Principles of Gaze and Arm Animation: Can We Apply Them to Animated Agents?

Tomislav Pejsa

1. Introduction

A central objective in character animation is to make characters believable in their behavior and clear in their communication with the viewer. Professional animators rely on empirical principles, derived through decades of practice, to animate their characters in a way that clearly conveys their intentions, emotions, and personalities. Even with these established principles as guidance, production of hand-authored animation is labor-intensive, time-consuming, and requires a great amount of skill.

However, there is a growing need for animation outside the domain of professional production. Availability of low-cost animation tools and devices [8, 24] is putting animation production within reach of novices. Potential applications of novice animation include producing instructional and demonstration videos with animated virtual agents, in the vein of similar live action videos on YouTube, but benefiting from the visual clarity and flexibility afforded by animation, as well as synthesis of movements of virtual agents and social robots which interact with human users as autonomous assistants [3], customer service representatives [19], and coworkers in settings such as manufacturing [1].

The character’s gaze behavior and arm movements are of particular importance in such applications of animation. Eye movements, head turns, and body posture shifts in gaze are important for communicating the character’s focus of attention and supporting a variety of conversational functions, such as facilitating better engagement and information recall in the viewer [22] and managing turn-taking in dialogue [20]. Arms movements are needed for gestures, which complement spoken dialogue and contribute to its efficiency [10], and they are also the agent’s means of interacting with objects in the virtual or physical world shared with the viewer.

Methods for computational synthesis of gaze and arm movements focus largely on simulating human movements or, in the realm of robotics, on synthesizing movements that are as efficient as
possible for the task at hand. However, in traditional animation these movements are often exaggerated beyond what is biologically plausible in order to enhance the expressiveness and clarity of the character’s actions. Computer animation researchers have proposed methods that automatically apply some animation principles to synthesized motion (e.g., adding anticipation [11, 26], ease-in and ease-out [13], and squash and stretch [26]), but these methods usually focus on a small subset of principles and do not explicitly consider gaze and arm movements.

Furthermore, animated agents often have simple designs and limited communication affordances. This is particularly pronounced in the realm of social robotics, where technological limitations and prohibitive cost mandate simple, minimalist designs. Many humanlike robots lack mobile eyes and have only limited mobility in the body, arms, and hands [18], while other robots have nothing but arms at their disposal (robotic arms used in manufacturing [25]). Yet even limited agents need to interact with humans, as collaborators, assistants, educators. Humans need to understand their goals and intentions, to feel comfortable in their presence, to feel engagement and even rapport in mutual interactions. This depth of social involvement can only be achieved if the agent’s behavior is believable and intuitively comprehensible.

Traditional animation offers solutions to the problem of saying more with less. Characters in animated films, even those with very stylized designs, are able to communicate clearly through the use of expressive movements and acting. Rather than try to accurately reproduce human movement, traditional animation abstracts away some of the detail of human movement and exaggerates other aspects in order to achieve motion that is both aesthetically more pleasing and more effective in conveying the character’s personality and state of mind.

To apply the principles of traditional animation to 3D computer animation, we must first know what those principles are. However, information about the practices of professional animators is often not readily available. Practices and traditions tend to vary from person to person and from studio to studio; animators and studios tend to have unique practices and tricks that support their distinctive style. The secrets of their craft, if publicly available at all, are often scattered across animation books, online tutorials, articles, blog posts, interviews, and other sources.

As part of this work, I have surveyed principles of character design, movement, and acting in traditional animation as they apply to the problem of animating gaze and arms of animated agents. In my research I have consulted a number of works of animation literature [9, 28, 2, 15, 16, 27] and ancillary resources such as animation blogs and video tutorials to obtain insights into how
animators achieve gaze and arm movements that look believable and communicate clearly. While synthesizing these insights into a paper I have tried to remain relevant to the problem of authoring animations for 3D characters and embodied agents, presenting the information in a form that is understandable and actionable to non-artist experts in the fields of computer animation and human-computer interaction. These experts could make use of the provided information to inspire and inform the development of computational animation techniques, enabling them to make their characters, virtual agents, and robots look and communicate better in a broad range of applications.

2. Principles of Gaze and Arm Animation

I classify principles of animation into three categories: (1) character design principles, (2) principles of movement, and (3) principles of acting. Character design principles specify how the character should be designed to achieve a particular personality and communication style with gaze and arm movements. Principles of movement are concerned with basic techniques of animating the character’s movements such that they look believable and visually pleasing, while acting principles specify how to use these movements together in order to communicate clearly the character’s personality, emotions, status, and intents.

2.1 Principles of Character Design

Gaze and arm behaviors engage not just the eyes and arms, but the entirety of the character’s body. The character conveys its focus of attention by turning its eyes, head, neck, and often whole body towards the target of interest. When the character gestures in dialogue or manipulates an object, the desired communicative effect is achieved through coordinated movements and attitudes of the hands, arms, and body. Therefore, when designing a character that will communicate using gaze and arm motion, it is important to consider the design of all these body parts.
A basic principle of designing appealing characters is to use curves and rounded forms [2]. The basic design of the character is composed of multiple solid, rounded forms connected together at socket points. Construction of the head—a key body part in gaze—typically begins with an oval ball or egg shape, with eyes drawn in the top hemisphere (Figure 1). This base shape is mostly rigid—it rotates and translates during movement, but it does not deform much. Additional features of the head, such as eyes, cheeks, snout, and mouth are more flexible and give expressivity to head movements through squash and stretch deformations and overlapping action.

Eyes are typically designed as circles or ovals that represent eyeballs, since that is a shape that looks appealing and offers the most opportunity for expressive movement [9]. The circle contains the pupil and gaze is communicated through movement of the pupil within the white of the eye. Some characters, such as Winnie Pooh, have simple eyes with just the pupils, but no whites. Such eye designs are restrictive, because eyes cannot wander across the face and the whole head must turn in gaze shifts. Despite this restriction, animators were able to make Winnie Pooh come alive through head movements, body attitudes, and hand gestures. The example of Winnie Pooh illustrates how an anatomically simple character is still capable of rich communication and it is particularly instructive to designers working with embodied agents without mobile eyes.

Similar principles apply to arm and hand design as well. The hand is composed of a base oval shape representing the palm, which has a thumb and fingers attached to it [2]. Classic Disney animal characters such as Mickey Mouse have four fingers on each hand (including the thumb), while more realistic human characters have five. Blair [2] recommends placing fingers unevenly to avoid monotony. Walt Stanchfield, former Disney animator and animation instructor, provided a set of guidelines for modeling arms [9]. He recommended using straight lines for bony parts of the arm that pull tight when bent, while curves should be used for fleshy, muscular parts that bend inward. The combined use of straights against curves makes for a more visually interesting design and it can also convey personality; e.g., wicked witch, an archetypical character in animated films, has gnarly, twisted hands with sharp, extruding nails.
Proportions of the body are a powerful cue to the character’s personality [2]. An archetypical “cute” character is proportioned like a human baby, with a large head and wide eyes on top of a smaller, plump body. Such use of proportions is an example of exaggeration, another important principle in character design. By exaggerating certain features of the character, the artist can bring out their essential qualities and emphasize personality traits [4]; for instance, a physically strong character would be designed with unnaturally muscular arms. Moreover, design needs to be kept simple, such that key features of the character are easily readable. The aforementioned example of four-fingered Disney character illustrates how simplification pays dividends in communicative clarity, but also in animation cost [4] – simpler characters are easier and cheaper to animate.

2.2 Principles of Movement

Fundamentally, animation is about timing and spacing [28]. In traditional animation, timing refers to how long each image depicting the movement is held on the screen, which determines how long the movement takes, while spacing refers to how far apart the images are frame to frame. Animators use timing and spacing to depict the character’s actions as fast or slow and to impart them with a sense of weight and force. From these fundamental principles, many additional ones arise. Ease in and ease out refers to the idea that the character should not move from one extreme pose to another monotonously, with constant velocity, but should accelerate into the movement and decelerate as it approaches the target pose. Movement should not start instantly, but it should be anticipated by a similar movement in the opposite direction in order to prepare the viewer for the upcoming action. Movements should not occur along straight lines, but they should trace arcs in order to look more appealing and give a continuous flow to the action. The character’s body parts should not remain rigid as they move, but they should deform and change shape – this style of animation is known as squash and stretch. Different body parts that partake in the action should not move all at the same time, but the timing of their movements should be staggered, a principle referred to as overlapping action.

The above principles play a role in animating believable and expressive gaze and arm movements. In the remainder of this section I survey the techniques for gaze and arm animation and I focus on how animators utilize the fundamental principles of movement for that purpose.
**Principles of Gaze**

The fundamental building block of gaze behavior is a *gaze shift* – coordinated movement of the eyes, head, and body towards a new focus of attention. The core issue of gaze animation is animating eye movements and the head turn. The approach to animating these movements in traditional animation is interesting in how it combines naturalistic principles that closely approximate real human movement and exaggeration principles that depart from it.

In human gaze the head tends to rotate toward the target along the shortest possible path. Copying this movement in 2D animation looks unappealing, because it appears that the head is stationary and facial features are sliding along it on straight lines – a consequence of the simplicity of character design and flat nature of the medium. To avoid this effect, it is necessary to apply the *arcs* principle and have the head lower as it turns, such that eyes and other facial features curve down over the course of the movement (Figure 2) [27, 28]. Furthermore, Art Babbit proposed [28] that head should be translated in space as it turns – an application of the *spacing* principle. Head translation implies that a gaze shift should involve coordinated shifts in neck and body posture, a feature incorporated in only a handful of computational models for gaze animation [21].

Gaze shifts are animated as *overlapping action* [27, 28]: eyes begin moving first, followed by the head (Figure 2). “Fleshy” parts of the face such as the cheeks might trail behind and follow through as the gaze shift ends. *Squash and stretch* is applied to all these parts to give them a more lifelike, organic quality; the face always appears to be contracting or expanding in motion [2]. The staggered timing of eye and head movements in gaze is an example of using timing to impart a sense of weight on different body parts. We observe the same principles in action in real human gaze – due to their small mass, eyes move much more quickly than the head and tend to lead the head in gaze shifts [29].

![Figure 2: Animating a gaze shift. Eyes lead, while the head follows. The head and all of its features move along arcs. Image source: [27].](image-url)
Where animation departs from reality is in the use of *anticipation* and *follow-through* for head turns [15]. Animators often anticipate large head turns with movements in the opposite direction. For instance, a large head turn to the right might be preceded by a slight turn to the left. Then at the end the main head turn, the head might overshoot the target orientation and turn slightly left to compensate, adding some follow-through to the motion. While this behavior is unusual in human gaze–human head generally moves directly to the target–it might be worth exploring in synthesized animation as a way of making the character’s movements easier to anticipate and read.

Eye movements during gaze are realized as animation of the pupils moving within the whites of the eyes, as well as *squash and stretch* deformations to visible parts of the eyeballs. Human eyes change shape; movements of the eyelids, eyebrows, and cheeks all exact forces upon the eyes and effect shape changes. Animators exaggerate these deformations for visual and communicative effect and introduce some non-realistic deformations – for instance, the eyeball gets distended where the pupil touches the rim [9].

Pupil position within the eye is key to determining how the viewer will interpret the character’s gaze [9]. Showing too little of the pupil makes it difficult to infer the expression and gaze direction. Having the pupil touching the rim and surrounded with a lot of white makes a more intense expression, but having the pupil completely surrounded with white (e.g., positioned in the center of the eye) makes the gaze direction uncertain and the expression vague. Pupils of both eyes must not look in divergent directions, otherwise the expression looks dead. Finally, pupil size matters: small pupil gives a dazed look to the character, while large pupil makes the character look attentive [9, 27]. These principles have received little consideration in computational synthesis of gaze and could serve as basis for novel techniques. Optimizing eye movement paths such that pupils always remain close to the rim and enabling parametric control over pupil size could introduce a new level of expressiveness into the gaze of animated agents.

An important human eye behavior is blinking – rapid closing of the eye effected by downward movement of the upper eyelid. Animators use eye blinks extensively and in tight coupling with gaze. To give the blink more weight, they move the pupil downward as the eyelid drops down [28]. The eyelid does not ease out of its movement – instead it accelerates and “impacts” the lower lid at full speed. *Squash and stretch* is used on the eye during blinking to give it a more lifelike quality [2] – the eye contracts when it blinks and expands as it opens wide. Another important
guideline is that the eyelid should never fully raise at the end of a blink [15]. Instead, it tends to raise about 80% of the way and only raises fully for expressions of surprise.

A typical use of eyelids is animating a blink during a long gaze shift, which adds flexibility as overlapping action of the gaze shift [28, 23] and it also calls attention to the change in gaze direction by introducing a lot of color change (as whites of the eyes are temporarily obscured) [9]. Maestri advises [15] using blinks only during long gaze shifts involving head turns and not short, saccadic eye movements that occur, for example, when the character is reading. This is consistent with human behavior, where gaze-evoked blinks serve as a mechanism for protecting and lubricating the eye during long gaze shifts [5].

Eye blinks can be an effective device outside of gaze shifts as well. When used during facial expression changes, they call attention to the change and add more flexibility [28]. Even when the character is not performing any large actions involving the face, blinks should still be employed occasionally to make the character look more alive. Blinks are an excellent device for moving holds – adding a blink into a held pose ensures that the character does not appear dead to the audience [2]. Timing of blinks is crucial to the effect they have on how the character is perceived [15]. A fast blink takes 4 frames, a normal one takes 6-8, while long blinks (up to 20 frames) make the character appear sleepy or dumb. Pixar animators are known [23] to offset the left and right eye blink by one frame to make them more visually interesting and reduce the effect of twins.

Principles of Arm Movement

Arm movements include gestures used in nonverbal communication and physically interacting movements such as touching, pressing, and picking up objects. Arm movements are actuated through coordinated movement of three joints in the shoulder, elbow, and wrist, respectively.

Natural, flowing arm movement in animation is achieved by having the wrists trace arcs through space [28]. This has basis in biomechanics of arm movements in primates, which are hypothesized to follow the so-called minimum jerk model [6]. Minimum jerk model also yields curved paths of the wrist and it has been applied to synthesis of smooth robot arm motion [7]. An additional principle with obvious biomechanical basis is that arm segments must maintain constant length throughout the movement [28].

Different arm segments do not move with the same timing, otherwise movement would appear very rigid and mechanical. Instead, either wrist or elbow lead the movement, while remaining
segments trail behind [28]. This means that joints must break as the trailing segments bend in the opposite direction of the movement. This technique is known as successive breaking of joints [28] (Figure 3) and it is a highly effective device for adding expressiveness to arm movements. It imbues the movements with a quality of curved action seen in rubber hose animation, while preserving the notion of a rigid skeleton.

Despite its effectiveness, there have been few attempts [12] to computationally model joint breaking and use it to automatically enhance synthesized motion. Yet there are numerous examples of creative and subtle applications of this technique in animation practice. For example, when the elbow leads the arm in a downward movement, the elbow joint breaks in a biomechanically implausible direction (Figure 3), yet the overall motion still looks plausible and appealing [28]. Even slight hand motion causes a break in the elbow – that is because any hand motion involves actuation in forearm muscles [16]. In pointing gestures, wrist leads the movement and breaks as it reaches the target pose, as the hand follows through slightly before correcting itself [28]. This break occurs over only one or two frames, which is too brief for the audience to consciously perceive, but enough that they feel a change in the quality of movement. Such imperceptibly fast action is called snap and animators use it to add more vitality to the character’s motion [28].

Actions involving arm movements tend to be large in terms of spacing and prominence and it is therefore important to anticipate them with smaller opposing movements. For example, pointing can be anticipated by moving the hand slightly back before pointing forward [28]. Similarly, in object manipulation movements such as writing the hand can slightly lift up before it moves down
and makes contact with the manipulated object [28]. Highly forceful actions, such as picking up a heavy object, are anticipated by the whole body – for instance, the character might bend its body backward and stretch before bending forward to pick up the object [28]. Such communication might be vital for an animated agent or robot demonstrating a task involving heavy objects, as a way of alerting the viewer of the object’s weight.

Picking up an object is an example of an arm movement that interacts with the physical world; other types of movements in that category are pressure-exerting movements such as touching or pushing and impactful movements such as punching. When animating these movements it is important to give the motion appropriate weight and convey how forceful the interaction is. For example, when animating a character picking up a heavy object, the whole body must bend to counterbalance the weight (Figure 4) [28]. In actions where the character exerts pressure on a physical surface with its hand, strength of the pressure is conveyed with verticality of the movement relative to the surface – for strong pressure, the arm is straight and vertical to the surface, while for light touch it remains curved and relaxed [28]. In actions that exert impact on the surface (e.g., punching), strength of the impact is enhanced by having the hand move in a straight line rather than an arc, by applying squash and stretch to the hand and the impacted surface, and by omitting the frame that shows the hand in contact with the target, instead showing just the result (e.g., the target being thrown back) [28].

**Figure 4: Picking up a heavy object: arm and body attitude convey the object’s weight.**

Image source: [28].
2.3 Principles of Acting

Acting in animation refers to the use of movements to convey the character’s emotions, goals, and motivations. Good acting makes the character’s feelings believable and real for the audience [9, 28] and creates empathy with the character [16]. Acting is a very broad concept and authoring animation with good acting requires both extraordinary skill and good intuition from the animator. In this section I will highlight some basic principles of acting in animation and how they apply to acting with gaze and arms.

Pioneers of acting in animation such as Bill Tytla and Art Babbitt have emphasized that the key to good acting is careful planning and simplicity [28]. In practice this means the character should only perform one action at a time and it should be the right action to convey what the character feels or wants. The need for simplicity implies that the character should have only a single focus of attention at a time [16]. Attention is primarily conveyed through gaze and body attitude, but it can also be conveyed using gesture (e.g., pointing at the object of interest). The character can switch its attention to other things and it can get distracted from the task at hand (which can be signaled, e.g., with a gaze aversion towards the distracting object), but its focus of attention at any given time must be unambiguously conveyed through its movements.

Clear communication of attention is the key to achieving clarity in the performance [16], such that the audience unambiguously knows what the character is doing and why. Other techniques for achieving clarity include using holds on important poses to give the audience a chance to absorb the character’s action and read their attention direction [16] and animating actions in profile such that they are readable just from the character’s silhouette [28]. The recommended duration of a hold needed to read a gesture varies, but even Tex Avery, known for his fast-paced animation, suggested a minimum of 5 frames [28].

Acting has the power to suggest the character’s status and role in the scene through movement [16]. Status can be conveyed through posture and gaze patterns – e.g., a character who hunches and gazes down is seen a submissive. Status does not merely refer to social status, but whether the character is currently in command of the scene. The ability to synthesize animated behaviors that convey a particular status would be useful for embodied conversational agents – for instance, an animated agent could use appropriate gaze patterns and body attitude to establish conversational roles in a multiparty interaction [17].
Eyes are the key affordance of the body in acting, a fact long recognized by animators; a rule of thumb at Disney was: “If you're short of time, spend it on the eyes. The eyes are what people watch.” Eye movements signal the character’s underlying thought process; a commonly cited principle is that each new thought triggers an eye movement [16]. Moreover, it is conventionally held that direction of the eye movement suggests the nature of the thought. While these conventions are not based on credible research in psychology, they are nonetheless well-established in acting and therefore warrant a consideration in designing character movements. (1) Looking up is believed to indicate visual thoughts; looking up and left suggests recalling an image from memory, while looking up and right suggests constructing an image, guessing, or making up a lie. (2) Looking level indicates thoughts about sound; looking left suggests remembering something that was said, while looking right suggests constructing a new sentence. (3) Looking down indicates emotion; looking down and left suggests internal dialogue, while looking down and right occurs when the character is expressing his or her emotions. Eye action in dialogue is typically an interplay of looking at the person the character is talking to and averted gaze in one of the six basic directions; increased frequency of such gaze aversions can suggest nervousness and dishonesty [15].

Other eye-based expressions include wide-open eyes, which suggest surprise, intelligence, and attentive listening [16], and squinting eyes, which suggest disbelief and puzzlement [9]. Eyes combined with other aspects of the character’s design can suggest their personality [9]. Cute characters often look up, because that expression is associated with children and small animals and therefore appealing to the audience. Eyes fixed into a stare are generally avoided, since they look lifeless and robotic. On the other hand, rapid eye darts appear exuberant [9, 23]. High frequency of eye blinks suggest shyness or dishonesty [15].

The most challenging aspects of acting are encountered in animated dialogue. Dialogue involves not only speech, but nonverbal actions that accompany it – gaze cues, facial expression changes, postural shifts, and a rich vocabulary of arm gestures. These actions are tightly coordinated with speech and with each other and animators take great care to ensure they are readable and unambiguous in communicating conversational meaning. A nonverbal action that emphasizes a particular word or phrase in dialogue is known as an accent [28]. One of the oldest and strongest accents is the pointing gesture [9], but animators use a variety of other gestures – for instance, so-called beat gestures that emphasize certain phrases and give rhythm to speech, or
emblems, symbolic gestures with culturally established meanings [10]. A guideline for accent timing developed at Disney is to have the accent precede the speech phrase by at least 3-4 frames, up to 20 frames for larger gestures or slower characters [9, 15]. Accents are always initiated by the eyes—because that is what the audience is watching—while the head, posture, and arms follow [15].

Entire tomes can be written (and indeed have been written, for example [14]) on the subject of using movement to act out internal states. Here I will mention just a few basic principles for arm movement oft cited in animation literature [16]. (1) Hands raised above the head are seen as an expression of intellectuality and anxiety. (2) Hands held near the chest suggest emotion. (3) Hands held lower (e.g., drooping by the sides of the body) are seen as more primitive. (4) Hands folded across the chest signify defensiveness or emotional closing-off. (5) Hands clasped behind lower back suggest respect or lower status.

Gestures occupy a large volume of space and are therefore more visually prominent than facial actions and speech during dialogue. This affords great opportunity for expression (Maestri asserts [16] that we can glean much more from dialogue involving gestures and no speech, than from dialogue involving just speech without gestures), but it also means they must be designed thoughtfully with regard to clarity and appeal. An important guideline enforced at Disney was avoiding twins [9]—animators were encouraged to avoid showing characters in symmetric poses, because symmetry is visually uninteresting. In computer animation we can achieve this by positioning the camera at an angle, rather than giving a perfect frontal view of the character, and by always having the arms perform different movements. While avoiding twins is a good general guideline, some animators [16] suggest they can be still be used when appropriate, for instance, in gestures that communicate order, balance, and harmony.

3. Conclusions

Experiences of professional animators over the decades have shown that in order to animate appealing and effective character movements, it is insufficient to simply mimic the anatomy and movements of real humans. While animation techniques and principles surveyed here are grounded in naturalistic human movements, they also simplify many aspects and exaggerate others in order to enhance the characters’ appeal, to bring out their essential qualities, and to make their feelings and intents clear to the viewer. Principles of good character animation have evolved out of the
need to create characters that are sufficiently appealing and interesting to hold the viewer’s attention over the running time of a feature-length film. If animated agents are going to find applications in people’s workplaces and everyday lives, where people will interact with them on a daily basis and for extended periods of time, they need to be designed and animated according to the same principles.

This survey is an attempt to synthesize traditional animators’ insights into what makes character animation good, specifically focusing on the important gaze and arm behaviors. One of the goals of the survey is to point the way to possible avenues for improving the appeal and communicative effectiveness of such behaviors in animated agents. Future work can build on the ideas synthesized herein by picking a subset of animation principles, building their quantitative and computational models, and evaluating their effectiveness in improving interaction with an animated agent.

References


