Design Considerations of a Robotic Head for Telepresence Applications

Wee Ching Pang, Burhan, Gerald Seet

BeingThere Centre, Institute for Media Innovation, Nanyang Technological University
BeingThere Centre, Institute for Media Innovation, Research Techno Plaza,
XFrontier Block Level 03-01, 50 Nanyang Drive, Singapore 637553
WEECHING@ntu.edu.sg; BURHAN@ntu.edu.sg

Robotics Research Centre, School of Mechanical and Aerospace Engineering
Nanyang Technological University, 50 Nanyang Avenue, N3-01a-01, Singapore 639798
MGLSEET@ntu.edu.sg

Abstract. This work attempts to enhance telepresence by empowering nonverbal communication i.e. head gestures communication with an implementation of a robotic rear-projection head device. The feature of this implementation which is considered to be novel is the containment of all the necessary components for rear-head projection within the volume of a typical human head. This provides a natural look to the rear-projection head and facilitates interaction. The resultant head is capable of accurately projecting human facial features.

Keywords: Rear-Projection, Robotic head, Telepresence, Face robot, Human-robot interaction

1 Introduction

Since humans began interacting with one another, direct face-to-face interaction has invariably been the interaction mode of choice. In addition to facilitating communication via speech, face-to-face interaction allows facial expressions or gestures to provide subliminal information that augment conscious communication efforts [1] [2]. When gestures are used in conjunction with speech, information is conveyed via the visual as well as the auditory sensory channels. In this manner, communication can become more efficient and easily understood. Such use of the visual and auditory sensory channels to effectively convey information has also been applied in user interface design for human interaction with rescue robots [3].

While speech is one of the first things to come to mind when considering face-to-face interaction, it is worth noting that a typical human face provides an array of visual information such as age, gender, ethnicity, line-of-work, status as well as moods and emotions. Small variations in expressions and gestures can significantly affect communication. Furthermore, eye movements and gazes [4] can also signal our perceptions of the people we are interacting with.

With the prevalence of the internet culture and the advances of technology, there is resurgence in the desire to replace direct face-to-face interaction with one that can make do without the need for time consuming international travel. Telepresence [5] [6] refers to a set of technologies that enables geographically separated individuals to interact effectively with all the sensations and advantages of actually being at the remote site. This has been demonstrated by the use video-mediated communications such as video conferences [7] [8] within boardrooms [9] and desktop video interaction tools such as Skype and Google Talk [10]. Although such communication via video screen provides a richer interaction experience, it lacks a number of spatial and perceptual cues. It is not easy to notice peripheral cues, control the floor when interacting with a large audience, have side conversations, point to things and manipulate real-world objects when communicating via video [11].

Fig. 1. Rear-Projection Robotic Head for MAVEN

Our work attempts to further enhance telepresence by empowering two aspects of nonverbal communication, i.e. head and hand gestures communication, as shown in Fig. 1. In this paper, the focus is on the implementation of a robotic head device capable of expressing basic human emotions and head gestures. Whilst a flat video image can convey facial gestures, viewing of a physical 3D head with the addition of textual features can be expected to provide a more rewarding experience. The ability to view the sides of the head coupled by its physical motion appears to be favored over a flat video image. In other words, a robotic head can provide a flexible solution for facial representations and expressions as well as head gestures during communication, taking into advantage the advancements of image processing techniques over mechanically controlled head.

The scope of this paper includes the engineering realization of hardware - that is a "complete" rear-projection robotic head for telepresence communication, as well as the reconstruction of face texture from a 2D image. The development of the robotic head is part of a larger effort for developing a humanoid robot for telepresence applications. Currently, a mobile holonomic platform called MAVEN (Mobile Avatar for Virtual Engagement by NTU) has been developed [12]. Navigation with the

MAVEN platform is possible via tele-operation as well as with autonomy. MAVEN can display the human operator (or inhabitor) on a flat screen TV mounted on the robot. The robotic head would provide an alternative to this if mounted on MAVEN as it would then be possible to project the inhabitor's facial features onto the head's face as shown in Fig. 1.

2   Related work

Among the various attempts to build a robotic head or face, Ishiguro [13] and Hanson [14] have tried to use solid silicon elastomers to build robotic heads that emulate human faces realistically. Facial expressions and emotions are articulated via pneumatic actuators and electro-actuated polymers respectively. However, such robotic android head is not generic, it has to be custom-made to look like a particular person and others will not be able to use the same head to represent themselves during a conversational communication. Furthermore, the implementation cost, as well as the maintenance cost of these realistic robotic heads is exorbitant.

Therefore, an alternative to robotic android head is to project an imagery of a face onto a surface that shaped like a human head. Disneyland™ [15] has pioneered the use of a front projection technique in its Haunted Mansion™ attraction to animate the Madam Leota figure. The front projection technique involves projecting the film imagery of a face directly onto the front of the head bust from a projector in front of the face. Academically, Lincoln et al. [16] has implemented an animatronics shader lamps avatar head using the front projection technique as well. Rather than projecting an animated or filmed imagery of a human, their system uses cameras to capture the appearance of a human user and map onto a Styrofoam head. The head is mounted on a pan-tilt gimbal such that the head's movement is driven by the actual head movement of the human user.

The disadvantage of a front-projection robotic head is that any movement of the head with respect to the projector will cause the projection to be out of focus or mismatched. Another drawback of a front projection head is that there must be a clear path between the projector and the head because any object that block the path of projection will cast a shadow on the head. This makes it impractical for communicating or interacting purposes. Several groups have tried to implement the projection head technique by designing a rear projection head such that the imagery of a face is projected onto the back of a translucent head. Typically, to project an image to a small projection surface (about 33cm width), the throw distance between the projector and screen surface would be very long depending on throw ratio. However, installing the projector at a large distance from the back of the face will not solve the shadow casting problem faced by frontal projection; it also results in an unnatural and irregular shaped head. Therefore, there is a motivation to hide the projector within the head or as near to the face as possible.

One of the earliest rear projection head is the "Talking Head Projection" [17] [18] implemented as an experimental work in telepresence in 1980. A large plane mirror is mounted behind the head, adapting to keep the projector near the head while projecting an image on the face. The head movement, as well as the audio video information, of a human user is first recorded on film, before being played back onto

the head that has a face-shaped screen mounted in a pan-tilt unit. Instead of using mirrors, Delaunay et al. [19] discussed the potential of using a small pico-projector with shorter throw ratio while the Disneyland™ [15] and Mask-bot [20] have installed a wide angle lens and fish-eye lens within their rear-projection heads to adjust the focusing as well as the throw distance of the projection respectively. All of the above mentioned rear-projection heads used the head as a device to play back recorded film (e.g. [17] [18]) or to display an animated cartoon (e.g. [15], [19]) or texture face (e.g. [20]). All these rear-projection heads, except that by Disneyland, are still constrained by a couple of outstanding issues needing resolution. The projector and lens configurations invariably require the rear of the robot head to be extended beyond the physical limits attributed to a human head. The rear of the head is incomplete and typically masked by scarf, providing for an unnatural setting.

3   Hardware configuration of the rear-projection head
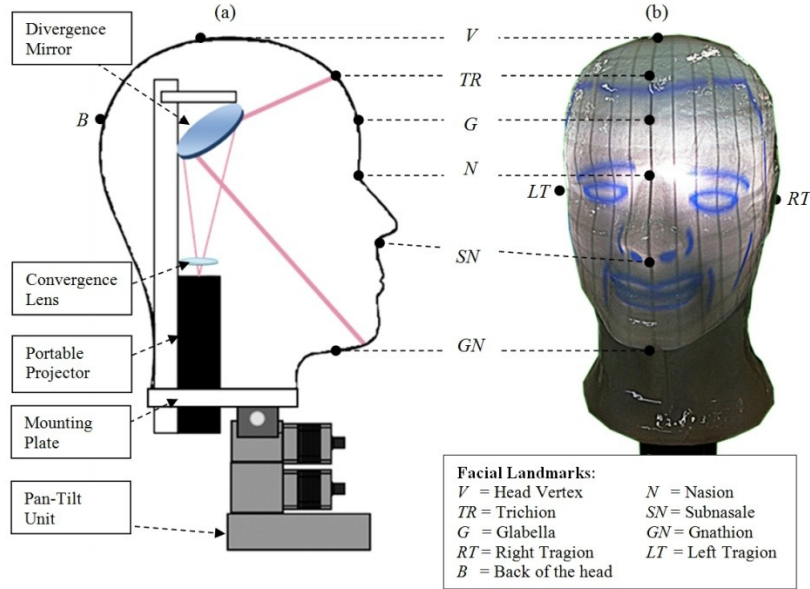

Fig. 2. (a) Hardware configuration of the rear-projection head system (b) Resultant head when a reference image is projected on the head screen

There is a motivation to produce a natural and realistic head that can be used to represent any human user during a telepresence interaction session. This would mean that the head is "completed" with the entire projection module encapsulated within the head and have no protruding components. This is expected to make the head look more natural. An overview of the hardware configuration can be seen in Fig. 2.

The hardware consists of three main components: the head screen, the projector system, and the motor-controlled base. Fig. 2(b) illustrates the resultant head when a reference image is projected on the head screen. The various head landmarks are also labeled on the figure.

3.1   Head Screen

The head screen has been developed using the vacuum forming technique on transparent APET (Amorphous Polyethylene Terephthalate) sheets. After the vacuum forming process with the mould of a face, a coat of matt paint has been applied to the inner surface of the head. This is to diffuse the light from the projector. The rear of the head was made by vacuum forming a black APET sheet with a mould of the back of a head. Because the APET sheets are thin, a frame in the shape of the head has been cut out of sheet plastic of approximately 2 cm in thickness to provide rigidity. Both the head screen and the back of the rear projector head have been mounted onto this frame.

Table 1. Anthropometric measurements of the rear-projection robotic head

Symbol	Description	Tool	Measurement (mm)
G-B-G	**Head circumference:** Surface distance from above the ridges of the eyebrows and around the back of the head	Tape	635
V-GN	**Head Length:** Axial distance between the top of the head and the chin	Caliper	211
B-G	**Head Depth:** Axial distance between the back of the head and the glabella landmark.	Caliper	230
LT-SN-RT	**Bitragion subnasale arc:** Surface distance between the left and right tragion landmarks across the bottom of the nose	Tape	277
LT-GN-RT	**Bitragion chin arc:** Surface distance between the left and right tragion landmarks across the anterior point of the chin	Tape	343
LT-G-RT	**Bitragion frontal arc :** Surface distance between the left and right tragion landmarks above the ridges of the eyebrows	Tape	280
TR-GN	**Face Length:** Axial distance between the hairline to the chin	Caliper	167
LT-RT	**Face Width:** Axial distance between the left and right tragion landmarks	Caliper	146

The summary of the anthropometric measurements of the head screen is as illustrated in the Table 1. A pair of calipers is used to measure the shortest distance between two facial landmarks, while a plastic measuring tape is used to measure surface distance between two landmarks. The depth of the head, measured from the back of the head (*B*) to the glabella landmark (*G*), is 23 cm. It is 4 cm longer than that of an average male head because of the additional plastic frame. The circumference of the head is around 63 cm while the length of the head is about 21 cm. The head length is measured from the top of the head (*V*) to the chin (*GN*) along the y-axis of the canonical coordinate as shown in Fig. 2. All equipments must be aligned suitably within the volume defined by the circumference, depth and length of the head. The length of the face, on the other hand, is measured from the hairline (*TR*) to the chin (*GN*). It is 16.7 cm, implying that the height of the projection area should be around 17 to 21cm.

The width of the projected face image must cover the width of the face. The face width (bitragion length) is measured as 14.6 cm along the x-axis of the canonical coordinate, while the various bitragion arc lengths such as the bitragion subnasale arc, bitragion frontal arc and bitragion chin arc are 27.7 cm, 28 cm and 34.3 cm respectively. Therefore, it is concluded that the width of the projection area should be around 27 cm to 35 cm.

3.2   Projector System

The projector used in this effort is the MP160 pocket projector from the 3M Corporation. It is capable of displaying at 32 ANSI lumens and weighs 0.3 kg. The dimension is 15cm by 6cm by 3cm. The throw distance of the projector ranges from 20 cm to 240 cm, and the diagonal of the projected image can range from 17 cm to 200 cm. Hence, to project an image within a short distance to suit the current application, some modification of the optics was necessary.
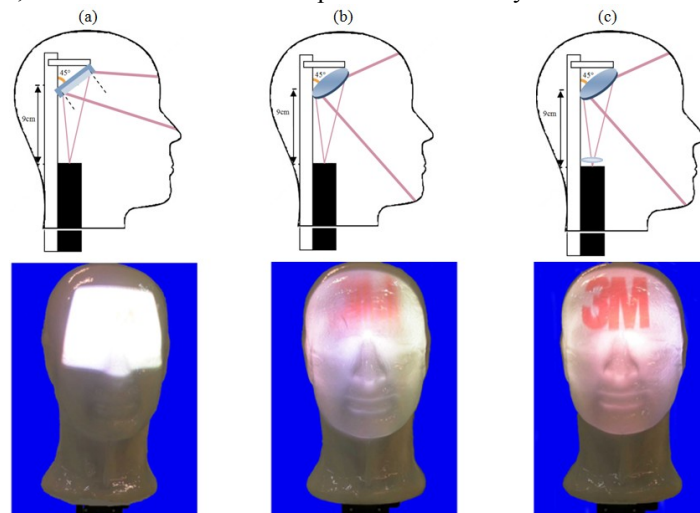

Fig. 3. Projector Setup and the projected image on the screen when (a) a normal mirror is used (b) a divergence mirror is used and (c) a divergence mirror is used with convergence lens

In this effort, the modification was performed by first increasing the projected image's distance from the projector to the head screen with the aid of a small mirror as shown in Fig. 3(a). The mirror is inclined at an angle of 45° from the vertical support bar. The distance between the center of the mirror and the projector lens is 9 cm and the distance between the center of the mirror and the head screen is 17 cm. In this manner, the throw distance is 26 cm and this would result in a projected area where its width is equal to 21 cm and its height is 13 cm as shown in Fig. 3(a).

As the projected area does not fit the entire face, a divergence spherical mirror is used, instead, to assist in magnifying the projected image so that it fits the screen. The mirror is of round shape and its diameter is 10 cm. The curvature is approximately 5 cm and the focal length of the mirror is estimated to be 5.5 cm. The result is as shown in Fig. 3(b).

The magnification of the mirror, M, is calculated to be 1.57, where $M = -\frac{image\ distance}{object\ distance} = -\frac{D_i}{D_o} = -\frac{-14.14}{9} = 1.57$ since $\frac{1}{D_i} = \frac{1}{D_o} - \frac{1}{f} = \frac{1}{9} - \frac{1}{5.5} = -0.07$ and therefore $D_i = -14.14$. The height of the projected screen is about 20 cm and the width of the projected screen is about 33 cm. However, magnifying the image diminishes its quality as shown in Fig. 3(b). To address this, a convergence lens has been added to focus the image. The convergence lens is approximately 3 cm wide and has a focal length of 25 cm. It is mounted 1 cm above the projector lens. The final projection result is as illustrated in Fig. 3(c).

3.3   Motor-controlled Base (Pan-Tilt Unit (Model PTU-D47))

The above head assembly has been installed on a pan-tilt unit (PTU) to allow for motion and gesturing. The PTU used is the PTU-D47 from the Directed Perception Corporation. Control of the PTU is achieved with a laptop computer via a serial connection.

4   Construction of the Human Face Texture

A texture of a human face has been rendered for display from the rear of the projection head. As the screen on the head is in the shape of a human face, it is not a flat 2D screen and there is a need to warp the texture of any human face image to fit the screen's contours. The process to this involves two key stages.

The first stage involves the generation of a reference image. This requires three steps. Firstly, markings are made on the outer surface of the rear projection head's screen. These markings will note locations on the screen that correspond to facial features such as the hairline, eyebrows, eyes, the cheekbones, the nose and the mouth. The second step involves the display of a grid from the projector within the rear-projection head onto the head's screen. The third step in generating the reference image is to note the corresponding locations of the facial features on the displayed grid. To do this, an image editor is used to make markings on the grid. The aim is to match the markings on the grid to correspond with those made previously on the head's screen. The marked grid then becomes the reference image. Fig. 2(b) illustrates the result when the reference image is projected on the head's screen.

The second stage is the morphing of the source image using the reference image for accurate display on the head's screen. A source image is basically the texture image of a person's face. Features on a source image will be corresponded to those on the reference image. The source image is then warped, via linear mapping, to produce the resultant image. This resultant image is then displayed on the head's screen, as shown in Fig. 1 and Fig. 4. To display another person's face, only stage two has to be repeated. Therefore, features on the new source image are then to be corresponded to those on the reference image and a new resultant image is produced.

The display of the human face texture described so far only allows for the display of a static image. That means that movements on the face such as those for eyelids and lips when a user blinks or talks are not shown on the screen. As such, the morphing of specific parts of the resultant image has also been developed. Such

morphing of the eyes and mouth of the resultant image produces movements on the screen to mean blinking or speech. At this point, work is being conducted to track facial movements on the inhabitor so that blinks and mouth movements made by him or her in real time can be correspondingly displayed with the face texture.

5   Result and Discussion

A rear-projection robotic head has been constructed for the purpose of telepresence communication, as shown in Fig 4. This head would be used to represent any male human user during a telepresence interaction session. The main strengths of this rear-projection head are its cost of manufacture as well as its flexibility. The cost of designing and building the rear-projection face is considerably low, due to rapid prototyping and vacuum forming production techniques used for making the translucent head mask, as well as the inexpensive commercially-off-the-shelf portable projection technology. The face of the robot head can be changed flexibly to represent any human user by changing the face texture projected on the head. Figure 4 illustrates the comparison of the implemented rear-projection robotic head with other robotic heads.
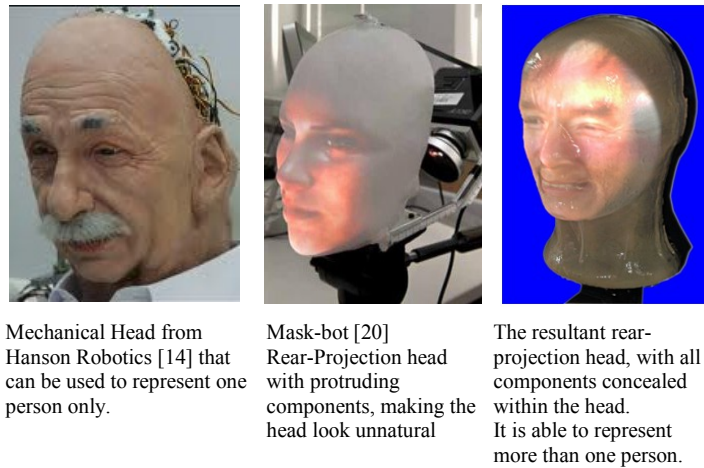

| Mechanical Head from Hanson Robotics [14] that can be used to represent one person only. | Mask-bot [20] Rear-Projection head with protruding components, making the head look unnatural | The resultant rear-projection head, with all components concealed within the head. It is able to represent more than one person. |

Fig. 4. The resultant rear-projection robotic head in comparison to the other robotic heads

5.1   Issues and potential problems

One potential issue with this rear-projection robotic head is that the projector used is a small portable projector which uses LED technology. Although LEDs are commonly used as a lighting solution, the brightness of LEDs constrains most of their applications to indoor environments. Furthermore, the light power of the projector is of 32 lumens which is relatively low when compared to a normal projector which typically have a light power of over 1000 lumens. This causes a dimmer projection and details of the face, such thin wrinkles, freckles or any small facial markings, to be lost during projection. This level of detail can be argued to be unnecessary in most applications.

The technology and design of this rear-projection head would require free space between the head screen and the projector. No opaque objects, such as sensors, can be used or fitted within this space as these would block the projected beam and cause a shadow on the head screen.

Lastly, as the shape of the head is rigid, accurately representing different faces could be difficult. It may be a challenge to recognize people whose faces are more rounded or thinner through the projection on the robotic head. Nevertheless, it is believed that with some training, users should be able to recognize people more easily through the projected face.

6   Conclusion

In this paper, the design and development of a rear-projection head for the purpose of telepresence has been presented. The feature of this implementation which is considered to be novel is the containment of all the necessary components for rear-head projection within the space of the head. This provides a natural look to the rear-projection head and facilitates interaction. The result has shown that the developed head is capable of accurately projecting human facial features. The implemented design is expected to be useful for designers of telepresence robots seeking a way to accurately display the features of a human head. The future work includes experimental evaluation to determine the accuracy in using the rear-projection robot head to represent a human face as well as to determine the effectiveness of using the head to enhance interaction. Future work also includes implementation of a software system to extract facial expressions rather than the user's entire facial texture for display over a generic facial texture on the screen. This will do away with the need for a camera to always point straight towards the user's face.

Acknowledgment

Reference

[1] S. Goldin-Meadow, "Beyond words: The importance of gesture to researchers and learners," *Child Development*, vol. 71, pp. 231--239, 2000.

[2] M. L. Knapp and J. A. Hall, *Nonverbal communication in human interaction*. Wadsworth Pub Co, 2009.

[3] C. Y. Wong, G. L. Seet, S. K. Sim, and W. C. Pang, "A hierarchically structured collective of coordinating mobile robots supervised by a single human," in *Mobile Ad Hoc Robots and Wireless Robotic Systems: Design and Implementation*. IGI Global, 2012.

[4] M. Argyle and M. Cook, *Gaze and mutual gaze.* Oxford, England: Cambridge U Press, 1976.

[5]   M. Minsky, "Telepresence," *Omni*, vol. 2, pp. 45--52, 1980.

[6]   T. B. Sheridan, "Musings on telepresence and virtual presence," in *Presence: Teleoperators and virtual environments*, 1992, pp. 120-126.

[7]   T. Turletti and C. Huitema, "Videoconferencing on the Internet," *IEEE/ACM Transactions on Networking (TON)*, vol. 4, pp. 340--351, 1996.

[8]   W. Buxton, "Telepresence: integrating shared task and person spaces," in *Proceedings of Graphics Interface*, vol. 92, 1992, pp. 123--129.

[9]   T. Szigeti, K. McMenamy, R. Saville, and A. Glowacki, *Cisco Telepresence Fundamentals*. Cisco Systems, 2009.

[10] B. Sat and B. W. Wah, "Analysis and evaluation of the Skype and Google-Talk VoIP systems," *IEEE International Conference on Multimedia and Expo*, pp. 2153--2156, 2006.

[11] E. A. Isaacs and J. C. Tang, "What video can and cannot do for collaboration: a case study," *Multimedia Systems*, vol. 2, pp. 63--73, 1994.

[12] G. G. L. Seet, W. C. Pang, and Burhan, "Towards the Realization of MAVEN - Mobile Robotic Avatar," in *The 25th International Conference on Computer Animation and Social Agents*, Singapore, 2012.

[13] H. Ishiguro, "Android science--Toward a new cross-interdisciplinary framework," in *CogSci-2005 Workshop: Toward Social Mechanismsof Android Science*, vol. 28, Stresa, Italy, 2005, pp. 1--6.

[14] J. H. Oh, et al., "Design of android type humanoid robot Albert HUBO," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 1428--1433.

[15] G. E. Liljegren and E. L. Foster, "Back Projected Image Using Fiber Optics," USA Patent US Patent# 4,978,216, Dec. 18, 1990.

[16] P. Lincoln, G. Welch, A. Nashel, A. Ilie, and H. Fuchs, "Animatronic Shader Lamps Avatars," in *8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2009*, Orlando, FL, 2009, pp. 27--33.

[17] I. Lobel. (2010, Dec.) Inna Lobel. [Online]. http://www.innalobel.com/projects/talkingHeads.html

[18] M. Naimark. (1980) Talking Head Projection. [Online]. http://www.naimark.net/projects/head.html

[19] F. Delaunay, J. de Greeff, and T. Belpaeme, "Towards retro-projected robot faces: an alternative to mechatronic and android faces," in *The 18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2009*, Toyama, Japan, 2009, pp. 306--311.

[20] T. Kuratate, Y. Matsusaka, B. Pierce, and G. Cheng, ""Mask-bot": A life-size robot head using talking head animation for human-robot communication," in *11th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Bled, Slovenia, 2011, pp. 99-104.