Int J Soc Robot (2013) 5(4):627-639
This is a draft of DOI 10.1007/s12369-013-0210-2 - Please check the published version for more accurate information.

Effects of 3D Shape and Texture on Gender Identification for a Retro-Projected Face Screen

Takaaki Kuratate · Marcia Riley · Gordon Cheng

This copy is a draft version of DOI 10.1007/s12369-013-0210-2

Abstract Retro-projected face displays have recently appeared as an alternative to mechanical robot faces, and stand apart by virtue of their flexibility: they are able to present a variety of faces varying in both realism and individual appearance. Here we examine the role of both 3D mask structure and texture image quality on the perception of gender in one such platform, the Mask-bot. In our experiments, we use three specific gender face screens as the 3D output – female, male and average masks – and display various face images that are gradually morphed between female and male on these screens. Additionally, we present three cases of morphed images: high quality texture, low quality texture, and averaged face texture from low quality data. Experiments were carried out over several days. 15 subjects rated the gender of each face projected on the female mask screen, and 10 subjects rated the gender of faces on the male and average screens. We found that even though the 3D mask screens have strong gender specific face features, gender identification is strongly determined by high-quality texture images. However, in the absence of strong texture cues or the presence of ambiguous information, the influence of the output structure may become more important. These results allow us to ascertain the ability to faithfully represent faces on these new platforms, and highlight the most important aspects – in this case texture – for correct perception.

T. Kuratate, M.Riley, G. Cheng
Institute for Cognitive Systems, Technische Universität München, Karlstraße 45 / II, München 80333 Germany
E-mail: kuratate@tum.de
M. Riley
E-mail: mjriley7@gmail.com
G. Cheng
E-mail: gordon@tum.de

Keywords face robot; facial animation, 3d face, retro-projected face; gender identification

1 Introduction

Faces are an essential element in human communication, and as such are a critical component in developing robots who are interactive and socially aware. Our sensitivity to faces is supported by studies that show even newborns can detect faces almost instantly (29; 8). Robot face options recently include platforms in the form of retro-projected systems (12; 6; 24; 20), which can project computer graphics animation on a non-flat screen. This approach is becoming increasingly popular because it can provide a richer 3D experience than 3D faces shown on 2D screens.

The curved display approaches use abstract (cartoonish) face models: simple eyes, eyebrows, a nose and a mouth are projected onto a sphere, (12), or FACS (9)-based simple face models (6; 7) are projected onto an abstract version of a face mask. The Furhat system (24) uses a 3D printed face screen of a facial animation character instead of a real face. It also has one of the best auditory-visual speech conversation systems to date. The Mask-bot system (20) is a retro-projected platform that can project realistic talking head animation, and can easily replace its face model starting from a single camera shot (22) or a scanned 3D face.

The hybrid approach used by Bazo et al. (1) is able to display different facial features – eyes and mouth, for example – on a display embedded as part of a contoured robotic face shell, augmenting the flexibility of computer graphics with a 3D physical structure. This solution facilitates changing the face in subsequent design cycles, and is a good solution for designing "robotic-looking" as opposed to "realistic-looking" humanoid robot faces. However, the overall

shape of the robot head will be limited by the shape of the 2D computer display.

These platforms exist alongside more traditional mechanical options, which also encompass specialized material such as Frubber^{TM}by Hanson to create flexible skin solutions (11). In a similar vein, Ishiguro (14) also builds realistic heads comprised of a mechanical structure covered with flexible skin that can realistically reproduce various complicated facial expressions and motions. Ishiguro is perhaps best known for creating a convincing replica of himself. In (15) Jaeckel and colleagues present their realistic robot head approach complete with a flexible outer skin where expressions are driven by learning from performance-based animation.

The upper body robot ROMAN is equipped with an expressive mechanical head covered by a silicon skin (2). Interestingly, ROMAN has been used to interact with a user by playing the Tangram game (13). However, its jaw motion and head motion are only weakly coupled with its speech, which may lower speech intelligibility (30; 25), and may possibly leave people with an unnatural impression reminiscent of the Uncanny Valley effect (23; 28). This effect might also arise from the mismatch between the head and torso, as the head has skin, but the torso is left uncovered, combining visible mechanical structures with a more natural head.

From the viewpoint of building an expressive head, both realistic and more abstract mechanical humanoid robot heads require much effort, as shown by Kędzierski et al. with their emotive head EMYS(16). Furthermore, as explained by Cheng et al.(5), improving the expressiveness of realistic robotic heads also requires careful observation of human face motion and application of the analysis results to the mechanical head.

These advances in robotic heads are impressive, but suffer from an important disadvantage in comparison to retro-projected approaches. Because their appearance is fixed, redesign is costly. Researchers must rebuild not only the mechanically sophisticated structures, but also replace the flexible skin, which requires extra care around lip and eye corners.

In contrast, the retro-projected systems stand out for their flexibility: they are able to present a variety of faces varying in both realism and individual appearance, which means the face can easily change to fit the application or the user preference. Additionally, unlike most mechanical robot faces, they are able to express nuanced, subtle gestures. They can also easily and iteratively improve the underlying software display algorithms as better methods are uncovered.

Lastly, the systems are generally lighter and less complicated than their mechanical counterparts, being built primarily from a small projector, optics and a 3D face screen. It is this 3D output device that improves the 3D presence of the platforms when compared to the more common 2D flat screens. (Version 1, or the original Mask-bot version, for example, weighed 1.4 kg, but the recent system is 0.4 kg, largely due to a smaller projector and lenses. See Sec. 2 for more details.)

However, these systems do share the same disadvantages of projectors: they are impractical in strong illumination conditions, including daylight. Also, the presence of a projector reduces the illusion of a lifelike head, although some heads use a small projector that can be encapsulated and thus hidden from view (26; 27). One last limitation is the possible perceptual mismatch caused when the face is animated, but the mask is stationary. Because of these aspects, retro-projected heads need to be evaluated for general aspects such as likability to gain a better understanding of how they are perceived by people during interactions.

Despite these limitations, retro-projected systems provide flexible platforms for a myriad of applications involving face-to-face encounters, including communication studies, video conference interfaces, and various human-robot applications.

In this paper we present our work to ascertain the ability of retro-projected platforms to faithfully portray faces. Our motivation is to understand and evaluate the components needed to maximize the effectiveness of these platforms for human interaction. We can then use this knowledge to focus on aspects that best influence the faithful perceptibility of the projected faces, as well as work to mitigate any undesirable, unintended effects.

Specifically, we present a study that explores the influence of both 3D screen structure and texture in a gender identification task using realistic faces displayed on the retro-projected face screen system Mask-bot, shown in Fig. 2. We describe our expanded work to explore gender identification not only on a female face screen (21), but on a male and an average display as well. We explain our method for creating a neutral, or average, 3D mask to test along with our male and female masks (or screens), and describe how we create a number of 3D face stimuli by morphing between pairs of male and female faces, resulting in faces from 100% female to 100% male. Stimuli is presented to subjects on all 3 screens. Additionally we test three types of texture images: high quality, low quality, and averaged low quality texture. We ask subjects to rate the gender of each face. Our main goal is to ascertain if and how the 3D mask shape and the texture image quality influence the perception of gender in realistic faces projected on Mask-bot.

Fig. 1 shows a few samples that helped motivate this work. In this figure we see that the Mask-bot system can project a variety of calibrated face models – male or female – on a given screen. In the first row, faces of both genders are projected onto a screen with a female shape. A male face (middle column) is still identifiably male, even when projected onto a female

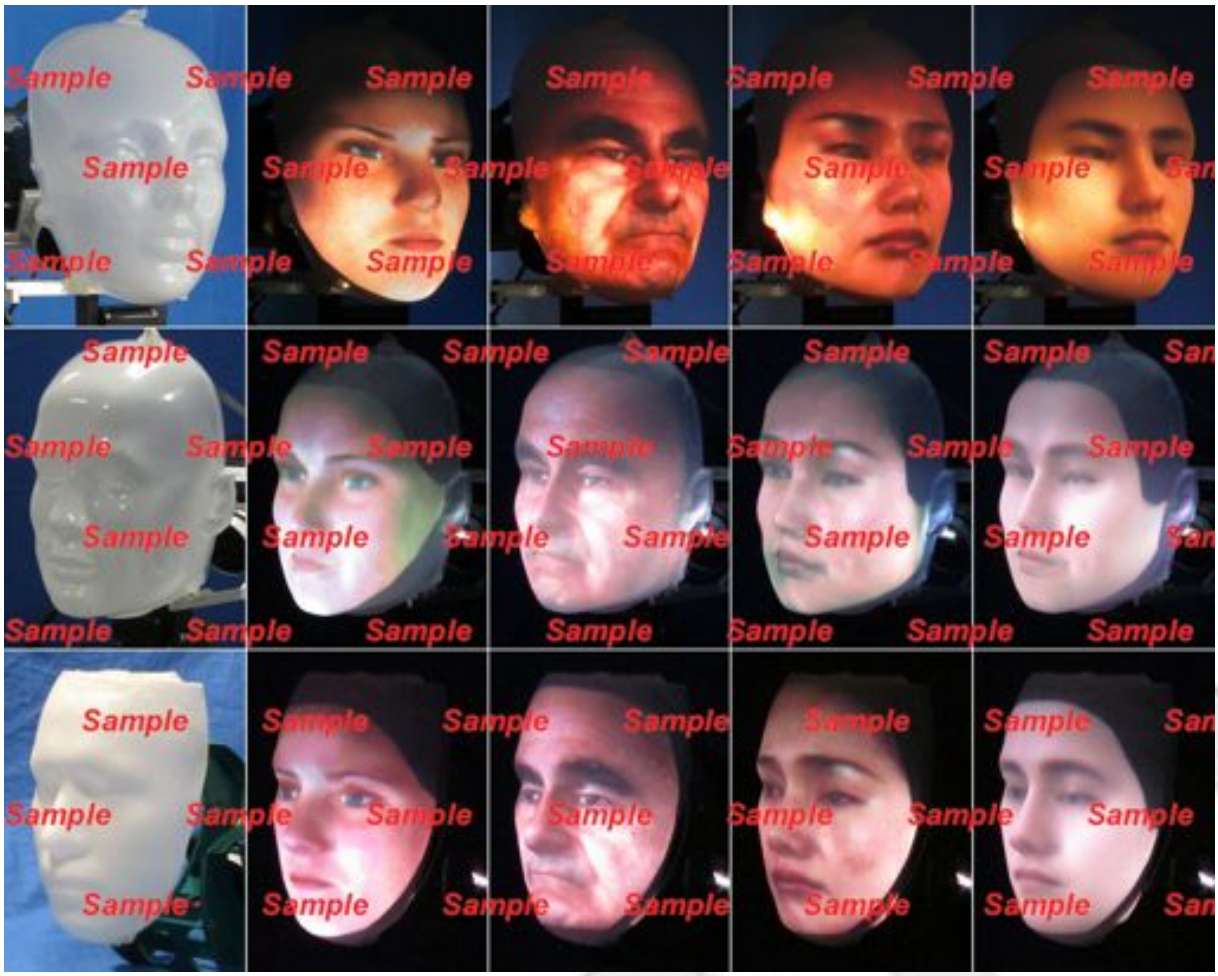

Fig. 1 The Mask-bot platform equipped with different mask screens. The female mask (top row (21)), male mask (middle) and average mask (bottom) are shown without a projected face, and then with different faces: from left to right, without rear projection, Caucasian female (high quality texture), Caucasian male (high quality texture), Asian female (low quality texture), and average face (low quality texture).

screen. Similarly, female faces projected onto a male screen (shown in the second row, second and fourth columns) maintain a female impression.

How, then, is gender perceived on these retro-projected platforms, and what factors are critical in determining this identification? We know from earlier studies on gender perception that an observer uses a number of cues to classify gender, including facial features, skin textures and 3D face structure. In Bruce et al. the authors show the importance of nose and chin protuberance in 3/4 views (3), while in (4) it is the the eye and brow region that provide key visual cues in front views. These perception experiments used conventional media such as photographs, television, or computer screens to present stimuli. With the retro-projected systems, a physical 3D presence functions as the output device, necessitating a closer look not only at gender perception, but also, in a broader sense, at how to effectively use these platforms as social tools in human-machine interaction (19).

Overall, our gender identification results on Mask-bot show that the shape of the output screens do not overtly bias gender perception, and that texture provides stronger cues than the 3D face surface structure, especially in the case of high-quality texture stimuli. However, in the absence of strong texture cues or in the presence of ambiguous information, the influence of the output structure may become more important.

Our results tell us to direct our efforts and attention to the more influential components of the system. In this case we show that attention to the rendering task, especially the quality of the texture maps, will maximize effectiveness. In the future we will explore additional tasks where 3D face mask shape may become more important, such as in identification of a specific individual.

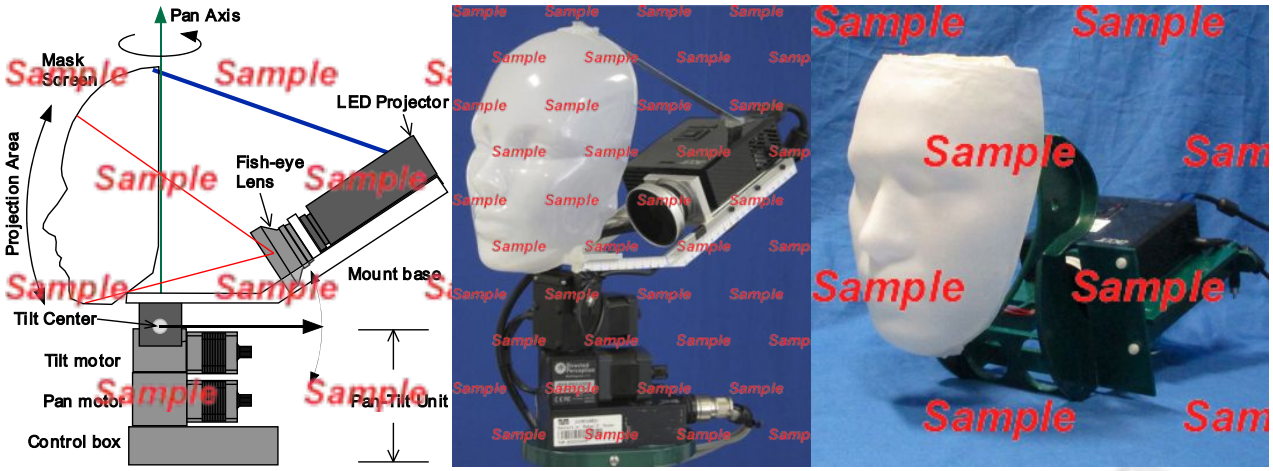

Fig. 2 Mask-bot, an example of a retro-projected face: the design diagram (left); the original system with a pan-tilt unit (PTU) (center); and the desktop version without a PTU (right).

2 Experiment setup

2.1 Mask-bot display and texture image

Our retro-projected system, Mask-bot, is a life-size, 3D face display system built especially for interactive social applications. It communicates via speech and face and head motion, and can change its appearance to support both abstract and realistic faces. In these experiments we use Mask-bot's realistic face capabilities. (20; 19) (Fig. 2). The first version of Mask-bot is shown in the center of Fig. 2. This version uses a heavier (0.61 kg) and bulkier LED projector compared to various pocket projectors, but it can project brighter images (200 ANSI lumens as compared to 70 lumens). The total weight without the pan-tilt unit and cable is 1.44 kg. This includes a projector, a fish-eye lens, a macro lens, a 3D mask and a supporting frame. The desktop version of Mask-bot shown on the right in the same figure weighs 1.29 kg, and uses acrylic plates instead of an aluminum frame.

Mask-bot 2i (26; 27), a newer version with a slightly different design, is lighter, weighing about 0.4 kg, as it incorporates a smaller but darker projector (70 ANSI lumens) and smaller lenses used with a mirror. This lighter model could be very useful for face-to-face communication studies when a newer, brighter projector is available in a smaller size.

The current Mask-bot system uses 3D face models that are carefully calibrated to compensate for distortion from the fisheye lens and the 3D screen shape. When displaying new faces, we can reduce the preprocessing time by replacing only the texture of a calibrated 3D face model. However, in so doing we sacrifice some of the accuracy of the resulting face model. Specifically, facial features resulting from texture changes may not exactly match the 3D face model structure. We know, though, that there is always some error between the projected face and the mask unless the mask is an exact match for the 3D face being projected. However, we discovered that for most observers these errors are so subtle as to be barely noticeable unless they are carefully pointed out.

Finally, we do not use Mask-bot's ability to display auditory-visual speech since this face motion may contribute to perceptual mismatch when used with different face models. Thus we present only still faces in the experiments.

2.2 Face images from 3D face data

We use two 3D face databases to obtain source faces for creating our stimuli. As described below, one database contains data with higher spatial resolution but lower texture resolution, and the other has higher texture resolution relative to the other database. These different data sets allow us to test the effects of high versus low quality texture resolution in our gender identification experiments.

The ATR database (lower texture resolution) contains 3D face data collected at the Advanced Telecommunications Research Institute International in Kyoto, Japan. It contains approximately 500 adult subjects, and each subject has either 9 or 25 scanned postures. The capture technology is a Cyberware 4020 and 3030 RGB/PS color digitizer [1], which stores the data in Cyberware ECHO format. See (17) for details. Both the range data and the associated texture image were scanned with a resolution of 480x450 in most cases, with the effective face area containing roughly 260x220 pixels. This resolution is sufficient for capturing detailed 3D shape information, but insufficient for good surface texture. (In Cylindrical coordinates, this translates to a resolution of 0.70 mm

[1] Cyberware, Inc., www.cyberware.com

in the polar axis direction, and 0.75 degrees in the angular direction.) Data from 200 subjects were preprocessed for use in subsequent analyses. A key step in this procedure was face feature annotation. From these data we selected 40 faces for stimuli generation, with 10 faces from each of the following subgroups: Caucasian male, Caucasian female, Asian male and Asian female.

The MARCS 3D face database (higher texture resolution) was collected at MARCS Auditory Laboratories, University of Western Sydney, Australia. A 3dMDface system [2] was used to collect head and torso data from approximately 200 subjects, from babies to adults. Most adults had between 25 to 50 postures scanned. The higher number of postures is made possible by the improved speed of this newer technology. Each of the two camera heads used in scanning yielded data with 1200x1600 pixel texture image resolution. The final data containing both the 3D structure and the combined texture images yield a face area of roughly 500x400 effective (non-background) pixels. The texture quality is very high, and the reconstructed 3D surface can adequately capture details of the face structures, allowing for 3D face geometry analysis. However, the spatial resolution is usually not as dense as that of Cyberware data.

The problem of using data from databases with different formats was solved by re-sampling and converting the TSB format (MARCS data) to the Cyberware ECHO format, resulting in compatible data with a final resolution of 960x900. The same set of facial features were annotated in all data. This step was important, as it allowed us to use the same processing method on all data. From this high texture resolution data we selected 5 adult faces for stimuli, specifically: 2 Caucasian male faces, 2 Caucasian females and 1 Asian male. (Because each face from the MARCS database requires significant preprocessing, fewer high quality faces were used in the current study. The preprocessing for the ATR face data was completed prior to this study.)

Fig. 3 shows sample data of the same subject from each database, with ATR data on the left (higher 3D spatial resolution, lower texture resolution), and MARCS data on the right (lower 3D spatial resolution, higher texture resolution). Comparing images across the top row shows that texture quality is much better in the MARCS database, while the bottom row shows that the number of 3D surface polygons and points is much higher in the ATR database. Note that the left-bottom image shows only 1/4 of the original 3D points used to visualize the model's 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:

[2] 3dMD, www.3dMD.com

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

1. convert face data to a common mesh structure by adapting it to a generic mesh model (18)
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 the new images on Mask-bot.

Figure 4 shows an overview of this procedure for preparing morphed texture of different quality for display on Mask-bot. Since we use still images as stimuli for this experiment (no movement nor speech), we modify the normal operation of the Mask-bot system as follows. We replace the last step of the pipeline, where control normally falls to the animation pipeline of the Mask-bot system, with either an image browser or DMDX, a standard psychological experiment presentation tool used for detailed response measurements (10). (We replaced the image browser with the DMDX control tool as soon as it was ready. Thus, the first 5 subjects viewed stimuli using the image browser, and all 20 later subjects used the DMDX interface. Viewing conditions were identical for the two response conditions.)

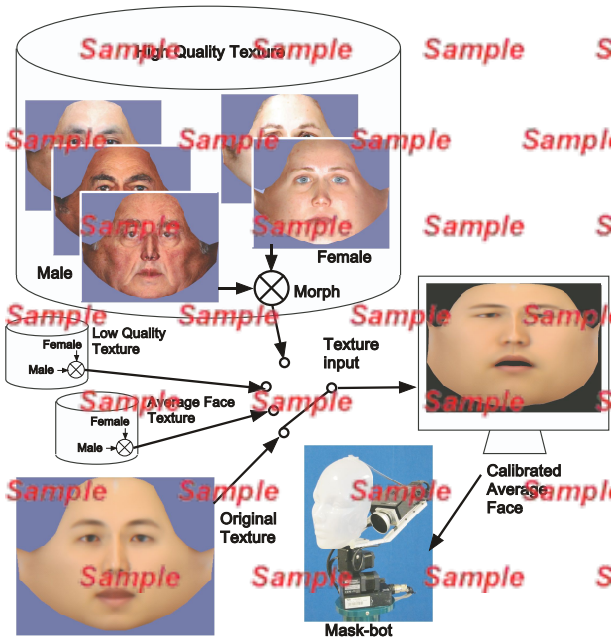

Fig. 4 Overview of the Mask-bot display with morphed face image texture from different texture groups used as input.

2.3 Stimuli Synthesis

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

- A high-quality texture face group
- A low-quality texture face group
- An averaged face group using low-quality texture

For the high quality group, we select single faces from each gender to use in synthesizing the morphed images: each female face is paired with each male face, and morphs are created with ratios of 0.00, 0.25, 0.50, 0.75 and 1.0 where 0.0 is 100% female, and 1.0 is 100% male. In this manner we created 30 (3 male x 2 female x 5 morphs) high quality stimuli.

We built the low quality group in the same manner, but with 20 female and 20 male faces for a total of 2000 stimuli (20x20x5). In both the female and male case, half of the faces are Caucasian and half are Asian.

The low quality averaged stimuli is created from 6 averaged face groups using the 40 faces from the low quality group above. The 3 female groups are: Asian female faces created by averaging 10 faces; Caucasian female faces created by averaging 10 faces; and all female faces (20 faces). The male groups are created in the same manner, i.e., Caucasian, Asian, and both from 10, 10 and 20 faces respectively. Thus, a total of 45 averaged stimuli are created (3x3x5).

Presenting all of the above stimuli (30, 2000, and 45 images) requires too much time for subjects, so we reduced the 2000 low quality image set to a reasonable amount by randomly selecting a smaller set of stimuli, ensuring that the same number of images from each morph percentile were represented. We thus used 185 images for the female mask experiment, and 65 for the male and average mask experiments in the low quality group. (We decreased the number from 185 to 65 in response to user feedback after running the first experiment, the female mask experiment.) Thus subjects rated a total of 260 images for the female mask experiment, and 140 images for the male and average mask experiments. More details are found in Sec. 2.5.

These images (260 and 140) were randomly separated into multiple blocks consisting of 10 faces each (26 blocks and 14 blocks). Also, 10 faces were randomly chosen as a practice block from the same total image pool. Practice answers were excluded from the subsequent analysis.

2.4 Mask types

The transparent 3D female and male masks were supplied by the same vendor. Both masks have strong visible gender characteristics (Fig.1, left column). To observe any differences between these strong gender geometrical cues and a more gender-neutral face, we built an average face mask using face data from the ATR 3D face database as described below.

2.4.1 Average Face Mask

To build an average 3D mask, we first selected 124 faces from the ATR database. These consisted of 31 faces from each of the following groups: Caucasian males, Caucasian females, Asian males, and Asian females, with an age range of 18 to 50 years old and a mean age of 29.4 years. Selection was based on two criteria: (1) the presence of neutral facial expression data without large amounts of noise; and (2) the basic facial features were already annotated.

Using facial features of the eyes and nose, we aligned all 124 neutral faces in the same head orientation, and then re-sampled them to the same resolution in cylindrical coordinates, resulting in a longitude and height equal to 960 x 900, with an effective face area of approximately 460 x 460. We then averaged these data to obtain the mean face, and cut it to fit the 3D printing volume by selecting the best window to print the entire face. In the last step we converted the selected data to volumetric data with a support structure acceptable to the 3D printer.

We used a RapMan 3.2 3D Printer [3], which uses Fused Deposition Modeling (FDM) technology, because this printer is affordable and able to print large volumes. The average face model was printed with 0.5 mm resolution, and a thickness of approximately

[3] Bits from Bytes Ltd., www.bitsfrombytes.com

Fig. 5 From left to right: rendered image of average 3D face data with texture; average 3D face data without texture; 3D printed results from a Fused Deposition Modeling (FDM) printer after filling gaps with lacquer-based modeling putty followed by surface smoothing with sandpaper; and a vacuum formed mask with rear-projection paint made from the 3D printed results.

3 mm, using PLA (polylactic acid) plastic in 8 to 12 hours. (ABS (acrylonitrile butadiene styrene) may also be used.) The printed 3D face includes extra support structures and visible lines between each layer that are created when the printer deposits melted plastic material as thin tubes. These create a ditch between each layer resulting in the lines. To counteract this undesirable effect, we roughly filled the ditches with lacquer-based modeling putty, and then smoothed the surface with fine sandpaper. Using a home-made vacuum former, we then made a mask using the 3D printed face as a mold. We used a transparent PETG (polyethylene terephthalate glycol) plastic sheet with a 1.0 mm thickness to apply vacuum forming. By following this same procedure, we can produce any face mask, with source data derived from 3D scanning devices, to any 3D face estimated from photographs using the 3D face database (17). Figure 5 shows the step-by-step production stages for making the average mask.

2.5 Experiments

We conducted experiments over two stages: first, we held female mask experiments as a preliminary test using 260 images with $N = 15$ subjects (age 23 to 53, average age = 30.0, gender = 12 males, 3 females) as we reported in (21). We then prepared and carried out the male and average face mask experiments using 140 images for each mask with $N = 10$ subjects (age 25 to 53, average age = 36.9, gender = 8 males, 2 females). In the second stage experiments, we provided two Mask-bot systems to evaluate two masks sequentially: the male mask with the pan-tilt unit base, and the average mask with a desktop version of the platform.

In each mask experiment subjects were asked to evaluate the gender of faces on a Likert scale from 0 to 4 (0=female, 1=may be female, 2=middle/ambiguous, 3=may be male, 4=male). We decided to use this scale rather than a binary male or female decision because it provides more information on the subjects' impressions of the faces, allowing us to better ascertain if subjects can identify synthesized morphed faces correctly.

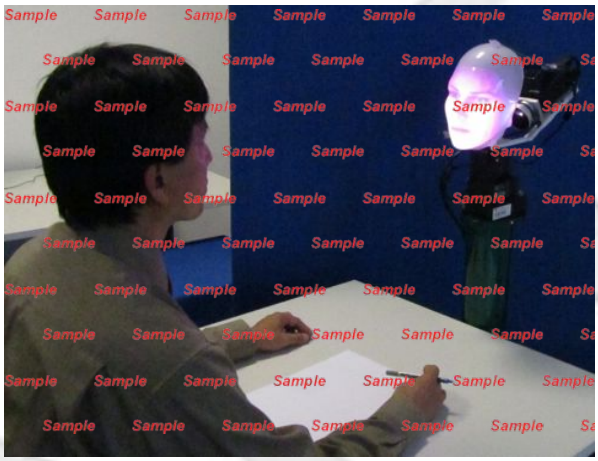

Fig. 6 Experiment setup: the Mask-bot display is located in front of a seated subject at a similar height to the subject's face.

As reported in (21) the Mask-bot display was located in front of a subject seated at a small desk at a height roughly matching the position of the subject's face and at a distance of about 1 m. Fig. 6 shows the setup for the experiment using the female mask. The desktop version (Fig. 2, right) was located at a similar position.

The first 5 subjects were asked to tick responses on evaluation sheets. For all other subjects, the integration of DMDX was ready, and input responses proceeded via a keyboard with graphical icons. The scale of 0 to 4 is mapped to keyboard 1 to 5 as shown in Fig. 7.

After 1 block of practice, a total of 26 blocks of 10 faces were presented for the female mask experiment, and the revised 14 blocks were presented in the same

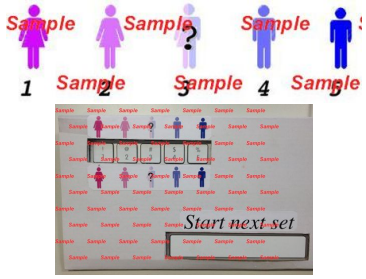

Fig. 7 Icons used to guide user input, top, and on a labeled keyboard, bottom.

manner for the male and average mask experiments. For each block of stimuli, the procedure is:

1. 2.5 seconds of text fixation (to focus subjects at the middle of the region where the face appears)
2. the presentation of face images for a block of 10 faces (with no blank interval):
   DMDX: a maximum of 2.5 seconds is allowed to type a response; (a new face appears when a subject presses a key or after 2.5 seconds)
   Evaluation sheets: a new face appears every 2.5 seconds
3. an interval:
   DMDX: a message asks the subject to press the SPACE key to start the next block
   Evaluation sheets: a fixed 7.5 second break is automatically generated; no projected image is shown at this time.

3 Results

Fig. 8 shows mean gender identification results with standard deviations from all subjects where the rows are: (a) high-quality texture stimuli, (b) low quality texture stimuli and (c) averaged texture stimuli obtained from low quality texture. From left to right the columns show results from: the female mask, the average mask and the male mask. The solid blue line shown in each graph functions as a guide to an ideal response.

In each mask case, that is, going across each row, if the responses are similar, the 3D mask structure has little effect on the overall impression of gender. Going down a column shows us the effect of texture quality and texture face averaging on the impression of gender. In broad strokes, we can see that gender identification responses show different responses when texture quality changes, but for the most part are similar within a texture group regardless of mask. A closer look follows.

(a) High-Quality Texture
For all morph ratios except 0.0 (100% female) the results follow a pattern similar to the ideal response case, but with a slight offset toward maleness. This tendency is present for all three masks conditions. This tells us that subjects can identify gender correctly almost always, regardless of 3D mask shape, and that the subjects can also correctly identify the in-between faces generated by morphing. Also, these results provide evidence that texture cues can override the 3D mask shape cues in the high-quality texture case.
However, questions remain concerning the slightly sub-par performance for female face categorization. (For the 100% female case, responses indicate an average of slightly female.) What is noteworthy is that this sub-par performance is similar for the female case in all 3 texture conditions. Thus, it is indicative of female cues perhaps missing from the type of stimuli used, or of some other influence. We discuss this further in Sec. 3.1.

(b) Low-Quality Texture
For the female mask, 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. That is, as maleness increases, we see an increase, although less than ideal, in male identification.
For the neutral mask, answers hover slightly above neutral to roughly 75% male. In all mask cases the 100% male case is under-identified, and has a larger standard deviation than in the high quality texture case. Also, the male mask does not improve the identification of male gender, as we see similar results for the male stimuli on both the female and male mask. Overall, performance is worse for both the female and male stimuli than in the high quality texture case. However, a pattern emerges where answers lie closer to neutral, as if the missing details from the low quality texture force a more ambiguous response.
Female gender, which is hard to identify in the high quality case, is even harder to identify in the low quality texture 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, and that the mask has at most only a minor (if any) influence on gender identification in the presence of low quality texture cues.

(c) Averaged Face Texture
For faces using averaged face texture, the responses show a suppressed response that hovers more toward neutral, although the general trend of the data still follows the ideal pattern for the female mask and average mask displays. (That is, we see a slight increase from female to male response as the morphed images become increasingly male.) Female faces are slightly below 2 (the neutral case), and male faces slightly above 2, with the 50% case falling almost exactly on neutral. This suppressed pattern once again points toward missing gender

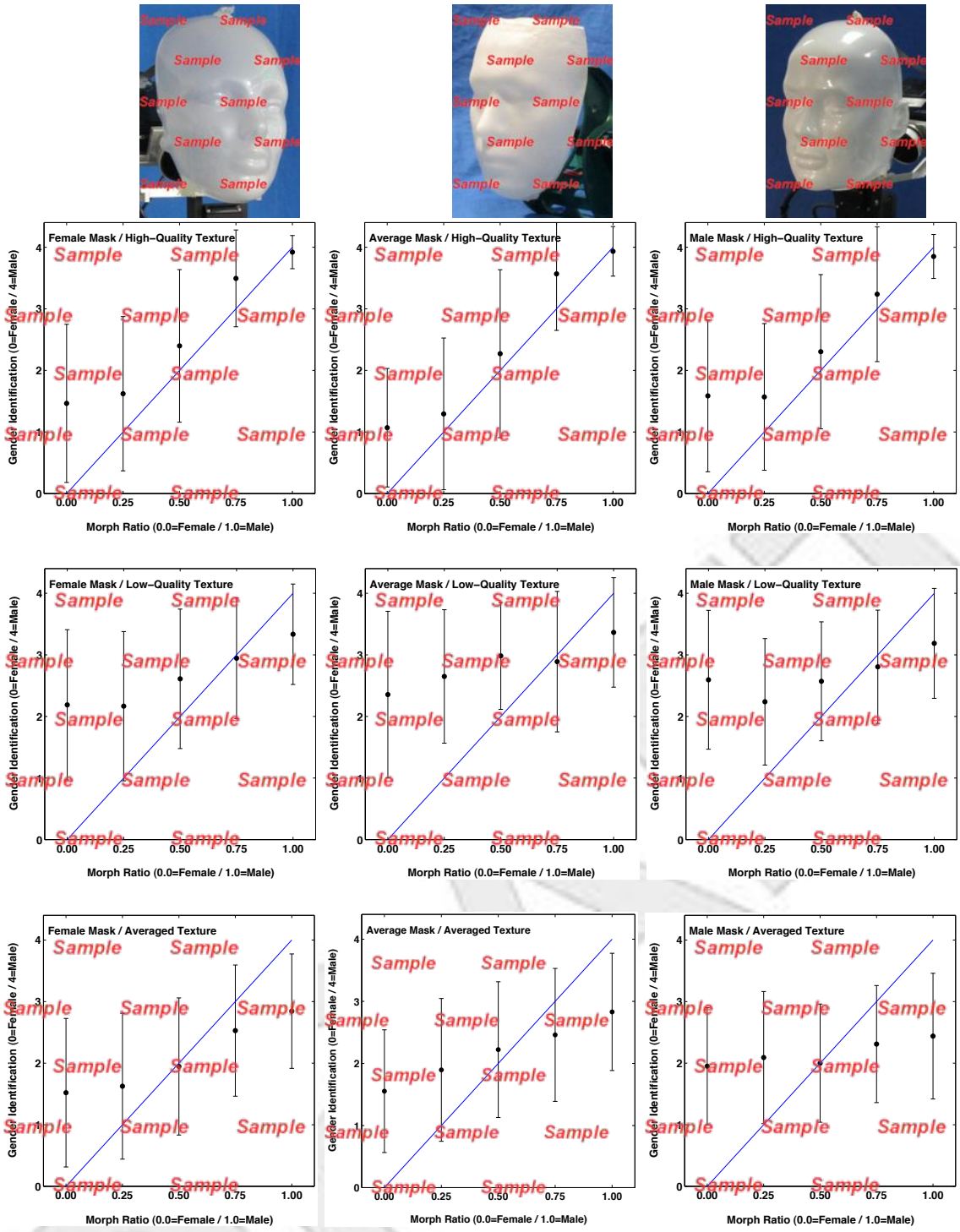

Fig. 8 Gender identification results (averaged values with standard deviation) for (a) high-quality texture, (b) low-quality texture and (c) average face texture obtained from low quality texture faces. Solid lines indicate the ideal response (closer to this line means gender is identified correctly). The left column corresponds to data for the female mask (results from previous work (21)), the middle column for the average mask, and right column for the male mask.

cues compared to the high quality texture. Also, the male face mask may have a subtle influence on the perception of female faces in this case, as the male mask pushes the average response to 100% female more toward neutral. Note, however, that female categorization using average face morphs shows a slight improvement over the low-quality individual case.

3.1 Discussion

From Fig. 8, we can see that there are some similarities between masks. To evaluate how similar these average values are, we applied a t-test between each pair of masks (Female - Average, Average - Male, Male - Female) for each morph level and texture quality condition. Prior to applying a t-test, we checked variances of the two samples with an f-test to verify whether they can be assumed equal or not. We then applied a t-test for unequal sample sizes either assuming equal variances or unequal variances based on these f-test results.

Table 1 shows the t-test results (p values) for all cases. Bold p values in this table indicate that p is smaller than 0.05, which rejects the null hypothesis that the mean values of target pairs are the same: i.e. for these bold values, average results are different between the two masks.

For the High-Quality texture case, where subjects' responses are closest to the ideal response for morphed images (Fig. 8), most cases are $p > 0.05$ except for two: the morph ratio of 0.0 (100% female face texture), between the female mask and the average mask, and the male mask and the average mask. This infers that the identification of the 100% female case is better on the average mask than on either of the gendered masks, even though other morph ratio results did not differ across mask shapes. Additionally, there is no significant difference for the female and male masks in the 100% female case. However, considering that we have a recurring problem in identifying the female stimuli, we need more information about what is happening before drawing conclusions.

For the Low-Quality texture case, the average mask shows differences compared to others for morph ratios of 0.25 and 0.50, but this time the average responses shift towards a higher score, meaning more strongly male. Thus, there is no clear indication that any mask is helping in a consistent way. Rather, the ambiguity of the texture stimuli may lead to variation across user response.

For average texture, the male mask shows differences for morph ratios of 0.0 and 1.0 when compared to the other two masks. If we look at the graphs we see, not surprisingly, that the responses for both the extreme cases (100% female and 100% male) are pushed towards neutral on the male mask. What is surprising is that this neutral effect seems stronger on the male mask for these cases, meaning that even average male faces are identified as less male on the male mask.

Overall, even though there are some differences of mean values between masks for certain conditions, these differences do not appear to be a major influence on accuracy of user response, as they lack consistency in how the responses vary. Also, many of the observed differences in t-test results from Table 1 occur in the slightly more challenging 100% female case. In fact, the female case represents the only noteworthy differences in the high quality case.

Table 2 provides additional support for the finding that texture is more important than shape in assessing gender on Mask-bot. With only a few exceptions, we see that differences exist in responses to the texture cases in almost all instances, with a few notable exceptions. We see an instance on the female mask where response to the 100% female case is not significantly different between the high quality texture case and the averaged texture case. However the averaged texture may soften male cues, resulting in a more accurate response, similar to the high quality case. In the case of the male mask, the performance of the high quality versus the averaged texture case for 100% female is just over the 0.05 threshold for similarity, with the average response being only slightly more accurate in the high-quality case.

We see some similarities between high quality texture results and the other two cases for the 50% morph ratio which, being an equal blend of a female and a male face, should be hard to classify. Such ambiguity could be interpreted as neutral, male or female, so we expect mixed results in this case.

Surprisingly, there is no clear evidence that 3D mask shape affects gender identification on the Mask-bot system, although there is some support that it makes a minor contribution in the absence of strong gender cues. This can be seen mostly in the case of the male mask with the low-quality and averaged texture images, where female faces are perceived as slightly more male when shown on the male face mask. However, results also show that the male faces are seen as less male on the male mask in the averaged texture case: in other words, the mask shape may have an influence in some cases, but it is not clear if this helps or hinders perceptual accuracy when low quality texture stimuli are used. (Averaged texture faces are also low quality since they originate from low quality data.) Perhaps, once again, ambiguous data merely illicit a more neutral response from the subjects regardless of mask shape.

Most evidence does, however, support a clear case for high quality texture providing stronger cues for accurate gender identification than mask shape and than low quality texture images.

Table 1 The t-test results (p value) for High-Quality texture (top), Low-Quality texture (middle) and Averaged texture (bottom). Bold numbers indicate $p < 0.05$, which signifies that average values (means) are different between these pairs.

	(100% Female)	Morph Ratio			(100% Male)
	0.00	0.25	0.50	0.75	1.00
High-Quality					
Female - Average	**0.032448**	0.125205	0.563585	0.606578	0.841520
Average - Male	**0.014018**	0.226627	0.895770	0.073803	0.236750
Male - Female	0.563804	0.805472	0.660885	0.126770	0.194469
Low-Quality					
Female - Average	0.190677	**0.000036**	**0.000113**	0.626637	0.739952
Average - Male	0.121755	**0.002957**	**0.000639**	0.539433	0.127809
Male - Female	**0.000634**	0.495887	0.679874	0.157250	0.096401
Averaged Texture					
Female - Average	0.856199	0.140013	0.114427	0.680492	0.917971
Average - Male	**0.006331**	0.239392	0.152363	0.323898	**0.007894**
Male - Female	**0.013436**	**0.010435**	0.744749	0.180296	**0.009197**

Table 2 The t-test results (p value) between different texture qualities for the Female mask (top), the Average mask (middle) and the Male mask (bottom). Bold numbers indicate $p < 0.05$, which signifies that average values (means) are different between these pairs.

	(100% Female)	Morph Ratio			(100% Male)
	0.00	0.25	0.50	0.75	1.00
Female Mask					
Hi-Q - Low-Q	**0.000000**	**0.000108**	0.133610	**0.000000**	**0.000000**
Hi-Q - AVG.	0.771965	0.982473	**0.017291**	**0.000000**	**0.000000**
Low-Q - AVG.	**0.000012**	**0.000367**	**0.000002**	**0.002680**	**0.000060**
Average Mask					
Hi-Q - Low-Q	**0.000000**	**0.000000**	**0.000515**	**0.000021**	**0.000000**
Hi-Q - AVG.	**0.004430**	**0.002722**	0.818839	**0.000000**	**0.000000**
Low-Q - AVG.	**0.000004**	**0.000002**	**0.000000**	**0.004905**	**0.000062**
Male Mask					
Hi-Q - Low-Q	**0.000000**	**0.000401**	0.162636	**0.011293**	**0.000000**
Hi-Q - AVG.	0.057488	**0.007046**	0.104646	**0.000000**	**0.000000**
Low-Q - AVG.	**0.000006**	0.314531	**0.000030**	**0.000252**	**0.000000**

The underperformance of female gender identification across the board is seen in all mask conditions, albeit with subtle differences noted above. This tells us there may be a global factor at work that makes it more difficult for subjects to clearly identify faces as female. Possible explanations are the presence of strong male texture cues and the absence of strong societal female cues. For example, strong male features (sideburns, beard or moustache shadows) persist in the face texture despite asking male subjects to shave prior to scanning. This is coupled with the absence of what may be strong societal female cues: all subjects were instructed to forego makeup, and so women's faces may look less feminine than in usual social circumstances. Hair is also not visible in the stimuli. (Here we discuss the societal norms of Europe, where the studies were carried out.) These issues may account for the poorer performance in categorizing female gender seen in case (a), and more strongly in case (b). When low quality or ambiguous stimuli are presented, these social cues may become increasingly important as gender markers.

Another question is whether this slight male bias is specific to the 3D retro-projected platform, or is more general to the stimuli. This could be tested by running the gender experiment using the same images, but displayed on a traditional 2D computer screen. If we obtain similar results, we know that the stimuli itself is causing the slight male bias in the gender response to female faces, rather than the platform. We hypothesize that the bias is most likely in the stimuli, as the texture cues are shown to be dominant.

Besides this global effect, we see that there are differences between the higher quality texture image case (a) and the other two cases. Missing texture details most likely account for poorer performance in case (b) as compared with case (a) (high quality texture images). Low quality images contain less information, and the nature of the information that persists is not sufficient to clearly identify gender, resulting in a mixed response.

Case (c) presents more ambiguous information in the form of averaged face images, but performs

slightly better than individual low quality faces for the female case. In averaging face texture, a slight blurring occurs, which accounts for smoother-looking skin. This smooth skin is perhaps associated with a stronger female presence, accounting for the better performance in the female case. Additionally, strong male texture cues may be softened by the averaging process. However, across the female-to-male spectrum of faces, the responses are suppressed, moving more toward neutral when compared with case (a).

In conclusion, texture images and texture quality are stronger cues in gender identification than 3D mask shape. But there is minor support for possible 3D shape influence when important cues are missing, although influence should not be confused with accuracy.

4 Conclusions

In testing a retro-projected platform for its ability to faithfully represent a variety of realistic faces, we show that texture is more important than 3D screen shape in gender identification, and that high quality texture images outperform low quality. However, there may be applications which require us to pay careful attention to the 3D structure used with a particular face, such as in individual identification and personalized models, where more than just the gender needs to be faithfully represented.

However, the current results imply that retro-projected systems can faithfully display a variety of faces with minimal need to vary the 3D face mask. This result is important for efficiency, for it allows us to exploit the flexibility in these systems with less effort, time and cost, as less 3D face screens need to be produced.

Acknowledgements 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. Bazo, D., Vaidyanathan, R., Lenz, A., Melhuish, C.: Design and testing of a hybrid expressive face for a humanoid robot. In: Intelligent Robots and Systems (IROS 2010), IEEE/RSJ International Conference on, pp. 5317–5322 (2010). DOI 10.1109/IROS.2010.5651469
2. Berns, K., Hirth, J.: Control of facial expressions of the humanoid robot head roman. In: Intelligent Robots and Systems (IROS 2006), IEEE/RSJ International Conference on, pp. 3119–3124 (2006). DOI 10.1109/IROS.2006.282331
3. Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., Linney, A.: Sex discrimination: how do we tell the difference between male and female faces? Perception **22**, 131–152 (1993). DOI 10.1068/p220131
4. Campbell, R., Benson, P., Wallace, S., Doesbergh, S., Coleman, M.: More about brows: how poses that change brow position affect perceptions of gender. Perception **28**(4), 489–504 (1999). DOI 10.1068/p2784
5. Cheng, L.C., Lin, C.Y., Huang, C.C.: Visualization of facial expression deformation applied to the mechanism improvement of face robot. International Journal of Social Robotics **5**(4), 423–439 (2012). DOI 10.1007/s12369-012-0168-5
6. Delaunay, F., de Greeff, J., Belpaeme, T.: Towards retro-projected robot faces: An alternative to mechatronic and android faces. In: Robot and Human Interactive Communication (RO-MAN 2009), 18th IEEE International Symposium on, pp. 306–311 (2009). DOI 10.1109/ROMAN.2009.5326314
7. Delaunay, F., de Greeff, J., Belpaeme, T.: Lighthead robotic face. In: Human-Robot Interaction (HRI 2011), 6th ACM/IEEE International Conference on, p. 101 (2011). URL `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6281244`
8. Dering, B., Martin, C.D., Moro, S., Pegna, A.J., Thierry, G.: Face-sensitive processes one hundred milliseconds after picture onset. Frontiers in Human Neuroscience **5**(93) (2011). DOI 10.3389/fnhum.2011.00093. DOI 10.3389/fnhum.2011.00093
9. Ekman, P., Friesen, W.V.: Manual for the Facial Action Coding System. Consulting Psychologists Press, Inc., Palo Alto, CA (1978)
10. Forster, K., Forster, J.: Dmdx: A windows display program with millisecond accuracy. Behavior Research Methods, Instruments, & Computers **35**(1), 116–124 (2003). DOI 10.3758/BF03195503. URL `http://dx.doi.org/10.3758/BF03195503`. DOI: 10.3758/BF03195503
11. Hanson, D.: Exploring the aesthetic range for humanoid robots. CogSci-2006 Workshop: Toward Social Mechanisms of Android Science (2006)
12. Hashimoto, M., Morooka, D.: Robotic facial expression using a curved surface display. Journal of Robotics and Mechatronics **18**(4), 504–510 (2006). URL `http://www.fujipress.jp/finder/xslt.php?mode=present&inputfile=ROBOT001800040017.xml`
13. Hirth, J., Schmitz, N., Berns, K.: Playing tangram with a humanoid robot. In: Robotics; Proceedings of ROBOTIK 2012; 7th German Conference on, pp. 1–6 (2012). URL `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6309474`
14. Ishiguro, H.: Understanding humans by building androids. In: Proceedings of the SIGDIAL 2010 Conference, pp. 175 (W10–4330). Association for Computational Linguistics, Tokyo, Japan (2010). URL `http://aclweb.org/anthology/W/W10/`
15. Jaeckel, P., Campbell, N., Melhuish, C.: Facial behaviour mapping - from video footage to a robot head. Robotics and Autonomous Systems **56**(12), 1042–1049 (2008). DOI 10.1016/j.robot.2008.09.002
16. Kędzierski, J., Muszyński, R., Zoll, C., Oleksy, A., Frontkiewicz, M.: EMYS - emotive head of a social robot. International Journal of Social Robotics **5**(2), 237–249 (2013). DOI 10.1007/s12369-013-0183-1
17. Kuratate, T.: Statistical analysis and synthesis of 3D faces for auditory-visual speech animation. In: Proceedings of AVSP'05 (Auditory-Visual Speech Processing), pp. 131–136 (Vancouver Island, Canada, July 24-27, 2005). http://www.isca-speech.org/archive_open/avsp05/av05_131.html
18. Kuratate, T., Masuda, S., Vatikiotis-Bateson, E.: What perceptible information can be implemented in talking head animations. In: Robot and Human Interactive Communication (RO-MAN 2001), 10th IEEE International Workshop on, pp. 430–435 (Baurdeux and Paris, France, Sep.9-13, 2001). DOI 10.1109/ROMAN.2001.981942
19. Kuratate, T., Matsusaka, Y., Pierce, B., Cheng, G.: "Mask-bot": A life-size robot head using talking head animation for human-robot communication. In: Humanoid Robots (Humanoids 2011), 11th IEEE-RAS International Conference on, pp. 99–104 (Bled, Slovenia, Oct.26–28, 2011). DOI 10.1109/Humanoids.2011.6100842
20. Kuratate, T., Pierce, B., Cheng, G.: "Mask-bot" - a life-size talking head animated robot for av speech and human-robot communication research. In: Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP 2011), pp. 107–112 (Volterra, Italy, Aug.31–Sep.3, 2011). http://www.isca-speech.org/archive/avsp11/av11_111.html
21. Kuratate, T., Riley, M., Pierce, B., Cheng, G.: Gender identification bias induced with texture images on a life size retro-projected face screen. In: Robot and Human Interactive Communication (RO-MAN 2012), 21st IEEE International Symposium on, pp. 43–48 (Paris, France, Sep.9–13, 2012). DOI 10.1109/ROMAN.2012.6343729
22. Maejima, A., Kuratate, T., Pierce, B., Morishima, S., Cheng, G.: Automatic face replacement for a humanoid robot with

3d face shape display. In: Humanoid Robots (Humanoids 2012), 12th IEEE-RAS International Conference on, pp. 469–474 (Osaka, Japan, Nov.29-Dec.1, 2012)
23. Mori, M.: The uncanny valley (in japanese). Energy **7**(4), 33–35 (1970)
24. Moubayed, S.A., Alexandersson, S., Beskow, J., Granström, B.: A robotic head using projected animated faces. In: Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP 2011), p. 69 (Volterra, Italy, Aug.31–Sep.3, 2011). http://www.isca-speech.org/archive/avsp11/av11_071.html
25. Munhall, K., Jones, J., Callan, D., Kuratate, T., Vatikiotis-Bateson, E.: Visual prosody and speech intelligibility: Head movement improves auditory speech perception. Psychological Science **15**(2), 133–137 (2004). DOI 10.1111/j.0963-7214.2004.01502010.x
26. Pierce, B., Kuratate, T., Maejima, A., Morishima, S., Matsusaka, Y., Durkovic, M., Diepold, K., Cheng, G.: Development of an integrated multi-modal communication robotic face. In: Advanced Robotics and its Social Impacts (ARSO), 2012 IEEE Workshop on, pp. 101–102 (Munich, Germany, May 21-23, 2012). DOI 10.1109/ARSO.2012.6213408
27. Pierce, B., Kuratate, T., Vogl, C., Cheng, G.: "Mask-Bot 2i": An active customisable robotic head with interchangeable face. In: Humanoid Robots (Humanoids), 2012 12th IEEE-RAS International Conference on, pp. 520–525 (Osaka, Japan, Nov.29-Dec.1, 2012)
28. Pollick, F.: In search of the uncanny valley. In: P. Daras, O. Ibarra (eds.) User Centric Media, *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 40, pp. 69–78. Springer Berlin Heidelberg (2010). DOI 10.1007/978-3-642-12630-7$_8$
29. Turati, C.: Why faces are not special to newborns: An alternative account of the face preference. Current Directions in Psychological Science **13**(1), 5–8 (2004). DOI 10.1111/j.0963-7214.2004.01301002.x
30. Vatikiotis-Bateson, E., Kroos, C., Kuratate, T., Munhall, K.G., Pitermann, M.: Task constraints on robot realism: the case of talking heads. In: Robot and Human Interactive Communication (RO-MAN 2000), 9th IEEE International Workshop on, pp. 352–357 (Osaka, Japan, Sep.27-29, 2000). DOI 10.1109/ROMAN.2000.892522

Takaaki Kuratate received both the Bachelor of Engineering and Master of Engineering degrees in Applied Physics from the University of Electro-Communications, Japan. He received his PA.D. degree in Information Science from the Nara Institute of Science and Technology, Japan in 2004. He worked at the TOSHIBA Research & Development Center, Japan (1991-2000), and the Advanced Telecommunication Research Institute (ATR), Japan (1997-2006) as a researcher working on topics in computer graphics, speech processing, image processing, and computer vision, with his primary emphasis on auditory-visual speech processing, and 3D face analysis and synthesis.
From 2006 to 2009, he was a postdoctoral research fellow at MARCS Auditory Laboratories, University of Western Sydney, Australia where he developed a text-to-auditory visual speech system based on extended face database research from ATR. Since 2010, he works as a senior researcher at the Institute for Cognitive Systems, Technische Universität München, where he heads development of the Mask-bot platform. His research explores topics in building machines able to communicate naturally with people, and includes both graphical and robotic solutions.

Marcia Jean Riley received a Master of Science degree in Computer Science from the University of Geneva, Switzerland at MIRALab, and a Bachelor of Science in Mathematics from the Johns Hopkins University, Baltimore, Maryland. She is currently a Ph.D. candidate at the University of Bremen, Germany. Her research interests lie in humanoids and humanoid robots, with specific interests in creating algorithms and approaches that profit from human-robot communication to efficiently effect transfer of and knowledge about skills and tasks. She is also interested in applications of machine learning to humanoid robotics, computer animation, human movement, human skill acquisition, and robot face representations and communication capabilities. She worked at the Institute for Cognitive Systems, Technische Universität München from 2010 to 2012 as a research assistant, and prior to this worked in Humanoid Robotics at ATR in Japan.

Gordon Cheng is the Professor & Chair of Cognitive Systems, Founder and Director of the Institute for Cognitive Systems, Technische Universität München. Formerly, he was the Head of the Department of Humanoid Robotics and Computational Neuroscience (2002-2008), ATR Computational Neuroscience Laboratories, Kyoto, Japan. He was the Group Leader (2004-2008) for the JST International Cooperative Research Project (ICORP), Computational Brain. He has also been designated as a Project Leader(2007-2008) for National Institute of Information and Communications Technology (NICT) of Japan. Additionally, he hold visiting professorships worldwide in multidisciplinary field comprising: Mechatronics (France), NeuroEngineering (Brazil) and Computer Science (USA).
He held fellowships from the Center of Excellence (COE), Science, and Technology Agency (STA) of Japan. Both of these fellowships were taken at the Humanoid Interaction Laboratory, Intelligent Systems Division at the ElectroTechnical Laboratory (ETL), Japan. He received a PhD in Systems Engineering (2001) from the Department of Systems Engineering, from The Australian National University, and Bachelor (1991) and Master (1993) degrees in Computer Science from the University of Wollongong, Australia. He was also the Managing Director of the company, G.T.I. Computing in Australia.
His research interests include, humanoid robotics, cognitive systems, brain machine interfaces, bio-mimetic of human vision, human-robot interaction, active vision and mobile robot navigation. Prof. Cheng is the co-inventor of approximately 15 patents and co-authored approximately 180 technical publications, proceedings, editorials and book chapters.

He is a senior member of the IEEE Robotics and Automation and Computer Society. He is on the editorial board of the International Journal of Humanoid Robotics.