Journal Pre-proof

Improving VIP viewer Gaze Estimation and Engagement Using Adaptive Dynamic Anamorphosis

Ye Pan, Kenny Mitchell

Please cite this article as: Ye Pan, Kenny Mitchell, Improving VIP viewer Gaze Estimation and Engagement Using Adaptive Dynamic Anamorphosis, *International Journal of Human-Computer Studies* (2020), doi: https://doi.org/10.1016/j.ijhcs.2020.102563

• We propose a dynamic adaptive anamorphosis method based on non-linear projections.
• Anamorphic rendering of a selective object with normal view rendering of the rest.
• The VIP guest results in an improved gaze and engagement estimation.
• This is performed without sacrificing the other guests' viewing experience.
• We discuss different viewpoints and the spatial relationship between objects.

Improving VIP viewer Gaze Estimation and Engagement Using Adaptive Dynamic Anamorphosis

Ye Pan[a], Kenny Mitchell[a,b,*]

[a]*Disney Research Los Angeles*
[b]*Edinburgh Napier University*

Abstract

Anamorphosis for 2D displays can provide viewer centric perspective viewing, enabling 3D appearance, eye contact and engagement, by adapting dynamically in real time to a single moving viewer's viewpoint, but at the cost of distorted viewing for other viewers. We present a method for constructing non-linear projections as a combination of anamorphic rendering of selective objects whilst reverting to normal perspective rendering of the rest of the scene. Our study defines a scene consisting of five characters, with one of these characters selectively rendered in anamorphic perspective. We conducted an evaluation experiment and demonstrate that the tracked viewer centric imagery for the selected character results in an improved gaze and engagement estimation. Critically, this is performed without sacrificing the other viewers' viewing experience. In addition, we present findings on the perception of gaze direction for regularly viewed characters located off-center to the origin, where perceived gaze shifts from being aligned to misalignment increasingly as the distance between viewer and character increases. Finally, we discuss different viewpoints and the spatial relationship between objects.

Keywords:
Gaze, Dynamic Anamorphosis, Display

*Corresponding author: Kenny Mitchell

Email addresses: ye.pan@disney.com (Ye Pan), k.mitchell2@napier.ac.uk (Kenny Mitchell)

1. Introduction

Gaze has been shown to matter for a number of aspects of social interaction, such as deictic referencing and viewers engagement [1, 2]. When presenting virtual agents or characters, modeling and delivering gaze targets accurately is crucial. From entertainment and art to advertisement and information visualization, 2D displays, such as flat monitors, are ubiquitously used. However, this approach introduces severe limitations for an accurate preservation of gaze direction. Notably, 2D displays are associated with several powerful effects and illusions, most important is the *Mona Lisa* effect, where the gaze of the projected head appears to follow the viewer regardless of viewpoint [3, 4].

As reviewed in the next section, a variety of light field, 3D display hardware, or virtual reality systems can eliminate the Mona Lisa effect, and provide accurate gaze direction, but at greater cost in hardware [5] or limited to a single user [6]. In this paper, we adapt and extend the previous dynamic anamorphosis approach [7, 8], which adapts itself to the changing position of the observer so that wherever the observer moves, he sees the same undeformed image. We propose a dynamic adaptive anamorphosis method, where we render selected objects from the selected viewer (VIP viewer)'s perspective, while the rest of the scene is rendered with a fixed direction normal perspective. The VIP viewer can be selected for the anamorphic experience using either face recognition or some other distinguishable tracked feature used for identification (e.g., [9]). The method can better preserve nonverbal cues, including gaze, and we show how it improves the viewer experience. Also, other viewers in the room can engage with the displayed scene, where only the selected objects will be slightly deformed from their viewpoint. Given that human perception has shown that the adjustment to oblique viewing is achieved before the contents of the image are interpreted [10], we propose that subtle variations in perspective do not interfere significantly for the engagement of non-selected viewers whilst improving the VIP viewer's experience.

We evaluated the effectiveness of our method by measuring the ability of viewers to accurately judge which target each character is gazing at, the ability to discriminate whether each character is looking directly into their eyes or not, and the level of engagement whilst watching characters performing movements. Results revealed that the accuracy of the estimation of object-focused gaze varies across different audience members (VIP viewer

and non-VIP viewer) and viewer position. It further varies across different character locations. The clear trends are that estimations for the VIP viewer are always superior. There is little obvious difference between the experience for VIP and non-VIP viewers at the central position. We also looked at when characters placed at off-center locations of screen, which is not explicitly addressed by previous work. The farther away a character is from the viewer, the more discrepancy in estimation appeared. For mutual gaze, we found the VIP viewer could always distinguish between being looked at and gaze to one side of them. The non-VIP performed badly in this test and again only when they were off the central position, but importantly not worse than traditional 2D display. For estimating viewer engagement, we found the VIP viewer rated the character rendered in anamorphic perspective higher than the rest of the characters, but it does not interfere with the other viewers' engagement. In a deployed entertainment scenario, the 'VIP' viewer may switch to another 'VIP' according to tracking of viewers and content of the experience. This demonstration and result thus motivates the further study in the context of an art installation or video conferencing.

The rest of this paper is organized as follows. In the next section, we review related work in the area of fish tank virtual reality, anamorphosis & non linear projection and eye gaze & engagement evaluation. Section 3 contains a description of the system implementation. The experiment is covered in Section 4. Finally, we present discussions of the results, conclusions and future work.

2. Related Work

2.1. Fish Tank Virtual Reality

Fish tank virtual reality (FTVR), where a stereo image of a three dimensional scene viewed on a monitor using a perspective projection coupled to the head position of the observer, was originally proposed with a single 2D display [11]. The important finding of the original FTVR studies was a comparison of different visual cues. For a variety of 3D interactions, they found that while head-tracking and stereo cues together were best, head-tracking alone resulted in better performance compared to stereo cues alone [12, 11]. This initial finding motivated many follow-on FTVR displays [6, 13], including multi-view FTVR displays [14, 15, 16] or non-planar displays [17, 18, 19] or indeed mobile hand held displays [20, 21], that omitted stereo display

hardware and so did not require any headset or glasses. However, these systems are more complicated than a simple 2D display or typically limited to one viewer.

To the best of our knowledge, there are no previous examples of single viewer eye contact (with improved experience) for 2D displays whilst providing an unnoticeable experience for other viewers.

2.2. Anamorphosis & Nonlinear Projection

Uses of anamorphosis or nonlinear projection can be found in art history, allowing artists to explore, understand, and subsequently express 3D shapes in 2D imagery. The painting, *The Ambassadors* , by Hans Holbein is probably the most famous example. On the bottom of this painting appears an indistinguishable diagonal blur which appears as a human skull when viewed from the upper right. The painting, the *Femme nue accroupie* , by Pablo Picasso in 1959, which typifies his style of composing different views of different parts of a scene into a single projection.

Artistic rendering also is an important research area in computer graphics. Of particular relevance to this paper is the prior work, dynamic anamorphosis [8] - dynamic changing of the anamorphic deformation in concert with the movement of the observer requires that the system tracks the 3D position of the selected observer's eyes and performs the recomputation of the anamorphic deformation in real time. This is achieved using computer vision methods which may consist of face detection and tracking the 3D position of the selected observer. However, this rendering method only supports one viewer, freeing him or herself from the Mona Lisa effect, but other viewers would see a distorted view.

On the other hand, non-linear perspective projections have been applied in computer generated imagery for a variety of purposes, that can be divided into the following main categories: image warping, 3D projections, and multi-perspective panoramas [22, 23, 24, 25]. Image warping [26] is a popular technique for manipulating digital images. Since this approach is inherently 2D, however, it limits the ability to explore different viewpoints and the spatial relationship between objects. As an alternative, 3D deformations [27] are widely used for manipulating 3D geometry. For some applications, however, it can be preferable to modify the camera transformation rather than to change the 3D shape of the object being depicted. Non-linear projections have also used in conjunction with multi-perspective panoramas. Inspired by the compelling illusion of depth in classic Disney animations, Wood et

4

al. [28] generated 2D panoramas for prescribed 3D camera paths, achieving the effect of 3D perspective as the camera panned across the panorama.

Our method is inspired by previous work, dynamic anamorphosis, and implemented by using nonlinear perspective projections concept: for selected objects, a perspective anamorphosis can be seen in its true or intended shape only from a particular viewpoint, for remaining objects or from other viewpoints it looks deformed or in the worst case not discernible.

2.3. Gaze Estimation & Engagement

Gaze is fundamental and probably the most studied resource in human social interaction, used for judging objects/people of interest, especially multiple viewers. The estimation of gaze has been studied broadly [29]. Gibson et al. [30] established that gaze direction may be perceived by both the direction in which the head is oriented and the eye's position relative to the head. Anstis et al. [31] investigated gaze estimation influenced by three orientations of a TV screen. They found a *TV screen turn* effect such that apparent displacement of the perceived direction in the same direction as the turn of the screen and suggested that the convex curvature of the screen probably caused the TV screen turn effect. They also reported an *over estimation* effect such that when gaze was to one side of the participant, the participant judges it to be further to that side than it actually was. They suggested that this overestimation became greater as the complexity of the viewing situation increased. Additionally, the *Wollaston* effect demonstrates that the perceived gaze direction of a portrait depends not only on the position of the irises but also on the orientation of the head [32]. For multiple viewers, the *Mona Lisa* effect was found when a 3D head is rendered in 2D, such as, painting, photograph, TV or video wall. Either every viewer or no viewer feels the 3D head is making eye contact with them. [3].

In applying human gaze models (incorporating features of human anatomy, including eye dimensions, inter-eye distance, and symmetry of oculomotor range) to stylized characters with with exaggerated or non-human anatomic features, a number of issues arise [33]. A detailed overview can be found in [34] and our characters' gaze model are developed by artists intent to retain naturalness and communicative accuracy.

Eye gaze is fundamental in showing interest levels between characters and as a means of anticipating events. By engagement, we refer to "the process by which participants establish, maintain and end their perceived connection", as defined in [35]. When audiences looked at the performer less than normal,
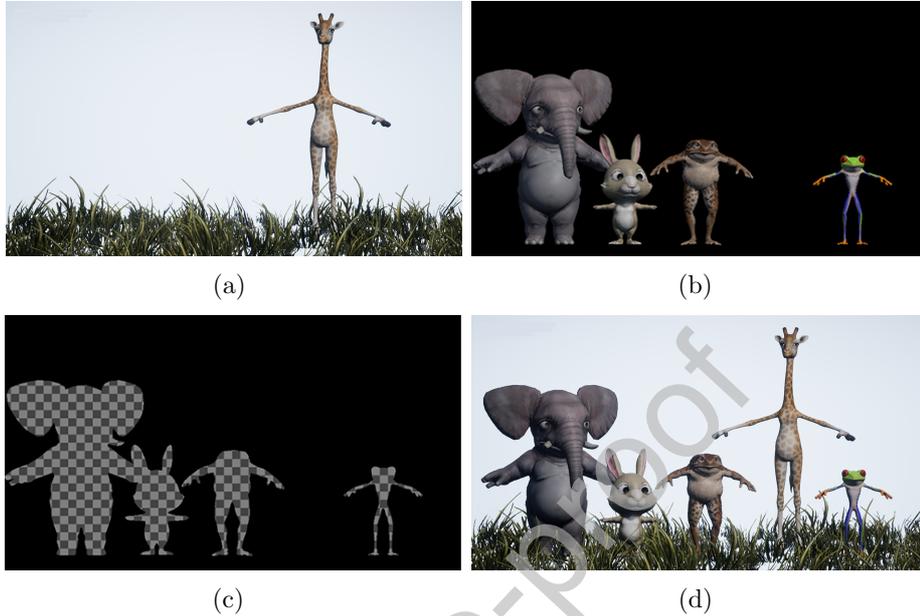
Figure 1: Composition of the adaptive dynamic anamorphosis method: (a) Anamorphic projected image corresponding to viewpoint C (60, -250) in Figure 3(a). (b) Normal perspective. (c) Mask used for composition. (d) Compositing (a) and (b) into a single resulting dynamic adaptive anamorphic perspective image.

the audiences rated less engagement [36]. Thus, the duration and frequency of glances directed towards the speaker will be considered indicative of the audience's engagement level.

Most of this previous empirical work is focused on evaluating viewer experience, such as gaze estimation or engagement level, when the character placed at the center of the screen (e.g., [12, 14, 15]). It is not so clear how viewer experience is affected when characters are placed at off-center locations.

3. The Adaptive Dynamic Anamorphosis System

3.1. Method

We adapt dynamic anamorphosis [8] technique with offset perspective projection, and propose a nonlinear projection method. The idea is to construct

a nonlinear projection of objects in a scene using multiple linear perspectives, particularly, a VIP viewer perspective (see Figure 1(a)) and a normal or orthogonal perspective (see Figure 1(b)).

In more detail, the off-axis perspective projection applies a skewed transform from the offset eye view point to the display's plane [37, 38, 39]. In our case, this plane is parallel and centered to the baseline of our experiment, and yields a simple offset $O_{-x,-y,-z}$ viewpoint translation matrix M_O applied with the shear components of the perspective projection matrix (1).

$$M_{proj} = \begin{bmatrix} \frac{2near}{width} & 0 & 0 & 0 \\ 0 & \frac{2near}{height} & 0 & 0 \\ -\frac{left+right}{width} & -\frac{top+bottom}{height} & \frac{far}{near-far} & 1 \\ 0 & 0 & -\frac{far*near}{near-far} & 0 \end{bmatrix}, M_{offset} = M_O M_{proj} \quad (1)$$

Our implementation is carried out in Epic's Unreal Engine 4.18. In this version, we found it necessary to disable motion vector related screen passes such as temporal anti-aliasing (TAA), motion blur, etc. Notice, the offset perspective projection matrix is a *reverse-Z* projection allowing an even quasi-logarithmic distribution of projection values [40] and is a common usage in video game engines including Roblox and Unity.

Each object in the scene is assigned to either the VIP viewer perspective camera or the normal perspective camera in the scene, and rendered based on the linear perspective of that camera. The VIP viewer perspective camera renders the anamorphic deformation in real time, in concert with the movement of the selected viewers' eyes (Note: eye position was calculated by the tracked head position). The rendering of both cameras in the scene are composited to generate the final image (see Figure 1(d)). A visibility ordering of objects can be created using a master camera and this can be used during the compositing stage.

Figure 2 demonstrates the result of our method. We zoomed in the giraffe to make the results clearer. Figure 2(a) to Figure 2(b) are benchmarks, showing the imagery of the traditional method, while the characters look at viewpoint B, viewpoint C, and viewpoint D, respectively. Figure 2(c) and Figure 2(d) show the characters look at the VIP viewpoint, but the giraffe rendered from anamorphic perspective, corresponding to the VIP viewpoint. To see the giraffe undeformed, look at it from the slightly right side of the paper for Figure 2(c) and more right for Figure 2(d).
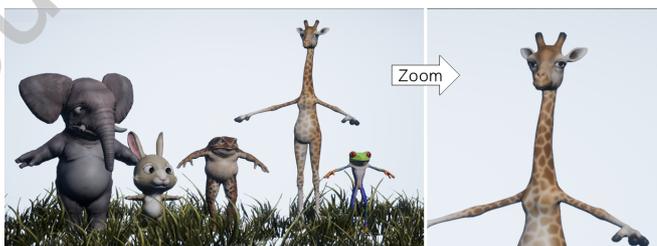
7

(a) All characters look at viewpoint B (0, 0), giraffe normally rendered

(b) All characters look at viewpoint D (120, 0), giraffe normally rendered

(c) All characters look at VIP viewpoint C (60, 0), giraffe anamorphically rendered

(d) All characters look at VIP viewpoint D (120, 0), giraffe anamorphically rendered

Figure 2: Comparison of results. (c) & (d) are the results of our methods. The giraffe rendered from anamorphic perspective, corresponding to the VIP viewpoint. To see the giraffe undeformed, look at it from the slightly right side of the paper for (c) and more right for (d).

8
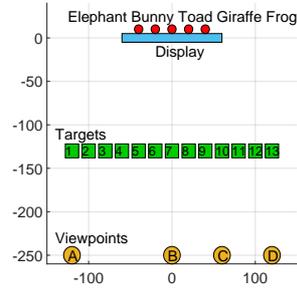
(a) Schematic layout. Units are in centimeters.

(b) Photo of the experiment room

(c) Giraffe rendered from the anamorphic perspective

(d) Frog rendered from the anamorphic perspective

Figure 3: Experiment setup. Note that subfigure (c) & (d) Photographs have been taken from the experiment subject viewpoint D.

3.2. Hardware Setup

Our hardware include a Windows 10 computer with an Intel Core i7 processor, 32GB RAM and a GeForce TitanX graphics card; a Samsung 55 inch 240Hz LED HDTV with a 1920×1080 resolution, and a Microsoft Xbox One Kinect to track VIP viewer's head position.

4. Experiment

The purpose of the study was to show that our implementation can provide one VIP viewer greater eye contact and higher engagement, without interfering with the other viewers' experience. We concentrated our investigation on three major aspects in the communication of attention: object-focused gaze, mutual gaze and user engagement. For each trial, while characters fixed their gaze, the participant was asked: Which object is being

9

Independent variable	Definition	Design	Level	Labels
characters	Participants will evaluate five characters presented on the display.	within-subjects	5	*elephant* *bunny* *toad* *giraffe* *frog*
anamorphic character	Only one of the five characters are selected to be rendered from anamorphic perspective for each group of participants.	between-subjects	5	*anamorphic elephant* *anamorphic bunny* *anamorphic toad* *anamorphic giraffe* *anamorphic frog*
viewers	Each group of four participants with different viewing positions are defined as viewers.	within-subjects	4	*viewer A* *viewer B* *viewer C* *viewer D*
anamorphic views	Only one of the four viewers are selected to see the undeformed image as the VIP viewer in each round.	within-subjects	4	*anamorphic view A* *anamorphic view B* *anamorphic view C* *anamorphic view D*

Table 1: A summary of the independent variables

looked at? Are you being looked at? Or asked to rate engagement scores, respectively. These three aspects were rolled into our experiment as three sections.

We included 5 characters (see Figure 3), including elephant $(-40, 10)$, bunny $(-20, 10)$, toad $(0, 10)$, giraffe $(20, 10)$ and frog $(40, 10)$. Note that characters are assigned at different locations of the screen, instead of being assigned to the center of the screen.

We looked at 5 anamorphic character conditions. Only one of the characters are rendered with anamorphic perspective, namely, *anamorphic elephant, anamorphic bunny, anamorphic toad, anamorphic giraffe* and *anamorphic frog.*

We explored four viewer conditions with different viewing positions (see Figure 3 A, B, C & D). We included viewer B with that viewer at the center position as a benchmark; viewer A and D where viewers sat at two extreme symmetric viewing positions; and viewer C where the viewer sat in between viewer B and viewer D.

We compared four anamorphic views, namely anamorphic view A $(-120, -250)$, anamorphic view B $(0, -250)$, anamorphic view C $(60, -250)$ and anamorphic view D $(120, -250)$, from which the VIP viewer can see the undeformed image. Thus, the VIP viewer would be the viewer position lined up with its corresponding view.

4.1. Hypotheses

4.1.1. Hypothesis 1 (H1)

For the anamorphic character, we expect the VIP viewer will be able to identify more correct targets compared to the non-VIP viewer.

4.1.2. Hypothesis 2 (H2)

For the anamorphic character, we expect that the VIP viewer will be able to better discriminate whether the characters are looking at them or not, than the non-VIP viewer.

4.1.3. Hypothesis 3 (H3)

For the anamorphic character, we further expect that the VIP viewer will result in a higher level of engagement than the non-VIP viewer

11

4.2. Method

4.2.1. Participants

We recruited 40 participants from Disney Research to take part as viewers in our user study. The median age was 28.2 ($SD = 8.27$). All participants had normal or corrected to normal eye sight.

4.2.2. Apparatus and materials

Figure 3 shows the layout of the experiment room. The distance from viewpoint B to screen and the target 7 were 250cm and 120cm, respectively. We arranged 13 table number cards as potential target. There was about 20cm separation between each target on the table. Each character was 20cm apart from one another in the scene.

For object-focused gaze task, we created 7 visual stimuli such that characters look at 7 targets out of 13 potential targets in a prearranged random order (Order: 7, 9, 6, 8, 5, 10, 4). Note that the range of potential targets was larger than the range of actual targets, enabling the quantitative investigation of bias in viewers perceived targets. For each of the stimuli, all characters looking at the same target. A new target was given every 30 seconds. Each target was gazed at only once, amounting to 7 visual stimuli.

For mutual gaze task, we further created 4 visual stimuli such that characters look at targeted participants in a prearranged random order (Order: viewer A, viewer D, viewer C, viewer B). Technically, the gaze performance was controlled by a Look-At blueprint in Unreal Engine orienting each character toward a target object every time the target object moving. The Look-At blueprint was created by our artists enabling stylized gaze shifts by a set of parameters (e.g., eye size and motor range) that adapt the target pose and gaze shift dynamics to the character's specific features, thus reducing artifacts and supporting a lifelike gaze motion performance.

For user engagement task, we also created four 120s animation clips that characters perform Zumba dance motions with the anamorphic character look at the VIP viewer in a prearranged random order (Order: viewer A, viewer D, viewer C, viewer B). We used Optitrack motion capture system capturing the body movements of a actor performing each dance. We retargeted the animations to animate the stylized characters and then layered the gaze performance to the body animation in Unreal.
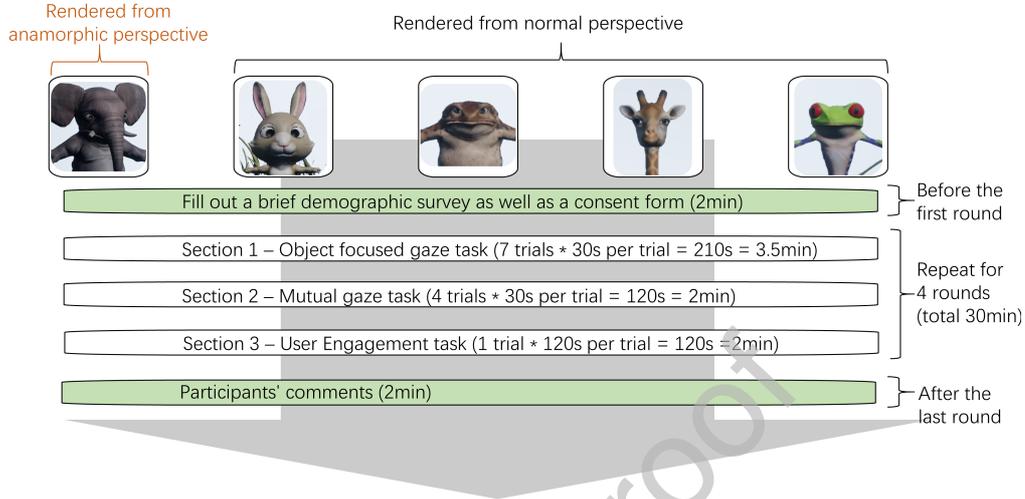
Figure 4: The timeline of the experimental protocol. Participants took part in groups of four. One of the five characters are selected to be rendered from anamorphic perspective for each group. One of four participants is selected to see the undeformed image as the VIP viewer in each round. The group performed four around. The VIP viewer stays the same on a per-round basis. On each repetition the VIP viewer changes to another viewer.

4.2.3. Design

The experiment had a 5 anamorphic character conditions × 4 views × 4 viewers × 5 characters mixed design, with a *within-subjects* design for views, viewers and characters, but a *between-subject* design regarding anamorphic character conditions. Using a counterbalanced measures design, we mixed the four view conditions in order to reduce any confounding influence of the orderings such as learning effects or fatigue.

4.2.4. Procedure

The timeline of the protocol is illustrated in Figure 4.

On arrival, each group of four participants experienced one of five anamorphic characters conditions. Each participant sat at one of the four viewpoints (see Figure 3). Participants stay in their spot throughout the study.

The following procedure was repeated 4 times with different anamorphic viewing conditions: Each participant was asked to fill out an answer sheet which consists of three sections, an empty grid of 35 squares (7 targets × 5 characters) for object-focused gaze, an empty grid of 35 squares (7 targeted × 5 characters) for mutual gaze and a questionnaire for user engagement.

13

- Section 1 Object-focused Gaze: The characters reoriented to a new target every 30 seconds. At the same time an audio prompt to the participants instructed them that this was a new target position. Then, participants would judge which target each character was gazing at and then write this in the relevant grid square. This process was repeated 7 times.

- Section 2 Mutual gaze: Similar to section 1, except that instead of gazing at each of the targets, the characters were gazing at the participants. This process was repeated 4 times.

- Section 3 User Engagement: Participants were asked to watch the characters perform the dance for 120 seconds, and answer the questionnaire.

There is no discussion allowed during the task, and the participant cannot see others' judgments. The experiment took about 30 minutes.

4.2.5. Data Collection & Scoring

Object-focused Gaze. The primary measurement in our results was the error in perceiving targets. We defined the error of each target (ϵ_i) to be the absolute value of a difference between the observer perceived target (t_{pi}) and the actual target (t_{ai}). Thus,

$$\epsilon_i = |t_{pi} - t_{ai}|$$

Mutual Gaze. For each character, viewers were asked, "Do you feel the character X is looking directly into your eyes?" We count the number of times ("Yes") a viewer replied positively as to whether or not they felt the character was looking directly into their eyes, and the number of times ("No") a participant replied negatively to that same question.

User Engagement. To assess the level of engagement, participants were presented with a questionnaire that consisted of ten items (see Table 2), each with an associated 1 - 7 Likert scale, where an answer of 1 indicated complete disagreement and 7 indicated complete agreement. The questionnaire is developed based on a previous work [41], capturing several dimensions of engagement, such as, focused attention, perceived usability, aesthetic appeal, and reward. Reverse code items include S2 and S10. The overall engagement score can be calculated by adding all of the items together and dividing by ten.

14

Table 2: User engagement scale (seven-point Likert scale)

No.	Questionnaire Item
S1	I lost myself while watching character X dancing.
S2	I felt frustrated while watching character X dancing.
S3	Character X always looked at me at the appropriate times.
S4	My experience of watching character X dancing was rewarding.
S5	Watching character X dancing was aesthetically appealing.
S6	The time I spent watching character X dancing just slipped away.
S7	I feel character X look at me while dancing.
S8	I felt interested in watching character X dancing.
S9	Watching character X dancing appealed to my senses.
S10	Watching character X dancing to learn samba dance was taxing.

Table 3: The mean error in perceiving targets.

	VIP viewer	Non-VIP viewer
Anamorphic character	$(M = 1.096, SD = .566)$	$(M = 2.881, SD = 1.166)$
Normal character	$(M = 2.112, SD = 1.003)$	$(M = 2.117, SD = .935)$

4.3. Results

4.3.1. Object-focused Gaze

Table 3 shows the mean error over the anamorphic character and the non-anamorphic ones for both the VIP viewer and the non-VIP viewers. Units are in one target difference. For overall (the anamorphic character + the non-anamorphic ones) mean error, a paired-samples t-test was used to determine whether there was a statistically significant mean difference between the VIP viewer and the non-VIP viewers. The assumption of normality was not violated, as assessed by Shapiro-Wilk's test (p = .780). Results revealed the mean error of the non-VIP viewers $(M = 2.27, SD = 1.031)$ are not significantly more than VIP viewer $(M = 1.908, SD = 1.039)$, $t(199) = .867, p = .387$. However, for the VIP viewer, a Wilcoxon signed-rank test was conducted to determine the effect of rendering character from anamorphic view vs. normal view on the mean error in perceiving targets. The difference scores were approximately symmetrically distributed, as assessed by a histogram with superimposed normal curve. Results revealed the character rendered from anamorphic view $(Mdn = 1.000)$ has significantly lower mean error than the character rendered with normal view $(Mdn = 2.196)$, $z = 5.331, p < .001$. Thus, hypothesis H1 is supported.

Figure 5 shows the mean error present differently for each character and each viewer position in each anamorphic character condition. Firstly, The
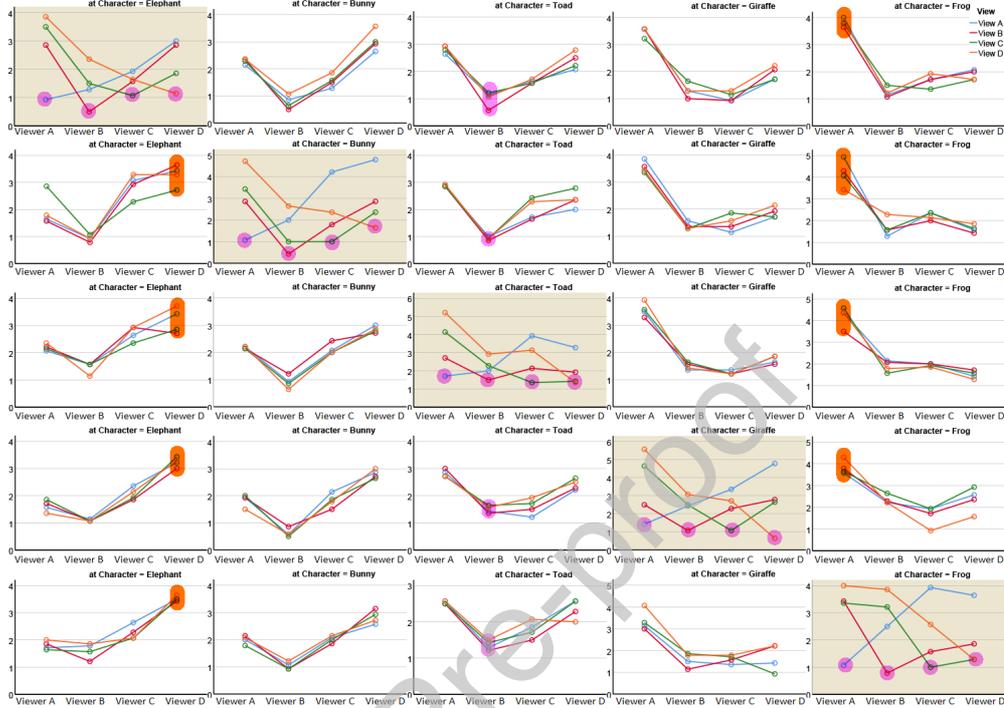
Figure 5: The mean error in perceiving targets for each character in each anamorphic character condition. Units are in one target difference. Each row represents each anamorphic character condition, and each column represents each character. The sub-figures with khaki background color show the mean error for the particular character rendered in anamorphic perspective.

interpretations of the results in these five sub-figures with khaki background color were similar, and the VIP viewer always achieved the lowest mean error (highlighted in pink) for the particular character rendered in anamorphic perspective. Secondly, for the toad located at the center of the screen, the viewer B achieved the lowest mean error and the error increased symmetrically as the viewer position diverged from the central (see the middle column of figure 5, highlighted in pink). This parallels the previous findings [14, 16, 12]. By contrast, when characters placed at off-center locations, the level of error will increase as the character location diverges from the viewer location (see the leftmost or rightmost column of figure 5, highlighted in orange).

A 5 anamorphic character conditions $\times$ 4 anamorphic views $\times$ 4 viewers $\times$ 5 characters mixed design ANOVA was conducted on the error in per-

ceiving targets. The data were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). There were two outliers, as assessed by inspection of a boxplot. The outliers were kept in the analysis because they did not materially affect the results as assessed by a comparison of the results with and without the outlier. There was homogeneity of variances, as assessed by Levene's test for equality of variances ($p > .05$). For all effects, Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p > .05$). Results revealed a significant main effect of viewer positions, $F(3, 15) = 43.178, p < .001$. Bonferroni post-hoc comparisons indicated the mean error for the viewer B ($M = 1.443, 95\% \ CI \ [1.039, 1.847]$) is significantly lower than the mean error for viewer A ($M = 2.906, 95\% \ CI \ [2.665, 3.148]$), $p < .001$,and viewer D ($M = 2.429, 95\% \ CI \ [2.221, 2.634]$), $p < .001$. However, the mean error for viewer A did not significantly differ from viewer D, ($p > .05$).

4.3.2. Mutual gaze

viewer	VIP viewer with	Non-VIP viewer			
Character	anamorphic character	A	B	C	D
Elephant	8/32	2/152	1/152	1/152	2/152
Bunny	7/32	3/152	1/152	1/152	1/152
Toad	9/32	37/152	39/152	39/152	38/152
Giraffe	8/32	1/152	0/152	0/152	2/152
Frog	9/32	0/152	0/152	1/152	2/152
Elephant	100.00%	5.26%	2.63%	2.63%	2.63%
Bunny	87.50%	2.63%	2.63%	2.63%	2.63%
Toad	87.50%	0.00%	97.37%	2.63%	5.26%
Giraffe	100.00%	2.63%	0.00%	0.00%	5.26%
Frog	87.50%	0.00%	0.00%	2.63%	2.63%

Table 4: A summary of the results for mutual gaze. Top: The number of times a participant replied positively ("Yes") as to whether or not they felt the character was looking directly at them/ the total number of stimuli. Bottom: Rate at which the participants answered correctly when the character was looking directly at them.

A summary of the results for mutual gaze are given in Table 4. The VIP viewer performed consistently well in this test with the viewer seldom unable to distinguish. In turn of the non-VIP viewers, for the toad (located at the center of the screen), only viewer B could distinguish between being looked
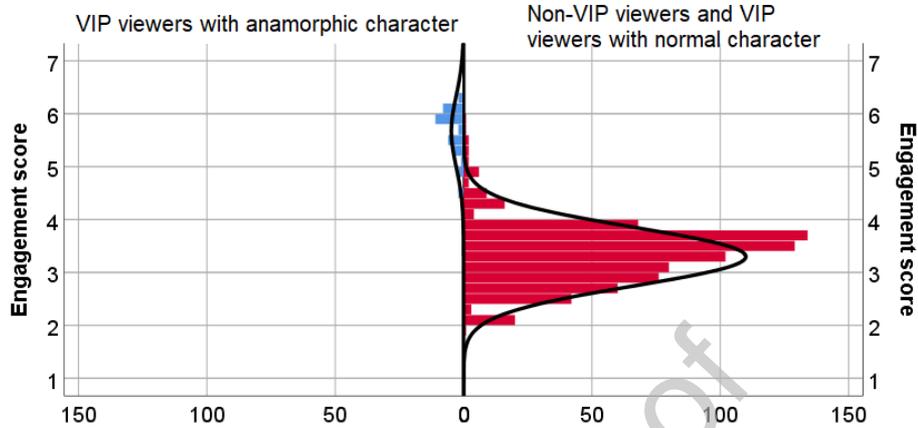
Figure 6: Population pyramid frequency engagement score by viewers with normal curve. VIP viewers viewing anamorphic character $N = 40$; Non-VIP viewers and VIP viewers viewing normal characters $N = 760$.

at and gaze to one side of them. Viewer A, C and D felt the toad was looking directly at them while the toad was looking at viewer B, due to 'Mona Lisa Effect'.

As viewers of the mutual gaze test gave a binary answer, the results were clear, and were in line with those of the more complex object focused gaze discussed above, we did not feel it necessary to analyze for significance.

4.3.3. User Engagement

The responses to each statement item given by each viewer for each iteration were averaged to create an aggregate response. Figure 6 shows the distributions of engagement score.

A Mann-Whitney U test was run to determine if there were differences in engagement score between the VIP viewer and non-VIP viewer. Distributions of the engagement scores for the VIP viewer and the non-VIP viewer were similar, as assessed by visual inspection. Engagement score was was statistically significantly higher in the VIP viewer for the character render from anamorphic view ($Mdn = 5.8$) than in non-VIP viewer ($Mdn = 3.3$), $p < .001$.

18

5. Discussion

5.1. Lesson learned & design recommendations

The most important lesson from this experiment is that the VIP guest results in an improved gaze and engagement estimation for one selected character. This is performed without sacrificing the other guests' viewing experience. Our method could be used for creating interactive experiences in a conventional 2D display. For example, Turtle Talk with Crush, an interactive show at Disney California Adventure Park where guests can chat live with crush the sea turtle. We could render the turtle (the main character) from a VIP guest perspective, while the rest of the scene from a normal perspective. Therefore, the VIP guest can be selected for the anamorphic experience. The VIP view enjoys one clear advantage over a conventional 2D show in the support of gaze: the main character's face is viewed from the perspective of the VIP viewer rather than that of a fixed camera. Also, other guests in the room could enjoy the projection.

Our results also showed that errors in judging a character's gaze direction were increasing with the increasing discrepancy between the character's location and the viewer's position. This suggested that presenting a character on a large screen/video wall has hot spots in which gaze can be easily and accurately discerned. It works well when the viewer is in the very center front position of the character, and stops accurately communicating gaze as soon as the viewer steps to the side of the character. Thus, we could line up viewers in front of the character to improve their gaze and engagement estimation.

5.2. The scope of the controlled experiment & limitation

In designing a controlled experiment we had to make several choices about the situation of the user. In our implementation, we tracked the position of the VIP viewer's head using Microsoft Kinect. As the VIP viewer moved around, the view of the image of the selected objects would adjust accordingly. It does however mean that one or more objects may be selected. In our experiment, to ensure accuracy and stability head tracking was disabled and the participants sat in fixed positions. An operator keyed in the correct position to provide anamorphic view. Only one of the characters was rendered from anamorphic view, and all characters in one line without depth difference. We only tested with one selected object and whether the same effect applies to multiple is yet to be known. Additionally, the character avatar

might have an effect on gaze perception, which could affect the results and are interesting avenues to explore. We also refer back to the related work on gaze estimation: a number of issues arise when applying human gaze models to stylised characters with eye geometry different from that of realistic humans (e.g. [34]). For gaze stimuli, we used the prearranged random order for the sake of convenience, instead of the counterbalanced order. We acknowledge there might be learning effects, as the task was repeated. Thirdly, we assigned characters 10cm "behind"$(+y)$ the screen, and it produced a simple pseudo-3D experience that the selected character is situated within the 2D display for the VIP viewer, by providing motion parallax cues via head position tracking. Lastly, the experiment was designed for four participants simultaneously as a very practical demonstration for a group viewing experience.

5.3. Potential Application of Adaptive Dynamic Anamorphosis & future work

5.3.1. Interactive art installation

We hope that our approach will motivate further discussion for interactive art installation and open the door to an interesting new type of computer generated imagery. Firstly, body posture, pointing and other view dependent viewer oriented content is applicable to this beyond eye contact for enriched viewer engagement. Secondly, with our approach, multiple individuals from each group of audience members can be provided their own gaze engagement with different subjects in the presented scene. Thirdly, wider engagement with small groups within a crowd is also possible, given close spatial proximity of those small group members to one-another, i.e. within a threshold viewpoint cone angle. Lastly, the approach is also suitable for 2D non-planar displays, such as spherical or cylindrical display surfaces. A further aspect of this approach is possible, where the display is reflected, refracted or warped the selective viewer engagement projection can be adapted to still maintain eye contact under these conditions, e.g. for pepper's ghost or anamorphic cylinder, or dynamic water refraction.

5.3.2. Enabling better videoconferencing or Mixed Reality experience

An interesting question is the potential support for video conferencing or mixed reality tele-collaboration. Previous 2D video conferencing fails to support many of these cues because a single camera perspective warps spatial characteristics in group-to-group meetings. These problems can be alleviated by our system. The potential possibility is constructing a live video as

20

a combination of multiple videos captured from different views. The idea comprising, setting up an array of cameras capturing the live green-screen video of remote people, selecting the correct video depending on the viewer's viewing location, chroma key processing, capturing or creating background real or virtual scene from normal view, compositing, and output of the final mixed-reality video for streaming. As a result, when a user who needs to constantly move in his work place is being addressed by somebody by means of his or her video image on a computer screen, anamorphic deformation can take care that this person is always turned towards the appropriate user. The rest of the group could still view the video, with anamorphic distortions remaining potentially imperceptible.

6. Conclusion

We have presented the design, implementation and evaluation of our adaptive dynamic anamorphosis technique. The contributions of this paper are twofold. First, we extend the idea of dynamic anamorphosis for multiple viewers simultaneously, rendered using non-linear projection as a selective combination of offset perspective anamorphic projection and normal perspective projection (section 3). To drive anamorphic projection we need to know the position of the VIP viewer's eyes, so that if the viewer looks at the anamorphic projection he or she sees the image on the selected objects undeformed. Tracking of the viewer can be achieved by different means. In the described system this is achieved using Microsoft Xbox One Kinect and the Kinect 4 Unreal middle-ware plugin to locate the viewer's head position in 3D space and in real time. Our method is technically quite simple to build and can be constructed very cheaply in comparison to volumetric displays, multiview displays or virtual reality systems. Second, supported through by empirical experiment we show that our method could improve gaze and engagement estimation for the VIP viewer, while other viewers' viewing experience (e.g., gaze and user engagement score) is no worse than traditional displays. One interesting lesson from this experiment is for the the perception of gaze direction for characters located off-center to the origin on the screen: such that errors in judging a character's gaze direction were increasing with the increasing discrepancy between the character's location and the viewer's position.

With the advent of interactive rendering techniques the question came into play how to create images and interfaces which better reflect and support

21

human perception and viewing behaviour. Our methods and findings have significant implications for art installation in general. By adopting dynamic anamorphosis, our method allow perspective-correct imagery to be seen for the VIP viewer, and hence avoid the problems we have observed with traditional flat displays. By using nonlinear projection technique, we still provide reasonable imagery for other viewers. Altogether, the VIP viewer is able to maintain a consistently high level of engagement regardless of viewing position, while other viewers' viewing experience would not be sacrificed. Eye contact and viewer engagement are fundamental parts of human interaction and we intend to explore other important scenarios and natural interaction in future work.

References

[1] D. J. Roberts, J. Rae, T. W. Duckworth, C. M. Moore, R. Aspin, Estimating the gaze of a virtuality human, IEEE transactions on visualization and computer graphics 19 (2013) 681–690.

[2] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, M. Billinghurst, Revisiting collaboration through mixed reality: The evolution of groupware, International Journal of Human-Computer Studies 131 (2019) 81–98.

[3] S. Al Moubayed, J. Edlund, J. Beskow, Taming mona lisa: communicating gaze faithfully in 2d and 3d facial projections, TiiS 1 (2012) 25.

[4] M. Hancock, M. Nacenta, C. Gutwin, S. Carpendale, The effects of changing projection geometry on the interpretation of 3d orientation on tabletops, in: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, 2009, pp. 157–164.

[5] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, P. Debevec, Achieving eye contact in a one-to-many 3d video teleconferencing system, in: TOG, volume 28, ACM, 2009, p. 64.

[6] J. C. Lee, Head tracking for desktop vr displays using the wii remote, Published online at http://www. cs. cmu. edu/johnny/projects/wii (2007).

[7] K. W. Arthur, K. S. Booth, C. Ware, Evaluating 3d task performance for fish tank virtual worlds, ACM Transactions on Information Systems (TOIS) 11 (1993) 239–265.

[8] R. Ravnik, B. Batagelj, B. Kverh, F. Solina, Dynamic anamorphosis as a special, computer-generated user interface, Interacting with Computers 26 (2014) 46–62.

[9] A. Fathi, J. K. Hodgins, J. M. Rehg, Social interactions: A first-person perspective, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 1226–1233.

[10] D. Man, A. Vision, A computational investigation into the human representation and processing of visual information, 1982.

[11] C. Ware, G. Franck, Evaluating stereo and motion cues for visualizing information nets in three dimensions, ACM Transactions on Graphics (TOG) 15 (1996) 121–140.

[12] Y. Pan, A. Steed, Effects of 3d perspective on head gaze estimation with a multiview autostereoscopic display, International Journal of Human-Computer Studies 86 (2016) 138–148.

[13] J. C. Lee, Hacking the nintendo wii remote, IEEE Pervasive Computing 7 (2008) 39–45.

[14] D. Nguyen, J. Canny, Multiview: spatially faithful group video conferencing, in: SIGCHI, 2005, pp. 799–808.

[15] D. Gotsch, X. Zhang, T. Merritt, R. Vertegaal, Telehuman2: A cylindrical light field teleconferencing system for life-size 3d human telepresence, in: SIGCHI, ACM, 2018, p. 522.

[16] Y. Pan, A. Steed, A gaze-preserving situated multiview telepresence system, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2014, pp. 2173–2176.

[17] K. Kim, J. Bolton, A. Girouard, J. Cooperstock, R. Vertegaal, Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod, in: SIGCHI, ACM, 2012, pp. 2531–2540.

[18] Y. Pan, O. Oyekoya, A. Steed, A surround video capture and presentation system for preservation of eye-gaze in teleconferencing applications, Presence 24 (2015) 24–43.

[19] Y. Pan, W. Steptoe, A. Steed, Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction, in: SIGCHI, ACM, 2014, pp. 1397–1406.

[20] K. Mitchell, Interactive three dimensional displays on handheld devices, US Patent Office (2010).

[21] J. Francone, L. Nigay, Using the user's point of view for interaction on mobile devices, in: Proceedings of the 23rd Conference on l'Interaction Homme-Machine, IHM '11, Association for Computing Machinery, New York, NY, USA, 2011.

[22] P. Coleman, K. Singh, Ryan: rendering your animation nonlinearly projected, in: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering, ACM, 2004, pp. 129–156.

[23] K. Singh, A fresh perspective, in: Graphics interface, volume 2002, 2002, pp. 17–24.

[24] J. Yu, L. McMillan, A framework for multiperspective rendering., in: Rendering Techniques, 2004, pp. 61–68.

[25] N. Sudarsanam, C. Grimm, K. Singh, Non-linear perspective widgets for creating multiple-view images, in: Proceedings of the 6th international symposium on Non-photorealistic animation and rendering, ACM, 2008, pp. 69–77.

[26] T. Beier, S. Neely, Feature-based image metamorphosis, in: ACM SIGGRAPH Computer Graphics, volume 26, ACM, 1992, pp. 35–42.

[27] T. W. Sederberg, S. R. Parry, Free-form deformation of solid geometric models, ACM SIGGRAPH computer graphics 20 (1986) 151–160.

[28] D. N. Wood, A. Finkelstein, J. F. Hughes, C. E. Thayer, D. H. Salesin, Multiperspective panoramas for cel animation, in: Proceedings of the 24th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., 1997, pp. 243–250.

[29] P. Moors, K. Verfaillie, T. Daems, I. Pomianowska, F. Germeys, The effect of head orientation on perceived gaze direction: revisiting gibson and pick (1963) and cline (1967), Frontiers in psychology 7 (2016) 1191.

[30] J. J. Gibson, A. D. Pick, Perception of another person's looking behavior, The American journal of psychology (1963) 386–394.

[31] S. M. Anstis, J. W. Mayhew, T. Morley, The perception of where a face or television'portrait'is looking, The American journal of psychology (1969).

[32] D. Todorović, Geometrical basis of perception of gaze direction, Vision research 46 (2006) 3549–3562.

[33] T. Pejsa, B. Mutlu, M. Gleicher, Stylized and performative gaze for character animation, in: Computer Graphics Forum, volume 32, Wiley Online Library, 2013, pp. 143–152.

[34] K. Ruhland, S. Andrist, J. Badler, C. Peters, N. Badler, M. Gleicher, B. Mutlu, R. Mcdonnell, Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems, in: Eurographics state-of-the-art report, 2014, pp. 69–91.

[35] N. Glas, C. Pelachaud, Definitions of engagement in human-agent interaction, in: International Workshop on Engagment in Human Computer Interaction (ENHANCE), 2015, pp. 944–949.

[36] C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, I. Poggi, A model of attention and interest using gaze behavior, in: International Workshop on Intelligent Virtual Agents, Springer, 2005, pp. 229–240.

[37] M. Deering, High resolution virtual reality, SIGGRAPH Comput. Graph. 26 (1992) 195–202.

[38] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, Surround-screen projection-based virtual reality: the design and implementation of the cave, in: Proceedings of the 20th annual conference on Computer graphics and interactive techniques, 1993, pp. 135–142.

[39] R. Kooima, Generalized perspective projection, J. Sch. Electron. Eng. Comput. Sci (2009).

[40] N. Reed, Depth precision visualized, NVIDIA Gameworks (2015).

[41] H. L. O'Brien, P. Cairns, M. Hall, A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form, International Journal of Human-Computer Studies 112 (2018) 28–39.

CRediT author statement

Ye Pan: Conceptualization, Methodology, Software, User study, Writing- Original draft preparation. **Kenny Mitchell:** Supervision, Conceptualization, Methodology, Software, Writing- Reviewing and Editing.

AUTHOR DECLARATION TEMPLATE

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We further confirm that any aspect of the work covered in this manuscript that has involved either experimental animals or human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from k.mitchell2@napier.ac.uk.

Signed by all authors as follows:
Ye Pan and Kenny Mitchell
6/26/2019