



# A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception

K. Ruhland<sup>1</sup>, C. E. Peters<sup>2</sup>, S. Andrist<sup>3</sup>, J. B. Badler<sup>4,5</sup>, N. I. Badler<sup>5</sup>, M. Gleicher<sup>3,6</sup>, B. Mutlu<sup>3</sup> and R. McDonnell<sup>1</sup>

<sup>1</sup>School of Computer Science and Statistics, Trinity College Dublin, Ireland  
ngc6823@googlemail.com, ramcdonn@cs.tcd.ie

<sup>2</sup>High Performance Computing and Visualization Department, KTH Royal Institute of Technology, Stockholm, Sweden  
chpeters@kth.se

<sup>3</sup>Department of Computer Sciences, University of Wisconsin, Madison, WI, USA  
{sandrist, gleicher, bilge}@cs.wisc.edu

<sup>4</sup>Northeastern University, Boston, MA, USA  
jbadler@ski.org

<sup>5</sup>Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA, USA  
badler@seas.upenn.edu

<sup>6</sup>Department of Computer Sciences, University of Grenoble and INRIA, France

---

## Abstract

*A person's emotions and state of mind are apparent in their face and eyes. As a Latin proverb states: 'The face is the portrait of the mind; the eyes, its informers'. This presents a significant challenge for Computer Graphics researchers who generate artificial entities that aim to replicate the movement and appearance of the human eye, which is so important in human–human interactions. This review article provides an overview of the efforts made on tackling this demanding task. As with many topics in computer graphics, a cross-disciplinary approach is required to fully understand the workings of the eye in the transmission of information to the user. We begin with a discussion of the movement of the eyeballs, eyelids and the head from a physiological perspective and how these movements can be modelled, rendered and animated in computer graphics applications. Furthermore, we present recent research from psychology and sociology that seeks to understand higher level behaviours, such as attention and eye gaze, during the expression of emotion or during conversation. We discuss how these findings are synthesized in computer graphics and can be utilized in the domains of Human–Robot Interaction and Human–Computer Interaction for allowing humans to interact with virtual agents and other artificial entities. We conclude with a summary of guidelines for animating the eye and head from the perspective of a character animator.*

**Keywords:** facial animation

**ACM CCS:** I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

---

## 1. Introduction

The generation of realistic artificial entities, from virtual characters to physical humanoid robots, has become a requirement underlying many applications in human–machine interactions. Apart from the obvious applications in the movie and visual effects industries, they are needed for interactive applications such as games, virtual tourism and e-commerce, urban planning, medicine, surgery and training, among others. Realistic, plausible and engaging characters contribute to high levels of immersion, enjoyment and learning among human users. However, creating artificial entities which replicate humans in both appearance and behaviour remains one of the greatest

challenges in the field of Computer Graphics. While their graphical fidelity has improved rapidly due to advances in graphics processing hardware and its accompanying rendering techniques, the qualitative mismatch between the appearance and behaviour of characters appears to be worsening. This can cause perceptual disparities that may disrupt viewers' sense of immersion or may even contribute to unsettling emotional reactions [Mor70]. Behaviour synthesis has not yet reached the levels of realism required to convince users' perceptual systems that a virtual human is the real thing. This is particularly true when behaviour is generated automatically or semi-automatically in real time, where little or no corrective actions from outside the system are possible.

The human face is an important instrument for communicating and defining underlying emotions in human–human interactions [EF03]. Eyes are central in conveying emotional information, enabling humans to interpret the intentions and feelings of other humans by observing their eyes, which lends credence to the common proverb ‘the eyes are the window to the soul’. We learn this ability from a very young age, as infants have been shown to be efficient at distinguishing between emotional stimuli and other stimuli [Sim94]. Furthermore, using eye gaze to guide and interpret social behaviour remains a central facet of social interactions throughout life. The process of associating mental states with human behaviour is referred to as ‘theory of mind’ [AHO90]. Further studies lead to ‘the language of the eyes’ [BCWJ97], demonstrating that complex mental states, such as interest, can sometimes be seen by observing the eyes alone. These are only a few examples of the role eyes play in people’s daily lives, for further information please see the previous excellent surveys by Kleinke [Kle86] and Kopp *et al.* [KKM\*06]. It is not surprising, therefore, that much effort is required in the creation and animation of realistic virtual eyes.

Historically, implementations of eye-gaze models typically have been proprietary in nature, often borrowing, somewhat arbitrarily, concepts from a variety of sources across a range of disciplines. One reason for this may be the difficulty of the endeavour, which requires the integration of knowledge from a large number of disciplines, such as psychology, neuroscience and the social sciences. More recently, however, there has been an increase in the publication of comprehensive models that integrate knowledge from these disciplines, making it possible to compile an overview of the process of designing gaze mechanisms appropriate for different applications. In order to design effective social gaze behaviours, we must first gain a better understanding of low-level gaze cues (e.g. saccadic eye movements, mutual gaze, head motion and so on), including the patterns in which they are produced and utilized in various social contexts. We also need to develop an understanding of how these low-level cues relate to high-level social and cognitive processes. At this higher level, gaze and eye movements are vital for creating compelling social interactions between human users and artificial entities, such as virtual agents and robots. Such interactions often require an artificial entity to simulate eye-gaze behaviour, often as part of a range of synchronized multi-modal behaviours, in order to respond to and ultimately sustain meaningful interaction with a human user. Research developing the required competencies crosses a variety of domains, including Computer Graphics, Human–Computer Interaction (HCI), Human–Robot Interaction (HRI), affective computing, social sciences and many others. Character animators are also included in this effort of striving for more realistic eye and head movements. With the main goal of designing fictional characters which are appealing and convey thoughts, emotions and personality, their realization might somewhat differ from the natural human behaviour.

In this survey, we present relevant and significant findings from anatomy and physiology on the construction and low-level workings of the human eye (Sections 2 and 3). We also discuss research from the social sciences and psychology on how eye gaze is used to convey information about direction of attention and emotional and mental states (Section 4). Throughout the report, we discuss how this information has been exploited in the development of eye-gaze models in Computer Graphics and HRI, as well as a control

method in HCI or as input for agent interaction. We highlight some outstanding challenges regarding stylized characters and robots that result from their diverging proportions and handling in comparison to their real-life counterparts (Section 5). In addition, we summarize animation guidelines from character animators for animating eye, eyelid and head movements for fictional characters that must convey thoughts, emotion and personality (Section 6). Finally, we address how researchers to date have evaluated the plausibility and realism of their eye models and conclude with future directions in the domain (Sections 7 and 8).

We anticipate that this review article will act as a focal point for knowledge in the Computer Graphics community and will help to solidify research in this domain. The survey will be accessible to both new and established Computer Graphics researchers who wish to develop enhanced models of eye-gaze behaviours, in addition to practitioners in related domains.

## 2. Virtual Eye Creation

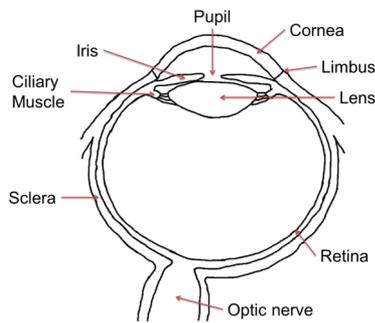
The synthesis of realistic artificial eyes requires the accurate modelling and rendering of the eyeball structure. In this section, we discuss the basic anatomical structure of the human eye and present the most relevant research that uses this anatomical knowledge to produce geometry for facial animation or medical applications. The iris in particular presents a difficult challenge, as it is a layered and highly complex structure. Different approaches have been used to recover the iris structure and scattering features, from layering simple painted textures to recovering important details from photographs of an eye. This section focuses on the modelling and rendering approaches needed in order to achieve high realism in the structure and appearance of eye geometry.

### 2.1. Anatomical structure

The basic anatomy of the human eye presented in this section is based on the book ‘Clinical Anatomy and Physiology of the Visual System’ by Remington [Rem11]. The interested reader is referred to this book for more detailed information.

The eye is one of the most complex organs in the human body. Each individual layer of the eye performs a specific task, such as capturing, focusing and transforming light into electrical signals, to ensure correct perception of the environment. The transparent *cornea*, located at the front of the eye, is the first refraction point for light entering the eye (Figure 1). A tear film moistens and smoothens the surface of the cornea to ensure minimal scattering and distortion of light. Separated by the *corneal limbus*, the cornea is embedded in the outer shell, the *sclera*. Both can be pictured as two, not perfectly symmetric, spheres. The white opaque tissue of the sclera preserves the eye’s shape and provides protection against harmful influences and substances. The radius of curvature for the sclera and cornea varies between humans. On average, the sclera and the cornea have radii of 12 and 8 mm with a horizontal and vertical deviation due to their elliptical forms.

The light passes through the visible *pupil* and is focused by the lens behind it. Depending on the prevailing light conditions, the muscles of the surrounding *iris* influence the shape and diameter of the pupil. The iris is a multi-layered conical tissue. The pigmentation



**Figure 1:** Schematic of the anatomy of the human eye.

and density of the iris define eye colour, a unique identifier of a person. The iris' primary function is to regulate the amount of light that enters the inner eye. Light passing through the pupil is focused by the *lens* onto the retina. The lens' shape changes through contractions of the *ciliary muscle*, thereby increasing optical power and accommodating the projection of objects nearby onto the retina.

The *retina* forms the inner, light sensitive part of the eye. The light captured at the retina is processed and transmitted as electrical signals to the brain for further processing.

## 2.2. Modelling and realistic rendering

Eyes presented in animations or virtual environments are commonly modelled as spheres or half spheres and use high-resolution pictures of human eyes as texture maps [ID03, WLO10]. Only a few studies deal with a more anatomically realistic model of the eye, which are discussed in the following paragraphs.

For example, Sagar *et al.* [SBMH94] used simple Gouraud shading on two spheres, representing the sclera and cornea, and texture maps to model and render the human eye for a surgical simulation application. They realistically generated the retinal blood vessels shining through the sclera using a fractal tree [Opp86]. Two layers of polygons, with the inner layer oriented towards the pupil and the outer layer towards the corneal limbus, represented the iris fibres. When the pupil dilates in the model, Gaussian perturbation makes the outer layer wavy.

Also in contrast to the standard procedure of modelling the eyeball as a sphere, Banf and Blanz [BB09] model the visible area of the eye as part of a continuous face mesh. A texture mapping approach is used to capture eye movements and occlusions by the eyelids. With this method, Banf and Blanz avoid the sharp edge between the eyeball and the surrounding skin and allow a smooth movement between them.

Other research has focused on replicating the visual appearance of the iris. Knowledge from ophthalmologists was used to synthesize the human iris in Lefohn *et al.*'s [LBS\*03] approach. Ophthalmologists design aesthetically pleasing ocular prostheses by applying multiple layers of paint onto a plastic shell to paint the iris. Starting with the most dominant eye colour as a base, layers consisting of combinations of dots, radial smears or radial spokes, representing iris components, are added. Lefohn *et al.* adapted this approach for a virtual

eye by mapping scanned textures of painted layers onto overlaying frustums of right circular cones.

The most influential method for iris modelling to date is the image-based model of the eye by François *et al.* [FGBB09]. In an automated process, the iris morphology and light scattering features on the cornea and iris are extracted from iris photographs. The iris subsurface map created serves as render input for the subsurface texture mapping method. François *et al.* also take into account reflection and refraction at the corneal surface based on ambient light.

Addressing the individuality of the human eye, Bérard *et al.* [BBN\*14] introduced a novel capture system allowing them to reconstruct the sclera, cornea and iris with its interpersonal differences. In successive processes, first the shape of the sclera is reconstructed, followed by the cornea and iris. In addition, the capture system is used to reconstruct the iris deformation during pupil dilation in detail.

In contrast to the previously introduced image-based methods, the *ILIT* (Iridal Light Transport Model) by Lam and Baransoki [LB06] is an iris model based purely on studies from the fields of physics and biology. A Monte Carlo-based rendering technique is implemented to replicate the interactions of the light within the iridal tissue and to determine the spectral response based on the known composition and structure of the iris and the properties of the light. The model allows the modification of biophysical parameters defining the composition of the iridal tissue and thereby affecting iris colour.

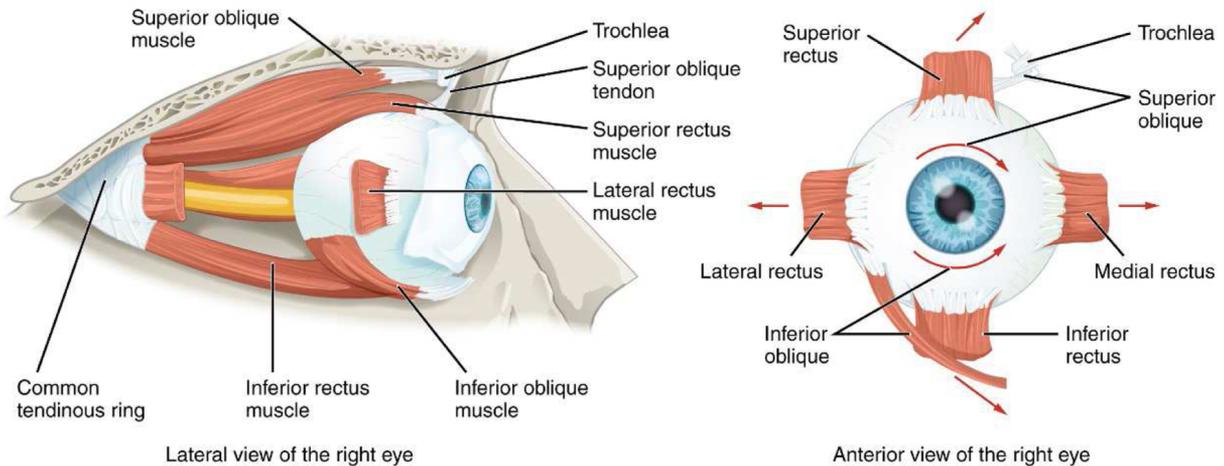
Other methods to synthesize iris images can be found in the field of iris recognition [WSG05, ZS06]. However, these methods are generated for verification of iris recognition algorithms and do not address the problem of modelling and rendering a realistic human iris.

An interesting area of research concerns the modelling of the pupil's reaction to the light. Pamplona *et al.* [POB09] presented a physiologically based model for the pupil light reflex and implemented an image-based iris deformation model. In this model, a delay-differential equation was used to define the pupil diameter as a function of environment light, enhancing results from mathematical biology and experimental data. The authors' measurements of feature points on the iris during deformation showed that these points move along a straight line on radial axes. Therefore, the animation could be simplified to a texture-mapped planar triangle-strip mesh with the inner triangle vertices mapped to the pupil border. When changing the pupil diameter as a result of incident light, all vertices at the pupil's border are relocated to a new position along a radial line, connecting the vertex with the pupil's centre.

An additional reason for modelling pupil diameter is that it changes as a function of emotional arousal [BMEL08], and that pupil dilation can function as an indication of attraction, when judging the opposite sex [TS04]. Obviously, emotion is expressed through eye gaze as well, which is the subject of Section 4.3.

## 3. Low-Level Eye Animation

The first step in convincingly modelling any effective emotional or social gaze behaviour for virtual agents is to gain a thorough



**Figure 2:** Arrangement of the extraocular muscles from 'Anatomy & Physiology', *Connexions*, June 19, 2013, <http://cnx.org/content/col11496/1.6/>.

understanding of the underlying behaviour in humans. Fortunately, eye movements have been studied by psychologists and physiologists for more than a half-century, so an abundance of data exists that precisely describes the kinematics of eye movements.

Located in the frontal part of the human head, the eyes perceive only a restricted view of the surrounding environment. Humans' field of view is further limited due to the fact that high visual acuity is available only in a small region of the retina. Saccades, the vestibulo-ocular reflex (VOR), smooth pursuit movements and vergence (terms defined in later subsections) serve as mechanisms for the eyes to maintain fixation on a moving object, to stabilize the location of this object on the retinas and to adjust both eyes to visual depth. In addition, the eyes do not move in isolation: they are intricately intertwined with blinks and eyelid movements, and are very frequently just one component of a larger gaze shift that includes head motion.

In this section, the physiology underlying eye movements, eyelid movements and combined eye-head movements are reviewed briefly. We also highlight relevant articles from computer science and animation literature. Numerous approaches, including parametric approaches based on physiological measurements and computation approaches derived from motion capture or tracking data, have been used to successfully simulate realistic gaze in agents performing a variety of actions.

### 3.1. Eyeball movement

The eyeball is a deceptively complex organ capable of executing a wide repertoire of movements and controlled by six separate muscles (Figure 2) [LZ99].

Fortunately, these movements have been studied extensively by neurologists, psychologists and neuroscientists. As a result, both their characteristics and the conditions under which they occur have been extremely well-described, providing valuable resources for the animator who wishes to reproduce them.

#### 3.1.1. Saccades

Saccades, if not the most frequent type of eye movement, are probably the most noticeable. They are the rapid shifts in eye position that centre the gaze on targets of interest. Saccade characteristics are quite stable across healthy individuals. Their duration increases approximately linearly and their peak velocity increases exponentially with magnitude [Bec89; LZ99, chapter 2]. This consistent relationship is referred to as the main sequence [BCS75, figure 2].

Initial acceleration and final deceleration are extremely rapid ( $>10\,000\text{ deg s}^{-2}$ ). For example, a very large saccade of  $30^\circ$  typically has a velocity of around  $500\text{ deg s}^{-1}$  and a duration of less than one-tenth of a second. Under natural conditions, such saccades are rare [BAS75]. More commonly occurring saccades of  $5\text{--}10^\circ$  have durations of  $30\text{--}40\text{ ms}$ , or approximately one frame at  $30\text{ Hz}$ . Saccade latency in response to a visual target is normally around  $200\text{ ms}$  [LZ99], but can reach approximately  $100\text{ ms}$  under certain conditions [FR84]. Under most conditions, the smallest possible intersaccadic interval (delay between consecutive saccades) is around  $150\text{ ms}$  [LZ99]. Other saccade properties such as slight target undershoots, slow drifts after saccade completion (glissades), slight curvature in the spatial trajectory of oblique saccades and torsional movements of the eyeball are all modest [LZ99] and are normally invisible to a casual observer.

Multiple approaches have been used to animate saccadic eye movements. Extremely faithful to the underlying muscle mechanics and innervation patterns driving the eyeball, Komogortsev *et al.* were able to generate extremely accurate saccadic velocity curves [KHJK13]. Lee *et al.* [LBB02] also attempted to faithfully reproduce a normalized velocity profile using a polynomial curve fit, and their results were used in some later work [GLBB07, QBM08, WLO10, Sha11, APMG12a].

Highly realistic movements can also be created using data-driven approaches; for example, by training a statistical model based on recorded movements [DLN05, DLN07, LMD12] or by simply replaying the records themselves (Figure 3) [HJO\*10, MBB12].



**Figure 3:** Eye movement capture using a head-mounted eyetracking device, and corresponding virtual avatar displaying the recorded eye movements [MBB12].

An opposing perspective is exemplified by Yeo *et al.* [YLN12], who implemented simple bell-shaped velocity curves with the rationale that subtle asymmetries of the velocity waveform were invisible to an observer. Similar methods use triangular velocity curves [VGSS04], uniform velocity [MH07] and instantaneous rotation [NBF\*13] to model saccadic eye movements.

The implementation of the frequency of saccades highly depends on the task being performed. However, as a general rule the time between saccades (i.e. fixations) can be modelled as an exponential distribution [HHA\*88], which for example was used by Normoyle *et al.* [NBF\*13].

### 3.1.2. Vestibulo-ocular reflex

The VOR acts to stabilize the eyes while they are fixating on an object during head motion. The VOR is modulated by a direct reflex via inner-ear vestibular neurons and thus occurs with extremely short latency, on the order of 7–15 ms [LZ99, chapter 3]. It therefore can be considered as effectively simultaneous with head movement. It is easy to implement from a modelling perspective: if the head rotates with some angle  $\theta_x, \theta_y, \theta_z$ , the eyeballs should counter-roll at  $-\theta_x, -\theta_y, -\theta_z$ . If the eyeballs are radially uniform, their vergence angle is zero (see Section 3.1.4), and their  $\theta_z$  component (corresponding to lateral head tilt) can be ignored as an additional simplification, see Anderson [And81] for an example of roll in a non-uniform (cat) eyeball. Note that the VOR is partially suppressed during large gaze shifts involving head movement (see Section 3.3). Most animation papers that parametrically simulated eye movements also included a VOR component [LBB02, GLBB07, MH07, WLO10, NBF\*13, APMG12a].

### 3.1.3. Smooth pursuit

The human fixation system is not constrained to static objects. For example, it is able to make the complex series of saccades required

to read an advertisement on the side of a moving bus. The smooth pursuit system is responsible for stabilizing moving images on the retina [LZ99, chapter 4]. It has a response latency intermediate to that of VOR and saccades, approximately 80–130 ms, and breaks down at high target speeds of  $30 \text{ deg s}^{-1}$  or more [TL86]. Unlike VOR and saccades, smooth pursuit is more situational and therefore is not often observed. It can be seen, however, by watching spectators during a sporting match, or by watching a companion looking out the window of a moving train (technically this is optokinetic nystagmus, but it is visually similar to smooth pursuit), for example. From a simulation perspective, pursuit is considerably harder to implement than saccades, since it requires a velocity calculation in addition to the positional one. It is also harder to decide *what* to pursue, since motion needs to be added to any underlying attentional model (e.g. Chopra-Khullar and Badler [CKB99], as well as Section 4). Due to its latency and limited velocity, a simulation of pursuit usually requires the computation and execution of one or more catch-up saccades, as well. For these reasons, it is currently of limited interest to animators, although the object-catching simulation of Yeo *et al.* [YLN12] included an elegant implementation.

### 3.1.4. Vergence

Normally the two eyes are yoked; that is, if one eyeball rotates in a particular direction, the other will rotate in exactly the same direction. Vergence [LZ99, chapter 8] is the exception to this rule. If an object lies on or near the visual midline, the two eyes must rotate in opposite directions to ensure that the object image appears at the corresponding position on both retinas. This process, called fusion, can accompany both saccades (e.g. if gaze changes from a far peripheral target to a close central one) and pursuit (e.g. if a moving target changes position in depth). Any animation system that calculates rotation angles for both eyes separately has *de facto* implemented vergence, but it is not strictly necessary for non-targeted gaze, such as that during speech (e.g. Masuko and Hoshino [MH07]) or emotional expression (e.g. Queiroz *et al.* [QBM08]). Issues in vergence are exacerbated when the character has large or stylized eyes (see Section 4.7).

## 3.2. Eyelid movement

The eyelids are not part of the oculomotor system, but they do interact with it. This, as well as their proximity to the eyeballs, necessitates a brief treatment of eyelid movement here. Normal eye blinks can be broken into spontaneous, voluntary and reflexive subclasses, all of which have slightly different eyelid dynamics [VBR\*03]. Spontaneous blinks, also called endogenous blinks, are the most interesting, since their frequency is linked to cognitive state and activity [SWG84, SNJ\*07]. Various studies have linked blink occurrence to attentional processes [NKM\*13], fatigue [JTC\*07, AWH10], lying [BBW96] and speech production [NK10]. Blink rates are highly variable, however. A meta-study found ranges of  $1.4\text{--}14.4 \text{ min}^{-1}$  during reading,  $8.0\text{--}21.0 \text{ min}^{-1}$  during primary gaze and  $10.5\text{--}32.5 \text{ min}^{-1}$  during conversation [Dou01].

The occurrence of individual blinks can be modelled as a Poisson process [Gre86]. However, blinks very often occur almost simultaneously with the onset of eye and eye-head gaze movements,

particularly large ones over  $30^\circ$  [EMP\*94]. It is also important to note that during a single blink, eyelid motion is *not* uniform: the downphase velocity is approximately twice as fast as that of the upphase velocity, and their respective durations are non-linear [EMS91, figure 2; GSC91]. Importantly, eyelid displacements, called lid saccades [BF88], always accompany vertical eye saccades. Lid saccades do not exhibit as much marked asymmetry between down- and upphases as do blinks [EMS91, figure 7; GSC91].

An unblinking face is visually disconcerting enough that nearly all graphics papers implement some form of blink production; however, the blinks are usually not described in detail, leaving their precise implementation to be handled by the facial animation package used. It is possible to animate blinks by calculating eyelid trajectories based on recorded video [DLN05, DLN07, WLO10, LMD12]. More sophisticated blink models also exist [SOS10, TCMH11], in which eyelid dynamics are modelled according to physiological data and account for both endogenous blinks and eyelid saccades. Other eye blink models focus on higher level aspects, such as the timing and synchronization of blinks during head movements and conversations [LBB02, GLBB07, MH07]. Peters [Pet10] compared the realism of different methods for simulating blink timing, such as constant frequency versus saccade-linked. Although the results did not show explicitly which blinking strategies were perceived as more natural, the experiment confirmed the importance of regular blinks in the general perception of blinks.

Normoyle and colleagues [NBF\*13] used a heuristic approach based on physiology: eyelid displacement was proportional to eyeball rotation except for the downward blink phase, which was modelled as a step displacement. Eyeball movement not only influences eyelid displacement, but also the eyelids' shape. This correlation was analysed and synthesized by Elisei *et al.* [EBCR\*07]. Based on a training data set, the eyelids shape was modified by taking the eye-gaze direction into account.

### 3.3. Combined eye-head movement

Most natural gaze shifts employ both eye and head movements to replicate human gaze behaviours [LR86, figure 1; LZ99, chapter 7]. The threshold for evoking an eye-head gaze shift as opposed to an eye-only saccade is approximately  $15\text{--}20^\circ$  [Sta99], although this varies between individuals [Ful92]. When reacting to a target, eyes normally move first with typical saccadic latency ( $\sim 200$  ms), while head motion begins  $20\text{--}50$  ms later. However, when gaze is shifted to a predictable target, the head movement begins around  $100$  ms before the eye saccades [LZ99]. Like eye saccades, head movements have a consistent velocity-magnitude relationship [GV87]. Unlike saccades, subjects can voluntarily modulate head velocity with a corresponding impact on eye velocity [LR86, figure 1B]. The calculation of gaze shifts with initial head and eye positions different from looking straight ahead, has to take the midline effect into account [Ful92, HM06]. With the midline defined as the torso midline, Fuller [Ful92] found that a gaze shift towards the midline results in larger head movements.

A common strategy employed in graphics literature is to define a cutoff displacement of  $10\text{--}15^\circ$  above which targets are acquired by an eye-head gaze [NBF\*13, MH07]. Once the desired head

displacement is known, there are several possibilities for computing the rest of its movement parameters. Its velocity can be estimated using the main sequence relation [GV87, YLNP12] or from a desired timing regime [MH07], or it can be computed automatically using inverse kinematics of the head and neck joints [NBF\*13]. Head animation can also be modelled by fitting physiological [Pet10] or motion capture [VGSS04, LMD12] data. The model of Andrist *et al.* [APMG12a] attempts to provide parametric control over many of the aspects of movements that vary based on actor and situation while maintaining the low-level kinematics common to all eye-head movements.

### 3.4. Summary

Eye movements, blinks and head gaze have all been extensively studied by physiologists, so their parameters are well understood. As a result, animators have access to a wealth of data that can be used to increase the realism of their own simulations. An important consideration is to decide which details are important for imparting realism, and which details are too subtle or unnoticeable to be worth incorporating into virtual agents.

Changes to low-level features of gaze motion, such as direction, velocity and amount of head recruitment, can produce substantial effects on high-level outcomes of the interaction. For this reason, it is important to parametrize the low-level features in such a way that designers of human-virtual agent interactions can easily target specific high-level outcomes through the manipulation of low-level parameters. For example, Andrist *et al.* [APMG12b] have shown that manipulation of the amount of head alignment in a gaze shift towards users and objects can serve to create gaze behaviour that is perceived as more affiliative—leading to higher feelings of rapport—or more referential—leading to learning outcomes—in an educational scenario.

In this section, a great deal of attention was paid to movement kinematics. However, there are other characteristics of eye movements that remain to be resolved; namely, when they occur, where they are directed, and the cognitive state of the agent executing the movement. These higher level issues will be addressed in the following section.

## 4. High-Level Aspects of Gaze Behaviour

This section is founded on low-level aspects of combined eye-head animation described in Section 3 for synthesizing more complex information seeking and interaction behaviours involving gaze. A key challenge is to automate a process often conducted manually by artists. Models must be capable of fully automating expressions appropriately given the context of the situation (i.e. environment, interaction type, culture and so on). In this respect, automation efforts for characters [CV99] share many crossovers with efforts made in HRI (Section 5.2), especially when characters are to be endowed with behaviour generation competencies that are associated with behavioural and cognitive animation approaches in computer animation [TT94, FTT99]. These competencies, often based on research related to the functions of gaze during interactions (Section 4.1), involve sensory investigation through computational visual

attention models (Section 4.2); gaze synthesis for expressing emotions (Section 4.3) and providing a window into the ‘mind’ of characters [LMT\*07] to better communicate internal states, attitudes, attentions and intentions [MGR04]; for regulating conversations, as speaker or attentive listener (Section 4.4); and for use with other modalities, including speech, to enable characters to engage more naturally in discourse (Section 4.5). This section also provides an overview of how eye gaze detected from humans is being used in a variety of diverse HCI applications (Section 4.6); and how gaze models may be applied to stylized characters (Section 4.7).

#### 4.1. Functions of gaze

Gaze is a powerful non-verbal cue that yields a number of positive effects in human–human interactions [Kle86]. It serves a number of functions, including information gathering, signalling interest and emotional state and regulating conversations through management of participant turn-taking [AC76, Ken90]. In conversations, gaze helps to mediate flow, including indicating the addressee, paying attention, displaying attentiveness, affecting turn transitions and signalling requests for backchannels. Gazing at transition-relevant places in a conversation can facilitate conversational turn-taking by reducing the length of pauses between speaker turns and reducing the amount of overlapping speech [Ken67], and a teacher who engages in mutual gaze with students is effective in improving learning outcomes [OO80]. Gaze is used to regulate intimacy [Abe86], facilitate references to objects in a shared visual space [BHAF09, GC11], signal dominance in multi-party conversations [FN11] and facilitate conversational grounding, which leads to improved collaborative outcomes [BHZS12]. Gaze behaviour is also crucial for facilitating mobile sensory investigation in complex dynamic environments. In this role, visual attention mechanisms may be used to control the target of gaze motions, as described in the following subsection.

#### 4.2. Visual attention

Visual attention refers to mechanisms allowing us to focus our processing capabilities on relevant information and to filter out information that is irrelevant. More specifically, it has been described by [TCW\*95] as referring to the selection of regions, features and values of interest in the visual field, the control of information flow through the visual system and the shifting of focus from one selected region to the next in time. For virtual agents and robots, models of visual attention allow the agent to become *situated*, which refers to an agent that is embedded in its environment in such a way as to be reactive to environmental objects, motions and/or the actions of the user [HJ96]. Visual attention models may drive gaze behaviours that facilitate sensory investigation when coupled with visual perception input systems (see Peters *et al.* [PCR\*11] for an overview). Based on perceptual inputs, visual attention models determine locations of interest to which gaze and eye movements are subsequently directed and also inform other aspects of gaze motions, such as the ratio of the head and eye movements or the quality of blink motions [PO03].

These approaches have been employed in the animation of idle looking behaviours [CKB99] of characters in a bottom–up manner, i.e. when there are no tasks at hand or to interrupt tasks, in

order to ensure that characters are perceived to be situated in their environment, dynamically responsive to events occurring in their environment. Thus, the approaches are useful for the generation of autonomous secondary behaviour [GD02] when characters are embedded in virtual environments [POS03]. More recent work [CGV09] also considers relationships between gaze targets and proxemics when generating idle gaze motions.

In other situations, the virtual character becomes situated in the user’s physical environment by interacting with the real world through the use of a web camera [ID03, PBER07] and makes eye movements and gaze motions towards salient locations in the environment. Such saliency-based approaches [IDP06, OSS09] are based on a neurobiological model of visual attention [Itt00] and have been popular for animating eye and gaze movements. More recent efforts have focused on specific aspects of visual attention; for example, the role of object relevance [OSP11] and task constraints [KOS11] on gaze specification. A key challenge is balancing bottom–up and top–down visual attention for gaze allocation [MHKS07]. This is a difficult issue due to the demands of real-time operation.

##### 4.2.1. Opening interactions and directing attention

The ability of characters to express interest [GP09] in users through gaze [PPB\*05] and associated modalities has led to research on less explicit and more natural ways in which interaction with humans is requested and human attention is directed. The ability to express interest through eye movements and gaze has been studied in relation to its effects on human perception of attention [Pet06] and human self-involvement during social interactions [MSH\*06] with characters. Other studies have considered the use of the eyes, head, body and locomotive direction of characters [Pet05] as a means for opening an interaction [Car78].

Gaze behaviours may be a powerful means for cueing the attention of human users [FBT07]. A number of studies have investigated the gaze cueing abilities of virtual characters [MSSSB10] in game scenarios [PAK10, QPA14] and in multi-task scenarios [KK13]. Further research is required to elucidate the factors that may underlie the ability of artificial systems to better direct the visual attention of human users (see Section 5.1.1 for similar research themes involving physical embodiments).

#### 4.3. Expression of emotion

Gaze is a powerful method by which emotions are expressed [Iza91]. It has also been noted that gaze affects how emotions are perceived by the addressee [PSA\*04]. Therefore, how characters perform gaze motions is an important consideration for successful conveyance of emotional states to users [FOM\*02].

While traditional approaches by skilled animators have been successful at creating expressive characters, automatic approaches are still an open challenge. Current approaches in the literature have therefore sought to analyse gaze motions in animated films to create animation models that can automatically map between emotions and gaze animation characteristics [LMK04, QBM07, LM10b].

Studies have also considered the expression of emotions through gaze shifts that involve movements of both the torso and the head [LM07]. For example, the SmartBody Gaze Controller [TLM09], released as part of the open-source SmartBody system, is capable of making full-body torso gaze shifts and aver-sions and also allows for the control of parameters, such as velocity and postural bias. Empirical studies have been performed in order to link low-level gaze attributes from non-verbal behaviour literature with observers' attributions of emotional states [LM10a, QBM08]. For example, Cig *et al.* [CKEMT10] conducted user studies to show that changes in gaze and head behaviour, achieved through the variation of gaze animation parameters, led to changes in impressions of the arousal and dominance levels of characters. Li and Mao [LM12] describe a rule-based approach to generate emotional eye movement based on the Geneva Emotion Wheel to enable virtual agents to convey different emotional expressions to users through eye movements. A data-driven approach was adopted by Busso *et al.* [BDNN07] to generate expressive head movements from speech data.

#### 4.4. Conversation

During conversations, non-verbal feedback behaviours relating to gaze, such as glances towards and away from others, in addition to other modalities, have many functions [CTP98] in mediating flow in conversational situations [CT99], including indicating the addressee, paying attention, displaying attentiveness, affecting turn transitions and signalling requests for backchannels [Hey06]. By deploying an agent's gaze strategically, a number of positive high-level outcomes can be achieved in conversations between humans and agents. Previous approaches modelling conversational gaze for characters include those considering communicative functions to pre-determine gaze suitable in a specific conversational context and later modifying gaze patterns using statistical information to generate directed and averted gaze for dyadic interactions [PB03]. Other approaches have used visual attention models to simulate social gaze, with accompanying engagement and distraction behaviours, during multi-party interactions [GB06, GLBB07]. People take another's eye gaze as evidence of what they are attending to and thinking about and use it to determine the course of their current utterance or action [CK03]. Eye gaze produced by a speaker can be used by an addressee to resolve conversational ambiguity [HB07]. When multiple interactions take place over a long period of time, patterns of gaze and other non-verbal behaviours have been shown to adapt as relationships evolve [SB12].

Wang and Gratch [WG10] found that a virtual agent exhibiting gaze attention and displaying cues of positivity and coordination to a user can create stronger feelings of rapport from the user. In an interactive storytelling scenario, a virtual agent that modulates mutual gaze by shifting its gaze in reaction to a user's gaze is able to improve user perceptions of social presence and rapport [BWA\*10]. In immersive virtual environments, a virtual agent has been shown to influence the amount of interpersonal distance a human user will maintain with the agent based on the amount of eye contact the agent makes with the user [BBBL01]. Appropriately timed shifts in gaze away from the user, in relation to the speech of the user, can lead to positive high-level outcomes in turn-taking and participant

disclosure [AMG13]. An agent can also use gaze (along with gesture and speech) to effectively shape a conversation with multiple participants according to its own intentions by directing gaze towards specific participants or averting gaze from participants in order to signal turn-taking intentions [BH10].

The effects of cultural differences and gender on the conversational behaviour of agents are also of importance. Jan *et al.* [JHM\*07] have simulated different cultural parameters related to gaze, overlap in turn-taking and proxemics for rating by native speakers of North American English, Mexican Spanish and Arabic. Studies in immersive and augmented reality environments have shown that users provide more personal space to agents that engage in mutual gaze with them [BBLB02], and users have a higher physiological arousal towards virtual agents not exhibiting behaviours of their cultural background [ODK\*12]. In Vala *et al.* [VBP11], gaze is considered in the creation of a model for varying the communication of agents based on gender.

Recent efforts have involved the use of comprehensive annotated databases of conversations between humans and characters, for example [MVC\*12]. Approaches also consider the animation of gaze aversion behaviours for virtual agents in conversation situations [AMG13].

##### 4.4.1. Artificial listeners

A notable area of research attention involves the definition of non-verbal signals for artificial listeners [HNP07] (see Bevacqua [Bev13] for an overview). In these situations, gaze has been studied in the context of backchannels during conversation, i.e. non-intrusive acoustic and visual signals provided to the speaker by listeners during their turn [YNG70]. The Rapport Agent, developed by Gratch *et al.* [GWG\*07], generates non-verbal backchannels for the listening virtual agent. More recently, Hjalmarrsson and Oertel [HO12] found that listeners are more prone to backchannels when the gaze of a speaking virtual agent is directed towards them; however, gaze alone cannot explain the timing of backchannels.

Such an approach is useful in situations where verbal utterances may not be desirable or practical, due to communication distance or for fear of negative social repercussions arising from failed attempts [Gof63]. Other work has focused on characters that initiate interactions in a favourable manner using gaze in order to promote subsequent interaction attempts [BAT09]. Cafaro *et al.* [CVB\*12] further investigate how the multimodal non-verbal behaviours of characters during the approach to interaction lead to the formation of impressions related to personality and interpersonal attitude based on smile, gaze and proximity, highlighting relationships between gaze and judgements of friendliness.

#### 4.5. Speech and multi-modal interaction

Gaze is an important component in multi-modal behaviours conducted by speakers in face-to-face interactions and is only one of a number of non-verbal channels employed in parallel to speech to efficiently request, take, hold and give interaction turns [Dun74] and request feedback from listeners [Cho92]. The generation of co-ordinated verbal and non-verbal multi-modal behaviours, including

gaze [PPDR00], has therefore been an important focus of research for virtual characters engaging in conversation with humans. Sometimes this involves artificial systems being able to detect engagement through gaze from real conversational partners [NI10, IONN13] in order to adapt to them.

#### 4.5.1. Speech driven gaze

A number of systems use speech as an input from which to generate facial expressions involving the mouth, head, eyes and eyebrows [AHS02]. More recently, Zoric *et al.* [ZFP11] automatically generated facial gestures in real time from the prosodic information obtained from speech signals. Nods, head movements, blinks, eyebrow gestures and gaze were generated using Hidden Markov Models and global statistics. Gaze level fell at the hesitation pause and rose at the end of the utterance to obtain listener feedback. Le *et al.* [LMD12] generated head motion, eye gaze and eyelid motion simultaneously from speech inputs. Non-linear dynamic

canonical correlation analysis was used to synthesize gaze from head motion and speech features. Mariooryad and Busso [MB12] focused on the generation of head and eyebrow motions from speech using three Dynamic Bayesian Networks (DBNs). In recent work, Marsella *et al.* [MXL\*13] utilized semantics based on a shallow analysis of the utterance text and prosody from the speech signal to generate head movements, eye saccades, gestures, blinks and gaze behaviour. Their method outperforms previous approaches that used prosody alone.

A number of research efforts have considered eye movements and gaze specifically during situations involving verbal communication between characters and humans [CCD00]. Vinayagamoorthy *et al.* [VGSS04] presented an eye-gaze model for user-controlled avatars involved in dyadic interactions in shared immersive virtual environments. Their model accounts for differences in saccade magnitude and the effect of the roles of listener or speaker on gaze and intersaccadic interval and also generates movements for other parts of the body. Breton *et al.* [BPG\*06] modelled the gaze behaviour of conversational characters in real-time multi-modal dialogue with a group of users. Ishii *et al.* [IMFN06] proposed a model for animating the gaze behaviour of an avatar for turn-taking in multi-party conversations using utterances and a probabilistic state transition model.

#### 4.5.2. Multi-modal behaviour generation

The automatic generation of conversational behaviour for characters [CPB\*94] involves the generation and coordination of verbal and non-verbal behaviour over multiple modalities, such as facial expressions, lip motions, eye gaze, head motion and arm gestures in order to create meaningful gestures (Figure 4).

For example, systems such as BodyChat [VC98], BEAT [CVB01] and Spark [Vil04] allowed characters to automatically animate their own conversational behaviours, including speech, gaze, turn-taking and backchannel feedback, with minimal user and animator specification through text input. Bee *et al.* [BPAW10] combined gaze and linguistic behaviours to create agents capable of expressing



**Figure 4:** Gaze as one of a number of coordinated modalities expressing panic and fear in an Embodied Conversational Agent (ECA) [NHP13].

social dominance. For an overview of multi-modal coordination, see Martin *et al.* [MDR\*11].

Initiatives such as SAIBA (Situation, Agent, Intention, Behaviour, Animation) [KKM\*06] are being developed to modularize the design of conversational characters. Behaviour Markup Language (BML) [VCC\*07], developed as one of three stages in the SAIBA model, defines multi-modal behaviours, such as gaze, head, face, body, gesture and speech, among others, in a human-readable XML mark-up format. BML allows the definition of multi-modal behaviours by specifying temporal details for primitive action elements (see Krenn *et al.* [KPPP11] for an overview). Another example is the Affective Presentation Mark-up Language (APML) introduced by de Carolis *et al.* [DCPPS04] as part of a Mind–Body interface, where specified communicative goals and their relations are translated into the agents behaviour using APML to express their meaning with the body.

#### 4.6. Eye gaze in HCI and beyond

Eye gaze has been used in a variety of diverse HCI applications, from video games [SG06] to drones [HAMM14] to allowing physically impaired users to control their home using their eyes [BCCDR11]. Gaze has been used to support basic user interaction tasks, such as keyboard input for pointing and selecting [KPW07]. Advancements in gaze tracking have focused on generic selection methods for gaze-only controlled interfaces based on gaze gestures [DS07, MLGH10], that is, gestures that consist of a single point-to-point movement in addition to studies involving eye closure [HR12] and dwell time [DAH12] for a variety of applications, including games [WEP\*08, EPM08, IHI\*10]. The difficulties encountered in the evaluation of these systems was addressed by Świrski and Dodgson [SD14], who generated physically accurate renderings of synthesized eyes as ground truth data for video-based gaze estimators.

Other research has taken multi-modal approaches seeking to augment eye gaze, for example, with simple head gestures [MHP12, vM12], voice [WEP\*08], hand gestures [YHC\*10], facial expressions [SII04] and touch [SSND11, SD12, TAB\*13] and EMG [CBCA07]. Numerous research efforts have focused on fast, dwell-free eye-typing [KV12] and text input through continuous movements of the eye [WRS\*07, BA08]. Other work has involved the use of eye gaze for drawing applications [YC06, Hei13a, JPMG13], in some cases involving the use of both gaze and voice [vdKS11]. For example, in Heikkilä [Hei13b] gaze

gestures enable moving and resizing, while the closure of the eyes stop actions in the drawing system.

#### 4.6.1. Detecting user eye-gaze behaviour

While gaze detection from users has been extensively studied in HCI for controlling and complimenting interfaces, core technologies and research endeavours are also relevant to those working in the domain of interactive intelligent agents, for example, Embodied Conversational Agents, or ‘ECAs’, and HRI. These systems are often tasked with engaging the user in dialogue, providing assistance or engaging in pedagogical activities [CBK\*14]. Sustained interactions require the agent to have some knowledge of user behaviour and potential state, such as their level of engagement [PCdF09]. Since gaze is a powerful and efficient communication modality, a number of efforts aim to detect gaze direction from users to fully or partially inform the behaviour of artificial agents. Some of these efforts use the position of a user’s head to inform an agent’s reactive gaze behaviour [KG08], while others recognize an individual’s gaze aversion behaviours given contextual information [MD08] or identify an addressee during multi-party conversations to ensure that an agent responds appropriately when it has been addressed [NBHH13].

The detection of user gaze and derived inferences of attentive state has also been used in HRI situations (see also Section 5.2) to inform and trigger robot behaviour. For example, Das *et al.* [DRKK14] classify human gaze behaviour as either spontaneous or scene-relevant looking. The robot interrupts the user to explain aspects of the scene only if it detects that they are looking at something of scene relevance.

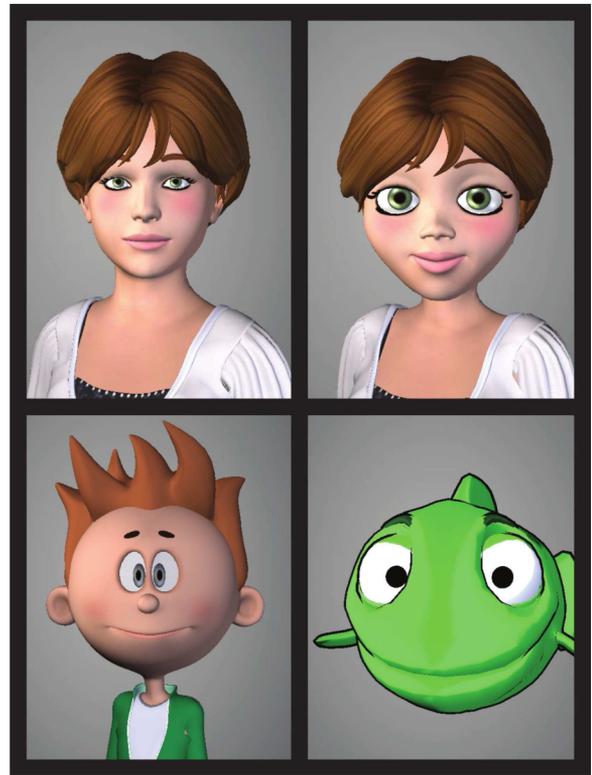
#### 4.6.2. Shared attention and engagement

Detecting user eye gaze is foundational to implementing more advanced artificial social competencies. Agent interaction capabilities relating to engagement detection [CPC13], and joint and shared attention may be supported by real-time gaze detection from human users utilizing computer vision techniques [PAK10] and eye tracking approaches [HPB\*07]. Attentive presentation agents [EPAI07], for example, are capable of utilizing the user’s visual attention, based on their tracked eye fixations, to detect objects of interest and guide the behaviours of two virtual presentation agents who describe the items in further detail.

A more recent study by Pfeiffer-Leßmann *et al.* [PLPW12] implements joint attention behaviours in the virtual agent ‘Max’ to compare the timings of initial referential gaze acts in human–human interactions to those in human–agent interactions. The results indicate that, at least in this situation, the dwell time in the human–agent interaction was comparable to the human–human interaction case. Shared attention has also been investigated in relation to physical embodiments (see Section 5.1.1).

#### 4.7. Stylized characters

Most gaze models are based on observations of humans or other animals that have eyes of a particular size and movements that are subject to physical properties and anatomical limits. Real-world



**Figure 5:** Four different characters with different eye geometry. The character on the upper left has realistic human proportions. Adapting gaze models to characters with stylized eye geometry requires the consideration of potential artefacts [PMG13].

gaze movement must be functional to be effective; however, when similar gaze models are applied to characters, such as cartoon or stylized characters, whose eyes are significantly different from those of their real-life counterparts, problems may arise. The question is whether or not it is appropriate to use models developed for humans and human eyes for characters. Lacking an alternative, this is the most common approach. The human brain has a particular affinity for faces and eyes (e.g. Bentin *et al.* [BAP\*96]), and peoples’ propensity to interpret even vaguely eye-like objects as eyes suggests adapting human models to characters is a viable approach.

In applying human gaze models to characters with varying eye geometry (see Figure 5), a number of issues arise. Pejsa *et al.* [PMG13] catalogue and propose solutions to several of these issues. The issues fall into three key areas. The first is that ‘normal’ human movements often look odd when examined closely. Human and animal eyes are sufficiently small that most people never see the details of their movements. When these movements are magnified, for example by using a character with proportionally larger eyes, the movements are unexpected. Secondly, stylized character eyes are often larger than realistic eyes and therefore require unrealistically fast movements to traverse the angular displacements made by real eyes. Thirdly, stylized eyes often are of unrealistic geometric configurations, such as asymmetric shapes, and movement ranges. Each of these issues leads to similar effects: when human eye models are applied to



**Figure 6:** *Implied eye-line (recreated from [Osi10, chapter 2]).*

stylized eyes, visual artefacts occur, such as movements that seem unusual or would not be seen in humans except under pathological circumstances (such as cross-eyedness or divergence).

To adapt a human gaze model to a range of stylized characters (see Figure 5), Pejisa *et al.* [PMG13] added constraints to ensure that pathological situations, such as cross-eyedness, do not occur. To meet these constraints, they exploited the fact that virtual agents are performative (see also Section 6.2 and Figure 6). They do not need to function correctly (e.g. enable the character to see); rather, they serve as a signal to the audience. The approach moves gaze targets and limits eye movements in order to avoid objectionable artefacts. While their model did not accurately point the eyes at the intended gaze targets, an empirical study confirmed that this inaccuracy did not limit viewers' ability to localize gaze targets; furthermore, the visible artefacts were significantly reduced.

Efforts to map gaze behaviours between virtual and physical embodiments, such as robots, are also of relevance to this work and are described in Section 5.

## 5. Social Robots

The modelling of eye-gaze behaviours in robotics presents a number of complimentary and interesting crossovers with similar efforts taking place in the domain of virtual characters. This section describes advancements in gaze generation for supporting enhanced interactions with social robots in conversational scenarios. An important issue that arises is the mapping of gaze behaviours between virtual and physical embodiments, such as robots, where there are challenges in retargeting to platforms with reduced embodiments. The investigation of reduced embodiments is significant to Computer Graphics applications where lower fidelity or more constrained characters are sometimes desirable while ensuring that core communicative qualities remain intact.

### 5.1. Human–Robot interaction

In the HRI community, a particular emphasis has been placed on examining the high-level outcomes achievable through gaze. This focus may be due to the HRI community's strong ties to social science communities and partially due to the fact that HRI researchers generally have less control over the low-level variables of gaze. For example, HRI researchers are limited by the velocities

achievable by physical motors, and affordances of their robots or when working with a robot that does not have articulated eyes and must express gaze through head movements alone. Srinivasan and Murphy [SM11] provide a survey of HRI gaze research that identifies a number of social contexts and discrete gaze acts considered so far by the community.

A number of high-level conversational outcomes achievable through gaze behaviour have been investigated in HRI research, including the use of gaze to signal participant roles [MSK\*09], facilitate turn-exchanges [KYH\*98] and affect the human user's physical and psychological closeness with the robot [MM11]. A robot that gazes responsively towards the human user, rather than randomly or statically, produces stronger feelings of being intentionally looked at [YSI\*06]. In conversations with multiple participants, a robot's scanning gaze behaviour serves the dual functions of indicating attention to all conversational partners and updating the robot's knowledge of the partners that are occasionally outside the range of its cameras [BFJ\*05].

#### 5.1.1. Shared attention and gaze cueing

Studies have also investigated how different types of gaze behaviour, short frequent glances and long, less frequent stares, may convey a robot's visual attention [AHFS\*13]. Similar to studies involving virtual agents (see Sections 4.2.1 and 4.6.2), shared attention behaviours and gaze cueing have also been studied with physical embodiments. Moon *et al.* [MTG\*14] demonstrate that human-like gaze cues helped to improve performance during a task involving the handover of water bottles. Subjects initiated reaching actions earlier when shared attention gaze behaviours were present than in other conditions. In another study involving a handover task, Kirchner *et al.* [KAD11] found gaze to be an effective individualizing cue for ensuring that a selected participant, in contrast to arbitrary group members, would take a salient object.

Previous research has studied how to deploy a robot's gaze as a cue for its attention, based on detected engagement levels of human participants in the interaction [XLW13] and timed in relation to object references made by the robot when speaking or by the user when the robot is listening [SC09]. Significantly, Admoni *et al.* [ABT\*11] conducted an experiment to investigate if robot gaze elicits the same reflexive cueing effect as human gaze. Their findings indicate that while robots conveyed directional information, they failed to elicit

reflexive attention cueing effects, implying that humans cognitively process robots differently from other humans in common social mechanisms.

### 5.1.2. Task performance and speech

In relation to gaze and speech behaviour, in a user study involving a map route drawing task, Skantze *et al.* [SHO13] show that appropriate robot gaze behaviours helped to improve task performance and reduce cognitive load by helping to disambiguate referring expressions to objects in a shared scene and manage the flow of interaction. Notably, the study also considers how gaze may be used as a cue to inhibit or encourage the user to act during pauses in speech. Users gave less feedback and looked at the robot more during the task when the preceding utterance segment was incomplete.

## 5.2. Mapping virtual and physical embodiments

The virtual agents and social robotics communities have separately investigated a number of gaze mechanisms and their conversational effects. However, it is often unclear how these mechanisms might be accurately translated between the virtual and physical realms. Robots and virtual agents have been shown to differ along a number of social dimensions, including realism, social presence, lifelikeness and physical proximity [PKFT07]. A number of studies have been performed that compare virtual agents with human-like robots in various contexts, such as health education and collaborative work [PKFT07, KB04, BHKS11].

A number of unique challenges arise when attempting to re-purpose behaviours from a virtual agent for a human-like robotic platform, including (1) the acceleration and speed of a robot's movements have both upper and lower limits, (2) due to physical inertia and communication latency, a robot will typically not react instantaneously to a command and (3) robot expression has fewer degrees of freedom [LvW12]. The third challenge is of particular importance when attempting to bring gaze mechanisms from a virtual platform, in which agents can use both their head and eyes, to a robotic platform without articulated eyes. In a similar vein, the physically co-located robot must actively direct its gaze to track users' faces, motions that are not required of a virtual agent due to the Mona Lisa gaze effect [DdGB10, AMEB12]. Conversely, this may also enable physical systems to better direct human gaze to targets of interest in the real environment. In general, retooling a behaviour for a completely different platform (e.g. translating a virtual agent gaze model for human-like robots) requires a consideration of what minimum representation that behaviour would need in order to evoke the desired response [CT07]. Recent work by Andrist *et al.* [ATGM14] showed how conversational gaze aversion behaviours, originally developed for virtual agents, could be effectively adapted for a human-like robot platform that lacks articulated eyes.

## 6. Animator's Eye Animation

In this section, we consider the animation of the eye and head from the perspective of a character animator. The creation of animated characters can be compared to a film or stage actor, where the

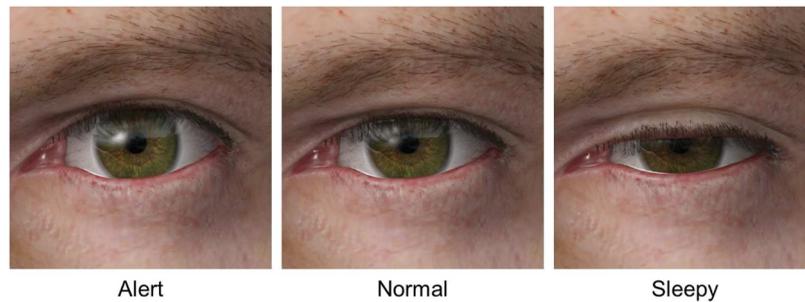
character animator designs a fictional character that must convey thoughts, emotion and personality. The main goal is to create a character which is appealing to the audience—one they can identify with, no matter if it is the hero or the villain. We will summarize animation guidelines from different animators, well known from the movie industry. These guidelines illustrate the requirements an animator has to meet when animating eye, eyelid and head movements, which often differ from the requirements of realistic approaches described in the previous sections.

### 6.1. Thoughts and emotions

Walt Disney recognized the importance of the eyes and advised that, 'the audience watches the eyes, and this is where the time and money must be spent if the character is to act convincingly' [TJ95]. Well-designed and animated eye and head movements inform the audience about the character's emotions and thoughts and are guided by the character's personality.

Before creating a character, the animator has to acquaint themselves with the personality of the character. Williams summarizes the job of an animator in three steps [Wil09]: first, to clearly communicate the point of the scene; secondly, to project themselves into the character's mind; and thirdly, to portray the thinking process of the character. To convince the audience that a character is alive, the animator matches the movements and actions with the thought process. Anticipation is one of the basic principle of animation [TJ95], and is especially suited for creating a character whose actions are guided by thoughts. A convincing action should always be led by the eyes or the head, starting with the eyes moving a few frames before the head and the rest of the body [Las01]. In specific cases where the character does not have eyes or where eye movement is unwanted, the head is used to lead the character's movement. The time frame of the eyes or head leading the action depends on the substance of the thought—does the character have to make a quick decision, or does the character have to choose between different options? Other tools to anticipate action are *takes* and *double takes* [Gle13]. In a take, the character reacts to a situation. First, the character recognizes a situation, followed by an anticipation action, the actual take and the final recovery. A double take increases the effect of the anticipation by delaying the reaction to the situation [Gle13]. The anticipation of eye movements enhances the character's credibility. An audience recognizing the thoughts of a character is aware of subsequent actions and can follow them easily [TJ95].

Some animators propose the use of Neuro-linguistic programming (NLP) as a guideline to map thoughts to specific eye movement [BG79]. While there is no true consensus between researchers as to whether or not these mappings are applicable, these rules can help animate the characters' eyes to convey a specific thinking process. Each eye movement and direction is mapped to an internal thought. A character looking upwards indicates visual thoughts, such as remembering or constructing an image, looking to the middle relates to sound and looking downwards relates to emotions. While these positions vary from person to person and depend on other stimuli, a more simplistic approach is to match the eye movement of a person looking upwards to thoughts regarding short-term memory and looking downwards to long-term and personal memory [Hoo13].



**Figure 7:** Relation eyelid to iris (recreated from [Osi10, chapter 2]).

## 6.2. Secondary movement—gaze, blink and pupil

The animation of conversing characters presents a specific challenge due to the production cycle of the animation movie and visual effects industry. After a script is written and voice actors act out the dialogue, the animator breaks down the dialogue to understand the character's state of mind and determines the thoughts behind the words [Hoo13]. This thought could be an intention, objective, motivation or emotion [Hoo13], indicated as a change in eye movement, concluding that the eye and head movement associated with this thought, anticipates the dialogue by several seconds [Mae06]. The animation must be simple and direct, communicating one single thought or emotion at a time, to catch and maintain the audience's attention [Mae06]. Excessive head movement could distract the audience and make it difficult to see the character's face.

The animator also must find the character's line-of-sight and the conversational partner's location to keep the character's focus fixed on the subject [Mae06]. In addition, the animator has to ensure that the audience is always aware of what is going on in the character's mind. A character turning the head completely towards a target, such as an object or a dialogue partner, conceals the character's facial expression from the audience. Osipa [Osi10] therefore recommends the use of an implied eye-line. The eye-line indicates where the eyes are pointing at and leads the audience toward the target. With an implied eye-line, more of the character's face and expression is visible. More white on the left side of the eye for example, implies that the character is looking to the right side of the screen towards a target. From a top-down view, the eye-line might not directly intersect with the target, but the audience still is able to follow the character's thoughts (see Figure 6). Furthermore, a character's constant attention at an object communicates life and attentiveness to the audience.

As with the eyes, the correct animation of eyelid movements and blinking contributes to the believability of the character and emphasize thoughts and emotions. The position of the eyelid in relation to the eye and pupil is a powerful method for changing entire facial expressions. Osipa's diagram (see Figure 7) shows the relationship between eyelid, eye and pupil [Osi10]. A half-open eye contributes to a sleepy expression, while a fully open eyelid gives the impression that the character is alert. Therefore, Maestri [Mae06] suggests to model a neutral, relaxed eyelid as approximately 80% open to avoid audience misinterpretations of

the character's appearance as over alert or sleepy (see Figure 7, middle).

To incorporate effective blinking into an animation, it is essential to correlate thoughts and eye blinks as well as depict anatomical movements correctly. An eye blink during significant events in the animation draws the viewers' attention and can be used to strengthen an expression. Murch's book 'In the blink of an eye' [Mur01] gives advice on film editing and the importance of eye blinks. He interprets eye blinks as the projection of internal thoughts and the internal struggle a human might have. A blink separates visual inputs into important portions for easier processing. For Murch, an eye blink supports the separation of thoughts to order them or to break down ideas. Eye blinks are described as in sync with our emotional state and internal thoughts. For example, a person who does not blink for a long time might hold onto one thought, while a person who blinks every few seconds might switch from one thought to another in an attempt to order them and to find the answer. The appropriate place to add an eye blink therefore would be when the character is processing a thought or when the character reaches a decision. In addition, a blink occurs almost simultaneously with the onset of large eye and eye-head gaze movements (see Section 3). Furthermore, Osipa [Osi10] advises against repetitive eye blinks following one another close in time, since this could indicate disbelief, shock or confusion and misleads the audience from the main point. Osipa further recommends not to add eye blinks 'in the first or last 5 to 10 frames of a shot' [Osi10], since an editor needs flexibility in the area to cut into and out of a shot.

Dilation of the pupils is often overlooked by animators, since pupils typically remain constant in size throughout a scene. However, changing their size adds expressiveness and liveliness to a character and underlines the character's state of mind [Rob07]. A child-like appearance is achieved with large pupils, while small pupils reflect fear, or could make the audience aware of an evil or untrustworthy character. Adapting the size of the pupil to surrounding light conditions further emphasizes the character's authenticity.

With all these tools, the animator designs a character's personality ensuring that its particular attitude [TJ95], supporting thoughts, emotions and words are communicated by the careful timing and animation of the eye, pupil and eyelid movements. In accomplishing this, the animator must be aware of the thoughts behind the character's action, where the character is looking and where the audience is looking at any given moment.

## 7. Perceptual Evaluation of Agents

Researchers strive to create virtual entities that evoke the same reactions as real humans do during social interaction. It is therefore important to take human perceptions into account in judging the plausibility and realism of virtual entity simulations. In this section, we discuss the ways in which low- and high-level gaze models have been evaluated in terms of plausibility, effectiveness and ability to achieve communicative goals. Various approaches have been adopted, including survey-based questionnaires, direct comparisons with state-of-the-art algorithms and ground truth eye data captured from a human, to achieve this goal. We highlight the importance of perception and how effective perceptual experiments can be designed.

### 7.1. Experimentation methods

The effectiveness of an experiment depends on the careful preparation of stimuli, the presentation of those stimuli and a correct analysis of the results. This section describes in detail these individual steps.

#### 7.1.1. Stimuli preparation

The development of an experiment begins with the acquisition of stimuli that will be shown to the participants. These stimuli are typically images or animations which focus on a specific factor of interest. There are a number of different media and methods for presenting stimuli, such as videos, interactive applications or virtual reality. In the case of evaluating the accuracy of a head-eye coordination model, the stimulus presented might be an animated character gazing towards objects on a desk, as found by Andrist *et al.* [APMG12b]. In this study, a series of videos was generated, and one of two videos was shown to participants. In one video, an animated virtual character performed gaze-shifts generated by the proposed model. In the other video, the state-of-the-art model and a human confederate looked at objects. Sixteen objects were placed in two rows on a desk with form or colour coding for easy identification. Communicative accuracy was measured based on how accurately the participants identified the object towards which the agent was looking. In work focusing on eye-representation, Steptoe and Steed generated a scene in which an avatar was seated behind a clock-face and viewed from different camera angles [SS08]. Only the avatar's eyes were directed towards a specific number on the clock-face. Conditions alternated with respect to eye socket deformation and vergence. The authors showed that realistic socket motion increased both perceived authenticity and the viewer's self-assessment of having correctly identified the point of regard, while vergence had a much more modest effect.

In order to investigate if a proposed model represents an improvement on previous methods, comparisons may be performed between sample stimuli. For example, Steptoe *et al.* [SOS10] generated a series of videos which animated blinks and lid saccades of two different virtual characters based on motion captured data, their proposed models and simple linear interpolation. Animations of each condition assembled in one video were then ranked regarding perceived realism and similarity to the recorded video. Similarly,

Deng *et al.* [DLN05] synthesized eye motions with different existing approaches and compared the results with their own model. This allowed them to test if their model showed an improvement in naturalness over previous methods. Peters and Qureshi [PQ10] focused on identifying the most appropriate parameters of an eye gaze and blinking system and created animations which altered only the blinking conditions, eye-head ratio or direction of head movements to assess which strategy would be better received by users. With a similar aim, Trutoiu *et al.* [TCMH11] studied the blinking mechanism in isolation by allowing participants to compare their model generated from real eye-blink data to other commonly used methods for blink animation, such as linear interpolation and ease-in ease-out curves.

Once stimuli have been generated, the order and way in which stimuli are presented must be decided. Most setups show each stimulus under all possible conditions in a random order to each participant (*within-groups* design). The randomized presentation of stimuli minimizes the risk of order effects, wherein participant judgements may be biased by their evaluation of preceding stimuli. An alternative setup was used by Andrist *et al.* [AMG13], however. In this setup, participants were randomly assigned to one of three gaze aversion conditions in which they performed tasks for each of the proposed hypotheses. This is referred to as the *between-groups* design, where two or more groups of subjects can be tested simultaneously under varied conditions with multiple modified variables. In a shared immersive virtual environment, Garau *et al.* [GSV\*03] paired two participants together as part of a role-play, one using a CAVE™-like system and the other using a head-mounted display. Each participant was assigned randomly to one of four conditions to investigate the impact of an avatar's eye-gaze behaviour on perceived realism. The setup allowed the authors to test subjects under various conditions with altered variables simultaneously to acquire an unbiased opinion from participants on behavioural realism and the quality of communication.

Some thought should also be given to where the experiment takes place. While most experiments usually take place in a laboratory due to the available infrastructure, online experiments offer an advantage in that they reach a broader range of subjects and hardware configurations [OSS09]. For researchers located in the United States, Amazon.com's Mechanical Turk [Ama14] gives researchers the opportunity to access a wide range of potential participants. However, care must be taken to ensure the experiment is set up carefully and that the results are examined critically to avoid selection bias and other confounds [PCI10, BKG11].

#### 7.1.2. Stimuli presentation and evaluation

Two types of measurements can be used to evaluate the participants response to the stimuli: subjective measurements and objective measurements. Subjective measurements include self-reported assessments of the characters by observers, such as how plausible or likable observers find a character to be. An example for objective measurements is the examination of how well participants remember information presented by a character.

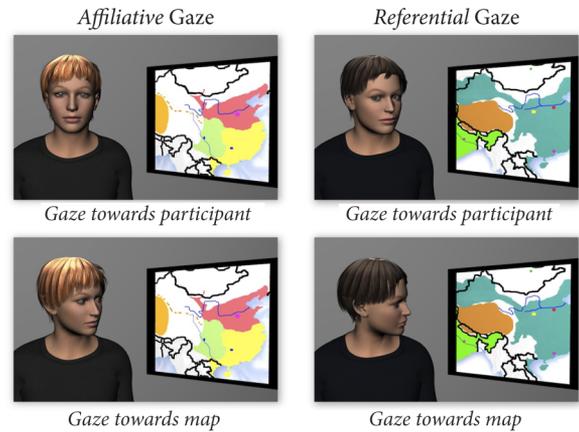
A user's subjective response to a stimulus can be evaluated by a rating scale, forced choices, a free description task or an

interview. The most common approach to capture personal impressions is a questionnaire employing a rating scale. One such scale is referred to as a *Likert* scale [Lik32], which captures the participant's level of agreement to a statement. The levels used in research literature questionnaires are typically five- or seven-point scales to ensure an equal number of positive and negative positions and a neutral middle position are incorporated into the scale. McDonnell et al. [MBB12] asked participants to rate different rendered models on a scale from 1 to 7, according to specific definitions (e.g. 'extremely abstract–extremely realistic', 'unappealing–appealing' and 'unfamiliar–familiar'). This enabled them to make a subjective judgement of the participants' experiences on well-defined scales.

An alternative to the Likert scale is the semantic differential method. On the commonly used seven-point scale, the connotative meaning of a concept is measured [OST57]. Adjectives of opposite meaning define extremes at both ends of the scale. A representative example of the semantic differential method is the evaluation of users' impressions on a presented gaze model by Fukayama et al. [FOM\*02], which used 20 pairs of polar adjectives adopted from psychological studies, such as 'extrovert versus introvert' or 'lazy versus diligent'. This technique is useful when the experimenter wishes to measure the attitudes of participants to the experimental stimuli. For example, a human motion could be presented and participants asked to judge it in terms of qualities such as happiness, energy level and so on.

The difficulty with rating scales lies in creating questions with answers that clearly map to all points on the scale. For example, on a one to seven-point scale, the participant needs to know what 1 and 7 mean when answering a question, such as 'How happy does this character look?' In practice, a number of researchers define new sets of questionnaires and mappings adapted to their requirements. A good resource for concepts to capture the impression of an avatar's personality or naturalness is the field of psychology. The *Ten Item Personality Inventory* (TIPI) [GRS03] was defined to capture the *Big Five personality traits* [Nor63] describing the human personality: openness, conscientiousness, extraversion, agreeableness and neuroticism. To test the effect on the perception of social dominance in synthesized conversational behaviours, Bee et al. [BPAW10] used TIPI as reference for extraversion and agreeableness. The same study also derived six questions from the *PAD emotional state model* to measure dominance. The psychological *Pleasure-Arousal-Dominance* (PAD) Emotional State Model, developed by Mehrabian, represents the fundamental human emotions in the dimensions: *Pleasure*, *Arousal* and *Dominance* [Meh80]. In connection with the evaluation of emotions, the use of *self-assessment manikins* (SAMs) [GK05] should be mentioned. SAMs is a non-verbal technique that has been used to capture the three PAD dimensions of emotions [BDG\*07]. Each dimension is depicted by stylized figures on a five-point scale. Used for the recognition of emotions in speech, this system has been shown to be simple and straightforward [BDG\*07].

In general, rating scales are an effective way of collecting data quickly. However, a disadvantage of the approach is that participants may give socially desirable responses, or they may develop a response set (e.g. giving consistently moderate answers). A forced choice task restricts the responses a user can give to a specific question. Presenting a limited number of cases from which the user



**Figure 8:** The characters in this study were delivering 3-min lectures. Assessment included both subjective questions about rapport with the agent and objective questions about how much of the lecture was remembered [APMG12b].

chooses avoids the problem of difficulty in interpretation. In the case of evaluating the naturalness of their head-and-eye motion model, Le et al. [LMD12] generated video clips using different methods, including their own, to animate head and eye motion and facial gestures. Two sequential clips were shown to the participant, who was asked to choose the more natural one. Excluding the motion capture animations, user votes were in favour of the speech-driven head and eye motion generator proposed by the authors.

After the experiment has taken place, it may be useful to include a free writing questionnaire or post-experiment interview in order to obtain additional information about participants' opinions. Following an experiment to test the effect of a robot's gaze behaviour on perceived believability and likability, Poel et al. [PHN\*09] carried out a semi-structured interview. More direct questions about the user's experience with the robot provided further qualitative information for comparison between the robot using the gaze behaviour system and the robot not using it.

In contrast to subjective measurements, the objective assessment of a method is possible by using different task designs. Such an objective approach was used by Murray et al. [MRS\*07] in an immersive virtual environment where the user had to pick the object from the screen at which he thought the avatar was looking. An objective measurement to verify whether or not the gaze model by Andrist et al. [APMG12b] improves learning took the form of a quiz taken by participants after they watched an agent present a lecture (Figure 8). A virtual agent gave a 3-min lecture on a Chinese city visible on a map under conditions employing affiliative gaze, referential gaze or a combination of both gaze types. The 10 questions on the quiz were divided into categories unknown to the user: recalling the knowledge associated with the map, the verbal content of the lecture and information linked between them. The objective assessment indicated that referential gaze improved learning performance, while the affiliative condition was rated better in the subjective measures.

Alternative assessment methods that can be employed to explore the design space in the formative stages of the process have been

proposed, although these methods have not yet found common use. An example is a ‘multivariate’ evaluation approach in which a number of design variables, such as gaze frequency at different targets, fixation length or alignment between the eyes and the head are simultaneously manipulated along a continuous scale, such as  $x_i \sim U(0,1)$ . Their joint effects on the measured outcomes are then modelled. Huang and Mutlu [HM13] used this approach to assess how different types of arm gestures of a storytelling robot affected how much participants recalled the details of the robot’s story, the participants’ perceptions of the robot as an effective storyteller and their ability to retell the story.

Mathematical- and computer-based evaluations offer an alternative to human participants. Testing the efficiency and performance improvement of a proposed approach compared to other approaches is one way [GT09]. An example of a computer-based evaluation was carried out by Itti and Dhavale [ID03]. The eye–head movements of their model were traced, demonstrating the basic modes and comparing them to possible human behaviours. Using various video segments as input, the horizontal and vertical eye and head coordinates were plotted on 2D graphs to illustrate the ‘correct’ (human) tracking behaviour.

## 7.2. Analysis

The final step in evaluating a proposed modelling technique is data analysis. Data reduction, descriptive statistics or inferential statistics, among other methods, are applicable depending on the task performed. Factor analysis is suited to results from semantic differential rating scales as in case of Fukayama *et al.* [FOM\*02]. For the evaluation of other rating scales, such as the Likert scale, Analysis of Variance (ANOVA) can be used [MB10]. A *post hoc* test on the ANOVA result further improves the reliability of the analysis. Other methods are the Tukey–Kramer HSD for experiments with unequal sizes of groups, Newman–Keuls tests for comparison of means [ZHRM13] or Bonferroni corrections for multiple comparisons [TCMH11].

Each of these tests is suited better for a conducted experimental study and has advantages and disadvantages. While the Tukey and Newman–Keuls tests are more suitable for simple comparison, Scheffé’s method or Bonferroni are useful when the experiment consists of subgroups. The choice of the appropriate test therefore depends on the research question, the given data set, what needs to be compared and the desired power of the analysis.

## 7.3. Perception of gaze behaviour

The aforementioned techniques form a basis for a wide range of user studies probing human impressions of artificial gaze under a variety of circumstances. Examples of studies are to be found in numerous papers described throughout this survey, and range from the perception of rendering and low-level animation (Sections 2 and 3), to influences on social interaction (Section 4) for both virtual and physical embodiments (Section 5). The approaches are applicable to evaluating and probing both automatically and manually created gaze animations (Section 6). Some further examples follow to demonstrate the variety of perceptual studies that can take place.

Peters and Qureshi [PQ10] conducted a perceptual study relating to the ratio of the head and eyes that contribute to a gaze motion, as well as the timing and amplitude of blinks that accompany gaze changes. The results of the study indicate that gaze movements composed of high eye to head movement ratios, i.e. those in which primarily either the head or the eyes move, are perceived to be less natural than when both the head and eyes contribute to the motion. Significant differences were also found regarding horizontal and vertical gaze shift conditions, suggesting the possibility of heightened viewer sensitivities to vertical gaze shifts.

Ruijten *et al.* [RMH13] considered interactions between gaze directions and for approach- versus avoid-oriented emotions expressed through facial expressions. Approach-oriented emotions (angry expressions) were rated as more credible when they were combined with a direct gaze, but this was not found to be the case for avoid-oriented (sad) expressions. Bee *et al.* [BFA09] studied the effects of eye gaze and head tilt on the perception of dominance.

Other studies have investigated the effects of agent gender and gaze. Kulms *et al.* [KKGK11] study used ‘The Rapport Agent’, varying the gender of the character in order to determine if female agents are evaluated more negatively than male embodiments when they engage in gaze motions with humans. The study found that female agents received more positive judgements than male agents. Participants provided significantly more verbal responses to the agent, even when not explicitly required to do so, when the agent engaged in behaviours that involved, among other factors, less aversive gaze.

More recently, Yu *et al.* [YSS12] found that, in joint attention tasks between humans and agents, humans were highly sensitive to momentary multi-modal behaviours generated by conversational partners, real or artificial, and rapidly adjusted their gaze behaviour patterns to these behaviours.

Experiments in the perception of gaze behaviour typically take place on a 2D display. This approach can lead to undesired side effects, such as the Mona Lisa gaze effect, where the gaze of a viewed object seems to follow the observer regardless of the perspective. al Moubayed *et al.* [AMEB12] investigated this effect by projecting embodied agents onto 2D screens and a new proposed 3D projection surfaces. As a result they found that the 3D projection surface eliminates the Mona Lisa effect.

## 8. Conclusions

In this report, we have presented the accomplishments to date on generating artificial entities that aim to replicate the motion and appearance of the human eye. From our perspective, there are a number of important future directions, such as how models can be used in practice between applications, how to create variety and individuality in gaze behaviour and how best to create more expressivity in cartoon characters.

One critical aspect of future gaze modelling research will be to better connect models to application-level outcomes so that agent developers can use models in practice. Achieving such a transfer will require three tasks: (1) cataloguing the kinds of effects that agents can achieve through proper use of the agents’ gaze, (2)

providing techniques that synthesize the appropriate gaze behaviours to achieve these effects, and validating that these effects can be achieved through the use of the techniques and (3) demonstrating potential benefits of the use of designed agent gaze behaviours. While some success has been achieved in illustrating these effects, to date these have been in small, controlled scenarios.

The majority of eye-gaze methods discussed in this report have attempted to create a generic model of gaze behaviour for all agents. In general, research in the area of eye-gaze models would hugely benefit from a fundamental foundation. One aspect that has been neglected is the idea of individuality and interpersonal differences in eye gaze. This may be particularly important when simulating groups or crowds of agents in order to increase the variety and individuality of the agents. Previous research has identified the importance of variety in both appearance and motion of agents [MLD\*08, MLH\*09], but has yet to investigate whether head and eye gaze could affect crowd variety also.

Computational gaze modelling has focused on reproducing the mechanisms seen in real humans and animals. However, animated characters are not subject to the same constraints as the real world and might be able to achieve more effective communication by using different mechanisms. Cartoon characters have long achieved communicative effects using eye movements that are impossible in the real world; for example, spinning eyes, eyes shooting rays at a target and eyes bulging. Understanding the potential for increased expressiveness and the trade-offs that arise once paradigms are broken is an interesting avenue for future work.

Although much work has been accomplished in this field, a number of challenges remain before we can replicate the complexity and individuality of human eye behaviour. We have indicated some of the future directions from our perspective, but anticipate a much wider range of future activities in this area.

### Acknowledgements

This work was sponsored by Science Foundation Ireland as part of the *Cartoon Motion* project. This research was also supported by the NSF award 1208632 and ERC advanced grant *EXPRESSIVE*, the U.S. National Science Foundation awards 1017952 and 1149970 and the U.S. National Institutes of Health award 1R01HD071089-01A1. This work was partially supported by the European Commission (EC) and was funded by the EU Horizon 2020 ICT-644204 project ProSocialLearn. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

### References

- [Abe86] ABELE A.: Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior* 10, 2 (1986), 83–101.
- [ABT\*11] ADMONI H., BANK C., TAN J., TONEVA M., SCASSELLATI B.: Robot gaze does not reflexively cue human attention. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society CogSci 2011* (Boston, MA, USA, 2011), Citeseer, pp. 1983–1988.
- [AC76] ARGYLE M., COOK M.: *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge, UK, 1976.
- [AHFS\*13] ADMONI H., HAYES B., FEIL-SEIFER D., ULLMAN D., SCASSELLATI B.: Are you looking at me?: Perception of robot attention is mediated by gaze type and group size. In *HRI '13: Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (Tokyo, Japan, 2013), IEEE, pp. 389–396.
- [AHO90] ASTINGTON J., HARRIS P., OLSON D.: *Developing Theories of Mind*. Cambridge University Press, Cambridge, UK, 1990.
- [AHS02] ALBRECHT I., HABER J., SEIDEL H.-p.: Automatic generation of non-verbal facial expressions from speech. In *Advances in Modelling, Animation and Rendering*. Springer London (2002), pp. 283–293.
- [Ama14] AMAZON.COM: Amazon Mechanical Turk, 2014. //www.mturk.com/mturk/. Accessed 15 January 2015.
- [AMEB12] AL MOUBAYED S., EDLUND J., BESKOW J.: Taming Mona Lisa: Communicating gaze faithfully in 2D and 3D facial projections. *ACM Transactions on Interactive Intelligent Systems* 1, 2 (2012), 11:1–11:25.
- [AMG13] ANDRIST S., MUTLU B., GLEICHER M.: Conversational gaze aversion for virtual agents. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents* (Edinburgh, UK, 2013), vol. 8108 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 249–262.
- [And81] ANDERSON J. H.: Ocular torsion in the cat after lesions of the interstitial nucleus of Cajal. *Annals of the New York Academy of Sciences* 374, 1 (1981), 865–871.
- [APMG12a] ANDRIST S., PEJSA T., MUTLU B., GLEICHER M.: A head-eye coordination model for animating gaze shifts of virtual characters. In *Gaze-In '12: Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction* (Santa Monica, CA, USA, 2012), ACM, pp. 4:1–4:6.
- [APMG12b] ANDRIST S., PEJSA T., MUTLU B., GLEICHER M.: Designing effective gaze mechanisms for virtual agents. In *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, TX, USA, 2012), ACM, pp. 705–714.
- [ATGM14] ANDRIST S., TAN X. Z., GLEICHER M., MUTLU B.: Conversational gaze aversion for humanlike robots. In *HRI '14: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany, 2014), ACM, pp. 25–32.
- [AWH10] ANDERSON C., WALES A., HORNE J.: PVT lapses differ according to eyes open, closed, or looking away. *Sleep* (2010), 3:197–3:204.
- [BA08] BEE N., ANDRÉ E.: Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze. In *Perception in Multimodal Dialogue Systems* (Kloster Irsee,

- Germany, 2008), vol. 5078 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 111–122.
- [BAP\*96] BENTIN S., ALLISON T., PUCE A., PEREZ E., MCCARTHY G.: Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience* 8, 6 (1996), 551–565.
- [BAS75] BAHILL A. T., ADLER D., STARK L.: Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science* 14, 6 (1975), 468–469.
- [BAT09] BEE N., ANDRÉ E., TOBER S.: Breaking the ice in human-agent communication: Eye-gaze based initiation of contact with an embodied conversational agent. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents* (Amsterdam, the Netherlands, 2009), vol. 5773 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 229–242.
- [BB09] BANF M., BLANZ V.: Example-based rendering of eye movements. *Computer Graphics Forum* 28, 2 (2009), 659–666.
- [BBBL01] BAIENSON J. N., BLASCOVICH J., BEALL A. C., LOOMIS J. M.: Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments* 10, 6 (2001), 583–598.
- [BBL02] BLASCOVICH J., BEALL A. C., LOOMIS J. M., BAIENSON J. N.: Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin* 29 (2002), 1–15.
- [BBN\*14] BÉRARD P., BRADLEY D., NITTI M., BEELER T., GROSS M.: High-quality capture of eyes. *ACM Transactions on Graphics* 33, 6 (2014), 223:1–223:12.
- [BBW96] BURGOON J., BULLER D., WOODALL W.: *Nonverbal Communication: The Unspoken Dialogue*. McGraw-Hill Higher Education, New York, USA, 1996.
- [BCCDR11] BONINO D., CASTELLINA E., CORNO F., DERUSSIS L.: DOGeye: Controlling your home with eye interaction. *Interacting with Computers* 23, 5 (2011), 484–498.
- [BCS75] BAHILL A., CLARK M. R., STARK L.: The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* 24, 3–4 (1975), 191–204.
- [BCWJ97] BARON-COHEN S., WHEELWRIGHT S., JOLLIFFE T.: Is there a ‘language of the eyes’? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition* 4, 3 (1997), 311–331.
- [BDG\*07] BUSSO C., DENG Z., GRIMM M., NEUMANN U., NARAYANAN S.: Rigid head motion in expressive speech animation: Analysis and synthesis. *IEEE Transactions on Audio, Speech, and Language Processing* 15, 3 (2007), 1075–1086.
- [BDNN07] BUSSO C., DENG Z., NEUMANN U., NARAYANAN S.: Learning expressive human-like head motion sequences from speech. In *Data-Driven 3D Facial Animation*. Springer-Verlag London Limited (2007), pp. 113–131.
- [Bec89] BECKER W.: The neurobiology of saccadic eye movements. *Metrics. Reviews of Oculomotor Research* 3 (1989), 13–67.
- [Bev13] BEVACQUA E.: A survey of listener behavior and listener models for embodied conversational agents. In *Coverbal Synchrony in Human-Machine Interaction*. M. Rojc and N. Campbell (Eds.). (Boca Raton, FL, USA, 2013), CRC Press, pp. 243–268.
- [BF88] BECKER W., FUCHS A. F.: Lid-eye coordination during vertical gaze changes in man and monkey. *Journal of Neurophysiology* 60, 4 (1988), 1227–52.
- [BFA09] BEE N., FRANKE S., ANDRÉ E.: Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents. In *Proceedings of 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009* (Amsterdam, the Netherlands, 2009), IEEE, pp. 1–7.
- [BFJ\*05] BENNEWITZ M., FABER F., JOHO D., SCHREIBER M., BEHNKE S.: Integrating vision and speech for conversations with multiple persons. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Edmonton, Canada, 2005), IEEE, pp. 2523–2528.
- [BG79] BANDLER R., GRINDER J.: *Frogs into Princes: Neuro Linguistic Programming*. Real People Press, Boulder, CO, USA, 1979.
- [BH10] BOHUS D., HORVITZ E.: Facilitating multiparty dialog with gaze, gesture, and speech. In *ICMI-MLMI '10: Proceedings of International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction* (Beijing, China, 2010), ACM, pp. 5:1–5:8.
- [BHA09] BARD E., HILL R., ARAI M., FOSTER M. E.: Referring and gaze alignment: Accessibility is alive and well in situated dialogue. In *Proceedings of CogSci 2009* (Amsterdam, the Netherlands, 2009), Cognitive Science Society, pp. 1246–1251.
- [BHKS11] BAINBRIDGE W., HART J., KIM E., SCASSELLATI B.: The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52.
- [BHZS12] BRENNAN S. E., HANNA J. E., ZELINSKY G. J., SAVIETTA K. J.: Eye gaze cues for coordination in collaborative tasks. In *Proceedings of 2012 ACM Conference on Computer Supported Cooperative Work DUET 2012 Workshop: Dual Eye Tracking in CSCE* (Seattle, WA, 2012), ACM.
- [BKG11] BUHRMESTER M., KWANG T., GOSLING S. D.: Amazon’s mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6, 1 (2011), 3–5.
- [BMEL08] BRADLEY M. M., MICCOLI L., ESCRIG M. A., LANG P. J.: The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 4 (2008), 602–607.

- [BPAW10] BEE N., POLLOCK C., ANDRÉ E., WALKER M.: Bossy or wimpy: Expressing social dominance by combining gaze and linguistic behaviors. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents* (Philadelphia, PA, USA, 2010), vol. 6356 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 265–271.
- [BPG\*06] BRETON G., PELÉ D., GARCIA C., PANAGET F., BRETIER P.: Modeling gaze behavior for a 3D ECA in a dialogue situation. In *Gesture in Human-Computer Interaction and Simulation* (Berder Island, France, 2006), vol. 3881 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 252–255.
- [BWA\*10] BEE N., WAGNER J., ANDRÉ E., VOGT T., CHARLES F., PIZZI D., CAVAZZA M.: Gaze behavior during interaction with a virtual character in interactive storytelling. In *Proceedings of International Workshop on Interacting with ECAs as Virtual Characters (AAMAS 2010)* (Toronto, Canada, 2010), AAMAS.
- [Car78] CARY M. S.: The role of gaze in the initiation of conversation. *Social Psychology* 41, 3 (1978), 269–271.
- [CBCA07] CHIN C. A., BARRETO A., CREMADES G., ADJOUADI M.: Performance analysis of an integrated eye gaze tracking/Electromyogram cursor control system. In *Assets '07: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility* (Tempe, AZ, USA, 2007), ACM, pp. 233–234.
- [CBK\*14] CORRIGAN L. J., BASEDOW C., KÜSTER D., KAPPAS A., PETERS C., CASTELLANO G.: Mixing implicit and explicit probes: Finding a ground truth for engagement in social human-robot interactions. In *HRI '14: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany, 2014), ACM, pp. 140–141.
- [CCD00] COHEN M. F., COLBURN R. A., DRUCKER S. M.: The role of eye gaze in avatar mediated conversational interfaces. *Sketches and Applications, SIGGRAPH '00* (2000).
- [CGV09] CAFARO A., GAITO R., VILHJÁLMSOHN H. H.: Animating idle gaze in public places. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents* (Amsterdam, the Netherlands, 2009), vol. 5773 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 250–256.
- [Cho92] CHOVIL N.: Discourse-oriented facial displays in conversation. *Research on Language & Social Interaction* 25 (1991/1992), 163–194.
- [CK03] CLARK H. H., KRYCH M. A.: Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 50, 1 (2003), 62–81.
- [CKB99] CHOPRA-KHULLAR S., BADLER N. I.: Where to look? Automating attending behaviors of virtual human characters. In *AGENTS '99: Proceedings of the Third Annual Conference on Autonomous Agents* (Seattle, WA, USA, 1999), ACM, pp. 16–23.
- [CKEMT10] CIG C., KASAP Z., EGGES A., MAGNENAT-THALMANN N.: Realistic emotional gaze and head behavior generation based on arousal and dominance factors. In *Motion in Games* (Utrecht, the Netherlands, 2010), vol. 6459 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 278–289.
- [CPB\*94] CASSELL J., PELACHAUD C., BADLER N., STEEDMAN M., ACHORN B., BECKET T., DOUVILLE B., PREVOST S., STONE M.: Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (Orlando, FL, USA, 1994), ACM, pp. 413–420.
- [CPC13] CORRIGAN L. J., PETERS C., CASTELLANO G.: Identifying task engagement: Towards personalised interactions with educational robots. In *Proceedings of 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)* (Geneva, Switzerland, 2013), IEEE, pp. 655–658.
- [CT99] CASSELL J., THÓRISSON K. R.: The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13, 4–5 (1999), 519–538.
- [CT07] CASSELL J., TARTARO A.: Intersubjectivity in human-agent interaction. *Interaction Studies* 8, 3 (2007), 391–410.
- [CTP98] CASSELL J., TORRES O. E., PREVOST S.: Turn taking vs. discourse structure: How best to model multimodal conversation. In *Machine Conversations* (1998), Kluwer, The Hague, pp. 143–154.
- [CV99] CASSELL J., VILHJÁLMSOHN H. H.: Fully embodied conversational avatars: making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems* 2, 1 (1999), 45–64.
- [CVB01] CASSELL J., VILHJÁLMSOHN H. H., BICKMORE T.: BEAT: The behavior expression animation toolkit. In *SIGGRAPH '01: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (Los Angeles, CA, USA, 2001), ACM, pp. 477–486.
- [CVB\*12] CAFARO A., VILHJÁLMSOHN H. H., BICKMORE T., HEYLEN D., JÓHANNSDÓTTIR K. R., VALGARSSON G. S.: First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents* (Santa Cruz, CA, USA, 2012), vol. 7502 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 67–80.
- [DAH12] DYBDAL M. L., AGUSTIN J. S., HANSEN J. P.: Gaze input for mobile devices by dwell and gestures. In *ETRA '12: Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, CA, USA, 2012), ACM, pp. 225–228.
- [DCPPS04] DECAROLIS B., PELACHAUD C., POGGI I., STEEDMAN M.: APMML, a markup language for believable behavior generation. In *Life-Like Characters*. Springer Berlin Heidelberg, 2004, pp. 65–85.

- [DdGB10] DELAUNAY F., DEGREEFF J., BELPAEME T.: A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *HRI '10: Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (Osaka, Japan, 2010), IEEE, pp. 39–44.
- [DLN05] DENG Z., LEWIS J. P., NEUMANN U.: Automated eye motion using texture synthesis. *IEEE Computer Graphics and Applications* 25, 2 (2005), 24–30.
- [DLN07] DENG Z., LEWIS J., NEUMANN U.: Realistic eye motion synthesis by texture synthesis. In *Data-Driven 3D Facial Animation*. Z. Deng and U. Neumann U (Eds.). Springer-Verlag London Limited (2007), pp. 98–112.
- [Dou01] DOUGHTY M. J.: Consideration of three types of spontaneous eyeblink activity in normal humans: during reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science* 78, 10 (2001), 712–725.
- [DRKK14] DAS D., RASHED M. G., KOBAYASHI Y., KUNO Y.: Recognizing gaze pattern for human robot interaction. In *HRI '14: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany, 2014), ACM, pp. 142–143.
- [DS07] DREWES H., SCHMIDT A.: Interacting with the computer using gaze gestures. In *Human-Computer Interaction – INTERACT 2007* (Rio de Janeiro, Brazil, 2007), vol. 4663 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 475–488.
- [Dun74] DUNCAN STARKEY J.: On the structure of speaker-auditor interaction during speaking turns. *Language in Society* 3, 2 (1974), 161–180.
- [EBCR\*07] ELISEI F., BAILLY G., CASARI A., RAIDT S.: Towards eyegaze-aware analysis and synthesis of audiovisual speech. In *ISCA: Proceedings of International Conference on Auditory-Visual Speech Processing, AVSP 2007* (Kasteel Groenendaal, Hilvarenbeek, the Netherlands, 2007), pp. 50–56.
- [EF03] EKMAN P., FRIESEN W.: *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Malor Books, Oxford, UK, 2003.
- [EMP\*94] EVINGER C., MANNING K. A., PELLEGRINI J. J., BASSO M. A., POWERS A. S., SIBONY P. A.: Not looking while leaping: The linkage of blinking and saccadic gazeshifts. *Experimental Brain Research* 100, 2 (1994), 337–344.
- [EMS91] EVINGER C., MANNING K. A., SIBONY P. A.: Eyelid movements. Mechanisms and normal data. *Investigative Ophthalmology & Visual Science* 32, 2 (1991), 387–400.
- [EPAI07] EICHNER T., PRENDINGER H., ANDRÉ E., ISHIZUKA M.: Attentive presentation agents. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 283–295.
- [EPM08] EKMAN I. M., POIKOLA A. W., MÄKÄRÄINEN M. K.: Invisible Eni: Using gaze and pupil size to control a game. In *CHI EA '08: Extended Abstracts on Human Factors in Computing Systems* (Florence, Italy, 2008), ACM, pp. 3135–3140.
- [FBT07] FRISCHEN A., BAYLISS A. P., TIPPER S. P.: Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin* 133, 4 (2007), 694–724.
- [FGBB09] FRANÇOIS G., GAUTRON P., BRETON G., BOUATOUCH K.: Image-based modeling of the human eye. *IEEE Transactions on Visualization and Computer Graphics* 15, 5 (2009), 815–827.
- [FN11] FUKUHARA Y., NAKANO Y.: Gaze and conversation dominance in multiparty interaction. In *Proceedings of 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction* (Palo Alto, CA, USA, 2011).
- [FOM\*02] FUKAYAMA A., OHNO T., MUKAWA N., SAWAKI M., HAGITA N.: Messages embedded in gaze of interface agents—Impression management with agent's gaze. In *CHI '02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, MN, USA, 2002), ACM, pp. 41–48.
- [FR84] FISCHER B., RAMSPERGER E.: Human express saccades: Extremely short reaction times of goal directed eye movements. *Experimental Brain Research* 57, 1 (1984), 191–195.
- [FTT99] FUNGE J., TU X., TERZOPOULOS D.: Cognitive modeling: Knowledge, reasoning and planning for intelligent characters. In *SIGGRAPH '99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (Los Angeles, CA, USA, 1999), ACM Press/Addison-Wesley Publishing Co, pp. 29–38.
- [Ful92] FULLER J.: Head movement propensity. *Experimental Brain Research* 92, 1 (1992), 152–164.
- [GB06] GU E., BADLER N. I.: Visual attention and eye gaze during multiparty conversations with distractions. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents* (Marina Del Rey, CA, 2006), vol. 4133 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 193–204.
- [GC11] GERGLE D., CLARK A. T.: See what I'm saying? Using dyadic mobile eye tracking to study collaborative reference. In *CSCW '11: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (Hangzhou, China, 2011), ACM Press, p. 435.
- [GD02] GILLIES M. F. P., DODGSON N. A.: Eye movements and attention for behavioural animation. *Journal of Visualization and Computer Animation* 13, 5 (2002), 287–300.
- [GK05] GRIMM M., KROSCHEL K.: Evaluation of natural emotions using self assessment manikins. In *Proceedings of 2005 IEEE Workshop on Automatic Speech Recognition and Understanding* (Cancún, Mexico, 2005), IEEE, pp. 381–385.

- [GLBB07] GU E., LEE S., BADLER J. B., BADLER N. I.: Eye movements, saccades, and multiparty conversations. In *Data-Driven 3D Facial Animation*. Springer-Verlag London Limited (2007), pp. 79–97.
- [Gle13] GLEBAS F.: *The Animator's Eye: Composition and Design for Better Animation*. Taylor & Francis, Abingdon, OX, UK, 2013.
- [Gof63] GOFFMAN E.: *Behaviour in Public Places: Notes on the Social Order of Gatherings*. The Free Press, New York, 1963.
- [GP09] GATICA-PEREZ D.: Modeling interest in face-to-face conversations from multimodal nonverbal behaviour. In *Multi-Modal Signal Processing: Methods and Techniques to Build Multimodal Interactive Systems*, J. Thiran, H. Bourlard and F. Marqués (Eds.). EURASIP and Academic Press Series in Signal and Image Processing. (Amsterdam, Netherlands, 2009), Elsevier B.V., pp. 309–326.
- [Gre86] GREENE P. R.: Gaussian and Poisson blink statistics: A preliminary study. *IEEE Transactions on Biomedical Engineering* 33, 3 (1986), 359–361.
- [GRS03] GOSLING S. D., RENTFROW P. J., SWANN W. B.: A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37, 6 (2003), 504–528.
- [GSC91] GUITTON D., SIMARD R., CODÈRE F.: Upper eyelid movements measured with a search coil during blinks and vertical saccades. *Investigative Ophthalmology & Visual Science* 32, 13 (1991), 3298–3305.
- [GSV\*03] GARAU M., SLATER M., VINAYAGAMOORTHY V., BROGNI A., STEED A., SASSE M. A.: The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, FL, USA, 2003), ACM, pp. 529–536.
- [GT09] GRILLON H., THALMANN D.: Simulating gaze attention behaviors for crowds. *Computer Animation and Virtual Worlds* 20, 2 (2009), 111–119.
- [GV87] GUITTON D., VOLLE M.: Gaze control in humans: Eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology* 58, 3 (1987), 427–459.
- [GWG\*07] GRATCH J., WANG N., GERTEN J., FAST E., DUFFY R.: Creating rapport with virtual agents. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 125–138.
- [HAMM14] HANSEN J. P., ALAPETITE A., MACKENZIE I. S., MØLLENBACH E.: The use of gaze to control drones. In *ETRA '14: Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, FL, USA, 2014), ACM, pp. 27–34.
- [HB07] HANNA J. E., BRENNAN S. E.: Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language* 57, 4 (2007), 596–615.
- [Hei13a] HEIKKILÄ H.: EyeSketch: A drawing application for gaze control. In *ETSA '13: Proceedings of the 2013 Conference on Eye Tracking South Africa* (Cape Town, South Africa, 2013), ACM, pp. 71–74.
- [Hei13b] HEIKKILÄ H.: Tools for a gaze-controlled drawing application—Comparing gaze gestures against dwell buttons. In *Proceedings of Human-Computer Interaction - INTERACT 2013* (Cape Town, South Africa, 2013), vol. 8118 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 187–201.
- [Hey06] HEYLEN D.: Head gestures, gaze and the principles of conversational structure. *International journal of Humanoid Robotics* 3, 3 (2006), 241–267.
- [HHA\*88] HARRIS C. M., HAINLINE L., ABRAMOV I., LEMERISE E., CAMENZULI C.: The distribution of fixation durations in infants and naive adults. *Vision Research* 28, 3 (1988), 419–432.
- [HJ96] HENDRIKS-JANSEN H.: *Catching Ourselves in the Act: Situated Activity, Interactive Emergence, Evolution, and Human Thought*. A Bradford book. MIT Press, Cambridge, MA, USA, 1996.
- [HJO\*10] HODGINS J., JÖRG S., O'SULLIVAN C., PARK S. I., MAHLER M.: The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception* 7, 4 (2010), 22:1–22:14.
- [HM06] HANES D. A., MCCOLLUM G.: Variables contributing to the coordination of rapid eye/head gaze shifts. *Biological Cybernetics* 94, 4 (2006), 300–324.
- [HM13] HUANG C.-M., MUTLU B.: Modeling and evaluating narrative gestures for humanlike robots. In *Proceedings of Robotics: Science and Systems* (Berlin, Germany, 2013), pp. 26–32.
- [HNP07] HEYLEN D., NIJHOLT A., POEL M.: Generating nonverbal signals for a sensitive artificial listener. In *Verbal and Nonverbal Communication Behaviours* (Vetri sul Mare, Italy, 2007), vol. 4775 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 264–274.
- [HO12] HJALMARSSON A., OERTEL C.: Gaze direction as a Back-Channel inviting cue in dialogue. In *Proceedings of IVA 2012 Workshop on Realtime Conversational Virtual Agents* (Santa Cruz, CA, USA, 2012).
- [Hoo13] HOOKS E.: *Acting for Animators*. Taylor & Francis, Abingdon, OX, UK, 2013.
- [HPB\*07] HOEKSTRA A., PRENDINGER H., BEE N., HEYLEN D., ISHIZUKA M.: Highly realistic 3D presentation agents with visual attention capability. In *Smart Graphics* (Kyoto, Japan, 2007), vol. 4569 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 73–84.

- [HR12] HEIKKILÄ H., RÄIHÄ K.-J.: Simple gaze gestures and the closure of the eyes as an interaction technique. In *ETRA '12: Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, CA, USA, 2012), ACM, pp. 147–154.
- [ID03] ITTI L., DHAVALÉ N.: Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proceedings of SPIE's 48th Annual Meeting Optical Science and Technology*, (San Diego, CA, USA, 2003), International Society for Optics and Photonics, SPIE Press, pp. 64–78.
- [IDP06] ITTI L., DHAVALÉ N., PIGHIN F. H.: Photorealistic attention-based gaze animation. In *Proceedings of IEEE International Conference on Multimedia and Expo* (Toronto, Ontario, Canada, 2006), IEEE, pp. 521–524.
- [IHI\*10] ISTANCE H., HYRSKYKARI A., IMMONEN L., MANSIKKAMAA S., VICKERS S.: Designing gaze gestures for gaming: An investigation of performance. In *ETRA '10: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, TX, USA, 2010), ACM, pp. 323–330.
- [IMFN06] ISHII R., MIYAJIMA T., FUJITA K., NAKANO Y.: Avatar's gaze control to facilitate conversational turn-taking in virtual-space multi-user voice chat system. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents* (Marina Del Rey, CA, USA, 2006), vol. 4133 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 458–458.
- [IONN13] ISHII R., OOKO R., NAKANO Y., NISHIDA T.: Effectiveness of gaze-based engagement estimation in conversational agents. In *Eye Gaze in Intelligent User Interfaces*. Y. I. Nakano, C. Conati and T. Bader (Eds.). Springer-Verlag London (2013), pp. 85–110.
- [Itt00] ITTI L.: *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of Technology, 2000.
- [Iza91] IZARD C.: *The Psychology of Emotions*. Emotions, Personality, and Psychotherapy. Springer US, Berlin, Heidelberg, Germany, 1991.
- [JHM\*07] JAN D., HERRERA D., MARTINOVSKI B., NOVICK D. G., TRAUM D. R.: A computational model of culture-specific conversational behavior. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 45–56.
- [JPMG13] JOWERS I., PRATS M., MCKAY A., GARNER S.: Evaluating an eye tracking interface for a two-dimensional sketch editor. *Computer-Aided Design* 45, 5 (2013), 923–936.
- [JTC\*07] JOHNS M., TUCKER A., CHAPMAN R., CROWLEY K., MICHAEL N.: Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie - Schlaforschung und Schlafmedizin* 11, 4 (2007), 234–242.
- [KAD11] KIRCHNER N., ALEMPIJEVIC A., DISSANAYAKE G.: Nonverbal robot-group interaction using an imitated gaze cue. In *HRI '11: Proceedings of the 6th International Conference on Human-Robot Interaction* (Lausanne, Switzerland, 2011), ACM, pp. 497–504.
- [KB04] KIDD C., BREAZEAL C.: Effect of a robot on user perceptions. In *Proceedings of IROS 2004* (Sendai, Japan, 2004), vol. 4, IEEE, pp. 3559–3564.
- [Ken67] KENDON A.: Some functions of gaze-direction in social interaction. *Acta Psychologica* 26, 1 (1967), 22–63.
- [Ken90] KENDON A.: *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Studies in Interactional Sociolinguistics. Cambridge University Press, Cambridge, UK, 1990.
- [KG08] KIPP M., GEBHARD P.: IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents* (Tokyo, Japan, 2008), vol. 5208 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 191–199.
- [KHJK13] KOMOGORTSEV O., HOLLAND C., JAYARATHNA S., KARPOV A.: 2D linear oculomotor plant mathematical model: Verification and biometric applications. *ACM Transactions on Applied Perception* 10, 4 (2013), 27:1–27:18.
- [KK13] KULMS P., KOPP S.: Using virtual agents to guide attention in multi-task scenarios. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents* (Edinburgh, UK, 2013), vol. 8108 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 295–302.
- [KKGK11] KULMS P., KRÄMER N. C., GRATCH J., KANG S.-H.: It's in their eyes: A study on female and male virtual humans' gaze. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents* (Reykjavik, Iceland, 2011), vol. 6895 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 80–92.
- [KKM\*06] KOPP S., KRENN B., MARSELLA S., MARSHALL A. N., PELACHAUD C., PIRKER H., THÓRISSON K. R., VILHJÁLMSOHN H. H.: Towards a common framework for multimodal generation: The behavior markup language. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents* (Marina Del Rey, CA, USA, 2006), vol. 4133 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 205–217.
- [Kle86] KLEINKE C. L.: Gaze and eye contact: A research review. *Psychological Bulletin* 100, 1 (1986), 78–100.
- [KOS11] KOKKINARA E., OYEKOYA O., STEED A.: Modelling selective visual attention for autonomous virtual characters. *Computer Animation and Virtual Worlds* 22, 4 (2011), 361–369.
- [KPPP11] KRENN B., PELACHAUD C., PIRKER H., PETERS C.: Embodied conversational characters: Representation formats for multimodal communicative behaviours. In *Emotion-Oriented Systems*. R. Cowie, C. Pelachaud and P. Petta (Eds.). Cognitive Technologies. Springer Berlin Heidelberg, 2011, pp. 389–415.

- [KPW07] KUMAR M., PAEPCKE A., WINOGRAD T.: EyePoint: Practical pointing and selection using gaze and keyboard. In *CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, CA, USA, 2007), ACM, pp. 421–430.
- [KV12] KRISTENSSON P. O., VERTANEN K.: The potential of dwell-free eye-typing for fast assistive gaze communication. In *ETRA '12: Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, CA, USA, 2012), ACM, pp. 241–244.
- [KYH\*98] KIKUCHI H., YOKOYAMA M., HOASHI K., HIDAKI Y., KOBAYASHI T., SHIRAI K.: Controlling gaze of humanoid in communication with human. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (Victoria, B.C., Canada, 1998), vol. 1, IEEE, pp. 255–260.
- [Las01] LASSETER J.: Tricks to animating characters with a computer. *SIGGRAPH Computer Graphics* 35, 2 (2001), 45–47.
- [LB06] LAM M. W., BARANOSKI G. V.: A predictive light transport model for the human iris. *Computer Graphics Forum* 25, 3 (2006), 359–368.
- [LBB02] LEE S. P., BADLER J. B., BADLER N. I.: Eyes alive. In *SIGGRAPH '02: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques* (San Antonio, TX, USA, 2002), ACM, pp. 637–644.
- [LBS\*03] LEFOHN A., BUDGE B., SHIRLEY P., CARUSO R., REINHARD E.: An ocularist's approach to human iris synthesis. *IEEE Computer Graphics and Applications* 23, 6 (2003), 70–75.
- [Lik32] LIKERT R.: A technique for the measurement of attitudes. *Archives of Psychology* 22, 140 (1932), 1–55.
- [LM07] LANCE B., MARSELLA S.: Emotionally expressive head and body movement during gaze shifts. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 72–85.
- [LM10a] LANCE B., MARSELLA S.: Glances, glares, and glowering: How should a virtual human express emotion through gaze? *Autonomous Agents and Multi-Agent Systems* 20, 1 (2010), 50–69.
- [LM10b] LANCE B., MARSELLA S.: The expressive gaze model: Using gaze to express emotion. *IEEE Computer Graphics and Applications* 30, 4 (2010), 62–73.
- [LM12] LI Z., MAO X.: Emotional eye movement generation based on Geneva emotion wheel for virtual agents. *Journal of Visual Languages & Computing* 23, 5 (2012), 299–310.
- [LMD12] LE B., MA X., DENG Z.: Live speech driven head-and-eye motion generators. *IEEE Transactions on Visualization and Computer Graphics* 18, 11 (2012), 1902–1914.
- [LMK04] LANCE B., MARSELLA S., KOIZUMI D.: Towards expressive gaze manner in embodied virtual agents. In *Proceedings of Autonomous Agents and Multi-Agent Systems Workshop on Empathic Agents* (Budapest, Hungary, 2004), AAMAS.
- [LMT\*07] LEE J., MARSELLA S., TRAUM D., GRATCH J., LANCE B.: The Rickel Gaze model: A window on the mind of a virtual human. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 296–303.
- [LR86] LAURUTIS V. P., ROBINSON D. A.: The vestibulo-ocular reflex during human saccadic eye movements. *The Journal of Physiology* 373, 1 (1986), 209–233.
- [LvW12] LOHSE M., VANWELBERGEN H.: Designing appropriate feedback for virtual agents and robots. In *Position paper at RO-MAN 2012 Workshop 'Robot Feedback in Human-Robot Interaction: How to Make a Robot "Readable" for a Human Interaction Partner'* (Paris, France, 2012), IEEE.
- [LZ99] LEIGH R. J., ZEE D. S.: *The Neurology of Eye Movements* (3rd edition). No. 55 in Contemporary Neurology Series. Oxford University Press, Oxford, England, UK, 1999.
- [Mae06] MAESTRI G.: *Digital Character Animation 3*. Pearson Education, New York, NY, USA, 2006.
- [MB10] McDONNELL R., BREIDT M.: Face reality: Investigating the uncanny valley for virtual faces. In *SA '10: ACM SIGGRAPH ASIA 2010 Sketches* (Seoul, Republic of Korea, 2010), ACM, pp. 41:1–41:2.
- [MB12] MARIOORYAD S., BUSSO C.: Generating human-like behaviors using joint, speech-driven models for conversational agents. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 8 (2012), 2329–2340.
- [MBB12] McDONNELL R., BREIDT M., BÜLTHOFF H. H.: Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics* 31, 4 (2012), 91:1–91:11.
- [MD08] MORENCY L.-P., DARRELL T.: Conditional sequence model for context-based recognition of gaze aversion. In *Machine Learning for Multimodal Interaction* (Brno, Czech Republic, 2008), vol. 4892 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 11–23.
- [MDR\*11] MARTIN J.-C., DEVILLERS L., RAOUZAIYOU A., CARIDAKIS G., RUTTKAY Z., PELACHAUD C., MANCINI M., NIEWIADOMSKI R., PIRKER H., KRENN B., POGGI I., CALDOGNETTO E. M., CAVICCHIO F., MEROLA G., ROJAS A. G., VEXO F., THALMANN D., EGGES A., MAGNENAT-THALMANN N.: Coordinating the generation of signs in multiple modalities in an affective agent. In *Emotion-Oriented Systems*. R. Cowie, C. Pelachaud and P. Petta (Eds.). Cognitive Technologies. Springer Berlin Heidelberg, 2011, pp. 349–367.

- [Meh80] MEHRABIAN A.: *Basic Dimensions for a General Psychological Theory*. OGH Publishers, Cambridge, MA, USA, 1980.
- [MGR04] MARSELLA S., GRATCH J., RICKEL J.: Expressive behaviors for virtual worlds. In *Life-Like Characters*. Cognitive Technologies. Springer Berlin Heidelberg, 2004, pp. 317–360.
- [MH07] MASUKO S., HOSHINO J.: Head-eye animation corresponding to a conversation for CG characters. *Computer Graphics Forum* 26, 3 (2007), 303–312.
- [MHKS07] MITAKE H., HASEGAWA S., KOIKE Y., SATO M.: Reactive virtual human with bottom-up and top-down visual attention for gaze generation in realtime interactions. In *VR '07: Proceedings of IEEE Virtual Reality Conference* (Charlotte, NC, USA, 2007), IEEE, pp. 211–214.
- [MHP12] MARDANBEGI D., HANSEN D. W., PEDERSON T.: Eye-based head gestures. In *ETRA '12: Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, CA, USA, 2012), ACM, pp. 139–146.
- [MLD\*08] McDONNELL R., LARKIN M., DOBBYN S., COLLINS S., O'SULLIVAN C.: Clone attack! Perception of crowd variety. *ACM Transactions on Graphics* 27, 3 (2008), 26:1–26:8.
- [MLGH10] MØLLENBACH E., LILLHOLM M., GAIL A., HANSEN J. P.: Single gaze gestures. In *ETRA '10: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas, 2010), ACM, pp. 177–180.
- [MLH\*09] McDONNELL R., LARKIN M., HERNÁNDEZ B., RUDOMIN I., O'SULLIVAN C.: Eye-catching crowds: Saliency based selective variation. *ACM Transactions on Graphics* 28, 3 (2009), 55:1–55:10.
- [MM11] MUMM J., MUTLU B.: Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *HRI '11: Proceedings of the 6th International Conference on Human-Robot Interaction* (Lausanne, Switzerland, 2011), ACM, pp. 331–338.
- [Mor70] MORI M.: The uncanny valley. *Energy* 7 (1970), 33–35.
- [MRS\*07] MURRAY N., ROBERTS D., STEED A., SHARKEY P., DICKERSON P., RAE J.: An assessment of eye-gaze potential within immersive virtual environments. *ACM Transactions on Multimedia Computing, Communications, and Applications* 3, 4 (2007), 8:1–8:17.
- [MSH\*06] MOJZISCH A., SCHILBACH L., HELMERT J. R., PANNASCH S., VELICHKOVSKY B. M., VOGELY K.: The effects of self-involvement on attention, arousal, and facial expression during social interaction with virtual others: A psychophysiological study. *Social Neuroscience* 1, 3–4 (2006), 184–195.
- [MSK\*09] MUTLU B., SHIWA T., KANDA T., ISHIGURO H., HAGITA N.: Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In *HRI '09: Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction* (La Jolla, CA, USA, 2009), ACM, pp. 61–68.
- [MSSSB10] MARTINEZ S., SLOAN R. J. S., SZYMKOWIAK A., SCOTT-BROWN K. C.: Using virtual agents to cue observer attention. In *Proceedings of CONTENT 2010: The Second International Conference on Creative Content Technologies* (2010), 7–12.
- [MTG\*14] MOON A., TRONIAK D. M., GLEESON B., PAN M. K., ZHENG M., BLUMER B. A., MACLEAN K., CROFT E. A.: Meet me where I'm gazing: How shared attention gaze affects human-robot handover timing. In *HRI '14: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany, 2014), ACM, pp. 334–341.
- [Mur01] MURCH W.: *In the Blink of an Eye: A Perspective on Film Editing*. Silman-James Press, Los Angeles, CA, USA, 2001.
- [MVC\*12] McKEOWN G., VALSTAR M., COWIE R., PANTIC M., SCHRODER M.: The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing* 3, 1 (2012), 5–17.
- [MXL\*13] MARSELLA S., XU Y., LHOMMET M., FENG A., SCHERER S., SHAPIRO A.: Virtual character performance from speech. In *SCA '13: Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Anaheim, CA, USA, 2013), ACM, pp. 25–35.
- [NBF\*13] NORMOYLE A., BADLER J. B., FAN T., BADLER N. I., CASSOL V. J., MUSSE S. R.: Evaluating perceived trust from procedurally animated gaze. In *MIG '13: Proceedings of Motion on Games* (Dublin, Ireland, 2013), ACM, pp. 119:141–119:148.
- [NBHH13] NAKANO Y. I., BABA N., HUANG H.-H., HAYASHI Y.: Implementation and evaluation of a multimodal addressee identification mechanism for multiparty conversation systems. In *ICMI '13: Proceedings of the 15th ACM International Conference on Multimodal Interaction* (Sydney, Australia, 2013), ACM, pp. 35–42.
- [NHP13] NIEWIADOMSKI R., HYNIEWSKA S., PELACHAUD C.: *Computational Models of Expressive Behaviors for a Virtual Agent*. Oxford Series on Cognitive Models and Architecture. OUP USA, 2013.
- [NI10] NAKANO Y. I., ISHII R.: Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *IUI '10: Proceedings of the 15th International Conference on Intelligent User Interfaces* (Hong Kong, China, 2010), ACM, pp. 139–148.
- [NK10] NAKANO T., KITAZAWA S.: Eyeblink entrainment at break-points of speech. *Experimental Brain Research* 205, 4 (2010), 577–581.
- [NKM\*13] NAKANO T., KATO M., MORITO Y., ITOI S., KITAZAWA S.: Blink-related momentary activation of the default mode network while viewing videos. *Proceedings of the National Academy of Sciences* 110, 2 (2013), 702–706.

- [Nor63] NORMAN W. T.: Toward an adequate taxonomy of personality attributes: Replicated factors structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology* 66 (1963), 574–583.
- [ODK\*12] OBAID M., DAMIAN I., KISTLER F., ENDRASS B., WAGNER J., ANDRÉ E.: Cultural Behaviors of Virtual Agents in an Augmented Reality Environment. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents* (Santa Cruz, CA, USA, 2012), vol. 7502 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 412–418.
- [OO80] OTTESON J., OTTESON C.: Effect of teacher's gaze on children's story recall. *Perceptual and Motor Skills* 50, 1 (1980), 35–42.
- [Opp86] OPPENHEIMER P. E.: Real time design and animation of fractal plants and trees. *ACM SIGGRAPH Computer Graphics* 20, 4 (1986), 55–64.
- [Osi10] OSIPA J.: *Stop Staring: Facial Modeling and Animation Done Right*. IT Pro. Wiley, Alameda, CA, USA, 2010.
- [OSP11] OYEKOYA O., STEED A., PAN X.: Exploring the object relevance of a gaze animation model. In *EGVE - JVRC'11: Proceedings of the 17th Eurographics Conference on Virtual Environments & Third Joint Virtual Reality* (Aire-la-Ville, Switzerland, 2011), Eurographics Association, pp. 111–114.
- [OSS09] OYEKOYA O., STEPTOE W., STEED A.: A saliency-based method of simulating visual attention in virtual scenes. In *VRST '09: Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (Kyoto, Japan, 2009), ACM, pp. 199–206.
- [OST57] OSGOOD C. E., SUCI G. J., TANNENBAUM P. H.: *The Measurement of Meaning*. University of Illinois Press, Champaign, IL, USA, 1957.
- [PAK10] PETERS C., ASTERIADIS S., KARPOUZIS K.: Investigating shared attention with a virtual agent using a gaze-based interface. *Journal on Multimodal User Interfaces* 3, 1–2 (2010), 119–130.
- [PB03] PELACHAUD C., BILVI M.: Modelling gaze behavior for conversational agents. In *Proceedings of the 4th International Conference on Intelligent Virtual Agents* (Kloster Irsee, Germany, 2003), vol. 2792 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 93–100.
- [PBER07] PICOT A., BAILLY G., ELISEI F., RAIDT S.: Scrutinizing natural scenes: Controlling the gaze of an embodied conversational agent. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 272–282.
- [PCdF09] PETERS C., CASTELLANO G., DEFREITAS S.: An exploration of user engagement in HCI. In *AFFINE '09: Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots* (Boston, MA, USA, 2009), ACM, pp. 9:1–9:3.
- [PCI10] PAOLACCI G., CHANDLER J., IPEIROTIS P. G.: Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 5 (2010), 411–419.
- [PCR\*11] PETERS C., CASTELLANO G., REHM M., ANDRÉ E., RAOUZAIYOU A., RAPANTZIKOS K., KARPOUZIS K., VOLPE G., CAMURRI A., VASALOU A.: Fundamentals of agent perception and attention modelling. In *Emotion-Oriented Systems*, Cognitive Technologies. Springer Berlin Heidelberg, 2011, pp. 293–319.
- [Pet05] PETERS C.: Direction of attention perception for conversation initiation in virtual environments. In *Proceedings of the 5th International Conference on Intelligent Virtual Agents* (Kos, Greece, 2005), vol. 3661 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 215–228.
- [Pet06] PETERS C.: Evaluating perception of interaction initiation in virtual environments using humanoid agents. In *Proceedings of the 2006 Conference on ECAI 2006: 17th European Conference on Artificial Intelligence* (Riva del Garda, Italy, 2006), IOS Press, pp. 46–50.
- [Pet10] PETERS C.: Animating gaze shifts for virtual characters based on head movement propensity. In *Proceedings of 2010 Second International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)* (Braga, Portugal, 2010), IEEE, pp. 11–18.
- [PHN\*09] POEL M., HEYLEN D., NIJHOLT A., MEULEMANS M., VAN BREEMEN A.: Gaze behaviour, believability, likability and the iCat. *AI & Society* 24, 1 (2009), 61–73.
- [PKFT07] POWERS A., KIESLER S., FUSSELL S., TORREY C.: Comparing a computer agent with a humanoid robot. In *Proceedings of 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Arlington, VA, USA, 2007), IEEE, pp. 145–152.
- [PLPW12] PFEIFFER-LESSMANN N., PFEIFFER T., WACHSMUTH I.: An operational model of joint attention—Timing of the initiate-act in interactions with a virtual human. In *Proceedings of KogWis*, D. Dörner, R. Goebel, M. Oaksford, M. Pauen and E. Stern (Eds). (Bamberg, Germany, 2012), University of Bamberg Press, pp. 96–97.
- [PMG13] PEJSA T., MUTLU B., GLEICHER M.: Stylized and performative gaze for character animation. *Computer Graphics Forum* 32, 2 (2013), 143–152.
- [PO03] PETERS C., O'SULLIVAN C.: Attention-driven eye gaze and blinking for virtual humans. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Sketches & Applications* (San Diego, CA, USA, 2003), ACM, pp. 1–1.
- [POB09] PAMPLONA V. F., OLIVEIRA M. M., BARANOSKI G. V. G.: Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Transactions on Graphics* 28, 4 (2009), 106:1–106:12.
- [POS03] PETERS C., O'SULLIVAN C.: Bottom-up visual attention for virtual human animation. In *CASA '03: Proceedings of the 16th International Conference on Computer Animation and Social*

- Agents (CASA 2003) (New Brunswick, NJ, USA, 2003), IEEE, pp. 111–117.
- [PPB\*05] PETERS C., PELACHAUD C., BEVACQUA E., MANCINI M., POGGI I.: A model of attention and interest using gaze behavior. In *Proceedings of the 5th International Conference on Intelligent Virtual Agents* (Kos, Greece, 2005), vol. 3661 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 229–240.
- [PPDR00] POGGI I., PELACHAUD C., DEROSIS F.: Eye communication in a conversational 3D synthetic agent. *AI Communications* 13, 3 (2000), 169–181.
- [PQ10] PETERS C., QURESHI A.: Graphics for serious games: A head movement propensity model for animating gaze shifts and blinks of virtual characters. *Computers and Graphics* 34, 6 (2010), 677–687.
- [PSA\*04] POURTOIS G., SANDER D., ANDRES M., GRANDJEAN D., REVERET L., OLIVIER E., VUILLEUMIER P.: Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *European Journal of Neuroscience* 20, 12 (2004), 3507–3515.
- [QBM07] QUEIROZ R., BARROS L., MUSSE S.: Automatic generation of expressive gaze in virtual animated characters: From artists craft to a behavioral animation model. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 401–402.
- [QBM08] QUEIROZ R. B., BARROS L. M., MUSSE S. R.: Providing expressive gaze to virtual animated characters in interactive applications. *Computers in Entertainment* 6, 3 (2008), 41:1–41:23.
- [QPA14] QURESHI A., PETERS C., APPERLY I.: How does varying gaze direction affect interaction between a virtual agent and participant in an on-line communication scenario? In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments* (Heraklion, Crete, Greece, 2014), vol. 8525 of *Lecture Notes in Computer Science*, Springer International Publishing Switzerland, pp. 305–316.
- [Rem11] REMINGTON L.: *Clinical Anatomy and Physiology of the Visual System*. Elsevier/Butterworth-Heinemann, Philadelphia, PA, USA, 2011.
- [RMH13] RUIJTEN P. A. M., MIDDEN C. J. H., HAM J.: I didn't know that virtual agent was angry at me: Investigating effects of gaze direction on emotion recognition and evaluation. In *Persuasive Technology* (Sydney, NSW, Australia, 2013), vol. 7822 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 192–197.
- [Rob07] ROBERTS S.: *Character Animation: 2D Skills for Better 3D*. Focal Press Visual Effects and Animation Series. Focal Press, Waltham, MA, USA, 2007.
- [SB12] SCHULMAN D., BICKMORE T.: Changes in verbal and non-verbal conversational behavior in long-term interaction. In *ICMI '12: Proceedings of the 14th ACM International Conference on Multimodal Interaction* (Santa Monica, CA, USA, 2012), ACM, pp. 11–18.
- [SBMH94] SAGAR M. A., BULLIVANT D., MALLINSON G. D., HUNTER P. J.: A virtual environment and model of the eye for surgical simulation. In *SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (Orlando, FL, USA, 1994), ACM, pp. 205–212.
- [SC09] STAUDTE M., CROCKER M.: The effect of robot gaze on processing robot utterances. In *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (Amsterdam, the Netherlands, 2009), Society, Inc.
- [SD12] STELLMACH S., DACHSELT R.: Look & touch: Gaze-supported target acquisition. In *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, TX, USA, 2012), ACM, pp. 2981–2990.
- [SD14] ŚWIRSKI L., DODGSON N.: Rendering synthetic ground truth images for eye tracker evaluation. In *ETRA '14: Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, FL, USA, 2014), ACM, pp. 219–222.
- [SG06] SMITH J. D., GRAHAM T. C. N.: Use of eye movements for video game control. In *ACE '06: Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Hollywood, CA, USA, 2006), ACM.
- [Sha11] SHAPIRO A.: Building a character animation system. In *Motion in Games* (Edinburgh, UK, 2011), vol. 7060 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 98–109.
- [SHO13] SKANTZE G., HJALMARSSON A., OERTEL C.: Exploring the effects of gaze and pauses in situated human-robot interaction. In *Proceedings of the SIGDIAL 2013 Conference* (Metz, France, 2013), Association for Computational Linguistics, pp. 163–172.
- [SII04] SURAKKA V., ILLI M., ISOKOSKI P.: Gazing and frowning as a new human-computer interaction technique. *ACM Transactions on Applied Perception* 1, 1 (2004), 40–56.
- [Sim94] SIMON B.-C.: How to build a baby that can read minds: Cognitive mechanisms in mindreading. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition* 13, 5 (1994), 513–552.
- [SM11] SRINIVASAN V., MURPHY R.: A survey of social gaze. In *Proceedings of 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (Lausanne, Switzerland, 2011), IEEE, pp. 253–254.
- [SNJ\*07] SKOTTE J., NØJGAARD J., JØRGENSEN L., CHRISTENSEN K., SJØGAARD G.: Eye blink frequency during different computer tasks quantified by electrooculography. *European Journal of Applied Physiology* 99, 2 (2007), 113–119.
- [SOS10] STEPTOE W., OYEKOYA O., STEED A.: Eyelid kinematics for virtual characters. *Computer Animation and Virtual Worlds* 21, 3–4 (2010), 161–171.

- [SS08] STEPTOE W., STEED A.: High-fidelity avatar eye-representation. In *VR '08: Proceedings of IEEE Virtual Reality Conference* (Reno, NV, USA, 2008), IEEE, pp. 111–114.
- [SSND11] STELLMACH S., STOBER S., NÜRNBERGER A., DACHSELT R.: Designing gaze-supported multimodal interactions for the exploration of large image collections. In *NGCA '11: Proceedings of the 1st Conference on Novel Gaze-Controlled Applications* (Karlskrona, Sweden, 2011), ACM, pp. 1:1–1:8.
- [Sta99] STAHL J. S.: Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research* 126, 1 (1999), 41–54.
- [SWG84] STERN J. A., WALRATH L. C., GOLDSTEIN R.: The endogenous eyeblink. *Psychophysiology* 21, 1 (1984), 22–33.
- [TAB\*13] TURNER J., ALEXANDER J., BULLING A., SCHMIDT D., GELLERSEN H.: Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Proceedings of Human-Computer Interaction - INTERACT 2013* (Cape Town, South Africa, 2013), vol. 8118 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 170–186.
- [TCMH11] TRUTOIU L. C., CARTER E. J., MATTHEWS I., HODGINS J. K.: Modeling and animating eye blinks. *ACM Transactions on Applied Perception* 8 (2011), 17:1–17:17.
- [TCW\*95] TSOTSOS J. K., CULHANE S. M., WAI W. Y. K., LAI Y., DAVIS N., NUFLO F.: Modeling visual attention via selective tuning. *Artificial Intelligence* 78, 1–2 (1995), 507–545.
- [TJ95] THOMAS F., JOHNSTON O.: *The Illusion of Life: Disney Animation*. Hyperion Press, New York, NY, USA, 1995.
- [TL86] TYCHSEN L., LISBERGER S. G.: Visual motion processing for the initiation of smooth-pursuit eye movements in humans. *Journal of Neurophysiology* 56, 4 (1986), 953–968.
- [TLM09] THIEBAUX M., LANCE B., MARSELLA S.: Real-time expressive gaze animation for virtual humans. In *AAMAS '09: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems* (Budapest, Hungary, 2009), International Foundation for Autonomous Agents and Multiagent Systems, pp. 321–328.
- [TS04] TOMBS S., SILVERMAN I.: Pupillometry: A sexual selection approach. *Evolution and Human Behavior* 25, 4 (2004), 221–228.
- [TT94] TU X., TERZOPOULOS D.: Artificial fishes: Physics, locomotion, perception, behavior. In *SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (Orlando, FL, USA, 1994), ACM, pp. 43–50.
- [VBP11] VALA M., BLANCO G., PAIVA A.: Providing gender to embodied conversational agents. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents* (Reykjavik, Iceland, 2011), vol. 6895 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 148–154.
- [VBR\*03] VANDERWERF F., BRASSINGA P., REITS D., ARAMIDEH M., ONGERBOER DE VISSER B.: Eyelid movements: Behavioral studies of blinking in humans under different stimulus conditions. *Journal of Neurophysiology* 89, 5 (2003), 2784–2796.
- [VC98] VILHJÁLMSSON H. H., CASSELL J.: BodyChat: Autonomous communicative behaviors in avatars. In *AGENTS '98: Proceedings of the Second International Conference on Autonomous Agents* (Minneapolis, MN, USA, 1998), ACM, pp. 269–276.
- [VCC\*07] VILHJÁLMSSON H. H., CANTELMO N., CASSELL J. E., CHAFAI N., KIPP M., KOPP S., MANCINI M., MARSELLA S., MARSHALL A. N., PELACHAUD C., RUTTKAY Z., THÓRISSON K. R., WELBERGEN H., WERF R. J.: The behavior markup language: Recent developments and challenges. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents* (Paris, France, 2007), vol. 4722 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 99–111.
- [vdKS11] VAN DERKAMP J., SUNDSTEDT V.: Gaze and voice controlled drawing. In *NGCA '11: Proceedings of the 1st Conference on Novel Gaze-Controlled Applications* (Karlskrona, Sweden, 2011), ACM, pp. 9:1–9:8.
- [VGSS04] VINAYAGAMOORTHY V., GARAU M., STEED A., SLATER M.: An eye gaze model for dyadic interaction in an immersive virtual environment: Practice and experience. *Computer Graphics Forum* 23, 1 (2004), 1–12.
- [Vil04] VILHJÁLMSSON H. H.: Animating conversation in online games. In *Entertainment Computing - ICEC 2004* (Eindhoven, the Netherlands, 2004), Rauterberg M., (Ed.), vol. 3166 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 139–150.
- [vM12] ŠPAKOV O., MAJARANTA P.: Enhanced gaze interaction using simple head gestures. In *UbiComp '12: Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, PA, USA, 2012), ACM, pp. 705–710.
- [WEP\*08] WILCOX T., EVANS M., PEARCE C., POLLARD N., SUNDSTEDT V.: Gaze and voice based game interaction: The revenge of the killer penguins. In *SIGGRAPH '08: ACM SIGGRAPH 2008 Posters* (Los Angeles, CA, USA, 2008), ACM, pp. 81:1–81:1.
- [WG10] WANG N., GRATCH J.: Don't just stare at me! In *CHI '10: Proceedings of the 28th International Conference on Human Factors in Computing Systems* (Atlanta, GA, USA, 2010), ACM, pp. 1241–1250.
- [Wil09] WILLIAMS R.: *The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators*. Faber & Faber, London, UK, 2009.
- [WLO10] WEISSENFELD A., LIU K., OSTERMANN J.: Video-realistic image-based eye animation via statistically driven state machines. *The Visual Computer: International Journal of Computer Graphics* 26, 9 (2010), 1201–1216.

- [WRS\*07] WOBROCK J. O., RUBINSTEIN J., SAWYER M., DUCHOWSKI A. T., UK L.: Gaze-based creativity not typing but writing: Eye-based text entry using letter-like gestures. In *Proceedings of the 3rd Conference on Communication by Gaze Interaction (CO-GAIN 2007)* (Leicester, UK, 2007).
- [WSG05] WECKER L., SAMAVATI F., GAVRILOVA M.: Iris synthesis: A reverse subdivision application. In *GRAPHITE '05: Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia* (Dunedin, New Zealand, 2005), ACM, pp. 121–125.
- [XLW13] XU Q., LI L., WANG G.: Designing engagement-aware agents for multiparty conversations. In *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France, 2013), ACM, pp. 2233–2242.
- [YC06] YEO A. W., CHIU P.-C.: Gaze estimation model for eye drawing. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montreal, Québec, Canada, 2006), CHI EA '06, ACM, pp. 1559–1564.
- [YHC\*10] YOO B., HAN J.-J., CHOI C., YI K., SUH S., PARK D., KIM C.: 3D user interface combining gaze and hand gestures for large-scale display. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, GA, USA, 2010), CHI EA '10, ACM, pp. 3709–3714.
- [YLN12] YEO S. H., LESMANA M., NEOG D. R., PAI D. K.: Eyecatch: simulating visuomotor coordination for object interception. *ACM Transactions on Graphics* 31, 4 (2012), 42:1–42:10.
- [YNG70] YNGVE V. H.: On getting a word in edgewise. In *Proceedings of Chicago Linguistics Society, 6th Meeting* (1970), Chicago, IL, USA, pp. 567–578.
- [YSI\*06] YOSHIKAWA Y., SHINOZAWA K., ISHIGURO H., HAGITA N., MIYAMOTO T.: Responsive robot gaze to interaction partner. In *Proceedings of Robotics: Science and Systems* (Philadelphia, PA, USA, 2006), IEEE.
- [YSS12] YU C., SCHERMERHORN P., SCHEUTZ M.: Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems* 1, 2 (2012), 13:1–13:25.
- [ZFP11] ZORIC G., FORCHHEIMER R., PANDZIC I.: On creating multimodal virtual humans—Real time speech driven facial gesturing. *Multimedia Tools and Applications* 54, 1 (2011), 165–179.
- [ZHRM13] ZIBREK K., HOYET L., RUHLAND K., MCDONNELL R.: Evaluating the effect of emotion on gender recognition in virtual humans. In *SAP '13: Proceedings of the ACM Symposium on Applied Perception* (Dublin, Ireland, 2013), ACM, pp. 45–49.
- [ZS06] ZUO J., SCHMID N. A.: A model based, anatomy based method for synthesizing iris images. In *ICB '06: Proceedings of the 2006 International Conference on Advances in Biometrics*. Springer Berlin Heidelberg, 2006, pp. 428–435.