Through steps such as these we can understand how it is, that as soon as some melancholy thought passes through the brain, there occurs a just perceptible drawing down of the corners of the mouth, or a slight raising up of the inner ends of the eyebrows, or both movements combined, and immediately afterwards a slight suffusion of tears.
Charles Darwin, 1875

Introduction

This paper is about how an animator may represent emotion. The approach taken is that of a practitioner-researcher; it begins by engaging with the related academic discourse, before progressing into a self-reflection of the artefact and animation in progress.

Exactly how much, how deeply or indeed how to express emotion remains a difficulty for the animator. They must somehow draw upon an innate understanding of how nonverbal communication and expression works in order to express it artistically, and in a way that can be recognisable to the viewer. While actors are able to tap into expression and gesture through, for example, the Stanislavsky technique of drawing on previously experienced emotions and letting them bleed naturally into the face and posture of the body, (Benedetti 2005) an animator has to register within themselves and then consciously construct such emotions, (Buchanan 2007) either by physically drawing them, or otherwise rendering them on a computer.

In contrast to live action, animators often resort to symbolic body movements and gestures; unambiguous, widely used and professionally accepted techniques and signs that are easily understood by the viewer. However, in the process we can sometimes by-pass the more subtle nuances of real human body language and expression (Buchanan 2007). As animators, we are often taught or indeed encouraged to exaggerate or almost over-emphasise the expression of emotions and gestures within our characters. Walt Disney was always asking his animators to make their work more “realistic” and yet when they did this, he complained and asked them to exaggerate the movements, again and again. One confused artist commented “I don’t think he meant ‘realism’ I think he meant something that was more convincing.” (Thomas, Johnston, 1981) The point being, that realistic animation is not always the primary aim of the (non-VFX) animator, and that perhaps being convincing and entertaining is the primary purpose, even if that involves exaggerated or symbolic gestures.

However, I wished to investigate the subtleties of how an animator might approach the expression of real and genuinely felt emotion, drawn out from within myself as auteur-animator-researcher, aiming to explore the practical difficulties in animating "pure" non-verbal communication without resorting to established symbols. While my research investigates though animation and personal reflection both body language and gesture and facial expression, for this paper, I will be focusing on facial expression and the personal exploration of two specific emotions, happiness and sadness.
Within this, a large part of my research would involve the practice of animation, involving 3D animation, motion capture, rotoscoping and hand drawn, more free-flowing animation. However, in this paper, I will discuss specifically the rotoscoped section of my research, which specifically involved rotoscoping live action filmed footage of myself experiencing genuine happiness and sadness. The methodology would be framed as a narrative enquiry; in particular, an auto-ethnographic analysis of the work in progress. “Narrative researchers [...] view themselves as narrators as they develop interpretations and find ways in which to present or publish their ideas about the narratives they studied.” (Chase 2005, p. 657). Though filming myself and then transposing my own facial expressions into rotoscoped animation, the work could not fail to take on a highly personal basis both as practitioner animating and creating the artefact, but also as the subject being filmed and analysed through animation. I was able to reflect and document each stage in sequence, as a route to analysing non-verbal communication through the non-verbal cues and gestures of my own filmed body and documentary-style account.

The advantages of such an approach are an “insider knowledge” (Adams, Jones et al, 2014) in this case, the insider knowledge of an animator practitioner, attempting to make sense of their work through the practice of animation, itself based on very person expressions of emotions drawn from personal memories. The disadvantages are that this auto-ethnographic (Denzin, 2013) approach is by nature extremely personal, my own ways of expressing emotion cannot necessarily encompass how all people might express emotion, and the paper that follows is particularly self-reflective from my own viewpoint and practice within my work, which itself is in part a performance piece. However, as one of many available approaches to studying the expression of emotion, it is hoped that this work can provide some interesting insights to promote further questions and exploration.

Emotion and Facial Expression within Psychology Research

Psychology researchers name six to seven basic emotions which can be seen expressed on the face; happiness, surprise, fear, anger, disgust, sadness and contempt. Other emotions are considered to be blends derived from mixing the features of these primaries. These base expressions of emotion are considered to be universal, cross-cultural expressions. (Russell, Fernández-Dols 1997; Ekman, Friesen 2009) Expressions can be spontaneous, genuine and truthful indications of an underlying emotion, but they can also be voluntary, false, and not necessarily with deliberate intent to deceive or lie. We are culturally conditioned to school our emotions in order to interact within our cultural social groups, to follow specific “display rules” (Ekman, Friesen 2009, p. 20). All babies cry when hurt, but little boys are encouraged to stifle tears, masking or inhibiting a tearful expression, a cultural display rule that stretches into adulthood. Ekman goes into further detail, describing “personal display rules” which are generally taught within the family unit and may be idiosyncratic to a particular upbringing.

Approaching this from the point of view of an animator, this opens up interesting points. It may not be enough to simply study facial expressions in order to express genuine emotion upon an animated character. To obtain a deeper level, the animator may need to consider the display rules of their character, their culture, their upbringing, and since a classic exercise is for the animator to use a mirror to study and replicate their own features, she needs to be aware of her own display rules, both cultural and personal. The point about display rules, is that as they are learned at such an early age, from being immersed into a culture or upbringing, that they are automatically and unconsciously made, an animator may not be aware they are expressing them when they use their own face for reference, or fail to add them into an animation through focusing on the base, unmasked and uninhibited emotion.

However, even within display rules, true emotions can leak out in the form of micro-expressions. While facial expressions can last for several seconds (very long lasting expressions tend to be more voluntary,) micro-expressions can be extremely fleeting, a quarter of a second to flash across a face or even less. Thus when you feel angry, you might be also aware (due to display rules) that showing your anger is inappropriate. You might try and hide your anger, clamping down on the expression into a more neutral pose, or switching to a different expression.
In that small moment between feeling emotion and realising you cannot show it, you may have flashed a micro-expression of that emotion (Ekman, Friesen 2009). Again, this has interesting implications for an animator. Should micro-expressions be worked into animations? And if they were, being so short lived and easily missed by a viewer, would there be any point to adding them in? What proportion of viewers would pick up on an animated micro-expression, and how might that change and split the perception of viewers between those sensing (albeit unconsciously) an emotion that other viewers would miss? Conversely, could sloppy animation that drops incongruous expressions between frames be influencing the viewer of the animation in ways the animator had not intended?

**How to approach the analysis of facial expressions, in such a way as to be turned into something an animator could explore**

Early research involved taking photographs of subjects in emotional situations (seeing the heads being twisted off rats among other things) followed by the photographs being shown to observers for interpretation (Landis 1924; Landis 1929). Taking into account the way that emotions can change and play across the face, blending and changing swiftly, not to mention micro-expressions, the use of still imagery could only show a partial, and incomplete snapshot of what was occurring. Later research made use of video footage, showing video clips of various emotion inducing footage while also filming the subjects as they watched, to minimise any possibility of playing up to being filmed, subjects would be filmed secretly, and only informed of this filming after the event, at which point they were given the option to withdraw consent and have the footage destroyed (Rosenberg, Ekman 2005). Studying the footage required a feasible methodology, for which Ekman and Friesen’s Facial Action Coding System (FACS) was developed (Ekman, Friesen 1978).

FACS is based on the anatomy of the human face, and codes expressions in terms of component movements called “action units” (AUs). Ekman and Friesen defined 46 AUs to describe each independent movement of the face. FACS measures all visible facial muscle movements, including head and eye movements, and not just those presumed to be related to emotion. When learning FACS, a coder is trained to identify the characteristic pattern of bulges, wrinkles, and movements for each facial AU (Bartlett et al. 2005, p. 393).

While FACS is certainly an exhaustive (in more ways than one) way of analysing facial data, as Bartlett and Marian point out, the process is time consuming. Video footage must be analysed by people who have undertaken about 100 hours of training in the FACS coding process, and who have additionally been tested for reliability. One minute of video can take more than 2 hours to analyse, and furthermore, to be rigorous, at least 2 trained FACS coders should comb over the same footage in order to compare results. Efforts have been made to computerize this process, however, Ekman’s website still recommends human training in order to get the best results. While undeniably a useful quantitative tool, I felt that my own research, might benefit from a more arts-based, (though in some ways no less arduous) qualitative approach.

**Self Documentary and Rotoscopying as an analytical and self reflective tool**

Taking on board the previous work with subjects in psychological research, I resolved to perform my own experiment. To film and motion capture myself experiencing emotion, and then to analyse this footage through the medium of animation. Firstly though a 3D motion captured avatar, and secondly through rotoscoping of the facial features, and finally, though a more free-form animation. It is the rotoscoped part of this research that I will outline here.

An interview situation was prepared, with a set of keywords assembled for a colleague to prompt me into talking about two subjects close to my heart that would induce emotion. In the course of a 20 minute “prompted” monologue, I was encouraged to describe a subject that made me genuinely cheerful and happy (a recent visit to India), followed by a subject that would induce sadness and distress, (my father’s ongoing chemotherapy and cancer treatment). At the time of recording, my father had just endured a major and intrusive operation in
hospital, so the distress I expressed during the filming was fresh and genuine, albeit tempered by display rules both cultural and personal.

In such a situation as this it might be considered difficult to engineer a setup where the subject could feel as natural as possible. Quite apart from the obtrusive kit, unflattering outfit and rather barren surroundings, I knew I would shortly have to speak on a difficult subject not just in front of my interviewing colleague but also in front of a technician and also one of my students who had offered technical support. Such difficulties, or contamination, have been raised as issues for consideration in facial study research (Ekman, Friesen & Simons 2005), where it is also an issue that display rules may come into play, in that a person may temper or mask their emotions (especially negative emotions) when in the presence of others.

To combat this, a small lead in or warm up chat was included as part of the footage. The technician and student retreated behind a screen where I could not see them and soon forgot about them. My colleague had an easy-going and friendly manner, such that after the initial and inevitable feelings of self-consciousness I felt I was able to relax and absorb myself in recounting my experiences. I was aware at the time of filming that I would be very likely to be expressing my own learned display rules, but as this was to be a self-reflective and auto-ethnographic exercise in many ways, recorded examples of inhibited emotion would be as interesting as an uninhibited, but uncharacteristic, display.

I was filmed by two cameras, one capturing the full body and the other set up to capture the face in close up. I also wore a motion capture suit so that my body movements could be captured. (See figure 1). Initial tests were run to try and capture facial expression. While I found the data capture points for the face fairly unobtrusive and quickly forgotten once I began to speak (see figure 2), the data captured proved too garbled and complicated to process with the software available to us at that time, so the final take was done without facial capture points.

I chose myself as a subject for several reasons, in that I wanted to induce real emotions and by using myself, I would be able to self-reflect on the (perceived) authenticity of the emotions. As these had to be real emotions (in this case, happiness and sadness/distress), it might be ethically difficult to ask a volunteer to relive unpleasant
memories, and conversely, I wished to avoid feigned, overly withheld or acted emotion as might be deliberately expressed by an actor.

Finally, since the intension was to use the both the motion captured data and the filmed live action footage as a basis for overworking and re-animating an artefact, there would be an opportunity to naturally blend in many layers of self-reflection and analysis; from expressing the emotions into the motion captured footage, being filmed, the written reflection of the emotions involved to the animation itself, tracing over my own face and studying my own facial expressions through rotoscoping. This was intended to be a deeply immersive, auto-ethnographic project.

Two clips from the happiness and sadness footage were selected to be rotoscoped. These clips were selected not just for clearly expressing the emotion required (first happiness, then sadness) but also as fragments of a short duration which might read well when viewed externally to the piece as a whole. While the final artefact was planned as a single animation, with rotoscoped clips cut with live action, motion captured 3D animated clips and hand drawn animation, an opportunity to view the rotoscoped segments (and potentially test them on viewers) as stand-alone pieces had to be considered. Ideally, larger segments would have been rotoscoped, but time restrictions proved a difficulty. While one minute of video footage can take around 2 hours for FACS trained coders to analyse, my own speed of rotoscoping ran to roughly 200 – 250 frames (on seconds) per day. This would be the equivalent of 16 to 20 seconds animated during a working day of roughly 8-10 hours duration with the result that each fragment took at least one week to produce. The end result was a “happiness” clip was of 1 minute and 20 seconds in duration, while the “sadness” clip came to 1 minute and 30 seconds duration.

The live action footage was imported into Flash software and a frame was hand drawn for every two frames of live action footage using a Wacom pen. It should be stressed that rotoscoping is not as “easy” an option as it might appear to the layman. It is not simply an exercise in “tracing” out the film footage. The animator must make judgment calls on which lines of the frame need to be traced, which to be ignored, and in what style. This can be very personal to the animator involved, and a good example of the art of rotoscoping as approached by a range of different animators can be seen in Sabiston’s Roadhead film (1998). In Roadhead, the same interviewee might be shared across several different animators, and while the face of the interviewee does not change within the filmed footage, the interpretation of that interviewee’s face varies from animator to animator. Each frame needs to be visually appraised, and sketched. The personality of the artist and their own stylistic approach and way of seeing and interpreting the visuals before them is difficult to avoid. As the footage is passed from animator to animator the style and interpretation of the footage changes. Sometimes scratchy, sometimes clean line, sometimes highly stylized, sometimes striving for realism. According to Sabiston,

with motion capture and the like, you are having the machine record something that becomes the spine, even the heart, of the animation derived from it. With Rotoshop, you are hand drawing the expressions and forms that you see. It is usually traced, yes, but even then you are starting with something hand drawn. There cannot help but be the smallest stamp of the artist in every line. From the very beginning, before it even enters the computer, the artwork is coming from someone’s head. The computer assistance happens afterward, and it springs from your artwork. That’s very different in my eyes. (Sabiston 2012, p. 79)

My own background was from 3D computer animation rather than traditional 2D techniques, but I also have a remoter background in comic-book illustration, where the thickness of a brushstroke is often used as a technique for emphasis; the thicker the line, the more emphasis the artist is trying to make on a part of an image. As this was my first exploration of the practice of rotoscoping, I had not anticipated how much I would need to study each frame of my face, trying to capture the expressions as accurately (albeit stylistically) as possible. I found myself reverting back to my old cartoonists training to use the thickness of line as much as possible as a medium for bringing out some elements or softening others (see figure 3).
Another unanticipated effect that I experienced as I worked, was that I subconsciously pulled my face into the same shape and emotion I was seeing on screen. Partly this is was a reflex action not uncommon with animators and illustrators to help them express the emotion they are trying to recreate artistically, but it could also be partly a mirroring effect, where two people in conversation will mirror their expressions to help establish rapport (Moore, Gorodnitsky & Pineda 2012). While working on the “happy” sequence, I felt upbeat and found myself repeatedly mimicking the happy expressions I was seeing. However, when I worked on the “sad” and “distressed” sequence, I found myself pulling the same unhappy faces, and found myself absorbing some of the unhappy mood I was feeling. Since this sequence took many days to complete, this made for an uncomfortable few days. Often, the act of having to study my own face, particularly in the sequences where I was fighting back tears, would make me mimic that face and start to feel tearful in response. It is unclear if part of this effect was due to having to see my own face on screen, but it was noticeable that seeing the unhappy expressions did remind me of the circumstances involved and lower my mood considerably.

On viewing the completed rotoscoped animations

On completing the rotoscoped fragments, I published them to my blog for a quick, initial reaction. Thirteen people responded and left comments, though one of them could not get the animations to work on his computer and
responded to let me know there was a viewing problem. All comments were screened so that commentors could not read each other’s comments (to prevent group-think or people feeling that their own comment might be incorrect compared to others (Fontana, Frey 2005). It should be stressed that at this stage the feedback was intended to be quick and fairly informal, an opportunity to pin-point difficulties and respond to feedback before the final piece was completed. A more rigorous feedback session was later implemented when the artefact was completed, and viewers would be able to view the rotoscoped sections of the work side by side with other animation approaches and cut with a control of the original live action footage. However, at this stage, the work on show was brief and fragmentary, involving only the two rotoscoped sections of the artefact.

Reviewing the completed sequences, I noticed two “errors” or possible problems. The first being that in rotoscoping the first sequence I appear to not move my mouth very much at all. It does happen that I do speak with a very closed mouth style, but the rotoscoping over-emphasised this to a degree that made “reading” the mouth movements difficult, so to combat this problem, I had deliberately traced the mouth slightly more generously for the second sequence, and this appeared to have a better result when the two pieces were shown to viewers. Three commentors complained that the “lip-synch was off” for the first animation, but fine for the second (though one of the three added that this seemed to be an artefact of the rotoscoping rather than actually being out of sync).

The second “error” was one that I noticed acutely during the practice of rotoscoping the individual frames but appeared to be less obvious on the completed animations. While drawing and viewing the footage as stills, I was able to perceive a crookedness, especially in the mouth area. This is a remnant of facial palsy that I had in 2005. The left side of my face became completely paralysed, though after some months treatment the paralysis wore off, though not entirely. This is generally not noticeable when my face is mobile (for instance when talking to someone) and I am assured that is it not visible to onlookers. However, it is noticeable on certain photographs, and was noticeable when working on the footage frame by frame. Since this is a personal reflection of an individual (warts and all) this is not necessarily a problem, but it bears mentioning, as permanent facial features should be measured and noted in case of any distortion to the data (Hager, Ekman 2005). The crookedness is not noticeable in the animations, but could be misread as a micro-expression, for instance, were the footage to be run through FACS testing.

All the commenters could clearly distinguish the two emotions, though as Ekman points out, the auditory signals are very powerful in influencing a viewer. He names three sources of information; the actual words, the tone of voice and the rapidity of the speech (Ekman, Friesen 2009), all of which were clear markers in both pieces. Interestingly, two commenters watched the animations without sound first. One thought that the “happy” animation was “sad and serious” for the first 48 seconds while the other detected no such emotion, though the head motions of the “happy” animation made them feel seasick until the sound was switched on.

Two of the commenters watched without sound after seeing the animations with sound, and found it hard to note any difference in the level of emotion seen.

Two commenters found the second “sad” animation “difficult to watch” due to the level of sadness portrayed, though these two both know me personally and would not expect to see such emotion from me normally.

**Conclusion**

Rotoshop films are not merely a computer-generated product, they not only bear the trace of the original filmed material, but also the imprint of the animator’s hand. These films, then, are doubly indexed, pointing to the presence of the interviewee in front of the camera, and the presence of the artist in the process of translating the video language to animation. This indexicality, however, as with all indexical signs, emphasizes the absence of the original. (Honess Roe 2012, p. 35)
By having the interviewee as animator, I hoped to explore deeper links between the artist and the subject, and the depth of the emotion within. Most rotoscoped films, for example, Roadhead film, (technically rotoshopped rather than rotoscoped though the underlying principle is the same) involve a separate interviewee to animator. The animator, or several animators, are given footage upon which they imprint their own artistic take. The results, in Roadhead in particular, are quite fascinating and varied, with the animator presumably reacting partly from their own interpretation of the footage they are seeing and hearing, and partly from their own style and artistic approach. However, they do not speak to the interviewee, they do not discuss with them their inner thoughts. Rather, they interpret from the footage, at a distance. Honess Roe (2012) mentions the double-edged sword of any animated documentary and of (in this case) rotoscoping in particular. In documentary, we judge not just from the spoken testimony, but the tone of voice and the visuals of the speaker, the tilt in their bodies, the expressions on their faces. While rotoscoping can highlight and intensify lines and gestures, often uncannily, it can omit, distort or misdirect a viewer, masking it. Ehrlich (2011) elaborates on the masking nature of animated documentaries; that the stylization of animation can conceal as well as expose, casting new perspectives as well as distancing them. Wells (1998) notes that animated documentary has a place as a different medium that helps “draw the viewers’ attention to significant, and sometimes unnoticed aspects of the character” (p. 28), while Ward (2005) points to animation as an “intensified route” to understanding, though the subjective eye of the animator (p. 91).

It is this exposure that fascinates me, likewise the possibility of loss of information, and how this might affect the viewer. By making myself the subject, in effect, I became the primary viewer, closely interconnected to the subject matter of the footage as well as seeing it though strange and distorted renditions; a motion-captured 3D avatar, as well as stylized rotoscopy. It is perhaps no coincidence that I deliberately chose to analyse a memory that would cause me some distress. I am not by nature a person given to exposing difficult emotions such as sadness and distress to other people, and yet exercising these difficult emotions through animation seems more palatable, more easily opened up to scrutiny, and potentially a way of exploring and coming to terms with difficult concepts, self-reflecting though animation.

I had begun this research using motion captured data to animate a 3D avatar, moving on to rotoscoping fragments of the footage. The work discussed here was always intended as part of a larger artefact, which was completed in December 2016 with the rotoscoped footage shown cut with live action, motion captured 3D and hand-drawn animation. A more formal feedback gathering process was instigated and the preliminary results hint at a further use for rotoscopy as a method of studying emotional expression. Viewers reported that the rotoscoped segments had been absorbing, helping the viewer to decipher expressions more vividly, particularly when juxtaposed with live action of the same character. The next stage would be to explore this reaction in more detail, with live action blending into rotoscoped footage and back, and to test this possible effect with a larger sample of animations and interviewed participants.

Note:

It might be interesting to see how a FACS coding might interpret the footage, being a more scientific, quantitative approach, whereas the method of analysing footage frame by frame with rotoscoping and the interpretation of an animator’s art would hint at a specifically artistic, qualitative approach. In some ways, the act of rotoscoping expressions shares a measure in common with FACS analysis, in that both methods require a human to personally go through footage frame by frame. Whereas a FACS coder must define and note each expression into coded (written summary) form, the animator defines each frame into a visual (drawn summary) form.

This article is based on a paper delivered at The Animator. The 26th Society for Animation Studies Annual Conference, Sheridan College, Toronto, Canada, 16–19 June 2014.

I want to express my acknowledgement to Prof Suzanne Buchan, Dr Magnus Moar and Dr Luke White for their constructive comments which led to this current form.
Sophie Mobbs is a senior lecturer and Joint Programme Leader in 3D Animation and Games at Middlesex University in London, where she specializes in teaching 2D and 3D animation, game art and animation, motion capture and virtual reality environments. Her research interests focus on animation with regards to body language. More specifically, she uses a creative practice research methodology to explore the relationship between non-verbal communication and animation. Prior to working in Higher Education, Sophie spent 10 years working as an animator in the games industry, where she took particular interest in character and monster animation and worked for companies that included Sony, Silicon Dreams and Rebellion and contributed work to 8 published games across various platforms.

References


© Sophie Mobbs

Edited by Amy Ratelle

This entry was posted in Sophie Mobbs - Intimate Scrutiny: Using Rotoscoping to Unravel the Auteur- Animator Beneath the Theory, Volume 11 and tagged emotions, expressions, motion capture, rotoscope, rotoshop. Bookmark the permalink.