can then be broken down into phonemes like the example he gives of the word "Animation" that becomes "a-n-ee-m-ay-sh-u-n". This method was used in 2D animation for lip sync and this tutorial shows us how to recreate it in 3D. The lowest amount of phonemes used is 9, but to get more realistic animation, 13 is a good amount and the general breakdown of these would be:

1. M B P	(Map, Bang , toP)
2. C K G	(Carry, looK)
3. CH SH J	(CHerry, SHout, Jump)
4. F V	(Fine, loVe)
5. A	(Apple, blAde, Ape)
6. I U	(If, Under, wOnder, pIck)
7. O	(Open, Oat, Over)
8. E	(swEEt, EAt, fEEt)
9. N D T L	(Name, Dove, abouT, faLL)
10. TH	(teeTH, forTH)
11. S Z	(Snow, Zoo)
12. R	(Roll, dooR)
13. W OO Q	(wOOd, dreW, fOOd)

From this we can create the shapes need for the animation. Additional blendshapes would be used for the eyebrows.

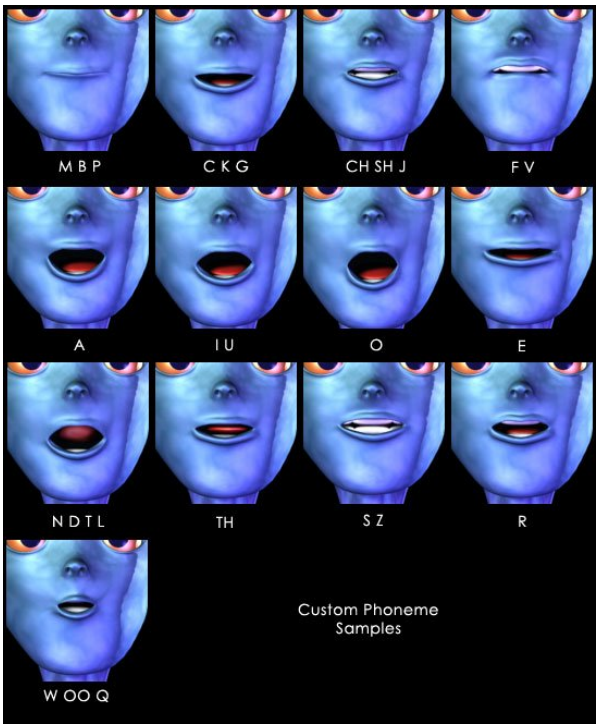

Fig 14: Breakdown of phonemes into blendshapes, ready for use in animation.

The next stage would be to break down the audio clip in phonemes and place the timing of each one on a dope sheet.

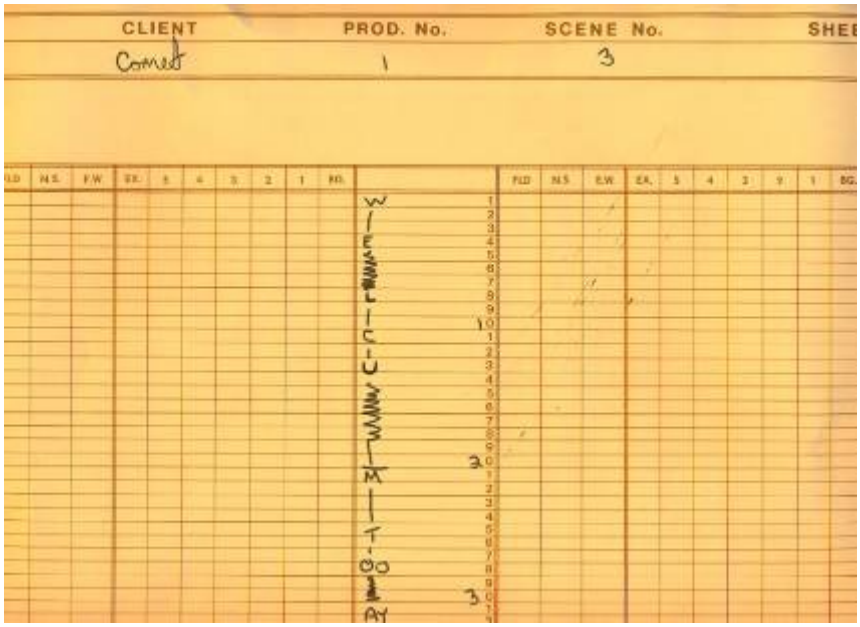

Fig 15: Phonemes are placed in the dope sheets against the audio clip to determine the keyframe number.

As I was reading this tutorial, I also started to read the book: “Stop Staring” from Jason Osipa [1] which I bought at last years 3D December. In it he discusses why a phonemes system is not the best for computer graphics. This led me to read further and to discover the problems of phonemes in a 3D lip sync environment.

After reading a few pages, I found this book to be more relevant than the website on the matter of creating a more natural looking animation, so I decided to continue with this book rather than the phonemes method.

Visimes:

In his book, Jason Osipa explains how the use of phoneme-based lip sync means that the exact theoretical sounds of the words are used from the breakdown and not the mood or shape of the mouth from the person speaking. He illustrates the idea behind a new concept called “visimes” with an example by using the letter “G” in different poses.

Fig 16: Different variation of the letter "G". The phoneme based method would have difficulty trying to create these target facial expressions.

This underlines the fact that by using the phoneme-based lip sync, the shapes become too habitual, something that looks unnatural and does not take into account of the way in which the character speaks. We should mention that most people do not articulate words precisely, if at all in some cases, making the phoneme-based method problematic. If we are to look at people speaking, we can see that the mouth follows, at its basics, a combination of open/close and narrow/wide movements. This coupled with some more advanced movement can be used to describe the whole mouth movement and therefore we can create all the shapes we need, and get a much more natural looking speech, rather than just a misfit mouth.

He bases this system on "visual phonemes" or "visimes" as he calls them. These are important shapes or visuals that are made by our lips. He clarifies it all by saying "If phonemes are sound, visimes are shapes." He goes onto to say that these visimes are linked to sound, rather than the exact letter in the text. This allows us to tailor the mouth to any voice tone, such as muttering or expressive mouth movement, something that phonemes have difficulty doing. This is largely noticeable in a lot of short animation tests, where different toned voice clips have the same animation, generally only differing in the head movement.

From these shapes or parts of shapes we can construct more customized and complete shapes.

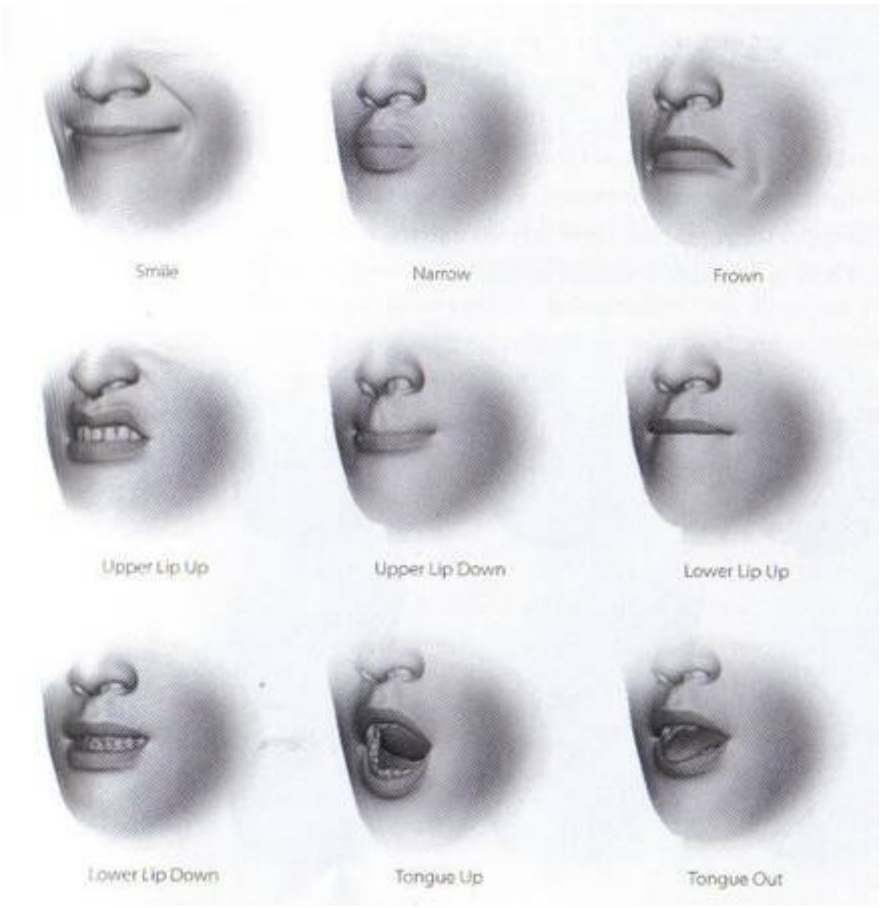

Fig 17: The 9 shapes used to create all other shapes.

These shapes are then split into Left and Right parts giving the animator more control over the shape of the final expression. These next images illustrate this powerful way of doing lip sync.

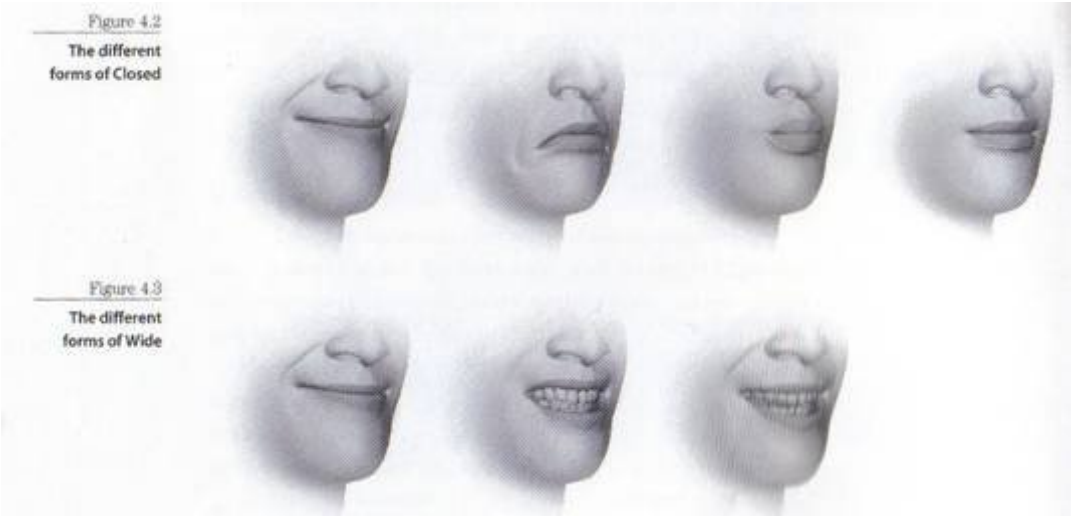

Fig 18: Different variations of one target expression allowing a more customizable facial rig.

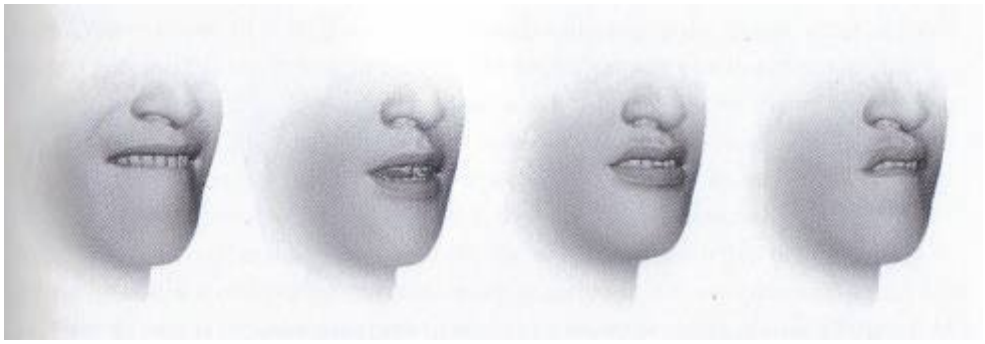

Fig 19: Different forms of F, making it extremely flexible for all kinds of voices.

After seeing these images, I started to create my blendshapes. I had eight blendshapes for the mouth, the split up into Left and Right parts to create the completed shapes. Six blendshapes were for the up and down eyebrow movement. The eyes lids were animated with joint rotating about a locator.

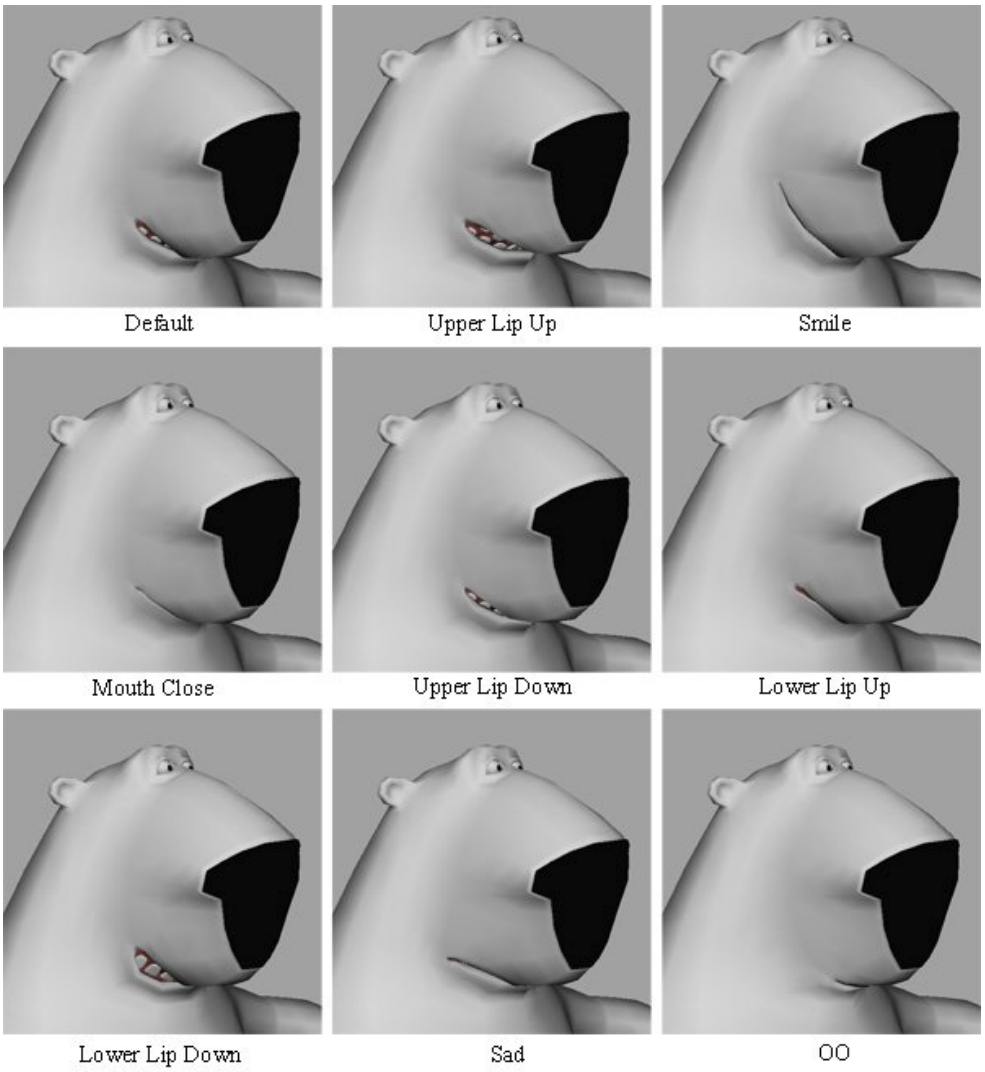

Fig 20: The blendshapes I created for my lip sync animation.

Even though I had never done lip sync before, I had used blendshapes in

past projects and had found that having a few extra default shapes along side will allow you to go back and add a shape that was either not envisaged before hand or correct other shapes.

A good tip also is not to delete your blendshapes. They may add geometry to your scene but will give you the possibility to change them at any time during the production.

We can reduce polygon count by having just the heads as the blendshapes rather than the whole body. This is done by extracting the head from the body, duplicating the head, then combining it back with the head as the first object. This means that the vertex ordering on the head is the same as the duplicated head even though they do not have the same geometry. The blendshapes will continue to work. On applying the shapes, make sure check topology is turned off.

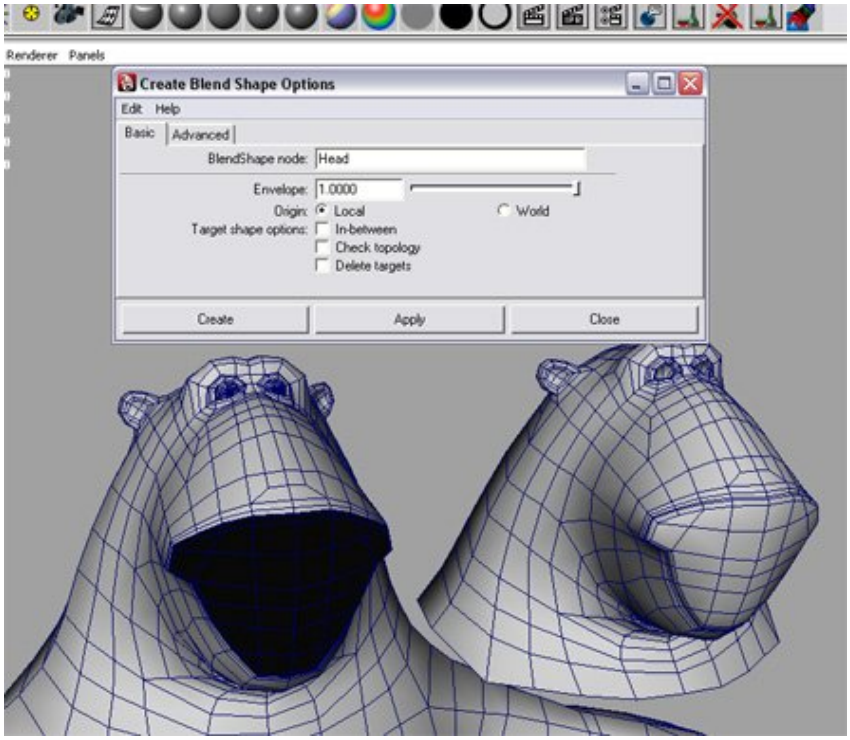

Fig 21: With Check topology turned off, we can have a shape that does not have the same geometry as a blendshapes.

Once the shapes setup, the blendshapes were assigned and the rest of the body rig was set up.

Even though I was concentrating on the facial animation, I wanted to add motion to the body, to avoid a static polar bear. Hand movement could also add more realism to the animation.

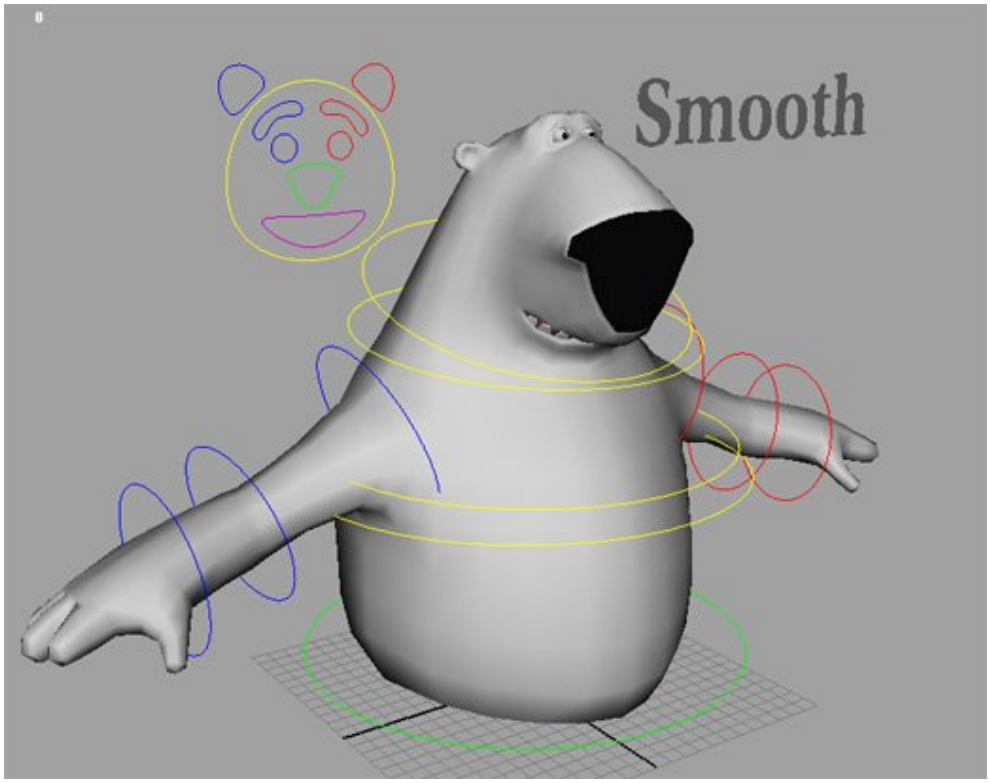

Fig 22: Simple body rig. Head controls allowing easy access to blendshape attributes.

One mistake I had made was to have the jaw's movement as a blendshape. My initial thought was that it would not move much but when I started trying out the blendshapes, I realised that a jaw joint would allow me to have the mouth open and close without the use of blendshapes and could animate in a natural arc rather than a linear line. I had luckily created a jaw joint to move the teeth, so I painted the weights and had a smooth moving jaw.

Animation:

The animation would be done on a trial and error basis, giving me the chance to realise my own mistakes and find a solution to them. I wanted to try a straight ahead approach to the animation as I have noticed that animations created in this way, appear more natural and less robotic than pose to pose. This would be done in stages though, first animating the mouth, then the head and eyes and finally the arms. As I started animating, I first wanted to test the "visimes" method out, to see how quick and easy it was. To help pose the polar bear's facial expressions, I used a small mirror as reference. The mirror allows you to act out the animation and therefore understand the motion better. The first few seconds of animation were done with relative ease, describing the movement of the mouth based on the voice was both enjoyable and hassle free. I discovered that by animating the movement of the mouth a few frames before the actual sound, the lip sync came across as more natural and convincing. Having the movement come late or on time gave the impression of

it being out of sync. By playing the animation with sound over and over again, small inconsistencies in the lip sync are quickly noticed, and I have found that things that look right are right, and things that do not need changing.

This way of animating became very intuitive and straight forward. Unlike the phonemes method where each word is broken down, mapped against a dope sheet and then keyed on those frames, I was able to just listen to sound and mimic the shape of it. This therefore allowed me to create a more seamless link between the polar bear and the voice, one of the criteria of my challenge.

Once the mouth had been animated, the head and eyes were next. Getting the polar bear to think convincingly and naturally was probably the hardest part of the animation. A good example would be on the first few words: "My greatest fear from childhood". The initial way I animated him, was to have his head tilt up thinking, bringing it back at every high note and then down on low notes. This resulted in him having his head tilt back too much and looked odd. After countless tries of getting the motion to look natural, I turned to the "Stop Staring" book for to look for any tips. At the end of his book, Jason Osipa has some overviews of animations he has done and describes the process of how he made them. In his "Pink or Blue" animation, where a girl can not decide between a pink or blue bow, he is confronted with the same problem. He says:

> *"If you haven't already noticed, when something fails to work for me, my first follow up attempt is to do the exact opposite of what I tried initially!" [Jason Osipa, 2001]*

On that note, instead of having him with his head up first, thinking, I had him tilted forwards in deep thought, then tilting his head upwards, searching for something to say. This proved to work well and was a more natural movement.

The next challenge was the timing of his shift in behaviour, when he realises his mistake. This was mostly based on trial and error and playing the animation back and forth, to get the right timing. Shifting the keyframe along by one frame can make all the difference in the animation. I also found that by giving the motion anticipation, the action had more significance. Before turning his head around, a quick counter movement emphasises the subsequent act. This technique was also used on the eyes, holding them closed the frame before.

To add more visual representation of a change in tone, such as the key word "fear", I decided open his eye lids on the up tone. The result gave more emphasis on the word and had more chance of registering with the viewer as important information.

. When I had initially acted out the part, my arms and hands became important in stressing some aspects of the action. I wanted them to complement the mood and facial animation but not distract the viewer's attention, while keeping it all natural. I took the knowledge previously gained from the eyes and applied it to the hands.

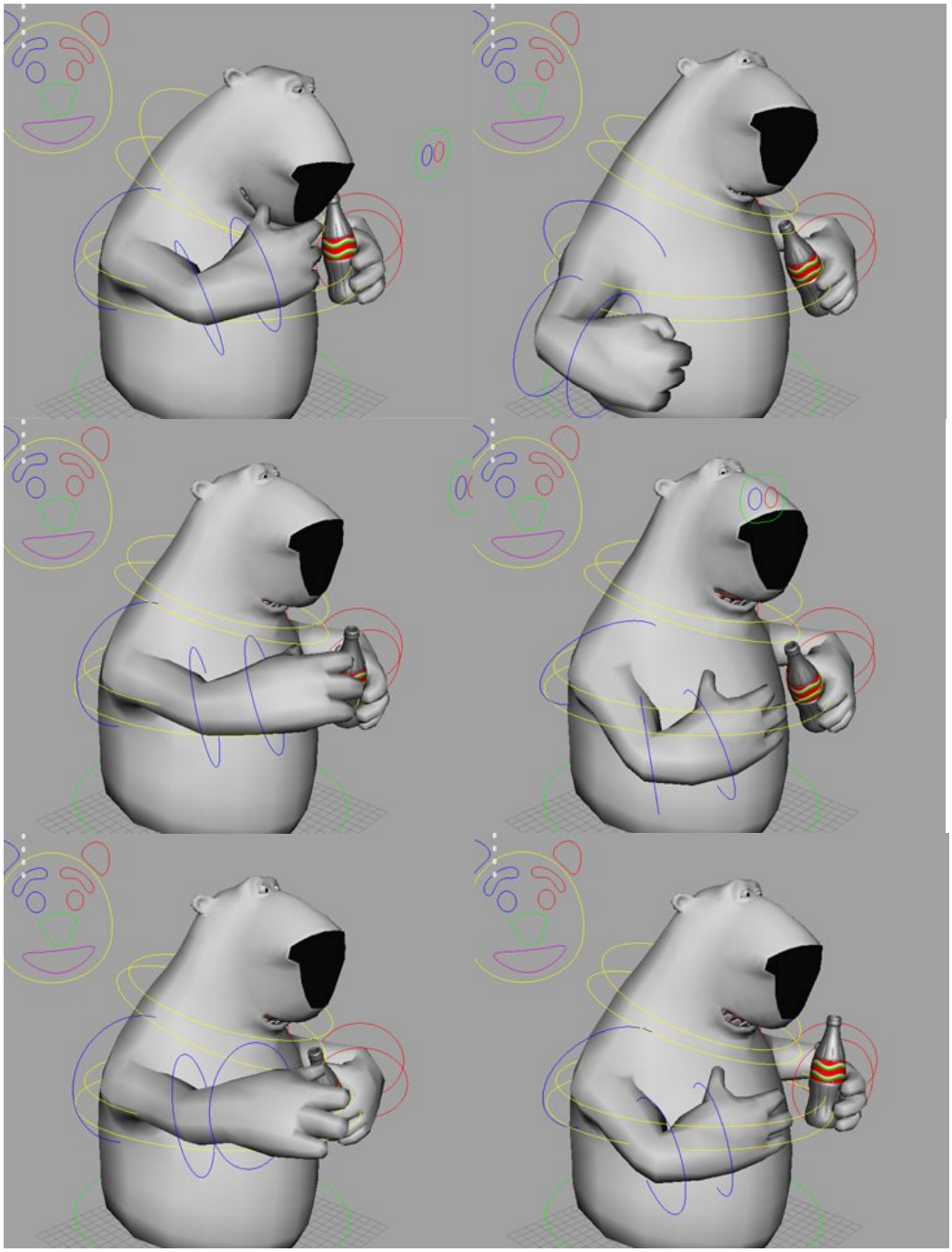

Fig 23: Animation of the arms to complement the facial lip sync.

The piece was now starting to feel as a whole, linked by subsequent motion. Stepping back from the animation and viewing as a whole will make sure that there is consistency and continuity throughout, so as not to give the impression of a modular or fragmented animation.

By applying what I had learnt throughout the animation phase, I was able to work quickly and efficiently and the project became much more of a fun learning experience. Although trying the same motion over and over and have it not work can be disheartening, but when it does finally “click”, the sense of achievement largely outweighs the hardship.

Conclusion:

Even though I was unable to animate using a joint-based facial rig as I first intended, the time spent on them would have meant that I would have been unable to learn about this new lip sync animation method of using visimes. However I have learnt a lot from this project and the challenges that I have faced throughout. This is the first time I have ever tried lip syncing and the knowledge I have gained from undertaking each method and discovering their pros and cons will help me in future projects as I now know a straightforward and effective method of character setup and animation. The ease of using the "visimes" method has led me to think about producing more facial animations in the future.

The main underlying problem I did have throughout the project was the animation playback. Maya had difficulty playing the animation of the low poly version of the bear to the audio file and kept lagging. I therefore had to listen to the animation more than a few times each time, to make sure the timing was correct in my animation. A possible solution would be to playblast the animation with the sound included and view it, but this is only possible on the windows version of Maya. The other disappointment was the curve based deformation rig with the clusters incapable of modifying the curves properly unless the mesh was black. With more time I would like to have solved the problem and tried to animate with it. However with the short time frame given, this led to me have a more pragmatic philosophy about the project, weighing the pros and cons of methods and finally finding one that would get the job done fast, an attitude that will be useful later on in the future.

If I was to do this project again, I would do the animation with a longer voice clip, but with the same amount of quality as this one had, maybe even have two characters. If the voice were more extreme, I would also make sure that I had more extreme poses for blendshapes, as mine were more subtle and closer to a human mouth, which maybe gave it a more natural look. I would also like to have had time to look into and try to implement a rig that can be directly modified on the face, a method that is commonly used in many CG companies.

Overall I am very happy with the outcome of the animation and even though it was tedious at first with the setup, the animation stage was especially satisfying and has led me to discover more about animation and to think about it even as potential career choice.

References:

Litterature:

[1] OSIPA, J, 2001. *Stop Staring*. USA: Sybex inc.

[2] RITCHIE, K, CALLERY, J, AND BIRI, K, 2005, *The Art of Rigging – Volume 1* .USA: Cg Toolkit

[3] MARAFFI, C, 2004, *Maya Character Creation*. USA: New Riders Publishing

Articles:

[4] FUTURE. 2005. The games industry's challenge. *3D World Magazine* Issue 70, p22

Websites:

[5] 2007. *Uncanny Valley* [online] Wikipedia. Available from: http://en.wikipedia.org/wiki/Uncanny_Valley

[6] 2005. *Uncaany Valley* [online] Karl F. MacDorman and Takashi Minato. Available from: http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html

[7] 2007. *Heavy Rain, PS3 Demo* [online] Wikipedia. Available from: http://en.wikipedia.org/wiki/Heavy_Rain_%28game%29

[8] 2005. *Anthropomorphism* [online]. Wikipedia. Available from: http://en.wikipedia.org/wiki/Anthropomorphism

[9] 2006. *Facial Animation Rig for Delgo* [online] Warren Grubb. Available from: http://www.highend3d.com/maya/tutorials/character/276-1.html

[10] 2007. *Lip Sync – Making a character Speak* [online] Michael Comet. Available from: http://www.comet-cartoons.com/3ddocs/lipsync/lipsync.html

[11] 2000. *Synthesizing Realistic Faces* [online] Frederic Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, David Salesin. Available from: http://grail.cs.washington.edu/projects/realface/

[12] 2005. *Cane-Toad Bone Facial Animation* [online] Andrew Silke, David Clayton. Available from : http://www.cane-toad.com/tuteRig_Facial.htm

[13] 2004: *Lip Tutorial* [online] Chris Fram. Available at: http://www.dirtyoldtoon.com/tutorials/LipTutorial_01.htm

Films:

[14] Creature Comforts. 1989, Film, Directed by Nick PARK UK: Aardman Animations

[15] The Polar Express, 2004, Film, Directed by Robert ZEMECKIS USA: Warner Bros. Pictures

[16] Final Fantasy – The Spirits Within, 2001, Directed by Hironobu SAKAGUCHI and Moto SAKAKIBARA Japan: Square Pictures

Images used in text:

Fig 1:
Image [online] *Uncanny Valley* [2007]. Available from: http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html

Fig 2:
Image [online] *Final Fantasy – The spirits Within 2001.* Available from: http://www.ocf.berkeley.edu/~amarisw/images/spiritswithin.photo.jpg

Fig 3:
Image [online] The Polar Express 2004. Available from: http://www.smh.com.au/ffximage/2005/11/14/Polar_051114104811699_wideweb__300x375.jpg

Fig 4:
"Key poses from the audio clip."

Fig 5:
"Early concept drawings."

Fig 6:
"A few concept design sketches of the polar bear."

Fig 7:
"Clean topology will help deform properly in the animation stage."

Fig 8:
Image [online] *Facial Animation Rig for Delgo* [2006] Available from: http://www.highend3d.com/maya/tutorials/character/276.html

Fig 9:
Image [online] *Cane-Toad Bone Facial Animation* [2005] Available from: http://www.cane-toad.com/tuteRig_Facial.htm

Fig 10:
Image [online] *Cane-Toad Bone Facial Animation* [2005] Available from: http://www.cane-toad.com/tuteRig_Facial.htm

Fig 11:
Image [online] *Cane-Toad Bone Facial Animation* [2005] Available from: http://www.cane-toad.com/tuteRig_Facial.htm

Fig 12:
Image taken from MARAFFI, C, 2004, *Maya Character Creation*. USA: New Riders Publishing

Fig 13:
Image taken from MARAFFI, C, 2004, *Maya Character Creation*. USA: New Riders Publishing

Fig 14:
Image from downloaded tutorial: *Lip Sync – Making a character Speak* [2007] Michael Comet. Available from: http://www.comet-cartoons.com/3ddocs/lipsync/lipsync.html

Fig 15:
Image from downloaded tutorial: *Lip Sync – Making a character Speak* [2007] Michael Comet. Available from: http://www.comet-cartoons.com/3ddocs/lipsync/lipsync.html

Fig 16:
Image taken from OSIPA, J, 2001. *Stop Staring*. USA: Sybex inc.

Fig 17:
Image taken from OSIPA, J, 2001. *Stop Staring*. USA: Sybex inc.

Fig 18:
Image taken from OSIPA, J, 2001. *Stop Staring*. USA: Sybex inc.

Fig 19:
Image taken from OSIPA, J, 2001. *Stop Staring*. USA: Sybex inc.

Fig 20:
"The blendshapes I created for my lip sync animation."

Fig 21:
"With Check topology turned off, we can have a shape that does not have the same geometry as a blendshapes."

Fig 22:
"Simple body rig. Head controls allowing easy access to blendshape attributes."

Fig 23:
"Animation of the arms to complement the facial lip sync."

Acknowledgements

Thanks go to Claudia Moore for guiding me through the project and Pete O'Campo for proof reading and to all others who have helped me with their opinions and critiques.