2012 IEEE RO-MAN: The 21st IEEE International Symposium on
Robot and Human Interactive Communication.
September 9-13, 2012. Paris, France.

Gender Identification Bias Induced with Texture Images on a Life Size Retro-Projected Face Screen

[1]Takaaki Kuratate, [2]Marcia Riley, [3]Brennand Pierce and [4]Gordon Cheng

***Abstract*— A retro-projected face display system has great advantages in being able to present realistic 3D appearances to users and to easily switch the appearance of the humanoid robot heads animated on the display. Therefore, it is useful to evaluate how effectively users can perceive various information from such devices and what type of animation is suitable for human-robot interaction – in particular, face-to-face communication with robots. In this paper, we examine how facial texture images affect people's ability to identify the gender of faces displayed on a retro-projected face screen system known as Mask-bot. In an evaluation study, we use a female face screen as the 3D output surface, and display various face images morphed between male and female. Subjects are asked to rate the gender of each projected face. We found that even though the output 3D mask screen has a female shape, gender identification is strongly determined by texture images, especially in the case of high-quality images.**

I. INTRODUCTION

Faces are one of the major modalities used in daily human communication, and as such are an essential topic for developing socially aware, interactive humanoid robots. Humans are instinctively and sensitively tuned to faces, and even newborns can detect faces almost instantly [1], [2]. Various realistic robotic heads have been developed, including those with articulated faces. Robot heads by Hanson [3] and Ishiguro [4], and the Jules robot at Bristol labs [5] achieved in collaboration with Hanson are among the best of these realistic articulated faces. However, robotic heads suffer from an important limitation. Because their appearance is fixed after instantiating the design, re-design based on new information is costly, as developers must re-build these mechanically-sophisticated heads.

In contrast to such traditional robotic approaches, various retro-projected face systems have been developed recently [6], [7], [8], [9], including our Mask-Bot shown in Figure 2. These systems bring with them several advantages. They can easily change the appearance of the head along two dimensions: 1) model realism; and 2) model selection. Thus, models can switch along one axis from abstract simplicity to detailed realism obtained by scanning human subjects; and for the same level of realism, different face models can be selected for display. Also, their communication abilities include the capacity to express more nuanced, subtle gestures often missing from today's robot faces. Lastly, the systems are generally lighter and less complicated than their mechanical counterparts, comprised of just a small projector, optics

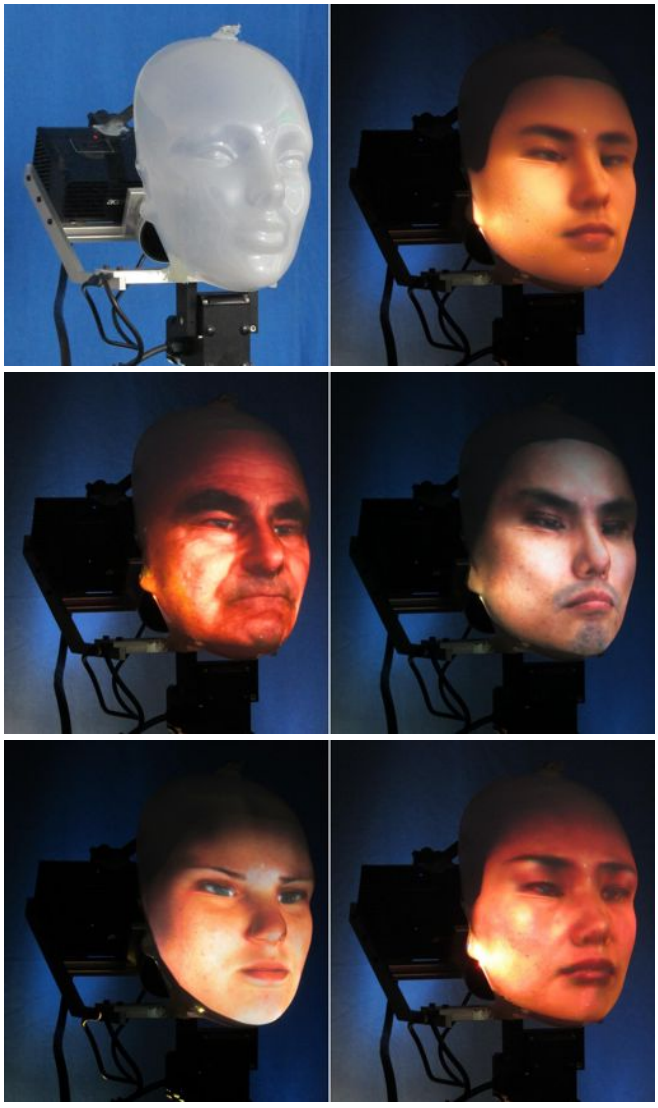


Fig. 1. Mask-bot with different face appearances: Mask-bot without rear-projection (top, left), average face obtained from low quality faces (top, right), male examples in the middle row (Caucasian – left, Asian – right, both in high-quality texture) and female examples in the bottom row (Caucasian with high-quality – left, Asian with low-quality – right).

Institute for Cognitive Systems, Technical University Munich, Karlstrasse 45 / II, 80333 Munich, Germany. http://www.ics.ei.tum.de [1]kuratate@tum.de, [2]mriley@tum.de, [3]bren@tum.de, [4]gordon@tum.de

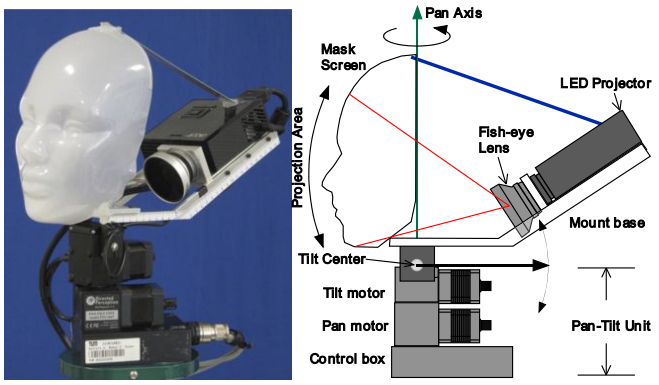


Fig. 2. Example of a retro-projected face - our Mask-bot.

and a face screen, with the face screen yielding better 3D appearance results than flat computer screens.

However, in the current versions of retro-projected face systems, there is one major hardware drawback: it is not easy to replace the display mask for 3D face screens, thus making it difficult to test different screen geometries for compatibility with the source image. We address this drawback in order to make full use of these systems' advantages by considering the balance between output face screen shape and facial appearance. Specifically, we wish to determine which type of 3D facial content is effective for a particular output shape.

Figure 1 shows sample Mask-bot faces. From these examples, we see that the system can project various calibrated face models with the same female screen, and, as seen from examples in the middle row, displaying male faces on a female screen still results in a strong male impression. These observations led us to explore more closely how people identify gender from such retro-projected faces. Various psychological studies on perception of gender report that facial features, skin textures and 3D structure of faces contribute to the classification of gender. Bruce et al. demonstrate that nose/chin protuberance is an important cue in 3/4 views [10], while the eye and brow region become particularly important in front views [11]). These perception experiments used conventional media such as photographs, TV or computer screens to present stimuli. As a new human-robot interface, the retro-projected screen systems need to be examined not only for gender studies, but also more generally on how they can be used effectively as a social tool in human-machine interaction [12].

In the work presented here, we investigate the question of gender identification in a 3D physical head using Mask-bot as our experimental platform. More precisely, we examine the role of texture as a cue for gender discrimination while maintaining a constant output mask shape. In this study, we utilise a female face screen as the 3D output geometry, and display various face images morphed between male and female. Subjects are asked to judge the gender of heads projected via the Mask-bot system. We ascertain the effectiveness of texture as a gender cue, and ask whether texture alone is sufficient to change the perception of gender in a projected 3D head.

II. Experiment Setup

A. Mask-bot display and texture image

The Mask-bot system is a life-size, retro-projected face shape display system with the ability to show realistic talking head animation and auditory and speech motion output [9], [12] (Figure 2). The current Mask-bot system uses pre-calibrated 3D face models for animation: each 3D face model is carefully aligned and calibrated for distortion resulting from a fisheye lens and projection onto a 3D mask surface. Replacing the texture image of one of these calibrated 3D face models will enable us to change face appearance quite easily without calibrating the face model each time, at the cost of possibly loosing an exact match between facial features of target 3D face models and the final output on the mask. In fact, the current output 3D mask shape is fixed, and there is always some error caused by a mismatch between between calibrated face model features and 3D mask features (unless the 3D face model is the same as the 3D mask). We discovered that in most cases, though, these errors are not perceived unless they are sought out with careful observation.

Mask-bot also currently uses specific text-to-speech output which may also cause a mismatch for different face models and morphed faces. For this reason, we use still images without head motion for stimuli in the experiment presented here.

B. Face images from 3D face data

We provide morphed face images obtained from a 3D face database to present on the Mask-bot screen. The following two types of 3D face databases are used to obtain 3D face data:

ATR 3D face database: This database contains ∼500 subjects (adults, some with 9 face postures each, and others with 25 postures) scanned with a Cyberware 4020 and 3030 RGB/PS color digitizer (Cyberware, Inc., www.cyberware.com) and stored in Cyberware ECHO format[13]. Most face data were scanned in 480x450 resolution in both range data and texture image. The effective pixels around the face area is roughly 260x220 pixels, which provides quite a good resolution for 3D shape (in Cylindrical coordinates, 0.7 mm of resolution for the polar axis direction, and 0.75 deg for the angular direction), but not enough for surface texture. Specific face features were annotated for 200 subjects for various statistical analyses and applications. We selected 40 faces from these annotated faces (10 faces from each subgroup: Caucasian male, Caucasian female, Asian male and Asian female) to create an average face.

MARCS 3D face database: This database was collected at MARCS Auditory Laboratories, University of Western Sydney and contains data from ∼200 subjects (babies to adults: from 1 to 50 postures, depending on the subject) scanned with a 3dMDface system with two camera heads (3dMD, www.3dMD.com). Most data were scanned in 1200x1600 pixel resolution for the texture image from each camera, and include the subjects' face and torso. Final 3D output data with combined texture images yield roughly

Fig. 3. Sample 3D face data from the ATR 3D face database (left) and the MARCS 3D face database (right): texture mapped image (top) and polygonal image (bottom) show differences in texture quality and 3D resolution. For the polygonal image from the ATR database (left-bottom), only 1/4 of the actual 3D points were used in this image.

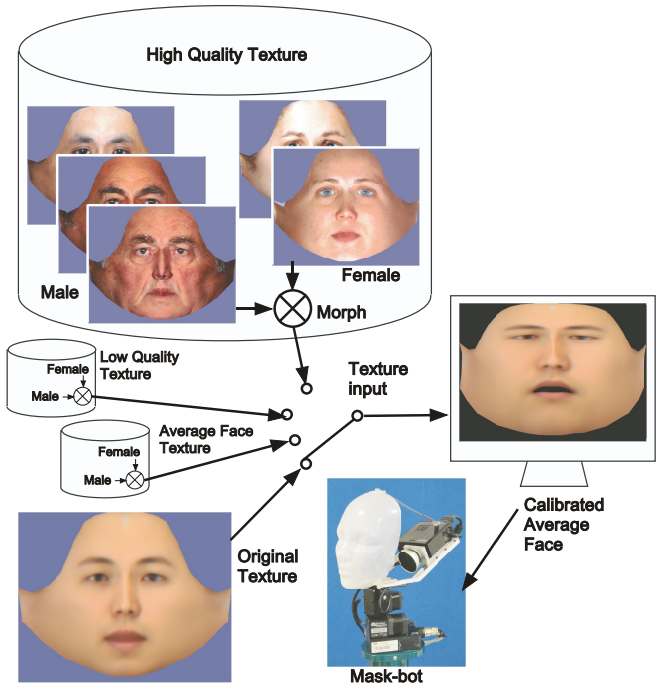


Fig. 4. Overview of the Mask-bot display with morphed face image texture from different texture groups

500x400 effective pixels for the face area. Although the texture quality is very high and the reconstructed 3D surface can adequately capture details of the face structures, its spatial resolution is usually not as dense as Cyberware data, but still high enough for 3D face geometry analysis.

In order to use the same processing method used on the ATR 3D face database, the 3D data stored in TSB format from this database were resampled and converted to Cyberware ECHO format in 960x900 resolution, and the same face features were annotated. From the data we selected 5 adult faces (2 Caucasian male, 2 Caucasian female and 1 Asian male) for this experiment.

Figure 3 shows sample data of the same subject from each database. The left column shows the sample from the ATR database, and the right column corresponds to the MARCS database data. As you can see from images in the top row, texture quality is much better in the MARCS database. On the other hand, the number of 3D surface polygons and points is much higher in the ATR database: the left-bottom image shows only 1/4 of the original 3D points to visualize triangle polygons, whereas the right-bottom image shows all original 3D points and polygons.

To obtain a face image for Mask-bot, we apply the following steps. For each face we:

1) convert to a common mesh structure by adapting it to a generic mesh model[14]
2) render the adapted face model in the generic mesh coordinates and create an image (800x640 pixels)
3) synthesize morphed texture images between two rendered face images using alpha blending
4) redefine the morphed texture image as an image applied to a pre-calibrated average face (made from 40 faces from the ATR face database)
5) display on Mask-bot.

Since we use still images as stimuli for this experiment (no movement nor speech), the last step, which is normally controlled by the talking head animation pipeline used for normal operation of the Mask-bot system, is instead controlled in one of two ways: by an image browser, or by DMDX, a standard psychological experiment presentation tool used for detailed response measurements[15]. (The DMDX control replaced the image browser when it was ready. Therefore, 5 subjects viewed the stimuli via the image browser, and 10 later subjects via the DMDX. Viewing conditions were identical for the two response conditions.)

C. Stimuli Synthesis

Using selected faces from two database, the following three groups of 3D face data are prepared:

- A high-quality texture face group from the MARCS 3D face database
- A low-quality texture face group from the ATR 3D face database
- An averaged face group using low-quality texture (6 average faces from: all male faces; all female; Caucasian male; Caucasian female; Asian male; Asian female).

In each group, single faces (or a single average face from the 3rd group) from each gender are selected, and used to synthesize morphed image with ratios of 0.00, 0.25, 0.50, 0.75 and 1.0 (0.0=female, 1.0=male). Finally, we create 30

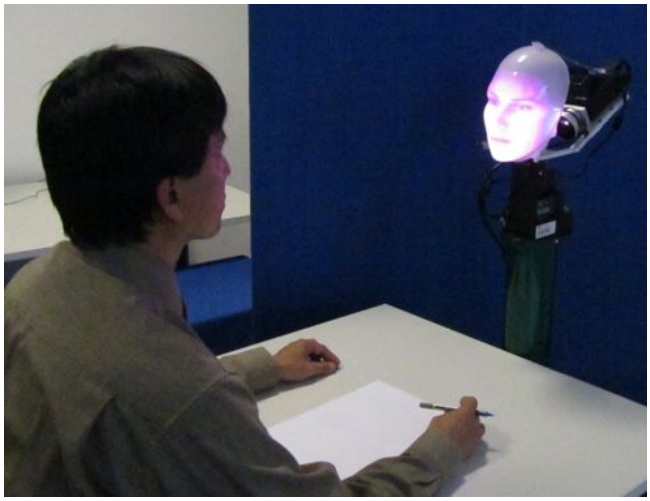


Fig. 5. Experiment setup: the Mask-bot display is located in front of a seated subject at a similar height to the subject's face.

(3 male x 2 female x 5 morphs), 2000 (20x20x5), and 45 (3x3x5) images for each group respectively. However, presenting all images would require a very long participation time for subjects, so we decided to use 30, 185 (randomly selected), and 45 images respectively,for a total of 260 images. These 260 images were randomly separated into 26 blocks consisting of 10 faces each. Also, 10 faces were randomly chosen as a practice block from the same 260 images.

D. Experiments

$N = 15$ subjects (age 23 to 53, average age = 30, gender = 12 males, 3 females) were asked to evaluate the gender of faces on a scale from 0 to 4 (0=female, 1=may be female, 2=Middle/Ambiguous, 3=may be male, 4=male). We decided to use such a scale rather than a binary male / female decision because we would like to know if subjects can identify synthesized morphed faces correctly. The Mask-bot display was located in front of a seated subject across a small desk at a similar height to the subject's face. The distance between the Mask-bot and the subject's face was $\sim$1m. Figure 5 shows the actual configuration of the experiment. The first 5 subjects were asked to tick responses on evaluation sheets. For the next 10 subjects, the integration of DMDX was ready, so input for identical visual stimuli was done via a keyboard.

A total of 26 blocks of 10 faces were presented after 1 block of practice for each subject. Each block of presented stimuli was designed as

1) 2.5 seconds of text fixation (to focus subjects at the middle of where the face appears)
2) sequential presentation of face images - 1 face every 2.5 seconds for a block of 10 faces (no blank interval)
3) 7.5 seconds of blank interval.

All communication functions of Mask-bot were disabled - it was used as a display without any head motion.

III. Results

Figure 6 shows averaged gender identification results with standard errors from all subjects for (a) high-quality texture stimuli, (b) low quality texture stimuli and (c) average face stimuli obtained from low quality texture faces. As a guide to ideal response to the morph ratio, a solid blue line is plotted in each graph. From these results, we can see that gender identification responses show slightly different properties in each case.

(a) High-Quality Texture
For all morph ratios except 0.0 (100% female) the results show a good trend matching the ideal response, but with a slight offset toward maleness. (Of course the 1.0 morph is capped to the highest value of 4.0.) These results indicate that subjects can identify the gender correctly almost always, including the in-between faces generated by morphing. This also means that male texture cues can override the female mask shape cues and be perceived correctly as male. The remaining questions which we discuss later revolve around the slightly sub-par performance for female face categorization. (For the 100% female case, responses indicate an average of slightly female.)

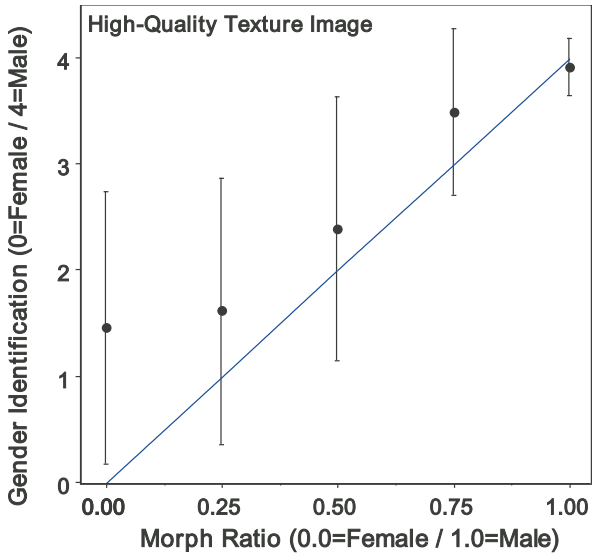
(b) Low-Quality Texture
The response is almost linear with respect to the morph ratio, excluding the 100% female case. However, the slope is less steep than the ideal response. As the maleness increases, we see a slight increase in male responses. Female gender is harder to correctly identify, even more so in the low quality texture images than the high quality images, with averaged responses near neutral for the 100% and 75% female morphs. These results indicate that more relevant gender texture cues are better preserved in the high quality images.

(c) Average Face Texture
Here, the responses follow the general trend of the data, but with a suppressed response that hovers more toward neutral. That is, there is a slight increase from female to male response as images become more male. However, responses have moved closer to neutral, with female faces slightly below 2 (the neutral case), and male faces slightly above 2, and the 50% case falling almost exactly on neutral. Note that female categorization with these average face morphs is slightly better than in the low-quality individual case.

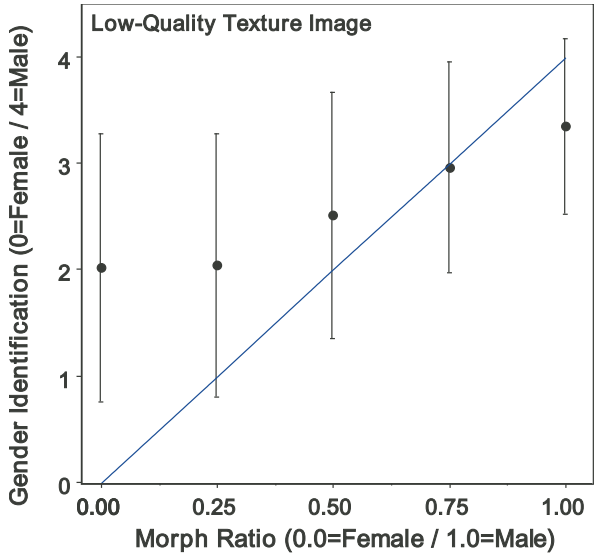
A. Discussion

These results could be explained by:

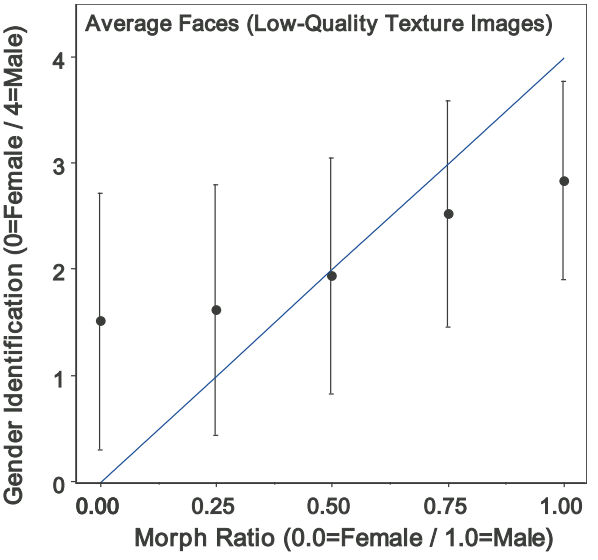
- Strong male texture cues: The scope of the face texture preserves strong male features (sideburns, beard or mustache shadows – even though male subjects were asked to shave prior to being scanned) that help in gender identification (case a,b).



(a)



(b)



(c)

Fig. 6. Gender identification results (averaged values with standard error) for (a) high-quality texture, (b) low-quality texture and (c) average face texture obtained from low quality texture faces. Solid blue lines indicate the ideal response (closer to this line means gender is identified correctly).

- Absence of strong societal female cues: Subjects in both databases are not wearing makeup. Therefore, female faces may look slightly less feminine than in usual social circumstances. Also, hair, another important social gender cue, is absent from the stimuli. This may result in underperformance for female gender categorization (slightly seen in case a, strongly seen in case b). In the case of low quality or ambiguous information, these social cues may increase in importance as gender markers.
- Missing details: Low quality images lose information that may serve as valuable cues to gender identification, thus causing a mixed response (case b).
- Ambiguous information: Averaged faces become blurred, leaving the skin looking smoother, giving a feminine impression. Also, in the absence of strong features, the shape of the output device may have a greater influence on the decision of gender identification. Or, the absence of strong features may lead to ambiguity across responses. Thus, faces tend to be identified as less male than they are, and females are less female (case c).

Surprisingly, there is no clear evidence that 3D mask shape affects the gender identification, although there is some support for it as a possible contribution to case (c). Smooth skin could still be the dominant cue, however. We need a larger subject group and more high quality texture with average face information to help discern what cues are influential. As a conclusion, texture images and texture quality are stronger cues in gender identification than 3D mask shape. But there is support for possible 3D shape influence when these other cues are minimized. To explore this further, we would like to run the experiment with a male mask as the output screen.

For applications using Mask-bot the 3D shape becomes important for more personalized models and personal identification purposes. These situations may require us to pay careful attention to the type of 3D representations used with a particular face.

Finally, we wish to investigate if there is a difference between male and female subjects for gender identification on the Mask-bot. We currently have tested only 3 female subjects, so we need to recruit more female subjects.

IV. CONCLUSIONS

We tested how people identify gender from various morphed and original face images which are presented on a Mask-bot display system. Even though the output 3D geometry is female, most people identified gender correctly in cases where the images contained good texture cues (e.g., high quality texture). Also high quality texture shows better identification results than low quality texture. We noted a slight underperformance for female gender identification, particularly in the low-quality case.

The current system uses a fixed female mask and requires significant calibration effort for each face model (e.g., 30

minutes per model) in order to make use of the Mask-bot's current talking head animation engine. To improve this, we are developing a new system which can replace face masks easily and can obtain calibration parameters for any new mask automatically in a few minutes. We are also developing a new animation engine which can apply calibration parameters to 3D face models directly.

We expect that this new system can help us explore not only further gender identification issues but also support other studies, such as personal identification, likability, uncanny valley effects, and so on.

ACKNOWLEDGMENT

This work was supported by the DFG cluster of excellence 'Cognition for Technical systems – CoTeSys' of Germany.

We also acknowledge ATR-International (Kyoto, Japan) and MARCS Institute (former MARCS Auditory Laboratories - Sydney, Australia) for accessing their 3D face databases for supporting this research.

REFERENCES

[1] C. Turati, "Why faces are not special to newborns: An alternative account of the face preference," *Current Directions in Psychological Science*, vol. 13, no. 1, pp. 5–8, 2004.

[2] B. Dering, C. D. Martin, S. Moro, A. J. Pegna, and G. Thierry, "Face-sensitive processes one hundred milliseconds after picture onset," *Frontiers in Human Neuroscience*, vol. 5, no. 93, 2011.

[3] D. Hanson, "Exploring the aesthetic range for humanoid robots," *CogSci-2006 Workshop: Toward Social Mechanisms of Android Science*, 2006.

[4] H. Ishiguro, "Understanding humans by building androids," in *SIGDIAL Conference*, R. Fernández, Y. Katagiri, K. Komatani, O. Lemon, and M. Nakano, Eds. The Association for Computer Linguistics, 2010, pp. 175–175.

[5] P. Jaeckel, N. Campbell, and C. Melhuish, "Facial behaviour mapping - from video footage to a robot head," *Robotics and Autonomous Systems*, vol. 56, no. 12, pp. 1042–1049, 2008.

[6] M. Hashimoto and D. Morooka, "Robotic facial expression using a curved surface display," *Journal of Robotics and Mechatronics*, vol. 18, no. 4, pp. 504–510, 2006.

[7] F. Delaunay, J. de Greeff, and T. Belpaeme, "Towards retro-projected robot faces: an alternative to mechatronic and android faces," *Robot and Human Interactive Communication (RO-MAN 2009)*, pp. 306–311, 2009.

[8] S. A. Moubayed, S. Alexandersson, J. Beskow, and B. Granström, "A robotic head using projected animated faces," *Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP 2011)*, p. 69, 2011.

[9] T. Kuratate, B. Pierce, and G. Cheng, ""Mask-bot" - a life-size talking head animated robot for av speech and human-robot communication research," *Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP 2011)*, pp. 107–112, 2011.

[10] V. Bruce, A. M. Burtonu, E. Hanna, P. Healey, O. Mason, A. Coombes, R. Fright, and A. Linney, "Sex discrimination: how do we tell the difference between male and female faces?" *Perception*, vol. 22, pp. 131–152, 1993.

[11] R. Campbell, P. Benson, S. Wallace, S. Doesbergh, and M. Coleman, "More about brows: how poses that change brow position affect perceptions of gender." *Perception*, vol. 28, no. 4, pp. 489–504, 1999.

[12] T. Kuratate, Y. Matsusaka, B. Pierce, and G. Cheng, ""Mask-bot" : a life-size robot head using talking head animation for human-robot communication," *Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2011)*, pp. 99–14, 2011.

[13] T. Kuratate, "Statistical analysis and synthesis of 3D faces for auditory-visual speech animation," *Proceedings of AVSP'05 (Auditory-Visual Speech Processing)*, pp. 131–136, 2005.

[14] T. Kuratate, S. Masuda, and E. Vatikiotis-Bateson, "What perceptible information can be implemented in talking head animations," *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2001)*, pp. 430–435, 2001.

[15] J. Forster, "DMDX (Display Software) Update Page," *http://www.u.arizona.edu/∼jforster/dmdx.htm (last accessed on Aug. 10, 2012)*.