Taming Mona Lisa: Communicating Gaze Faithfully in 2D and 3D Facial Projections

SAMER AL MOUBAYED, JENS EDLUND, and JONAS BESKOW,
KTH Royal Institute of Technology

The perception of gaze plays a crucial role in human-human interaction. Gaze has been shown to matter for a number of aspects of communication and dialogue, especially for managing the flow of the dialogue and participant attention, for deictic referencing, and for the communication of attitude. When developing embodied conversational agents (ECAs) and talking heads, modeling and delivering accurate gaze targets is crucial. Traditionally, systems communicating through talking heads have been displayed to the human conversant using 2D displays, such as flat monitors. This approach introduces severe limitations for an accurate communication of gaze since 2D displays are associated with several powerful effects and illusions, most importantly the Mona Lisa gaze effect, where the gaze of the projected head appears to follow the observer regardless of viewing angle. We describe the Mona Lisa gaze effect and its consequences in the interaction loop, and propose a new approach for displaying talking heads using a 3D projection surface (a physical model of a human head) as an alternative to the traditional flat surface projection. We investigate and compare the accuracy of the perception of gaze direction and the Mona Lisa gaze effect in 2D and 3D projection surfaces in a five subject gaze perception experiment. The experiment confirms that a 3D projection surface completely eliminates the Mona Lisa gaze effect and delivers very accurate gaze direction that is independent of the observer's viewing angle. Based on the data collected in this experiment, we rephrase the formulation of the Mona Lisa gaze effect. The data, when reinterpreted, confirms the predictions of the new model for both 2D and 3D projection surfaces. Finally, we discuss the requirements on different spatially interactive systems in terms of gaze direction, and propose new applications and experiments for interaction in a human-ECA and a human-robot settings made possible by this technology.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems; I.3.6 [**Computer Graphics**]: Methodology and Techniques—*Interaction techniques*

General Terms: Human factors

Additional Key Words and Phrases: Gaze perception, 3D projected avatars, embodied conversational agents, Mona Lisa gaze effect, robot head, situated interaction, multiparty dialogue

ACM Reference Format:
Al Moubayed, S., Edlund, J., and Beskow, J. 2012. Taming Mona Lisa: Communicating gaze faithfully in 2D and 3D facial projections. ACM Trans. Interact. Intell. Syst. 1, 2, Article 11 (January 2012), 25 pages.
DOI = 10.1145/2070719.2070724 http://doi.acm.org/10.1145/2070719.2070724

1. INTRODUCTION

The importance of gaze in social interaction is well-established. From a human communication perspective, Kendon's work on gaze direction in conversation [Kendon 1967] is

This work has been done at the Department for Speech, Music, and Hearing.
This work was funded by the EU project IURO (Interactive Urban Robot) No. 248314.
Authors' address: S. Al Moubayed, J. Edlund, and J. Beskow, Department of Speech, Music, and Hearing, School of Computer Science, KTH Royal Institute of Technology, Lindstedtsvägen 24, 10044, Stockholm, Sweden, email: samer.am@kth.se.

DOI 10.1145/2070719.2070724 http://doi.acm.org/10.1145/2070719.2070724

particularly important in inspiring a wealth of studies that singled out gaze as one of the strongest nonvocal cues in human face-to-face interaction [Argyle and Cook 1976]. Gaze has been associated with a variety of functions within social interaction [Kleinke 1986; Abele 1986]. Kleinke's review article [Kleinke 1986], for example, contains the following list: "(a) provide information, (b) regulate interaction, (c) express intimacy, (d) exercise social control, and (e) facilitate service and task goals."

With these findings, an increasing research has focused on the roles of the direction, movement and dynamics of gaze in nonverbal communication and social interaction, this research has shed the light on the significant function of gaze in communication and perception. For example, just to mention few, in a study by Bloom and Erickson [1971], it was found that infants establish purposeful eye-contact at an age as early as 7 mounts, which gives an idea on how important gaze in social interaction.

Waxer [1977] found that gaze movements correlate with emotions (anxiety levels) in patients. Bente et al. [1998] quantified significant differences in gaze movements relating to attention across sex and familiarity of dyads. Frieschen et al. [2007] provide a comprehensive review of gaze cues of attention in the infant, adult and clinical population.

These efforts and findings, in turn, were and are shadowed by an increasing effort in the human-computer interaction community, which recognized the importance of modelling gaze and its social functions such as expressing and communicating attitudes and emotions in artificial personas such as embodied conversational agents (ECAs) [Takeuchi and Nagao 1993; Poggi and Pelachaud 2000; Bilvi and Pelachaud 2003; Lance and Marsella 2010]. This effort comes natural and important while multimodal and facial communication becomes more advanced, plausible, and popular, since these multimodal interfaces are now able to provide a testing and manipulation frameworks for behavioural models of gaze and other non-vocal signals. These artificial agents have recently been effectively used to investigate and quantify the effects of gaze using controlled experiments [Lance and Marsella 2008; Gu and Badler 2006; Edlund and Beskow 2009; Nordenberg et al. 2005].

The bulk of that work just mentioned studies the production of gaze and its function, which is not the focus of the work presented here. Instead, in this work we are interested in a specific aspect of the *perception of gaze direction* which lends importance from the fact that an overwhelming majority of ECAs are either 2D or 3D models rendered on 2D displays. The perception of 2D renditions of 3D scenes is notoriously riddled with artefacts and illusions of many sorts (for an overview, see Gregory 1997). Perhaps the most important of these with respect to communicative gaze behaviors in ECAs is *the Mona Lisa gaze effect* or *the Mona Lisa stare*, commonly described as an effect that makes it appear as if the Mona Lisa's gaze rests steadily on the viewer as the viewer moves through the room (Figure 1).

Although the reference to the Mona Lisa is a modern invention, documentation of the effect dates back at least as far as Ptolemy in around 100AD "[. . .] the image of a face painted on panels follows the gaze of moving viewers to some extent even though there is no motion in the image itself" [Smith 1996]. The effect has since earned frequent enough mention, and a number of more or less detailed explanations have been presented from Ptolemy and onwards [Descartes 1637; Smith 1996; Cuijpers et al. 2010], but to our knowledge, there is no model that explains this effect in a manner that satisfies the requirements of a designer of gaze behaviors.

The fact that the Mona Lisa gaze effect occurs when a face is presented on a 2D display has significant consequences for the use and control of gaze in communication. Such a gaze does not point unambiguously at a point in 3D space. In case there are of multiple observers, each observer shares the same perception, such that the rendition either looks everybody in the eye, or nobody. This results in an inability to establish a

Fig. 1. Leonardo da Vinci's Mona Lisa. Mona Lisa appears to be looking straight at the viewer, regardless of viewing angle. (The painting is in the public domain.)

Table I. Faithful (+) or Unrealistic (−) Mutual Gaze under Different System Capabilities and Application Requirements

		Capabilities				
		Yes		No		**Mona Lisa effect resistant**
		Yes	No	Yes	No	**Head tracking**
Interlocutors	Single	+	−	−	−	
	Multiple	+	−	−	−	

situated eye contact with one particular observer without simultaneously establishing eye contact with all others, which will lead to miscommunication or at the very least risks causing unexpected results if the effect is not exploited deliberately.

Examples where the Mona Lisa gaze effect is relevant are plentiful, especially as the study of situated and multiparty interaction is attracting an increasing amount of interest. In a recent study, Bohus and Horvitz [2010] present a system capable of carrying a physically situated dialogue. In this system, a virtual human receptionist capable of engaging with multiple users is developed, and the system manages attention and dialogue flow using the gaze of an ECA. The system appears to employ a 2D flat screen to interact with the users. This system is presented interacting with two subjects, where subjects may have inferred to whom the ECA was talking from the rotation of the head or the eyeballs, but there is no evidence that exclusive eye contact could be established. To test this, perceived mutual gaze would have to be tracked, and the dialogue would have to involve more than two dialogue partners.

The supposed impact of the Mona Lisa gaze effect can be generalized quite readily. Table I shows how systems which are susceptible to the effect (2D displays) compare to those that are not (real people) regarding their ability to achieve exclusive mutual gaze under different circumstances: two-party conversation vs. multiparty conversation and whether the system has access to head tracking. A distinction is made between *faithful* and *unrealistic* mutual gaze, where faithful mutual gaze means that when an observer is the gaze target, the observer correctly perceives this. When the observer is not the gaze target, the observer correctly perceives this. In other words, the observer can correctly answer the question: Does she look me in the eye? One would expect that a

Fig. 2. I want you for the U.S. Army nearest recruiting station, commissioned by the US federal government and painted by James Montgomery Flagg (cropped in the left pane, and cropped and edited in the right pane). Most viewers feel that both the complete Uncle Sam in the left pane and faceless Uncle Sam in the right pane points straight at them, regardless of viewing angle. (The painting is in the public domain.)

system must be able to both counter the Mona Lisa effect and to know where in the room its interlocutors are in order to establish mutual gaze.

In this work, we model the Mona Lisa gaze effect with the help of a number of relatively uncontroversial assumptions. The model leads to several predictions, which we test using an experimental method which allows us to quantify the Mona Lisa gaze effect. The results show that we can counter the Mona Lisa gaze effect, as well as providing support for the model and validation of the experimental method.

2. WHAT'S BEHIND MONA LISA'S GAZE

The model we propose explains Mona Lisa stare effects as well as other observations with a minimum of complexity, and predicts additional effects which can be tested. The model is based on a number of assumptions, which are described and corroborated in the following, before the model in itself is presented.

Assumption 1

The first assumption is that what causes the Mona Lisa stare effect is more general than eye gaze. An image need not depict eyes or even a face for the effect to occur, as illustrated by Figure 2.

Assumption 2

The second assumption is that 2D images representing 3D objects or scenes are interpreted as having their own virtual 3D space, distinct from physical space, with axes oriented along the horizontal and vertical edges of the image (perceived as width and height, respectively) and the third axis perpendicular to it (perceived as depth). Although effort has been spent in the VR field to describe the relation between the physical and the virtual environment, this relation is by nature ambiguous. Figure 3 serves as an illustration.

Assumption 3

The third assumption is about shape constancy: viewers of 2D images perceive the shapes in the images as invariant, even when the viewing angle changes. The shapes in Figure 4 serve as examples. The phenomenon, known as *shape constancy*, was described

Fig. 3. Interiors of the Winter Palace. The Throne Room of Empress Maria Fiodorovna. The painting gives a clear impression of a large 3D space with a throne located at the far back. If the viewing angle and distance is varied, the throne's position in the virtual space is maintained, and its position in physical space remains unclear. Painting by Yevgraf Fyodorovich Krendovsky. The picture is in the public domain.

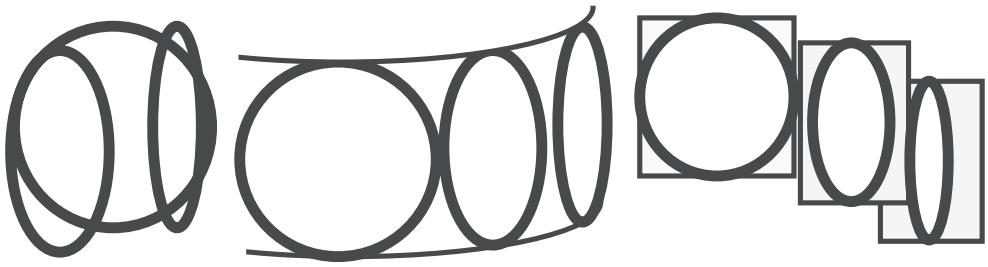

Fig. 4. Three groups of the same three shapes. The left group is easily interpreted as three different shapes drawn on top of each other, while the center and right groups are more easily perceived as identical circular shapes viewed at different angles.

by Descartes as follows: "[. . .]shape is judged by the knowledge, or opinion, that we have of the position of various parts of the objects, and not by the resemblance of the pictures in the eye; for these pictures usually contain only ovals and diamond shapes, yet they cause us to see circles and squares" [Descartes 1637, p. 107]. For a more detailed account of shape constancy, see Gregory [1997].

Assumption 4

The perceived gaze direction within the virtual 3D space of a person depicted within that space is a function of the perceived angle of the gazing person's head within that space, and the angle of her eyes, relative her head. This is based on the observation that in order to calculate (actual) gaze direction, it is not sufficient to know the angle of the eyes relative to the head, which can be estimated for example by means of relative pupil

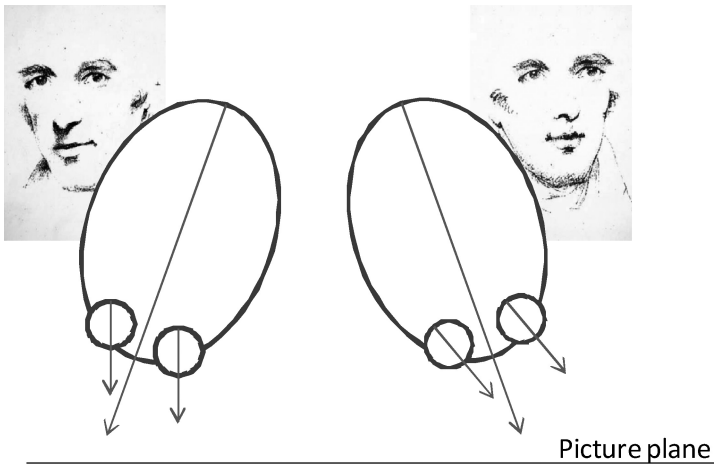

Fig. 5. The Wollaston effect is seen in the two drawings: gaze direction is perceived differently although the eyes are identical, and only the head shape differs. The ovals with two circles represent a possible interpretation of the drawings as seen from above in 3D space. The two drawings are from [Wollaston 1824], and are in the public domain.

position within the sclera [Cuijpers et al. 2010]. An estimation of the position and angle of the head is also required. Although seemingly a trivial observation, the Wollaston "effect" [Wollaston 1824] seems to result from an insistence to view our interpretation of depicted eyes as somehow isolated from the head in which they are lodged. If we, like Todorović [2006], instead assume that head and eyes are interpreted in relation to each other and to the virtual 3D space they are depicted in, the Wollaston effect is not only accounted for, but rather ceases being an effect, as illustrated in Figure 5. Our fourth assumption is somewhat less complex than Todorović's account, as it speaks exclusively of gaze direction in virtual space, whereas Todorović relates eyes and head pose directly to perceived gaze direction in physical space.

Assumption 5

Viewers of 2D images depicting 3D objects interpret their position in relation to the virtual 3D space as head-on, perpendicular to the surface plane of the image. 2D images, at least those that uses perspective to depict a 3D space, are created as seen from some vantage point in front of the objects seen in the picture. In the case of photographs or paintings created using camera obscura, this vantage point can be calculated exactly from the geometry of the image. Paintings allow for artistic licence and may leave more ambiguity, but are still generally interpreted as seen head-on. Again, this is an observation that may seem trivial, but it has bearing as to how we may connect the virtual 3D space depicted in an image to the physical space of our surroundings. It is worth noting that provided that we are standing in front of a picture, interpreting the general orientation and left-right position of the objects depicted in it is often relatively straightforward, whereas deciding the distance to the objects from the imagined vantage point of the creator can pose more of a problem, as illustrated by Figure 6.

Based on these assumptions, we can say although eye and pupil position clearly affects how we perceive gaze direction in general, and may have additional bearing on the Mona Lisa stare effect [Cuijpers et al. 2010], the fundamental causes for the Mona Lisa gaze effect are general to anything with a visible direction, and should be sought elsewhere (assumption 1). We know that the Mona Lisa gaze effect occurs when 3D

Fig. 6. Bremen Town Band, Bremen, Germany. Size variation of objects in 2D depictions is ambiguous and can be interpreted as deriving from at least three sources: the size of the depicted object, the distance and projection from the object to the position from which it is captured, and the size of the actual 2D image. The image in the left pane has been edited to remove the sculpture's surroundings. Without references, it is difficult to judge the size of the sculpture. The right pane contains enough clues that the viewer gets a fair sense of distance and size. Relations between the objects in the image are unaffected: in both panes, it is clear that the rooster is on top of the other animals and that the horse is at the bottom. (The photo is released to the public domain by Adrian Pingstone, who took it in 1990.)

objects are rendered in 2D, for example in painting or on monitors. Assumption 2 gives us that such objects are interpreted as inhibiting their own virtual space, with its own coordinate system, which aligns to our physical space such that up, down, right and left are mapped to the surface plane of the 2D surface and depth is perpendicular to it. Assumption 4 tells us that we should be able to correctly perceive gaze direction of heads projected within such a virtual space, and assumption 2, again, tells us how these perceived gaze directions are mapped in case the viewer is standing right in front of the 2D display. Finally, assumption 5 tells us that regardless of our position relative such a display, we reinterpret the display to what it would look like in case we were standing directly in front of it, and assumption 3 tells us that this can indeed be done, through the phenomenon known as shape constancy.

From this, we get the following predictions.

If a viewer perceives a head as being present in a virtual space displayed through a flat "window" (the frame of a painting or edges of a monitor), the viewer will perceive the gaze direction as if standing directly in front of the window, regardless of actual position in the room. This means that regardless of viewing angle, any gaze that is directed straight out of the picture at an angle perpendicular to its surface will be perceived as being directed at the viewer (as is the case with the Mona Lisa). If on the other hand the head is perceived as copresent—that is, present in the same space as the viewer—the effect will not take place.

A viewer of a head that is interpreted as present in a virtual space is not unable to judge gaze direction within that space. The depth of the virtual space is then aligned with a line perpendicular to the painting or monitor in physical space, and up/down with its left and right edges and right/left with its top and bottom edges. This means that regardless of viewing angle any gaze directed to the left or right in the picture will be perceived as being directed at something to the left or right of the viewer,

Fig. 7. The technical setup: the physical model of a human head used as a 3D projection surface, to the right; the laser projector in the middle; and a snapshot of the 3D talking head to the left.

respectively. It is also likely that viewers can judge how far to the left or right of them a gaze is directed, and that this judgement will be unaffected by viewing angle. However, given that the distance from the vantage point of the creator of the image to an object in the image is ambiguous, the absolute target gazed at is also ambiguous unless the gaze is directed straight out of the picture. Viewers of 2D renditions, then, would be able to judge gaze angle relative the monitor or frame (and interpret this as gaze angle relative themselves) than gaze target when the target is not the viewer, as angle is independent of distance.

3. METHOD

Our basic assumptions suggest that the Mona Lisa gaze effect is introduced by 2D projection surfaces, which are interpreted as a window onto a virtual space. We therefore look for an alternative to 2D projection surfaces, with which the Mona Lisa gaze effect would be avoided. Our approach is to use 3D projection surfaces, and for the present work we use a 3D physical, static model of a human head (as seen to the left in Figure 7). In order to compare this with a traditional 2D projection surface, we designed an experimental paradigm that tests for mutual gaze as well as for gaze direction in the physical space of the viewer. The method is used to test the predictions from our model of the Mona Lisa gaze effect.

3.1. Projecting an Animated Talking Head onto a 3D Surface

The technique of manipulating static objects with light is commonly referred to as the *Shader Lamps* technique [Raskar et al. 2001; Lincoln et al. 2009]. This technique is used to change the physical appearance of still objects by illuminating them using projections of static or animated textures, or video streams. We implement this technique by projecting the animated talking head (seen to the left in Figure 7) onto an arbitrary physical model of a human head (seen to the right in Figure 7) using a laser micro projector (SHOWWX^{TM} Pico Projector, seen in the centre of Figure 7).

The main advantage of using a laser projector is that the all of the image is in focus, even on curved surfaces. However, a limitation of using micro laser projectors instead of, for example, LED or DLP projectors, is that, until today, their brightness is still considerably low (the SHOWWX^{TM} has a brightness of 10 Lumens, where an average LED micro projector is $\sim$50 Lumens, and an average DLP projector is $\sim$2500 Lumens). However, micro projectors are very light weight compared to large projectors (the SHOWWX^{TM} weights 122 grams). Moreover, their small size allows for a setup where the projector is attached to the head, and would move with it if the head is attached to a robotic neck.

One issue that warrants clarification about this simple projection system is that it projects a moving image onto a physical model, which makes the physical model as if

Fig. 8. A physical model of a human head, without projection (left) and complete with a projection of the talking head, a furry hat, and a camera (right).

it is a moving sculpture that is physically placed, and independent from the observer (if the observer moves around the physical object, the viewed perspective will change accordingly). This setup is fundamentally different from stereoscopic 3D systems (or S3D), such as those popular in motion pictures, where an illusion of a 3D scene (an illusion of depth perception) is made using two 2D images, each projected on one eye. This typical stereoscopic projection still allows for an identical perception of the image by observers independently from their viewing point (which basically allows the audience to be seated anywhere in the theatre while they guarantee the same perception of the movie).

In the setup of the experiments described here, the projector was connected to a computer, and placed in front of the physical head model (as shown in the top-left corner of Figure 11). The image of the talking head was then transmitted via the projector and projected onto the physical head.

The talking head used in the studies is detailed in [Beskow 2003] and includes a face, eyes, tongue, and teeth. Figure 8 shows the 3D projection surface with and without a projection of the talking head.

3.2. Experimental Setup

The experiment setup employed a set of subjects simultaneously seated on a circle centred at the stimulus point—a 2D or 3D projection surface—and facing the stimuli point. Adjacent subjects were equidistant to each other and all subjects were equidistant to the projection surface, in such a manner that the angle between two adjacent subjects and the projection surface was always 26.5 degrees. The positions were annotated as $\{-53, -26.5, 0, 26.5, 53\}$, where 0 was the seat directly in front of the projection surface. The distance from subjects to the projection surface was 1.80 meters (x in Figure 9).

Two identical sets of stimuli were projected on a 2D surface in the 2D condition (2DCOND) and on a 3D surface in the 3D condition (3DCOND). The stimuli sets contained the animated talking head with 20 different gaze angles. The angles were equally spaced between -25 degrees and 13 degrees (horizontal eyeball rotation in relation to skull) in 2 degree increments, where 0 degree rotation was when the eyes were looking straight ahead. The angles between 13 degrees and 25 degrees were left out because of a programmatic error, but there were no indications that this asymmetry has any negative effects on the experimental results.

One set of five simultaneous subjects was employed in a within-subject design, where each subject judged each stimulus in the experiment. All five subjects had normal or corrected to normal eye sight.

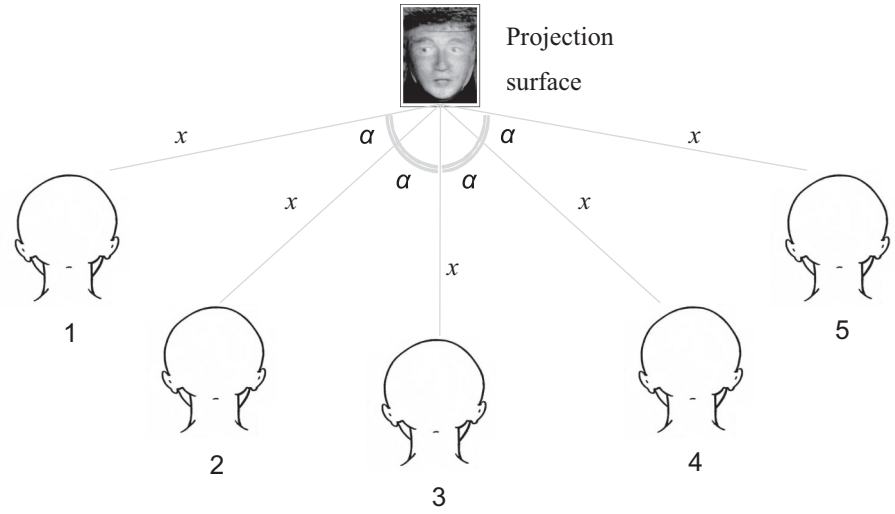

Fig. 9. Schematic of the experiment setup: five subjects were placed simultaneously at equal distances along the perimeter of a circle centred on the projection surface. x = 1.80 meters; $\alpha = 26.5$ degrees.

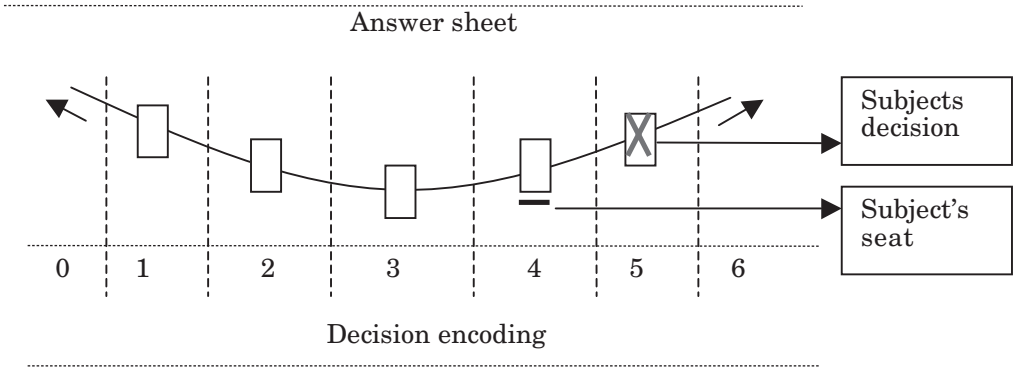

Fig. 10. Example of one line in an answer sheet.

3.3. Procedure

At the beginning of the experiment, the subjects were presented with an answer sheet, and the task was explained: to point out on the answer sheet, for each stimulus, which subject the gaze of the animated head is pointing at. For ecological validity, the subjects themselves were used as gaze targets instead of using, for example, a spatial grid as in Delaunay et al. [2010]. People are perceptually and communicatively relevant gaze targets.

For each set of 20 stimuli, each of the seated subjects were handed an empty answer sheet with 20 answer lines, which were to be answered sequentially. Figure 10 shows part of an example answer sheet. Each box in the sheet represented a subject. The underlined box indicated where the subject who was filling in the sheet was seated. The subject entered a mark in one of the boxes indicating the decision. If the subject believed the head was looking beyond the rightmost or the leftmost subject, the subject entered the mark at the end of either of the two arrows, to the left and right of the boxes. A training set was conducted to allow subjects some practice with the answer

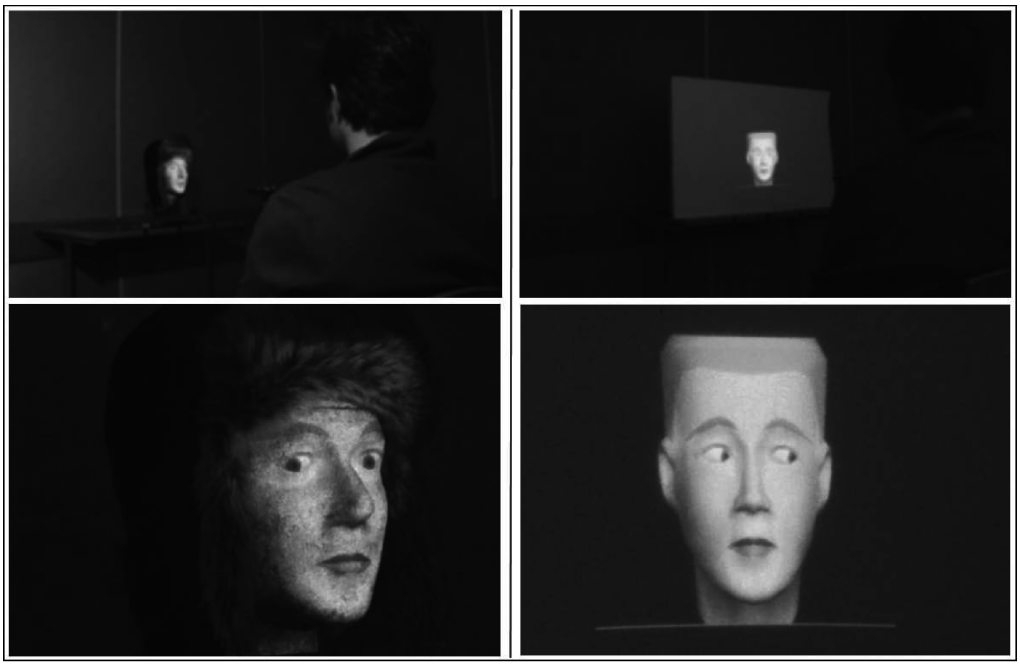

Fig. 11. Snapshots, taken over the shoulder of a subject, of the projection surfaces in 3DCOND (left) and 2DCOND (right).

sheet. The training set contained the randomized set of 20 angles in 2DCOND; these answers were disregarded.

The five subjects were then randomly seated on the five positions and the first set of 20 stimuli was projected in 3DCOND, as seen in the left of Figure 11. Subjects marked their answer sheets after each stimulus. When all stimuli were presented, the subjects were shifted to new positions and the process repeated, in order to capture any bias for subject/position combinations. The process was repeated five times, so that each subject sat in each position once, resulting in five sets of responses from each subject. As gaze stimuli were randomized per iteration, the sets of the randomized angles were also saved to enable matching of angles and answers.

After a break during which 2DCOND, with the talking head projected on a flat white board as seen in the right part of Figure 11, was set up, the entire procedure was repeated. In total, the experiment yielded results for 20 angels $*$ 5 positions $*$ 5 subjects $*$ 2 conditions $=$ 1000 stimuli responses.

3.4. Analyses

The answers were manually encoded into values between 0, representing the leftmost gaze target, and 6, representing the rightmost gaze target, and values between 1 and 5 representing the 5 subjects from left to right accordingly. Each sample of the complete resulting data set consists of the following.

—SUBJECTID: a number between 1 and 5, indicating which subject provided the answer.
—SEATING: a number between 1 and 5, indicating on which seat on which the subject providing the answer was seated.
—CONDITION: a value of 3DCOND or 2DCOND, indicating the condition the stimulus was presented under.

—SET: a number between 1 and 5 indicating which set this stimulus was presented in. The subjects are seated differently in each set.
—ANGLE: the internal angle of the eye balls rotation in the animated head.
—ABSOLUTERESPONSE: a number between 0 and 6 indicating the decision of the subject.

We then add a data point based on our view of the Mona Lisa stare effect: RELATIVERESPONSE. We assume that the subjects perceive the distance to the head as being the same as the distance to the projection surface—everything about the setup, from the life-sized 3D projection surface to the manner in which they were seated in a semicircle around the surface—pointed towards such an interpretation. With the distance ambiguity eliminated, our prediction would be that in 3DCONF, the subjects should be able to state, with a high degree of accuracy, in absolute terms at what subject the projection was looking. In the case of 2DCONF, however, the prediction is different. The accuracy of gaze perception would be similar to that of 3DCONF, as our model does not give us any reason to believe that we cannot estimate gaze direction within the 3D space represented by the 2D image. Subjects would perceive the gaze direction as if they were standing straight in front of the projection surface; in other words, we would expect them to see the gaze in a relative rather than an absolute manner. RELATIVERESPONSE, then, is ABSOLUTERESPONSE transcoded as if the subject would have been sitting straight in front of the projection surface. The result is a number varying from -5 to 5, with -5 representing gaze directed at a person sitting four steps to the left of the subject producing the answer, -4 a person sitting three steps to the left, and so on, with 0 representing a gaze directed straight at the subject. Our prediction is that this will produce a poor fit for 3DCONF data, and a good fit for the 2DCONF data.

4. RESULTS

Figure 12 shows plots the raw data for all the responses over gaze angles in the absolute and the relative interpretation. The size of the bubbles indicates the number of responses with the corresponding value for that angle; the bigger the bubble, the more subjects had perceived gaze in that particular direction. It is clear that in 3DCOND, the perception of gaze is precise (i.e., there are fewer bubbles per row) in absolute terms whereas in 2DCOND, the opposite is true, confirming our prediction.

A regression analysis on the data yields the numbers in Table II. We see that our prediction that viewers are quite able to accurately perceive gaze direction in relative terms within virtual space holds; the R square value for the relative interpretation in the 2DCOND is quite similar to that of the absolute interpretation in the 2DCOND condition.

There is a measure of dependency between the absolute and the relative interpretations. This is not predicted by the model: an absolute and a relative interpretation of angle would be independent if all degrees of freedom were permitted. The reason that they are not is to be found in the experiment design, which was not originally intended for comparisons of absolute and relative interpretations. The design involves five simultaneous, collocated, equidistant and immobile subjects. When the results of this setup are transformed to relative results, however, a methodological artefact causes the results to be correlated, so that the interpretation with the better fit taints the other interpretation. Another effect of the reinterpretation is that the expected distribution of relative votes for the relative case is uneven, with a greatest number of 0 votes the smallest number of -4 and 4 votes, and this is confirmed in the data.

When looking at spatial, situated and multiparty interaction, an accurate and absolute perception of gaze becomes crucial. That is to guarantee a global and objective perception of the intended gaze target among all the observers since the communication with the observers is taking place in their own 3D world. Due to the importance of

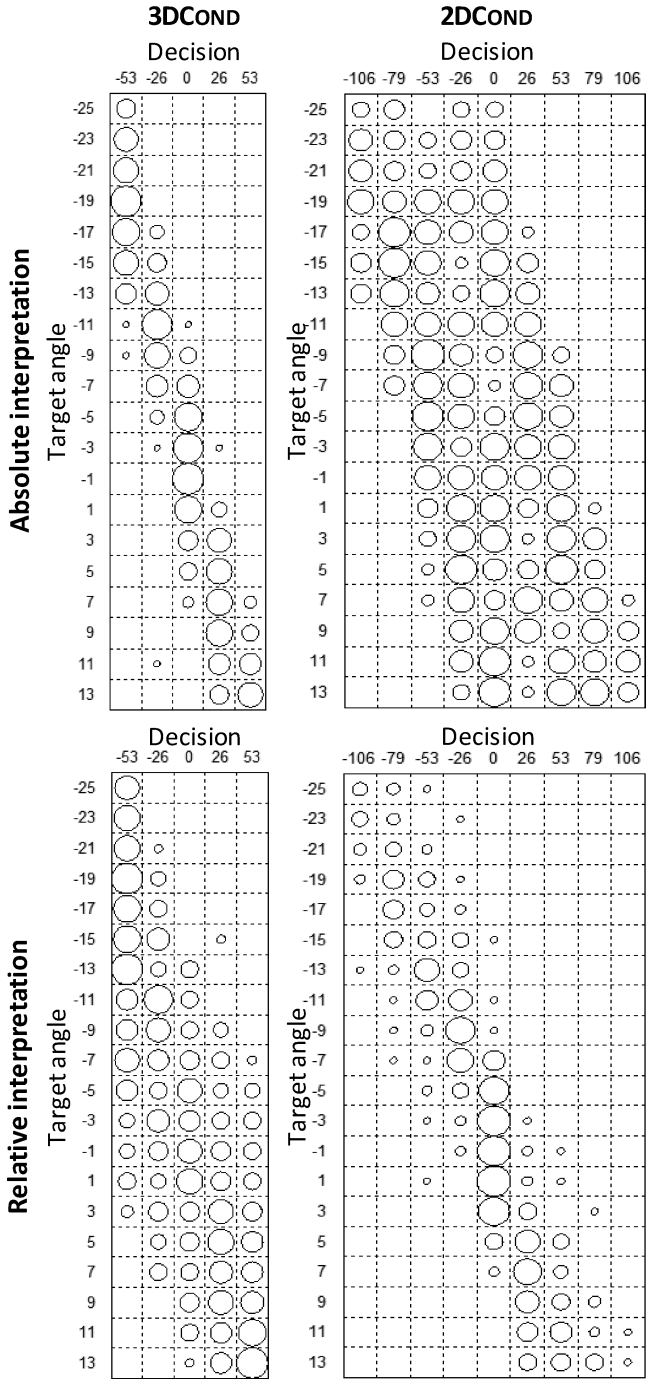

Fig. 12. The upper half shows subject-absolute targets. The lower half shows the same plots on the data reinterpreted as subject-relative. We see that the absolute interpretation yields a good fit on 3DCONF and the relative on 2DCONF, whereas the opposites do not hold.

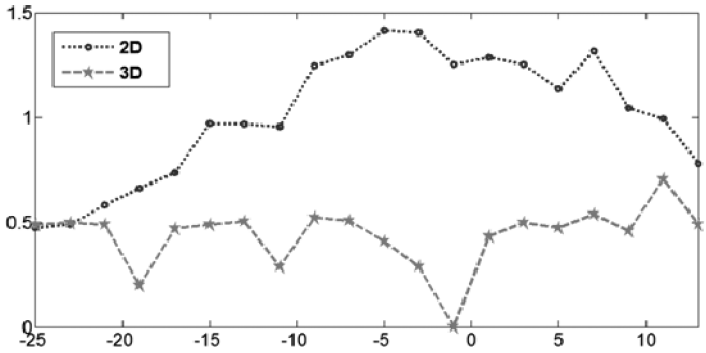

Fig. 13. Agreement between participants: standard deviation (Y axis) of the answers of all subjects over each angel (X axis) for both conditions.

Table II. R Square Values of the Fits for the Absolute and Relative Gaze Target Answers

	Absolute	Relative
2DConf	.537	.821
3DConf	.877	.371

gaze perception in absolute terms for situated interaction, the experiment warrants a detailed analysis of the results when interpreted in absolute terms. As a step towards understanding how accurately gaze is perceived in 3DCOND and 2DCOND, the variance of the answers among the subjects for each gaze angle was measured. Figure 13 shows the standard deviation of the answers per angle over all the subjects and sets. From the figure, it is clear that for all gaze angles, the standard deviation of the answers is always greater for 2DCOND. The 2DCOND variance increases with more frontal angles and decreases as the angle moves to the sides.

Figure 14 plots the raw data for all the responses over gaze angles. The size of the bubbles indicates the number of responses with the corresponding value for that angle; the bigger the bubble, the more subjects had perceived gaze in that particular direction. It is again clear that in 3DCOND, the perception of gaze is more precise (i.e., fewer bubbles per row) compared to 2DCOND.

We also calculated the agreement between subjects for observer position and condition. Table III and Table IV contain the results of a pair-wise Pearson correlation calculated over the stimuli answers, for 2DCOND and 3DCOND respectively. The mean correlation for 2DCOND is 0.607, and for 3DCOND is 0.9; the agreement among subjects is considerably higher in 3DCOND.

4.1. The Mona Lisa Gaze Effect in 2D and 3D Conditions

Figure 15 shows bubble plots similar to those in Figure 14, with responses for each stimulus. The figure differs in that the data plotted is filtered so that only responses where perceived gaze matched the responding subject, that is when subjects responded that the gaze was directed directly at themselves – what is commonly called eye-contact or mutual gaze. These plots show the location of and the number of subjects that perceived eye-contact over different gaze angles. In 2DCOND, the Mona Lisa gaze effect is very visible: for all the near-frontal angles, each of the five subjects, independently from where they are seated, perceives eye contact. The figure also shows that the effect is completely eliminated in 3DCOND, in which only one subject perceived eye-contact with the head over the different angles in the overwhelming majority of cases.

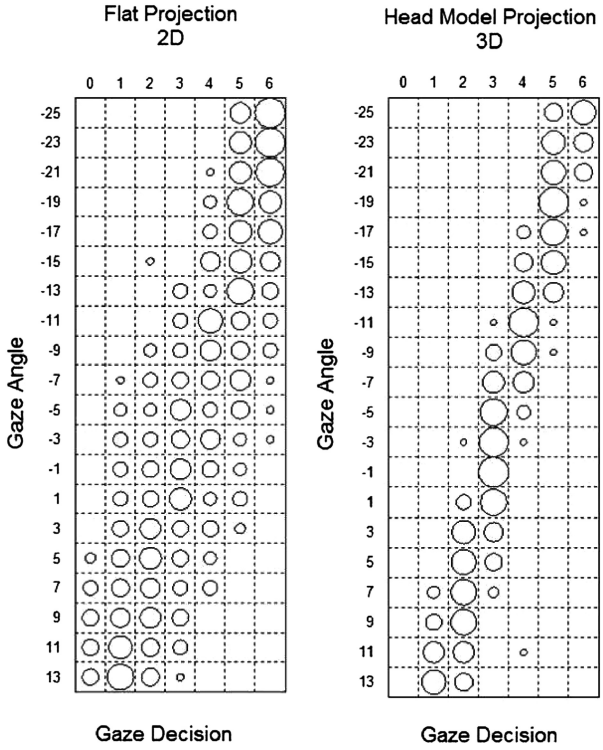

Fig. 14. Answers of all subject decisions (X axis) over all internal angles (Y axis) for each of the conditions: 2DCOND to the left and 3DCOND to the right. Bubble size indicates number of responses.

Table III.
An average subject pair-wise correlation for the answers of the subjects for each gaze stimulus in the 2D flat projection

	S1	S2	S3	S4	S5
S1		0.69	0.67	0.71	0.73
S2			0.60	0.69	0.50
S3				0.45	0.37
S4					0.61
S5					

Table IV.
An average subject pair-wise correlation for the answers of the subjects for each gaze stimulus in the 3D flat projection

	S1	S2	S3	S4	S5
S1		0.94	0.91	0.89	0.91
S2			0.89	0.89	0.89
S3				0.88	0.87
S4					0.90
S5					

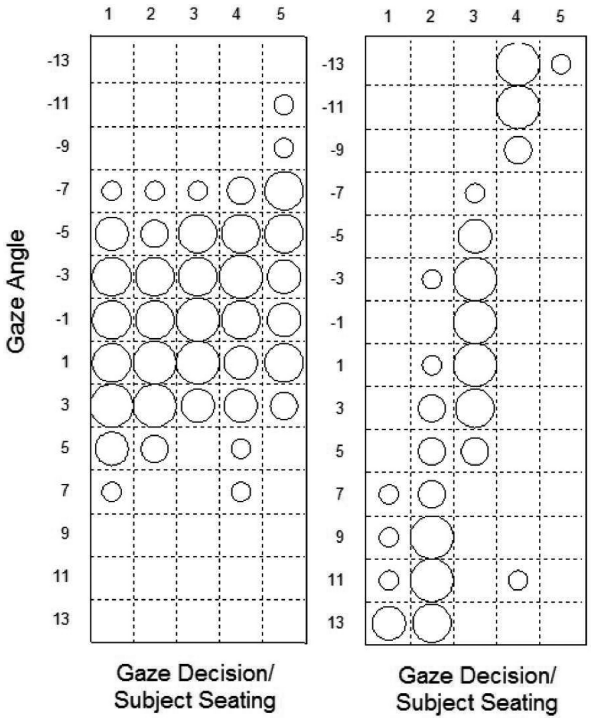

Fig. 15. Bubble plot showing only responses where subjects perceived eye-contact: subject position (X axis) over all internal angles (Y axis) for each of the conditions: 2DCOND to the left and 3DCOND to the right. Bubble size indicates number of responses.

Table V.
The variance of perceived target gaze (subject seats) for each angle, when subjects perceived eye-contact with the animated head

Angle	3D Variance	2D Variance
−13	0.1667	–
−11	0	–
−9	0	
−7	0	2.0662
−5	0	1.9447
−3	0.1667	1.8842
−1	0	1.9905
1	0.1667	1.9085
3	0.2667	1.3667
5	0.3333	4.5000
7	0.3333	2.0662
9	0.1667	–
11	0.8095	–
13	0.2857	–

Table V presents the variance of the responses where subjects perceived eye-contact. This table corresponds to the data plotted in Figure 15. It is notable that in 3DCOND, there were subjects that perceived eye-contact for all angles between −13 and 13, while in 2DCOND, sideway angles did not result in eye-contact with any subject (angles −13, −11, −9, 9, 11, 13). In addition, the variance of subjects who perceived eye-contact in 2DCOND was always higher than in 3DCOND.

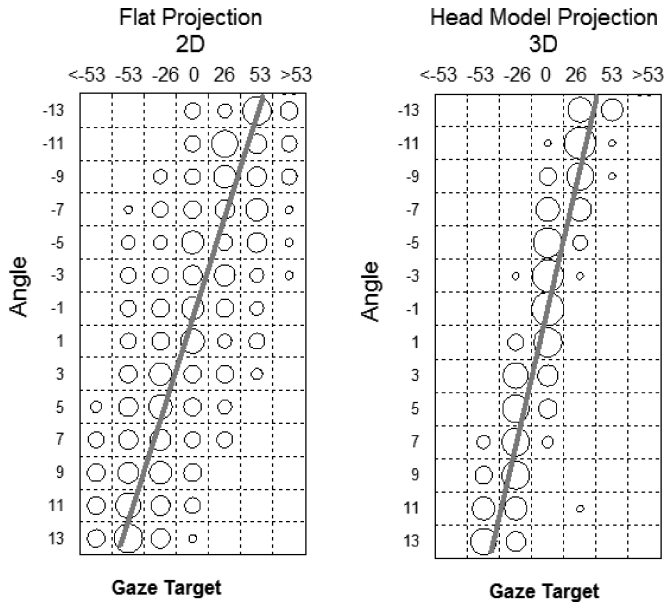

Fig. 16. A plot showing the gaze mapping function (linear fit to the data) for the 2D and 3D displays.

4.2. The Psychometric Function for Gaze

In addition to investigating the gaze perception accuracy of projections on different types of surfaces, the experimental setup employed allows us to measure a psychometric function for gaze which maps eye ball rotation in a virtual talking head to physical, real-world angles: an essential function to establish accurate delivery of gaze between the real and virtual world. We estimated this function by applying a linear fit to the data to get a mapping from the real positions of the gaze targets perceived by the subjects to the actual internal eyeball angles in the projected animated talking head, for each condition.

Figure 16 shows the estimated gaze psychometric function for both conditions. In 2DCOND, the estimated function that resulted from the linear fit is:

$$\begin{aligned} &\text{Angle} = -5.2 * \text{GazeTarget} \\ &\text{RMSE} = 17.66 \\ &\text{Rsquare} = .668 \end{aligned} \tag{1}$$

And for the 3DCOND:

$$\begin{aligned} &\text{Angle} = -4.1 * \text{GazeTarget} \\ &\text{RMSE} = 6.65 \\ &\text{Rsquare} = .892. \end{aligned} \tag{2}$$

Although the resulting gaze functions from the two conditions are similar, the goodness of fit is markedly better in 3DCOND than in 2DCOND.

5. DISCUSSION

Our experimental results show that given the absolute interpretation, the regression explains more than 80% of the variation in the 3DCONF observations. Given that the internal eye ball angles were moved in increments of 2 degrees, and that several angles would have been directed in between two of the seating positions, this is a high number. The relative interpretation of 3DCONF explains less than 40% of the variation: a low

number considering that the experiment configuration does not properly decorrelate the relative and absolute interpretations.

Conversely, in 2DCONF, more than 80% of the observations are explained by the regression based on the relative data, whereas just above 50% of 2DCONF observations are explained: again a low number given the dependency between the two interpretations. The numbers confirm that eye gaze direction in ECAs projected on a 2D surface and a 3D surface alike can be estimated with similar precision, only in observer relative terms in the former case and in absolute terms in the latter, which validates our model of the Mona Lisa gaze effect.

Armed with this model and with the distinction between relative and absolute gaze direction in 2D and 3D facial projections, respectively, we now turn to how different communicative gaze requirements are met by the two system types. Situated interaction requires a shared perception of spatial properties: where interlocutors and objects are placed, in which direction a speaker or listener turns, and at what the interlocutors are looking. Accurate gaze perception is crucial, but plays different roles in different types of communication, for example between colocated interlocutors, between humans in avatar or video mediated human-human communication, and between humans and ECAs in conversation with spoken dialogue systems or robots.

5.1. Gaze Faithfulness

We can now revisit the *gaze faithfulness* table from the introduction. The *observer* as the entity perceiving gaze and a *target point* is an absolute position in the observer's space.

—*Faithful / Unrealistic Mutual Gaze.* When the observer is the gaze target, the observer correctly perceives this. When the observer is *not* the gaze target, the observer correctly perceives this. In other words, the observer can correctly answer the question: Does she look me in the eye?
—*Faithful / Unrealistic Relative Gaze.* There is a direct and linear mapping between the intended *angle* of the gaze relative the observer and the observer's perception of that angle. In other words, provided that the observer is interpreted as standing in front of the gazing head, the observer can correctly answer the question: How much to the left of/to the right of/above/below me is she looking?
—*Faithful / Unrealistic Absolute Gaze*. A one-to-one mapping is correctly preserved between the intended *target point of gaze* and the observer's perception of that target point. In other words, the observer can accurately answer the question: What is she looking at?

These levels of faithfulness are largely hierarchical: a configuration meeting the requirements for faithful absolute gaze also meets the first two requirements; and one that shows relative gaze faithfulness meets mutual gaze faithfulness as well.

It would seem that whether a system can produce faithful gaze or not depends largely on four parameters. The first two represent system capabilities: the ability to be perceived as copresent in physical space and thereby negate the Mona Lisa gaze effect, limited here to whether the system produces gaze on a 2D surface or on a 3D surface, and whether the system knows where relevant objects (including the interlocutors head and eyes) are in the physical space; (e.g., through automatic object tracking or with the help of manual guidance). A special case of the second capability is the ability to know only where the head of the interlocutor is. The final two parameters are different and have to do with the requirements of the application: the first is what level of faithfulness is needed and the second whether the system is to interact with one or many interlocutors at the same time.

Table VI.
Faithful (+) or unrealistic (−) gaze behavior under different system capabilities and application requirements

			System capabilities					
		Projection	2D			3D		
		Object tracking	No		Yes	No		Yes
		Head tracking	No	Yes	Yes	No	Yes	Yes
Application requirements: Interlocutors	Faithfulness							
Single	Mutual		+	+	+	−	+	+
	Relative		+	+	+	−	+	+
	Absolute		−	−	+	−	−	+
Multiple	Mutual		−	−	−	−	+	+
	Relative		−	−	−	−	−	+
	Absolute		−	−	−	−	−	+

Single user systems with a traditional 2D display without object tracking are *faithful* in terms of *mutual gaze*, no matter where in the room the observer is the system can look straight ahead to achieve mutual gaze and can look anywhere else to avoid it. It is *faithful* in terms of *relative gaze as well*; regardless of where in the room the observer is, the system can look to the left and be perceived as looking *to the right of the observer*, and so on. It is *unrealistic* in terms of *absolute gaze*: the system can only be perceived as looking at target objects other than the observer by pot luck.

Single-user systems with a traditional 2D display with object tracking are generally the same as those without object tracking. It is possible, however, that object tracking can help *absolute gaze faithfulness* if the objects are targeted in relative terms, and the gaze of the agent is constantly changed depending on the viewing point of the observer.

Multiuser systems with a traditional 2D display and no object tracking fare poorly. They are *unrealistic* in terms of *mutual gaze*, as either all or none of the observers will perceive mutual gaze; they are *unrealistic* with respect to *relative gaze*, as all observers will perceive the gaze to be directed at the same angle relative to themselves; and hence, obviously, they are *unrealistic* in terms of *absolute gaze* as well.

Multiuser systems with a traditional 2D display and object tracking fare exactly as poorly as those without object tracking; regardless of any attempt to use the object tracking to help absolute faithfulness by transforming target positions in relative terms, all observers will perceive the same angle in relation to themselves, and only one, at best, will perceive the intended position.

Turning to the 3D projection surface systems, both *single* and *multiuser systems with a 3D projection surface and no object tracking* are *unrealistic* in terms of *mutual gaze*, *relative gaze*, and *absolute gaze*; without knowing where to direct its gaze in real space, it is lost. By adding head tracking only, the systems can produce faithful mutual gaze, and single users systems with head tracking can attempt faithful relative gaze by shifting gaze angle relative the observers head.

In contrast, both *single* and *multiuser system with a 3D projection surface and object tracking*, coupling the ability to know where objects and observers are with the ability to target any position are *faithful* in terms of all of *mutual gaze*, *relative gaze*, and *absolute gaze*.

Table VI presents an overview of how meeting the three levels of faithfulness depends on system capabilities and application requirements. Examining the table, we first note that in applications where more than one participant is involved, using a 2D projection surface will result in a system that is unrealistic on all levels (lower left quadrant of the table), and second, that a system with a 3D projection surface and object tracking will provide faithful eye gaze regardless of application requirements (rightmost column). These are the perhaps unsurprising results of the Mona Liza gaze effect being in place in the first case, causing the gaze perception of all in a room to be the same.

Third, we note that *if no automatic or manual object or head tracking is available, the 3D projection surface is unrealistic in all conditions*, as it requires information on where in the room to direct its gaze, and that head only tracking improves the situation to some extent.

Fourth, and more interestingly, we note that *in single-user cases where no object tracking or head tracking only is available, the 2D surface is the most faithful one* (upper left quadrant). In these cases, we can tame and harness the Mona Lisa gaze effect. This suggests that gaze experiments such as those described in [Edlund and Beskow 2009; Edlund and Nordstrand 2002] *could not have been performed with a 3D projection surface* unless sophisticated head trackers would have been employed.

In summation, it is worthwhile to have a clear view of the requirements of the application before designing the system. In some cases (i.e., single-user cases with no need for absolute gaze faithfulness), a simpler 2D display system without any tracking can give similar results, as a more complex 3D projection surface system with head or object tracking, at considerably lower cost and effort. On the other hand, if we are to study situated interaction with objects and multiple participants, we need to guarantee successful delivery of gaze direction at all levels with a 3D projection surface that inhibits the Mona Lisa stare effect and reliable object tracking, manual or automatic, to direct the gaze.

5.2. Practical Applications—Robotic Heads

We have demonstrated that the use of a 3D projection surface, a physical model of a human head, permits faithful communication of mutual gaze, observer relative gaze angles and absolute gaze targets. In addition, the 3D projection surface can be utilized as a robotic head, as an alternative to using a flat screen as a head and to other approaches in designing robotic heads.

The design of the 3D projection setup we employed in this study is similar to this developed recently in a study by Delauney et al. [2010], where the animated face is rear-projected onto the mask, and the projector and the mask are connected together, allowing for the control of the mask using a robotic neck.

When designing humanoid robots for the purpose of communicating with humans, the capacity for adequate interaction is a key concern. Since a great proportion of human interaction is managed nonverbally, through gestures, facial expressions and gaze, an important current research trend in robotics deals with the design of social robots. But what mechanical and behavioral compromises should be considered in order to achieve satisfying interaction with human interlocutors? In the following, we present an overview of the practical benefits of using an animated talking head optically projected on a 3D surface as a robotic head.

Optically based. Since the approach utilizes a static 3D projection surface, the actual animation is done completely using computer graphics projected on the surface. This provides an alternative to mechanically controlled faces, saving electrical consumption and avoiding much complex motor control. Computer graphics also offer many advantages over motor based animation of robotic heads in speed, animation accuracy, resolution and flexibility.

Animation using computer graphics. Facial animation technology has shown tremendous progress over the last decade, and currently offers realistic, efficient, and reliable renditions. The technology is currently able to establish facial designs that are very humanlike in appearance and behaviour compared to the physical designs of mechanical robotic heads.

Facial design. The face design is done through software, which provides the flexibility of having an unlimited range of facial designs for the same head. Even if the static projection surface needs to be re-customized to match a particularly unusual design,

this is considerably simpler, faster, and cheaper than redesigning a whole mechanical head. In addition, the easily interchangeable face designs offers the possibility to efficiently experiment with the different aspects of facial designs and characteristics in robotics heads, for example to examine the anthropomorphic spectrum.

Light weight. The optical design of the face leads to a considerably more lightweight head, depending only on the design of the projection surface. This makes the design of the neck much simpler and a more lightweight neck can be used, as it has to carry and move less weight. Ultimately, a lighter mobile robot is safer and saves energy.

Low noise level. The alternative of using light projection over a motor-controlled face avoids all motor noises generated by moving the face. This is crucial for a robot interacting verbally with humans, and in any situation where noise generation is a problem.

Low maintenance. Maintenance is reduced to software maintenance and maintenance of the micro laser projector, which is very easily replaceable. In contrast, mechanical faces are complicated, both electronically and mechanically, and an error in the system can be difficult and time consuming to troubleshoot.

Naturally, there are drawbacks as well. Some robotic face designs cannot be achieved in full using light-projected animation alone, for example those requiring very large jaw openings which cannot be easily and realistically delivered without mechanically changing the physical projection surface. For such requirements, a hybrid approach can be implemented which combines a motor based physical animation of the head for the larger facial movements, with an optically projected animation for the more subtle movements, for example changes in eyes, wrinkles and eyebrows.

In addition, the animations are delivered using light, so the projector must be able to outshine the ambient light, which becomes an issue if the robot is designed to be used in very bright light, such as full daylight. The problem can be remedied by employing the evermore powerful laser projectors that are being brought to the market.

Our future developments of the robot head are similar to the head design of Delauney et al. [2010]. This will include producing a translucent face mask suitable for back projection. Acquiring the mask will be done with the help of 3D printers, where a matching 3D print to the virtual 3D face will be established. This mask is then to be attached to the projector so that the face is back-projected on the translucent mask. The head is then to be placed on a mechanical neck, so that gaze and head pose can be integrally controlled and studied.

5.3. Types of Applications

As we have seen, the Mona Lisa gaze effect is highly undesirable in several communicative setups due to the manner in which it limits our ability to control gaze target perception. We have also seen that under certain circumstances, the effect—a cognitive ability to perceive a depicted scene from the point of view of the camera or painter—can be harnessed to allow us to build relatively simple applications, which would otherwise have required much more effort. A hugely successful example is the use of TV screens and movie theatres, where entire audiences perceive the same scene, independently of where they are seated. If this was not the case, the film and TV industries might well have been less successful. There are also situations where an ECA can benefit from establishing eye contact with either all viewers simultaneously in a multiparty situation, as when delivering a message or taking the role of a news presenter, and when it is required to establish eye contact with one person whose position in the room is unknown to the system, as is the case in most spoken dialogue system experiments to date involving an ECA.

Although the Mona Lisa gaze effect can be exploited in some cases, it is an obstacle to be overcome in the majority of interaction scenarios, as those where gaze is required to

Fig. 17. A traditional one-to-many video conferencing setup.

point exclusively to objects in the physical 3D space of the observer, or where multiple observers are involved in anything but the most basic interactions. In order to do controlled experiments investigating gaze in situated multiparty dialogues, the Mona Lisa effect must be overcome, and this can be done using the proposed technique. In other words, the technique opens possibilities for many applications that require the perception of absolute gaze direction, but would not have been possible with the use of a 2D display. In the following we present a short list of application families that we have recently begun to explore in the situated interaction domain, all of which require the levels of gaze perception afforded by 3D projection surfaces.

The first family of applications is *situated and multiparty dialogues with ECAs or social conversational robots*. These systems need to be able to switch their attention among the different dialogue partners, while keeping the partners informed about the status of the dialogue and who is being addressed. Moreover, in these systems, exclusive eye contact with single subjects is crucial for selecting an addressee. In such scenarios, a coherently shared and absolute perception of gaze targets is needed to achieve a smooth humanlike dialogue flow; a requirement that cannot be met unless the Mona Lisa gaze effect is eliminated.

The second family involves any application where there is a need for a pointing device to point at objects in real space; the space of the human participant. Gaze is a powerful pointing device that can point from virtual space to real space while being completely non-mechanic—as opposed to for example fingers or arrows—it is, as well, nonintrusive and subtle.

A third family of applications is mediated interaction and tele-presence. A typical application in this family is virtual conferencing systems. An example of a virtual conferencing setup is illustrated in Figure 17. In such a system, the remote partner cannot meaningfully gaze into the environment of the other partners, since the remote partner is presented through a 2D display subjected to the Mona Lisa gaze effect. Establishing a one-to-one interaction through mutual gaze cannot be done, as there is no ability to establish an exclusive eye contact. In addition to that, a very common problem in video conferencing is that, people look at the video presenting the other partners instead of looking into the camera, which is another obstacle for shared attention and mutual gaze, and no one can estimate reliably at what the remote participant is looking. This problem has been under interest for a long time, with solutions often involving complex hardware and software designs (for examples, refer to Gemmel et al. [2000] and Vertegaal [1999]).

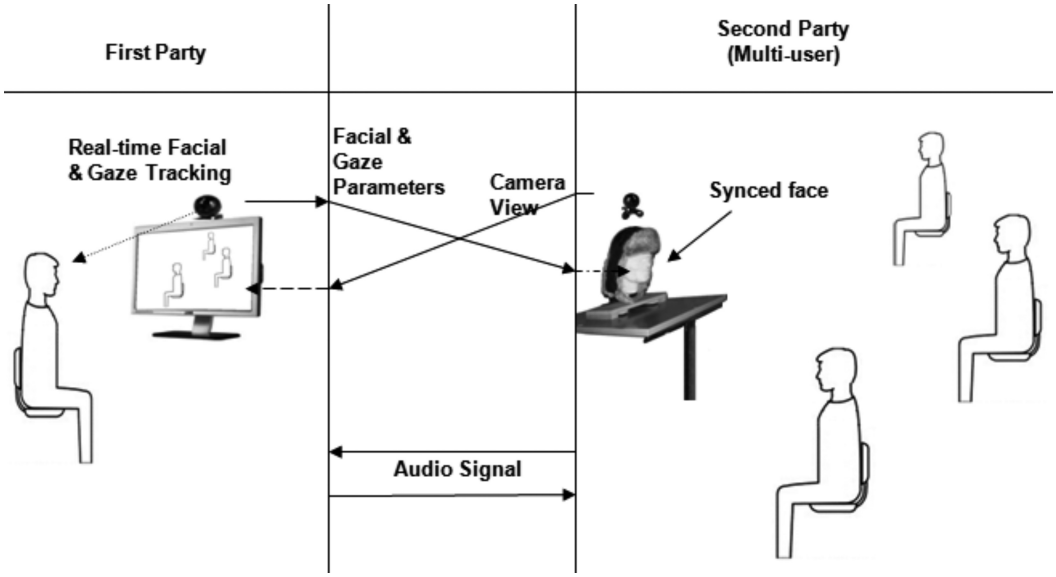

Fig. 18. A setup of a facial tele-presence application suitable for a multi-party setup.

If a 3D projection surface is used to represent the remote subject, who is represented through mediation as an avatar, these limitations to video conferencing can be resolved.

Figure 18 shows a flow chart diagram of a one-to-many video conferencing (telepresence) setup using a 3D projected head model as an impersonation of the remote subject. In this setup, the gaze and the facial parameters are tracked in real-time, and the gaze coordinates are mapped into physical gaze coordinates which point at the same coordinates the subject is looking at in the video of the other partners. In such a setup, users can perceive exactly where the remote subject is looking at in their own three dimensional space, and can establish a single and exclusive eye contact with him. Note that in this application, no automatic object tracking is needed; the remote participant and the gaze tracker acts as a manual object tracker and provides sufficient information to target objects and persons.

6. CONCLUSIONS

To sum up, we have proposed two ways of "taming Mona Lisa": first by eliminating the effect and second by harnessing and exploiting it.

En route to this conclusion, we have proposed an affordable way of eliminating the effect by projecting an animated talking head onto a 3D projection surface; a generic physical 3D model of a human head, and verified experimentally that it allows subjects to perceive gaze targets in the room clearly from various viewing angles, meaning that the Mona Lisa effect is eliminated. In the experiment, the 3D projection surface was contrasted with a 2D projections surface, clearly displaying the Mona Lisa gaze effect in the 2D case. The experiment setup employed five subjects seated simultaneously at equal distances from each other and from the presentation point. Twenty gaze angles communicated through an animated talking head were tested, and each subject was asked to mark who the animated talking head looked at. In addition to eliminating the Mona Lisa gaze effect, the 3D setup allowed observers to perceive with very high agreement the person that was being looked at. The 2D setup showed

no such agreement. We showed how the data serves to estimate a gaze psychometric function to map actual gaze target into eyeballs rotation values in the animated talking head.

Based on the experimental data and the working model, we proposed three levels of gaze faithfulness relevant to applications using gaze: *mutual gaze faithfulness*, *relative gaze faithfulness*, and *absolute gaze faithfulness*. We further suggested that whether a system achieves gaze faithfulness or not depends on several system capabilities: whether the system uses a 2D display or the proposed 3D projection surface, and whether the system has some means of knowing where objects and where its interlocutors are, but also on the application requirements: whether the system is required to interact with more than one person at a time and the level of gaze faithfulness it requires.

One of the implications of this is that the Mona Lisa gaze effect can be advantageous and put to work in some types of applications. Although perhaps obvious, it falls out neatly from the working model. Another implication is that the only way to robustly achieve all three levels of gaze faithfulness is to have some means of tracking objects in the room and to use an appropriate 3D projection surface. But without knowledge of objects' positions, the 3D projection surface falls short.

We close by discussing the benefits of 3D projection surfaces in terms of human-robot interaction, where the technique can be used to advantage to create faces for robotic heads with a high degree of human-likeness, better design flexibility, more sustainable animation, low weight and noise levels and lower maintenance, and by discussing in some detail a few application types and research areas where the elimination of the Mona Lisa gaze effect through the use of 3D projection surfaces is particularly useful, such as when dealing with situated interaction or multiple interlocutors.

We consider this work to be a stepping stone for several future investigations and studies into the role and employment of gaze in human-robot, human-ECA, and human avatar interaction.

REFERENCES

Abele, A. 1986. Function of gaze in social interaction: Communication and monitoring. *J. Nonverb. Behav. 10*, 83–101.

Argyle, M. and Cook, M. 1976. *Gaze and Mutual Gaze*. Cambridge University Press.

Beskow, J. 2003. Talking heads: Models and applications for multimodal speech synthesis. Doctoral dissertation, KTH Royal Institute of Technology, Stockholm, Sweden.

Bente, G., Donaghy, W. C., and Suwelack, D. 1998. Sex differences in body movement and visual attention: An integrated analysis of movement and gaze in mixed-sex dyads. *J. Nonverb. Behav. 22*, 31–58.

Bilvi, M. and Pelachaud, C. 2003. Communicative and statistical eye gaze predictions. In *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Bloom, K. and Erickson, M. 1971. The role of eye contact in social reinforcement of infant vocalizations. In *Proceedings of the Biannual Meeting of the Society for Research in Child Development*.

Bohus, D. and Horvitz, E. 2010. Facilitating multiparty dialog with gaze, gesture, and speech. In *Proceedings of the 12th International Conference on Multimodal Interfaces and the 7th Workshop on Machine Learning for Multimodal Interaction*.

Cuijpers, R. H., van der Pol, D., and Meesters, L. M. J. 2010. Mediated eye-contact is determined by relative pupil position within the sclera. In *Proceedings of the European Conference on Visual Perception (ECVP'10)*. 129.

Delaunay, F., De Greeff, J., and Belpaeme, T. 2010. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. *In Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*. 39–44.

Descartes, R. 1637. Dioptrics. In *Discourse on Method, Optics, Geometry, and Meteorology*, Hackett Publishing, Indianapolis, IN. 65–162.

Edlund, J. and Beskow, 1. 2009. MushyPeek: A framework for online investigation of audiovisual dialogue phenomena. *Lang. Speech, 52*, 2–3, 351–367.

EDLUND, J. AND NORDSTRAND, M. 2002. Turn-taking gestures and hour-glasses in a multi-modal dialogue system. *In Proceedings of the ISCA Workshop on Multi-Modal Dialogue in Mobile Environments*.

FRIESCHEN, A., BAYLISS, A., AND TIPPER, S. 2007. Gaze cueing of attention: Visual attention, social cognition and individual differences. *Psych. Bull. 133*, 694–724.

GEMMELL, J., ZITNICK, C., KANG, T., TOYAMA, K., AND SEITZ, S. 2000. Gaze-awareness for videoconferencing: A software approach. *IEEE Multimedia 7*, 4, 26–35.

GREGORY, R. 1997. *Eye and Brain: The Psychology of Seeing*. Princeton University Press.

GU, E. AND BADLER, N. 2006. Visual attention and eye gaze during multiparty conversations with distractions. In *Proceedings of the International Conference on Intelligent Virtual Agents*.

KENDON, A. 1967. Some functions of gaze direction in social interaction. *Acta Psychologica 26,* 22–63.

KUENKE, C. L. 1986. Gaze and eye contact: a research review. *Psych. Bull. 100*, 78–100.

LANCE, B. AND MARSELLA, S. 2008. A model of gaze for the purpose of emotional expression in virtual embodied agents. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*. 199–206.

LANCE, B. AND MARSELLA, S. 2010. Glances, glares, and glowering: How should a virtual human express emotion through gaze? *J. Auton. Agents Multiagent Syst. 20*, 50–69.

LINCOLN, P., WELCH, G., NASHEL, A., ILIE, A., STATE, A., AND FUCHS, H. 2009. Animatronic shader lamps avatars. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality*.

NORDENBERG, M., SVANFELDT, G., AND WIK, P. 2005. Artificial gaze: Perception experiment of eye gaze in synthetic faces. In *Proceedings of the 2nd Nordic Conference on Multimodal Communication*.

PELACHAUD, C. AND BILVI, M., 2003. Modelling gaze behavior for conversational agents. In *Proceedings of the International Conference on Intelligent Virtual Agents*. 15–17.

POGGI, I. AND PELACHAUD, C. 2000. Emotional meaning and expression in performative faces. In *Affective Interactions: Towards a New Generation of Computer Interfaces*, Paiva, A., Ed., Lecture Notes in Computer Science vol. 1814, 182–195.

RASKAR, R., WELCH, G., LOW, K.-L., AND BANDYOPADHYAY, D. 2001. Shader lamps: Animating realobjects with image-based illumination. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*. 89–102.

SMITH, A. M. 1996. *Ptolemy's Theory of Visual Perception: An English translation of the Optics with Introduction and Commentary*. American Philosophical Society, Philadelphia, PA.

TAKEUCHI, A. AND NAGAO, K. 1993. Communicative facial displays as a new conversational modality. In *Proceedings of the Conference on Human Factors in Computing Systems.*

TODOROVIĆ, D. 2006. Geometrical basis of perception of gaze direction. *Vis. Resear. 45*, 21, 3549–3S62.

VERTEGAAL, R. 1999. The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of the ACM CHI Human Factors in Computing Systems Conference*, Addison-Wesley/ACM Press, 294–301.

WAXER, P. 1977. Nonverbal cues for anxiety: An examination of emotional leakage. *J. Abnorm. Psych. 86*, 306–314.

WOLLASTON, W. H. 1824. On the apparent direction of eyes in a portrait. *Phil Trans. Royal Soc. London, B114*, 247–260.

Received January 2011; revised August 2011, October 2011; accepted October 2011