2012 12th IEEE-RAS International Conference on Humanoid Robots
Nov.29-Dec.1, 2012. Business Innovation Center Osaka, Japan

"Mask-Bot 2i": An active customisable Robotic Head with Interchangeable Face

Brennand Pierce, Takaaki Kuratate, Christian Vogl and Gordon Cheng
Institute for Cognitive Systems (http://www.ics.ei.tum.de), Technische Universität München

***Abstract*—This paper describes the development of "Mask-Bot 2i" (codenamed: *Kabuto*), a robotic head designed for research into Human-Robot-Interactions. The uniqueness of our new robotic head is that the appearance of its face can be alter on-the-fly. Different faces can be projected onto the active head system. The head is fully active with a 3-DOF neck, with support for biannual hearing as well as video camera for seeing. An interchangeable face is the main feature of this new Mask-Bot, the head can be equipped with an average face mask as well as a highly customised individualised face that can be easily exchanged. Additionally, the actuation of the head has been designed to match the natural head movements of an average human. Thus, enabling the head and face to be articulated synchronously to the speech production while the natural head movement matches that of the animated face. The design and realisation of this new system is presented in details in this paper.**

I. Introduction

During the last decades numerous humanoid robotic systems have come to be. One element that is consistent among them all is the fact they are all equipped with a head. As from the practical point of view, it is the most logical place to mount the visual and the auditory systems. But one might say that it is also an important part of human-robot interaction (HRI), especially for face-to-face communication, as the head provides a natural means of human interaction with a robot. This leads to the question, "what elements are important in the design of a robotic head for HRI?". For instance, we have studied how the quality of the projected image effects people's identification of the avatar's gender with the original "Mask-Bot" [1]. For interaction, the head should be able to communicate using auditory (verbal and non-verbal) as well as visual communication. But what about its appearance? To explore this question, we developed a robotic head that can be used to carry out research into how the appearance of an animated face displayed onto a robotic head effects HRI.

When a robotic head is designed it normally has one primary goal. For example, it could be designed with the aim of being a stable platform for activated cameras. Good examples of this are the head of the "iCub" humanoid robot [2]; or Karlsruhe humanoid head [3]. Another aim is to look as life-like as possible, where the texture of the skin, the mimicking of the muscles are most important. Hanson robotics developed a very realistic head used in the "Albert" version of HUBO humanoid robot [4], which utilises a high number of servo motors that were designed to mimic human muscles and deform its rubbery skin. This means the head is able to display emotions as well as to articulate the mouth when it speaks. Simpler heads compared to the Hanson's head are also used for displaying emotions. An early example of this type of head is MIT's KISMET robot [5], which was designed to display emotional expressions and had a simple mouth. These latter two examples rely heavily on complex mechanical structures, which need a high number of motors to be controlled in order to modify their facial expressions.

Fig. 1. The new "Mask-Bot 2i" (codenamed: Kabuto).

To overcome these problems, there is an emerging type of humanoid heads with the concept of displaying an avatar instead of relaying on a complicated mechanical mechanism. Examples of this would be the "LightHead" by Delaunay et al.[6], [7] and the "Curved Screen Face" from Hashimoto and colleagues [8], as well as our own robotic head "Mask-Bot" [9], [10]. This type of robotic heads have the advantage that the face can be changed and animated fairly easily. Also the articulation of the face does not rely on complex mechanical components. This means that the mouth can be animated and synchronised with the vocal system and it can also display emotions. But current research on the projected heads have mainly focused on the animated side and have neglected the complete modalities that a robotic head needs, for example the integration of stereo microphones for "hearing", cameras to "see" the world and the full articulation of head

978-1-4673-1369-8/12/$31.00 ©2012 IEEE

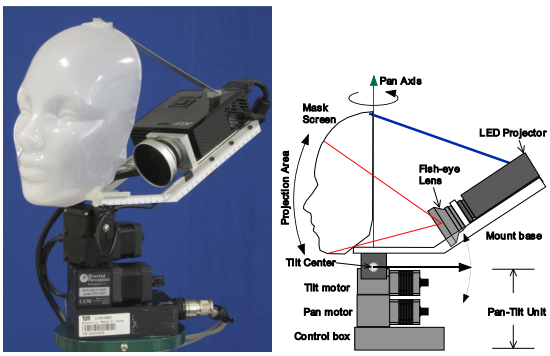

Fig. 2. 1st generation "Mask-Bot"

movements. To date, all the current projected heads have a fixed, very simple surface to project their avatar. From our experience this has a down side as different avatars work better with different shaped masks.

In this paper, we present our new humanoid head with a projected face, Mask-Bot 2i (see Fig. 1). This robotic head has been developed as a standalone platform for performing research into human-robot interaction (HRI). In the next section, we present the system requirements that lead to the specification for this robotic head. Section III and IV present the unique feature of our robotic head, which is its interchangeable projected face. In section V, VI and VII the implementation of the complete system is discussed in detail. Finally, the results and the conclusion are presented in section VIII and IX.

II. System Requirements

To build our new head we took into account our experience with the original prototype Mask-Bot (see Fig. 2), which was design and created to validate the feasibility of a projected face for HRI. With Mask-Bot 2i we decided to pay more attention to the complete system and not just focus on the displaying aspects of the avatar. We found a couple of short coming with the previous head: The overall system was noisy, which was distracting when using it for HRI experiments. Also, the mask we used had too much surface details in the eye and lip regions, which made it hard to align with the avatar being projected. This mask was also not interchangeable, which means we were unable to carry out experiments into how different masks affect the avatar being projected. There was also no way for Mask-Bot to truly interact with its surroundings as it had no way to "see" or the ability to know where the person, who it was interacting with, was located. This lead to an unnatural HRI, as the robot would not always be facing the person it was interacting with. This lead to the new requirements for a new robotic head:

- The robot head should have an interchangeable mask, so that we can experiment with different masks and determine how it effects the appearance of the avatar being projected.
- The robot head should model the major degrees of freedom (DOFs) of a human, so that there is adequate velocity and range to replay human-like motion data in a smooth and natural manner.
- The overall size should be as close as possible to an average adult human, so that the projected avatar can be displayed at a natural size.
- It should provide a vision system, so that it can interact better with the human, for example tracking the face of a person during interaction.
- It should have an audio system capable of 3D sound localisation, so that it can locate the person that it is interacting with.
- It should be quiet, so that the sound of the head's hardware does not create a distraction during HRI experiments.

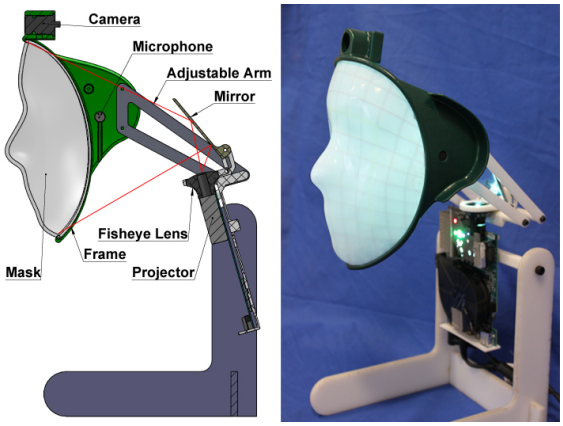

Fig. 3. Test-rig. Built to test projector position coupled with different fisheye lenses and mirror configurations.

Based on these system requirements, we derived our new full system.

III. Projector System

The main feature of this robotic head is the avatar animated onto a mask using a projection system. There are three main hardware components used to achieve this: 1) the projector; 2) the optics used to modify the light beam; and 3) the mask. So when designing our new head this projection system hardware dominated our design. Therefore, we had two main design goals for the projector system compared to the original. First we wanted to reduce its weight, thus we could reduce the size of the actuators, which in turn would reduce the noise. The second goal was to make the overall projection system more compact so that it fits into the footprint of an average adult human.

The most obvious and easiest way of making the projection system smaller and more compact was to use a small projector. Through numerous trials, we selected the LED projector C112 (Acer Inc.), which is 70 ANSI lumens and has a contrast ratio 1000:1. This has the advantage of being 421% lighter then the original projector e.g., 138g compared to 582g, but has the disadvantage of only having 35% the lumens. After experiments with this new projector we felt that the trade off was worth it - as it can still provide adequate brightness for the projection of the face. After the selection of the projector we needed to design the optic system. In

TABLE I
SYSTEM OVERVIEW; MOTOR, SENSORS AND COMPUTATIONAL SYSTEM.

Kinematics	3 DOF in the neck arranged as pan, tilt and roll.
Actuators	5W brushless motor - "351008" by maxon.
Gear Ratio	Pan 1:200; Tilt 1:120; Roll 1:120.
Encoders	14bits digital encoder - "Vert-X 13" by contelec.
Vision Camera	1920 x 1080 @ 30Hz - "C920" by Logitec.
Auditory	Stereo microphones.
Inertial Sensors	6 axis IMU, combined 3-axis gyroscope and 3-axis accelerometer - "MPU-6000" by invensense.
Control Module	FPGA, running onboard PID at 5Khz - "Sparten 6 XL45" by Xilinx.
MOSFET	3 x 5A integrated three phase motor driver, controlled at 24.4Khz- "L6234" by STMicroelectronics.
Communication	Ethernet; UDP packets, running at 2Khz.
Control PC	Intel i7 PC, running Ubuntu Linux OS.
Software Framework	Robot Operating System (ROS)

the original Mask-Bot we used a fisheye lens which required to be aligned along the same plane as the projector. This had one major disadvantage as it required the projection to be very far away from the mask, thus, making the overall system very long, with a large volume.

Therefore, for our new projection system we needed to validate different combinations of different lenses and mirrors to determine the smallest volume we could achieve. For this purpose we created an experimental test-rig, as shown in Fig. 3, which allows us to easily modify all the key variables, i.e. mirror, fisheye lens and alignment between components. We achieved this by building different adjustable arms with different sizes and shapes. To test the result we projected a grid pattern which made it possible to test the resulting image with respect to the area covered and the focal of the resulting image.

After our empirical studies we determined that a fisheye lens by "pixeet", designed for a mobile phone camera with an viewing angle of 180 degrees, the size 30mm x 17mm and weight 18g, gave the best results when size and weight were given the highest priority. The mirror, which was aligned the same as in the configuration shown in Fig. 3, gave us the best result in the smallest volume. The main problem we faced with the alignment was the trade off between overall size and covering the complete mask. Also, as the projector is LCD based and the focus is designed for a flat screen, which means it was difficult to get the focus to be perfect for the complete mask due to its complex 3D shape. Thus, we had to make a compromise. We made sure that the two most important features, the eyes and mouth, were in focus. That means the edge of the face and tip of the nose were slightly out of focus, which both can be said to be less significant for HRI.

IV. INTERCHANGEABLE MASK

In our previous prototype version of the Mask-Bot we simply used a commercially available manikin head. In this new version it would be desirable to be able to experiment with different shaped masks: from a very generic version, where we could project any avatar onto, all the way to a highly specific mask that matches the projected avatar

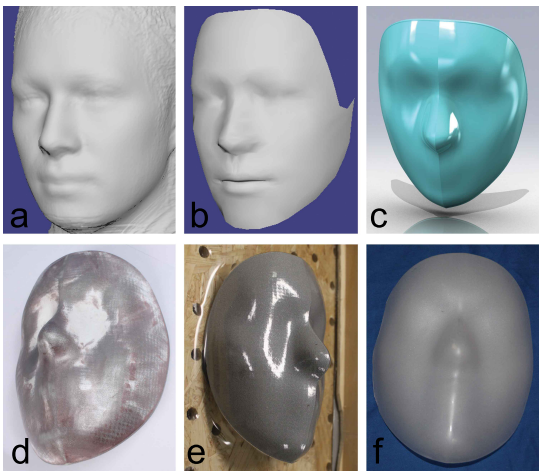

Fig. 4. Designing the interchangeable mask; a) Average face data; b) The cleaned up dataset; c) the CAD model of the mask with the correct outer shape to fit into the Mask-Bot frame; d) The 3D printed mold after being sprayed in white paint then sanded back to give a smooth surface finish; e) The vacuum forming process; f) The final sprayed mask.

created from 3D scan data of a specific human subject. This means that we needed to design the new head to have an interchangeable function, which resulted in a frame design where different masks can be attached.

In order to produce a version of an interchangeable mask, we first developed a face-model which uses a mean average of 124 faces (31 faces from each Caucasian male group, Caucasian female group, Asian male group and Asian female group), with an age range of 18 to 50 years old and with the average age of 29.4 years to create an "average" face (Fig. 4.a). Before we could use this data, it needed to be preprocessed to reduce the noise and to trim the excess data (Fig. 4.b).

Next we imported this data into CAD software, so that we can turn it into a solid object which can be manufactured. Afterwards, we applied a transformation function which turned this data into a planner surface that can be modified. At the same time we applied a smoothing function to reduce the fine details like in the lip and eye area. Due to our experience with the original Mask-Bot, these features are too detailed to match a large selection of general avatars. Then we trimmed the excess parts, like the ears and the back of the head, and added a rim around the edge so that it fits into the frame of the head (Fig. 4.c).

We needed a way to turn the CAD planner model into a plastic mask. For this we wanted to use the vacuum forming manufacturing method. Thus, we needed to make a mold, so we transformed the planner surface into a solid object by thickening it by 2mm. We then 3D printed this mold out of aluminium (Fig. 4.d), using selective laser sintering process (SLS). This means the mold can take temperatures of up to 172 °C. The 3D printed mold was then used on a vacuum table with a 1mm thick PETG clear plastic which produced a clear mask (Fig. 4.e). The last step was to paint the clear mask with a special rear projection paint by "Goo Systems", which gave the finished mask a silver finish (Fig. 4.f) and has

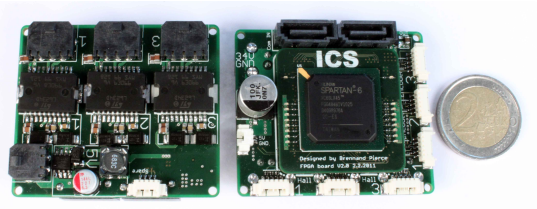

Fig. 5. The electronic (FPGA-based) module, capable of control three BLDC motors with 5A power rating per motor, as well as interfacing with three digital encoders.

shown to yield very good results when the avatar is projected onto its surface.

V. Control System

The control system is divided into two layers: low-level and high-level control. The high-level control consists of a standard linux PC running Robot Operating System (ROS), whereas the low-level control consists of a FPGA which in turn controls the MOSFETs. For a complete list of our control system please refer to Table I.

A. Low-Level

The low-level control is accomplished by a self-contained single control board (Fig. 5), which measures only 48mm x 56mm. The main purpose of this control board is to interface with the three encoders ("Vert-X 13" by contelec), control the three motors ("351008" by MAXON) and compute the PID loop. For the control of the motors we selected a three phase motor driver by STMicroElectronics "L6234", which has six TTL inputs to control the three half-bridge MOSFETs that has a maximum power rating of 5A. A single FPGA ("Sparten 6 XL45" by Xilinx) takes care of the low-level logic. This FPGA has a PID controller that is capable of doing position and velocity control at 5Khz, as well as communicating with the motor drivers at 24.4Khz.

B. High-Level

The high-level is controlled by a standard PC running Ubuntu Linux with the Robot Operating System (ROS) as the software framework. A single ROS node is used as the interface to the low-level controller. This node takes the desired position or velocity as input and publishes the current states of the DOF, which contain the position and velocity. This node also physically communicates with the control board via ethernet with standard UDP packets. The sent packet contains the desired position and velocity as well as the PID gains for all DOF and once this packet is received the control board responds with a UDP packet containing the current velocity and position of all DOF at 2Khz. The transmission latency for the complete loop has been measured at 26.66μs.

A number of control nodes was used to fully test the capabilities of our control module, including a node that can replay human motion captured data to test if our robotic head is capable of tracking the desired human movements (see Fig. 8). We also developed a node that generates desired positions of square waves to test for maximum velocity and acceleration (see Table III).

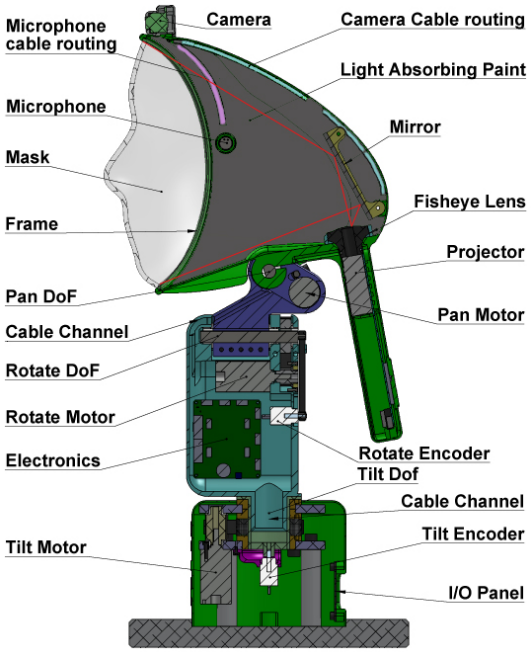

Fig. 6. Cut through of the CAD model of the new Mask-Bot.

VI. Mechanical Design

After determining the layout of the projection system from the test-rig (Fig. 3), with all the main components already positioned in CAD, we then created a mono-shell that covered all the components – transforming it into a self-contained head. The mono-shell is then 3D printed using SLS out of nylon. The total weight, including shell, projector system, interchange mask, camera and stereo system, is only 443g. In making sure that the stray light is not reflected by the inside of the head we painted it with black light absorbing paint by "Goo Systems". We also designed a neck by using brushless DC motors coupled to a shaft with pulley belts. A diagram of the key features can be seen in Fig. 6 and the resulting design can be seen in Fig. 1.

VII. Auto Calibration

As we use a 2D projector to project the avatar onto the 3D mask, we have to consider several important factors, which may disturb a proper fit of the projected face and their corresponding locations. Firstly, the optical system consisting of mirrors and lenses distorts the image being projected. Secondly, the mask itself is a complex 3D shape, which is difficult to represent by mathematical models. Our first approach to compensate for the optical distortion was to project a known pixel grid and to observe the distortion on the mask. We then used this to compute a distortion rule, which has been applied to the 3D face in the inverse manner [9]. However, this procedure required various manual adjustments and did not consider the particular 3D shape of the screen. Therefore, we wanted a way to automate the

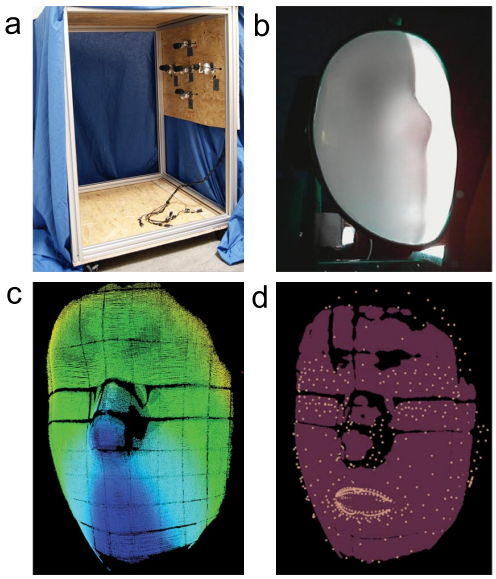

Fig. 7. Front top left working clockwise; a) the Auto calibration booth, note the 5 cameras; b) The pattern being projected on to the mask, this picture was taken by one of the cameras; c) The 3D cloud map generated from the booth; d) The 3D cloud map (red) from the booth fitted to the CAD model (yellow).

calibration procedure and provide high robustness to any optical system. The auto calibration presented in this section will give us more opportunities to alter optics easily, as it is based only on the known input and the measurable output of the whole optical system.

A. Calibration Booth

To calibrate distortions caused by the optical system and the 3D mask, we built a calibration booth (Fig. 7.a). This calibration booth consists of five cameras, arranged into four stereo pairs, where the centre camera is paired with each outer camera. We had to use multiple cameras so that the field of view (FOV) covers the whole mask. Then the Mask-Bot can be placed inside and scanned under ideal conditions. As a result, it can automatically generate the best mapping between the projected avatar and the 3D mask. The stereo systems are calibrated using Zhang's method [11] to determine the intrinsics of each camera and epipolar geometry calculus [12] to compute the stereo transformation matrices.

B. 3D Data Recording

The most common way to acquire 3D shapes with a stereo camera system are structured-light techniques. Usually, a monochrome or colored pattern is projected onto an object, from which both cameras take a picture. Using the projected pattern, we are able to detect corresponding points on the object, which can be projected back to 3D space using their binocular disparity. Actually, we do not aim to reconstruct the 3D shape of the mask, as it is already known. Instead, for our auto calibration, we need linkages between single pixels on the projector and their position after projection on the 3D screen. This is why we chose Gühring's line-shift method

TABLE II
NOISE OF THE HEAD IN DECIBEL, MEASURED 1 METER AWAY.

Peak Noise at different speeds dB			
	None	Slow	Fast
Original Mask-Bot	32	61	64
Mask-Bot 2i	31	53	57

[13], which allows an accurate reconstruction of a 3D object with high-contrast patterns and which is robust towards ambient light. An example of the pattern been projected on the mask can be seen in Fig. 7.b. We modified the method for thicker lines with a lower frequency, which resulted in a longer recording time, but also a more robust detection of points due to improved contrast. Gühring's method also allows us to store the position of the pixel source of any projected point during recording, which helps us to find the required pixel/3D correlation. After recording, corresponding pixels in the camera observations can be identified by evaluating their sequence of underlying gray-coded patterns. The corresponding 3D point of a pair of points is computed from their lateral disparity obtained during camera calibration.

C. 3D Data Postprocessing

After reconstructing all of the 3D points, we have four 3D point clouds which are supposed to coincide. Due to calibration and reconstruction uncertainties, the reconstructed clouds may be rotated and shifted with respect to each other. The post-processing of the clouds consists of stitching and joining the obtained points, so we get one single cloud of unique linkages. For the stitching, we apply Horn's correspondence-based rigid transform method [14] with simultaneous outlier filtering using the RanSaC Algorithm. The result is a smooth 3D cloud with up to four 3D correspondences for each projector pixel, which can be averaged to one single point. The resulting cloud is shown in Fig. 7.c.

D. Face Calibration

With the resulting cloud, we have received a kind of look-up-table (LUT), which tells us which projector pixel we have to use to produce a certain 3D point and vice versa. That is, if we use a 3D cloud alignment strategy like the iterative closest point (ICP) method, we can align an avatar 3D mesh from a database to the reconstructed cloud (Fig. 7.d). Now considering one of the mesh vertices and taking that point from the reconstructed 3D cloud, which has a minimum distance to it, we can project the avatar vertex back to the projector 2D space by taking the linked projector pixel. As we can not ensure that all reconstruction errors in the cloud have been eliminated, we additionally do an averaging around a small neighbourhood of each vertex.

This method transforms an avatar according to a measured LUT holding the positions of the projector pixels and their location in 3D space after projection. The result is a distorted 2D face, which is scaled and positioned in the correct manner so that when it is projected onto the 3D mask it appears non-distorted as well as the eyes and lips are located correctly.

TABLE III
DOF PERFORMANCES, FROM TESTS ON "MASK-BOT 2I"

Motor Results			
	Range [°]	Velocity Max [°/s]	Acceleration Max [°/s^2]
Pan	±65	153.3	2168.6
Roll	±55	258.2	2273.5
Tilt	±30	266.9	2443.1

VIII. RESULTS

Several tests where made to evaluate the functionality of the head. They can be split into two sections: the projection system and mechanical system.

A. Mechanical System

To validate the mechanical hardware and the control software we need to make sure the system was able to track human motion data. To test this, we replayed data recorded from a real human, the resulting trajectory can be seen in Fig. 8. From this test we were able to show that the head was able to replay natural human motions. We also ran square waves through each DOF to quantify their maximum velocity and acceleration. These results can be seen in Table III.

We also quantified the reduction in noise of the overall system compared to the original Mask-Bot. To test this we recorded the sound level at a distance of one meter using a sound level meter. We ran three tests. The first was with the head still, thus recording the base level. The second test consisted of replaying slow human data which can be characterised by the normal movement the head is expected to perform. Finally, very fast and erratic data was replayed, which consisted of sudden fast changes in direction and pushed the motors to their performance limit.

The results can be seen in Table II. As you can see we managed to half the noise of the system for both slow and fast movements, which translates into big improvements for HRI, as the user was not distracted by the overall system sound. This reduction in the system noise was as a result of using smaller and higher quality motors.

B. Projection System

It is difficult for the projection system to provide quantifiable results as the appearance is subjective and prone to individual opinion. Therefore, it is beyond the current scope and we leave this work for further studies. Nonetheless, Fig. 1 provides an example of a final calibrated avatar projected on the mask where the calibration process matches the 3D shape with the avatar. Using the software described in our previous paper [9], the animated avatar is better displayed onto the mask as well as the overall appearance has improved compared to the original system.

IX. CONCLUSION

We have developed a compact active robotic head system that is capable of dynamically updating its appearance. Furthermore, this head has the ability to interchange different shaped 3D masks so that we can carry out experiments into how different masks affect the appearance of the avatar being projected. We can also explore what features are important in the appearance of the avatar for HRI interaction by experimenting with different avatars. We believe this system is easily integrated into any full-sized humanoid robot. We also presented an effective method that can automatically calibrate the 3D mask to compensate for the lens distortion using Gühring's line-shift method.

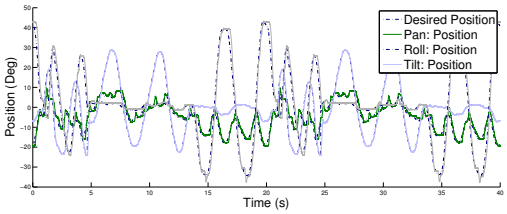

Fig. 8. Plot of the head replaying human motion data.

X. ACKNOWLEDGMENT

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

REFERENCES

[1] T. Kuratate, M. Riley, B. Pierce, and G. Cheng, "Gender identification bias induced with texture images on a life size retro-projected face screen," in *RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.*, IEEE, 2012, pp. 43 – 48.

[2] R. Beira, M. Lopes, M. Praga, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Saltaren, "Design of the Robot-Cub (iCub) Head," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, 2006, pp. 94–100.

[3] T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann, "The Karlsruhe Humanoid Head," pp. 447–453, 2008.

[4] J.-h. Oh, D. Hanson, W.-s. Kim, Y. Han, J.-y. Kim, and I.-w. Park, "Design of Android type Humanoid Robot Albert HUBO," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Oct. 2006, pp. 1428–1433.

[5] R. A. Brooks, C. Breazeal, M. Marjanović, B. Scassellati, and M. M. Williamson, "The cog project: building a humanoid robot," pp. 52–87, Jan. 1999.

[6] 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-MAN2009)*, pp. 306–311, 2009.

[7] ——, "Lighthead robotic face," *Proceedings of the 6th International Conference on Human-robot interaction (HRI'11)*, p. 101, 2011.

[8] 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.

[9] 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 *2011 11th IEEE-RAS International Conference on Humanoid Robots*. IEEE, Oct. 2011, pp. 99–104.

[10] T. Kuratate, B. Pierce, and G. Cheng, ""Mask-bot" - a life-size talking head animated robot for AV speech and human-robot communication research," *AVSP*, 2011.

[11] Z. Zhang, "A flexible new technique for camera calibration," Microsoft Research (MSR), Tech. Rep. MSR-TR-98-71, 1998.

[12] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer-Verlag, 2005.

[13] J. Gühring, "Dense 3-D surface acquisition by structured light using off-the-shelf components," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 4309. Universität Stuttgart; Fakultät Luft- und Raumfahrttechnik und Geodäsie. Institut für Photogrammetrie, 2001, pp. 220–231.

[14] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour, "Closed form solutions of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America*, vol. 5, no. 7, pp. 1127–1135, 1987.