MorphFace: A Hybrid Morphable Face for a Robopatient

Thilina Dulantha Lalitharatne, *Member, IEEE*, Yongxuan Tan, Liang He, *Member, IEEE*, Florence Leong, Nejra Van Zalk, Simon de Lusignan, Fumiya Iida, *Senior Member, IEEE*, and Thrishantha Nanayakkara, *Senior Member, IEEE*

***Abstract*—Physicians use pain expressions shown in a patient's face to regulate their palpation methods during physical examination. Training to interpret patients' facial expressions with different genders and ethnicities still remains a challenge, taking novices a long time to learn through experience. This paper presents MorphFace: a controllable 3D physical-virtual hybrid face to represent pain expressions of patients from different ethnicity-gender backgrounds. It is also an intermediate step to expose trainee physicians to the gender and ethnic diversity of patients. We extracted four principal components from the Chicago Face Database to design a four degrees of freedom (DoF) physical face controlled via tendons to span $\sim 85\%$ of facial variations among gender and ethnicity. Details such as skin colour, skin texture, and facial expressions are synthesized by a virtual model and projected onto the 3D physical face via a front-mounted LED projector to obtain a hybrid controllable patient face simulator. A user study revealed that certain differences in ethnicity between the observer and the MorphFace lead to different perceived pain intensity for the same pain level rendered by the MorphFace. This highlights the value of having MorphFace as a controllable hybrid simulator to quantify perceptual differences during physician training.**

***Index Terms*—Modeling and Simulating Humans,Medical Robots and Systems, Gesture, Posture and Facial Expressions,**

I. Introduction

ENCODING and decoding facial expressions is a complex process of communicating non-verbal messages, often associated with emotional states. Being able to receive and interpret facial expressions accurately is an important ability in communicating between one another [1], especially in the medical field as patients provide emotional cues during consultations, which may be verbal or non-verbal [2]. Studies have shown that one's ability to interpret facial expressions is conditioned by cultural differences [3], especially when using avatars [4], [5]. Other demographic features such as age [6], [7] and gender [8] also underpin the perception of facial expressions. Therefore, it is desirable to expose medical students to patients of various demographic features in order to minimise the perceptual subjectivity of facial expressions.

Traditional clinical medical education is less effective than the relatively new simulation-based education (SBE) approaches in achieving specific clinical skill acquisition goals [9]. SBE allows students to learn in a safe and effective environment, and helps them confront and learn from mistakes and errors [10], [11]. Recent advancements of robotic patients have shown promising new trends in SBE systems. Some of the commercial robotic patient manikins such as SimMan 3G [12] and Paediatric HAL [13] can respond to physical inputs with movement, verbal, and haptic feedback. The Paediatric HAL and dentistry training robots such as SIMROID [14] are also able to render facial expressions such as pain and discomfort, which further improves the realism of the training experience.

Many robotic systems in domains such as social robots and human-robot interaction (HRI) are also able to render human facial expressions to a high degree of accuracy, such as Erica [15] and Sophia [16]. These robots use a large number of actuators to facilitate the mechanical movements to simulate the expressions. However, it is challenging for a purely physical system to simulate faces of different demographics. In contrast, virtual human face simulation systems frequently used in computer graphics (CG) such as FACSHuman [17] can render high-fidelity human avatars of different demographics, but cannot respond to physical inputs. The benefits and limitations of using these two approaches were evaluated in greater detail in our recent review in facial expression rendering in medical training simulators [18].

Physical-virtual (hybrid) systems are able to respond to physical inputs and accurately render virtual human avatars, blending advantageous features from both physical and virtual systems. Many of the current approaches use a fixed physical shell and have the virtual human projected either from the front [19], or from the rear [20], [21] onto the shell. Projection mapping of virtual human avatars of different demographic features onto a fixed shell can result in mismatches in the projection caused by the variability in the human face anthropometry. This issue is particularly noticeable in medical training systems, where the rendered face of the patient serves as

Manuscript received: October, 15, 2020; Accepted December, 18, 2020. This paper was recommended for publication by Editor P.Valdastri upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the Robopatient project funded by the EPSRC Grants No EP/T00603X/1, EP/T00519X/1 and EP/T004509/1. *(T.D Lalitharatne and Y. Tan contributed equally to this work) (Corresponding author: T.D Lalitharatne.)*

T.D Lalitharatne, Y. Tan, L. He, F. Leong, N. Van Zalk and T. Nanayakkara are with Dyson School of Design Engineering, Imperial College London, London SW7 1AL, U.K. (e-mail: t.lalitharatne@imperial.ac.uk, yongxuan.tan16@imperial.ac.uk, l.he17@imperial.ac.uk, f.leong@imperial.ac.uk, n.van-zalk@imperial.ac.uk, t.nanayakkara@imperial.ac.uk)

S. de Lusignan is with Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, OX2 6GG, U.K. (e-mail: simon.delusignan@phc.ox.ac.uk)

F. Iida is with Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: fi224@cam.ac.uk)

Digital Object Identifier (DOI): see top of this page.

a major feedback modality, and the distance between the user and the rendered face is relatively short. The solution to this problem would be to have a system where the physical face can morph to match the projected virtual face to accommodate different facial dimensions.

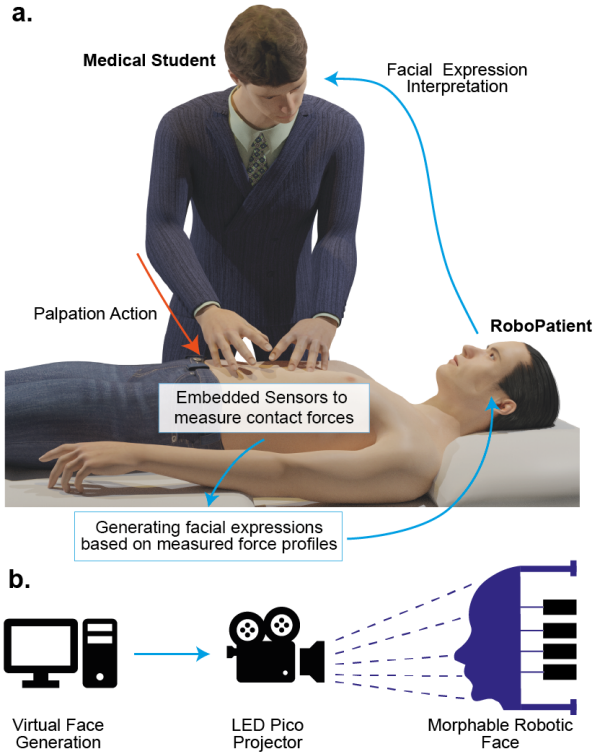

Fig. 1. **(a)** The student receives real-time facial expressions from the RoboPatient platform when they palpate the sensorised abdominal region. **(a)** MorphFace consists of a tendon-driven morphable physical face and a virtual face projected onto the physical face using a LED projector.

This paper proposes a first-ever novel hybrid morphable face capable of synthesising faces of different gender and ethnicity. Our proposed system uses a physical morphable robotic face with a virtual projection system. Vertical positions of three main facial features (eyes, nose and mouth), and the face length can be changed in the physical face. These feature changes are synchronised with the front projection of the virtual faces which adds colour, texture and more facial details to the rendered output. This development work focused on the facial expressions in response to physical examination as shown in Fig. 1 a, we have not at this stage included physical signs that may be clues to their intrabdominal pathology, or coordinate breathing with examination. An overview of the proposed system is shown in Fig. 1 b. More specifically, our contribution to this paper are three-fold:

1) We propose a novel data-driven approach for designing a morphable robotic face.
2) We develop and test a hybrid morphable face which can replicate the faces of three ethnic groups (White, Black and Asian) and two genders (Male and Female).
3) We show how human participants perceive the same pain level projected on the different ethnicity-gender backgrounds of morphface.

The rest of the paper is organized as follows: section II presents related work. Section III discusses the design and development of the physical face and system for rendering the virtual faces. Experiments and results are presented in section IV. Finally in section V, important conclusions are made, while suggesting possible future directions.

II. Related Work

One of the popular methods in implementing hybrid face rendering systems is front or rear projection. Furhat [20] is a social robot which is capable of interacting with humans via its face. It consists of a human head-like physical structure and face-like shell. Virtual human faces of different gender and ethnicity are projected onto a semitransparent shell using a rear-mounted projector. The physical face mask can be customised and replaced for synthesising faces of different sizes and shapes. Mask-bot [22] uses similar approaches to synthesize the faces with fixed face-like shells. Daher et al. [23] used an interchangeable translucent plastic shell as the physical structure of the patient body and face, and a virtual patient was rear-projected onto the shell using two LED Pico projectors to provide imagery for the patient's head and body. Bermano et al. [19] proposed an augmented physical avatar using front-projector-based illumination. To make their model synthesize more expressions, they introduced a system that decomposed the motion into low-frequency motions that were physically reproduced by a physical robotic head which consists of 13 actuators, and high-frequency details that were added using projected shading. While above-mentioned studies used either fixed or interchangeable faces like shells or physical robotic heads, none of them have considered or attempted to accurately represent the variations of facial features due to gender and ethnic diversity.

Hayashi et al. [24] proposed a dynamically modifiable soft mask face which was able to modify a life-size face mask to different subjects and different facial expressions. This design was fully mechanical, relatively large in size and used a large number of actuators and motors. Using a different approach, R. Schubert [25] proposed an adaptive filtering-based method for removing artifacts and mismatch between the projected face and the face-like shell in physical-virtual designs. However, their adaptive filtering provided a tuneable trade-off, sacrificing some visual detail to avoid distracting artifacts that may provide unwanted shape cues for the underlying physical display surface. Inspired by these two studies, MorphFace uses a data-driven approach to select the most significant features to be rendered mechanically, while simulating the less significant details virtually using front projection. This approach reduces the mechanical complexity while rendering the face to a good level of accuracy and realism.

III. Methods

MorphFace is consisted of a motor-actuated physical robotic face shell and a front-mounted projector for rendering virtual faces and facial expressions. The two systems are synchronised using a MATLAB program to match the dimensions of the facial features of the physical face with the projected virtual

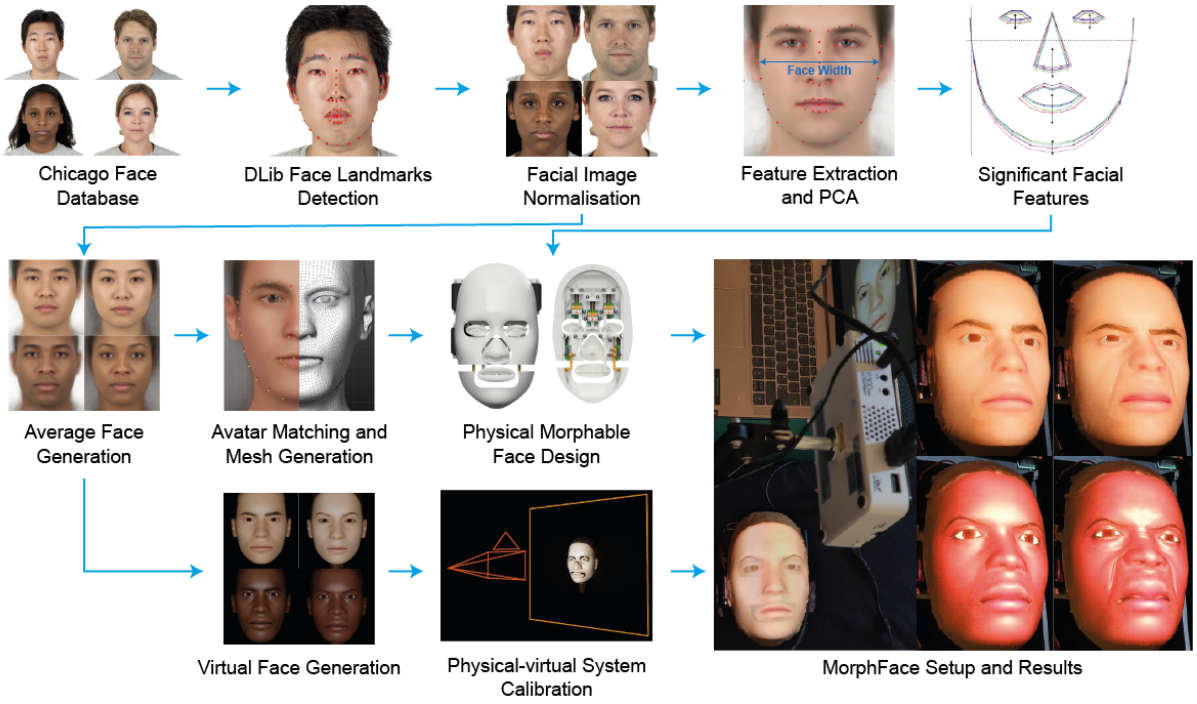

Fig. 2. Data-driven design process of MorphFace. Facial features were extracted from CFD (facial images are reproduced with permissions from CFD, Center for Decision Research, The University of Chicago) to identify the principal features to actuate physically, and the physical face was informed by the dimensions of the averaged faces. MorphFace changes the positions of the principal facial features on the physical face to match with the projected virtual faces.

facial details and expressions. The design and development of MorphFace is a data-driven design process using principal component analysis (PCA) on the anthropometric measurements of the frontal images from the Chicago Face Database (CFD) (version 2.0.3) [26]. The facial features were extracted using DLib facial landmark detection algorithm [27], and the physical and virtual faces were developed based on averaged faces of the CFD images. An overview of the design process is shown in Fig. 2.

A. Facial Image Processing and Principal Component Analysis

Facial images from the CFD were used to extract the facial anthropometric differences between people of different gender and ethnicity. We used 50 frontal natural faces from three ethnic groups: White, Black, Asian, and both males and females in our analysis. 68 facial landmarks of the 300 faces were extracted using Python with DLib, OpenCV [28] and a pre-trained face landmark detector [29]. Then we normalized all face images to have a resolution of 1500×1500 pixels, with a fixed face width by anchoring the pixel locations of landmark 2 and 16 ($|x2 - x16|$) across all images relative to the image size using similarity transformation. A Principal Component Analysis (PCA) was conducted to identify the key differing features between the facial landmarks of the image set.

14 facial measurements based on the normalised images were calculated by extracting the 68 landmarks ($L = \{(x1, y1), (x2, y2)...(x68, y68)\}$). We chose landmark 2 $(x2, y2)$ as the origin as all face images were standardized along it. The 14 measurements consisted of size-related measurements of facial features (right eye width ($|x37 - x40|$), right eye height ($|y39 - y41|$), left eye width ($|x43 - x46|$), left eye height ($|y44 - y48|$), nose height ($|y28 - y34|$), nose width ($|x32 - x36|$), mouth width ($|x49 - x55|$), mouth height ($|y52 - y58|$), bigonial breadth ($|x5 - x13|$), interocular breadth ($x40 - x43$), face height ($|y28 - y9|$)) and facial-feature-position-related measurements (eye position ($|y2 - y40|$), nose position ($|y2 - y28|$), mouth position ($|y2 - y52|$)) with respect to the origin. The PCA was implemented in MATLAB and the input vector to PCA consisted of 300 observations (300 selected faces) of the 14 measurements. As shown in Fig. 3, we chose PCs with percentage of explained variance $> 5\%$ and set a threshold > 0.4 for selecting the principal facial features. Based on this analysis, we identified two face-size-related features: face height and bigonial width; three face-part-position-related features: eyes, nose and mouth positions; and three face-part-size-related features: nose height, mouth width and mouth height. These features represent the highest variance among all features.

An average face for each gender and ethnicity was generated with OpenCV. The normalised images were divided into Delaunay triangles and affine transformations were performed on these triangles to match the position of the pixels inside each triangle for all images. The averaged image was then generated by dividing the sum of the RGB intensities of each pixel in the transformed image set by the number of images. Average face width data from Wen et al. [30] was used to convert the units of the average faces from pixels to millimeters. The converted dimensions of other facial features also match the result in

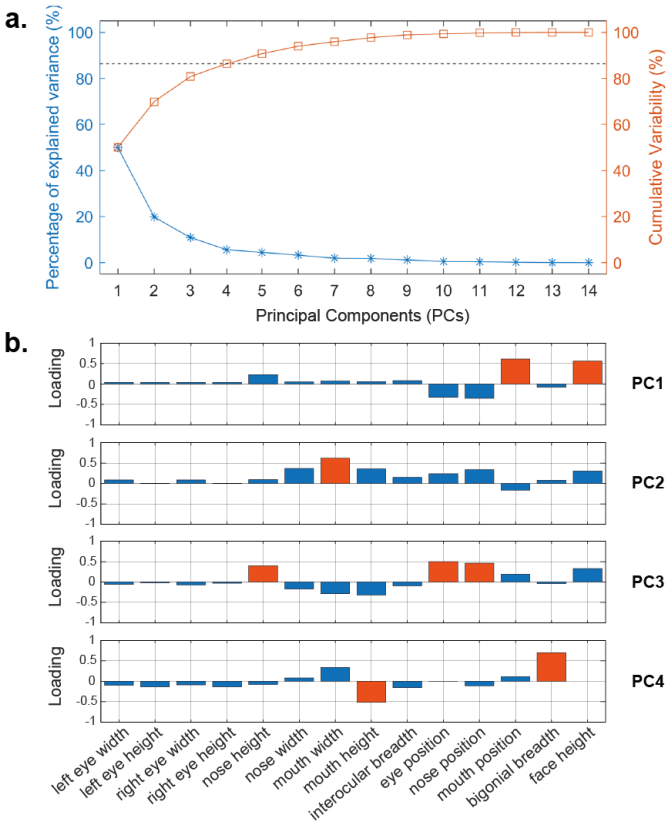

Fig. 3. PCA results: **(a)** Percentage of explained variance and cumulative variability vs. PCs. We chose PCs with percentage of explained variance $> 5\%$ for our design which resulted in cumulative variability of 86%. **(a)** Loading values of respective facial features contributed to PC1-PC4. We set a threshold $abs(Loading) > 0.4$ for selecting the principal facial features [colored in orange].

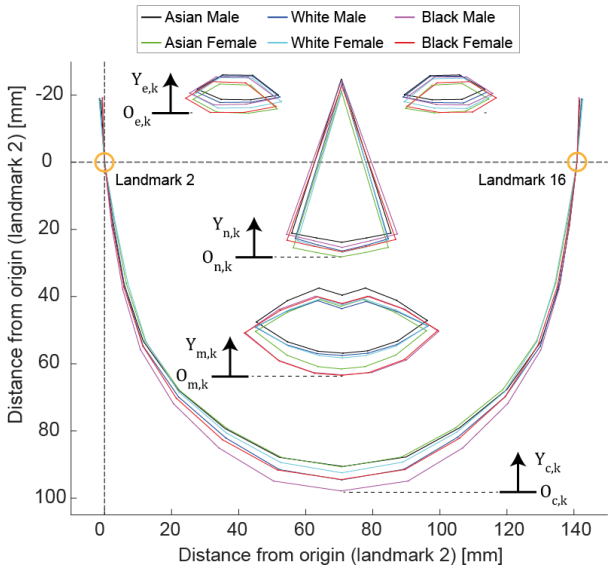

Fig. 4. Superimposed landmark plots of all six average faces with actual facial dimensions. Landmark 2 was set as the origin of the measurements. Based on this result with the PCs, we concluded that we could represent any given face as a function of eyes, nose and mouth position, and face height. To realise this in the physical face, we calculated the relative variations of eyes, mouth and nose positions ($Y_{e,k}$, $Y_{m,k}$, $Y_{n,k}$) and face height ($Y_{c,k}$) where $k = [Whitemale, Whitefemale, Blackmale, Blackfemale, Asianmale\ Asianfemale]$. Initial positions ($O_{e,k}$, $O_{m,k}$, $O_{n,k}$, $O_{c,k}$) were set as maximum deviation from the origin along the y-axis.

TABLE I
LOWER FACE HEIGHT BETWEEN THE AVERAGE FACES AND FACIAL MEASUREMENTS FROM WEN ET AL.

	Asian Male	Asian Female	Black Male	Black Female	White Male	White Female
Mean (95% Crl) from Wen et, al.	67.4 (59.7, 75.1)	64.6 (60.2, 69.4)	72.2 (64.4, 80.0)	65.5 (59.9, 72.1)	69.4 (64.4, 73.9)	63.0 (60.0, 66.2)
Average faces	66.8	62.5	72.5	67.8	68.2	65.7

[30], hence validating this approach, as shown in Table. I.

Fig. 4 depicts the supper-imposed landmark plots of the six average faces with actual facial dimensions resulting from the previous step. We compared this result with the principal features found using PCA to decide which primitives should be controlled in the physical face. Considering the physical setup implementation, We decided to use a fixed-size nose and mouth in this first version of MorphFace, even though the nose height, mouth width and height were found to be principal features. In our design, we kept the width of the face at constant for all gender and ethnicity. This allowed us to represent any given face as a function of the face height and the position of the principal features along the y-direction on the physical face to replicate the variances in different faces of different gender and ethnicity. Therefore, we could represent any given face by $f(Y_{e,k}, Y_{m,k}, Y_{n,k}, Y_{c,k})$.

B. Implementation of the Morphable Physical Face

The average images of the six demographic groups were used as references for generating 3D meshes representing each group. A MakeHuman [31] plugin was developed to iteratively changing dimensions of facial features of a base human face mesh, rendering the frontal view of the mesh, identify the landmarks, and calculating the errors between the landmarks of the current mesh and the reference image until the errors of all landmarks are smaller than a threshold of 20 pixels, or the maximum number of iterations (80) was reached. Fig. 6 shows the matching for the average White male mesh using this algorithm.

The Asian female face was selected as the base mesh for the physical face because it had the smallest face length. The MHX2 file of the Asian female MakeHuman mesh was imported into Blender, thickened to 4 mm, exported as an STL mesh, and imported into Autodesk Fusion 360. The STL mesh was separated into five parts: upper face, jaw, eyes, nose, and mouth. The upper face was used as the mechanical ground and cutouts were made to accommodate the vertical movements of the eyes, nose, mouth, and jaw. The initial positions of all face parts were set as their lowest vertical positions. Each component is controlled by a modular linear actuation mechanism as shown in Fig. 5 a.

The modular linear actuation unit consists of a linear guide (LWL7B Miniature Linear Rail Guide 7mm(W) $\times$ 40mm(L)), a movement piece (eyes, nose, mouth and jaw), a mechanical ground bracket for securing the linear rail to the upper face and cable sheath with PTFE tube insert for guiding the Bowden

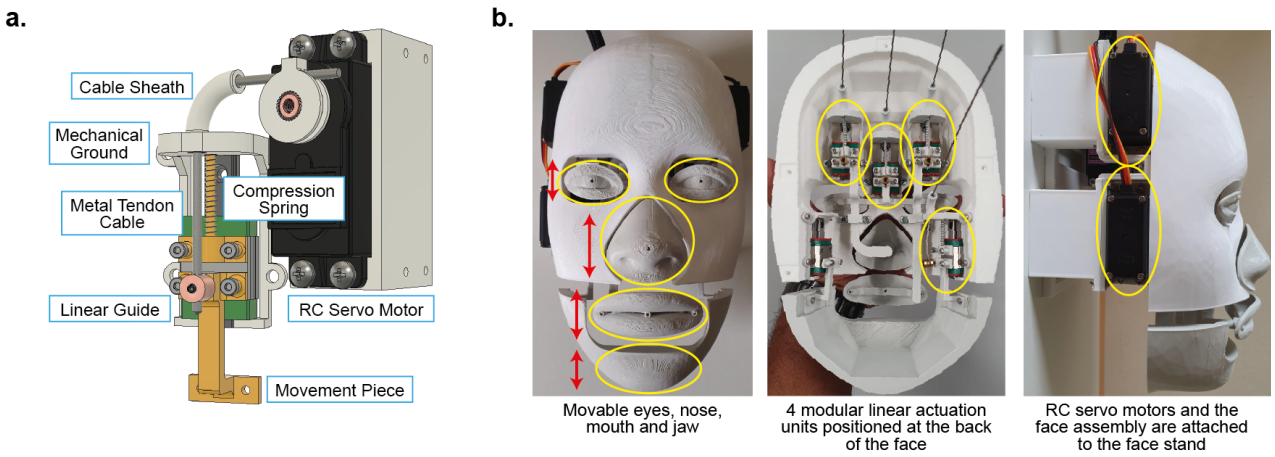

Fig. 5. Design and implementation of the physical face. **(a)** Autodesk Fusion 360 3D model of a modular linear actuation unit. **(a)** 3D printed physical face with movable face parts (movements indicated by red arrows), Bowden cables and RC servo motors installed.

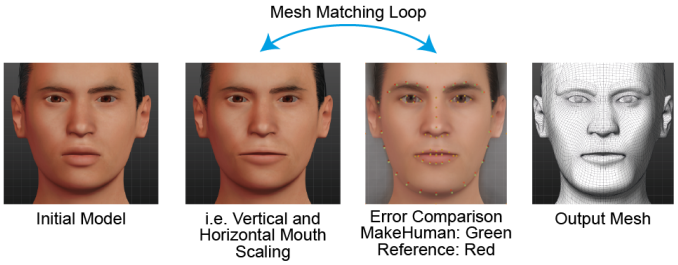

Fig. 6. Average face to MakeHuman 3D mesh matching process. The iterative error-minimising algorithm repeatedly changes a facial feature, identify the landmarks and compares the landmark positions against the reference photo, (second and third image).

cable, a compression spring to restore the movement piece position, and a RC servo motor (MG996R) to actuate the whole unit. Clockwise-rotation of the pulley attached to the metal gear of the servo motor tightens the Bowden cable which pulls the linear guide and moves up the attached movement piece, and anticlockwise-rotation of the pulley loosens the Bowden cable and the compression spring restores the position of the movement piece. All parts of the modular actuation unit, the upper face, and movement pieces were 3D printed with PLA. The layout of the modular linear actuation units and RC servo motors are shown in Fig. 5 b.

Servo commands were updated as a servo angle vector for each face, given by $[\theta_{e,k}, \theta_{n,k}, \theta_{m,k}, \theta_{c,k}]$. We calculated this vector for each face based on the simple relationship between servo pulley motion and linear motion of the guide as $[Y_{e,k}, Y_{m,k}, Y_{n,k}, Y_{c,k}]/r$ where r is the radius of the servo pulley. All pulleys were designed to have a diameter of $12mm$. A 12-channel Pololu Maestro servo driver was used to control the motors. Two layers of skin-color Nylon fabric were used as the skin of the physical face. Its elasticity and opacity helped to blend the margin space around the movable face parts and gave the model a 3D texture visually resembles the human skin.

C. Virtual Face Generation

Pain expressions were generated using a modified version of FACSHuman [17], by setting the activation intensities for action units (AUs) $4, 7, 9$, and 10 at $24\%, 21\%, 72\%$ and 100% based on our survey-based preliminary study on the effective weightings of AUs on perceived pain intensity. The resultant expression from the combination of these AUs was applied to the matched MakeHuman models. Two meshes for each gender-ethnicity face were generated: a neutral one (no pain) where the activation intensities are all 0, and a painful one (100% pain) with the activation intensities as described. The MakeHuman models were imported into Blender, and texture and colour maps were applied to each mesh. The camera position was found using the same mesh matching algorithm and each mesh was rendered at 1080×720 pixels, resulting in a total of 12 virtual faces (neutral and painful faces for 2 genders and 3 ethnicities).

D. Integration of Physical and Virtual system

We used an LED projector (AAXA P300 Neo Smart Android Mini Pico Projector, AAXA Technologies) to project the images synthesized by the virtual system, as shown in Fig. 7 a. The system was calibrated following three steps: projector distance, throw ratio, and feature positions, as shown in Fig. 7 b. The Blender Projector plugin was used to calculate these parameters before they were set for the physical model. The CAD model of the face with face parts was imported into Blender and a 2-D plane was positioned adjacent to the back of the face. The jaw was moved up to its maximum position so that the face length matches an Asian female's face length. The virtual face of an Asian female was then projected onto the face CAD, and the matching between the projection and the CAD model was inspected by eye. The position of the projector and its throw ratio was tuned until the face contour and mouth width of the projection match with the CAD model. The vertical position of the face parts was then moved to match with the projection.

A MATLAB program was used to synchronise the virtual projection and the robotic face variations. The program

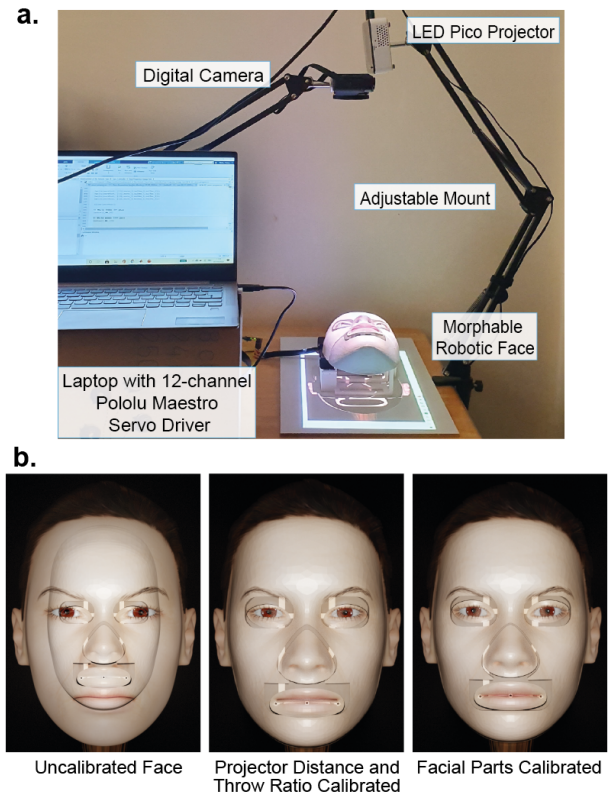

Fig. 7. **(a)** Overall set-up after system integration. The projection distance of the LED Pico projector (throw ratio: 1.4) was set using an adjustable mount. A digital camera was mounted close to the projector to capture results. **(a)** System calibration in Blender. Projector distance and throw ratio were set by checking the matching accuracy at the contour of the face and the width of the mouth.

changes the projected virtual faces and simultaneously calculates the respective servo commands as the angle vector $[\theta_{e,k}, \theta_{n,k}, \theta_{m,k}, \theta_{c,k}]$, and sends them to the servo motors to morph the physical face.

E. Validation experiments with human participants

Undergraduate students and teaching staff from Imperial College London (n = 11, 5 females, 6 males, aged 21 to 37 (M = 23, SD = 5.56)) were invited to participates in this the experiment (protocol approved by Imperial College Research Ethics Committee, study number 20IC6295). 3 participants self-identified as White, and 8 self-identified as Asian. The experiment consisted of three parts; a perceived pain intensity calibration process, followed by two pain intensity rating trials. 5 of the subjects were tested with 2D virtual projection in their first pain intensity rating trial, and 6 were tested with MorphFace to reduce the sequential effects.

1) Perceived Pain Intensity Calibration: The participants were asked to stand about 0.5m away from the projector, and were informed that the faces to be shown were considered as 0 pain intensity on a numeric pain scale from 0 to 10, and each image shown in the whole experiment lasts 10 seconds. The participants were shown 6 neutral faces projected onto a flat tabletop (2D virtual face). They were informed those 6 faces were considered as ”no pain” faces, and were asked whether they thought ”the projected face looked realistic and resembles a real person” once all faces were displayed, reporting a score from 1 to 5 where 1 corresponds to ”strongly agreed” and 5 corresponds to ”strongly disagree”. The participants were then shown 6 neutral faces generated using MorphFace, also informed that those were ”no pain” faces and asked the same question. This step calibrated the participants pain intensity perception by acknowledging the ”no pain” (0 pain intensity) faces of all of the virtual avatars in both face rendering methods.

2) Pain Intensity Rating Trials: Each participant completed two trials. Each trial involved observing 3 sets of 12 faces (pre-randomised, as shown in Fig. 8), either as a 2D projected virtual face or shown with MorphFace. They were asked to write down their perceived pain intensity of each face using a numeric scale from 0 to 10 where 0 means ”no pain” and 10 means ”worst pain possible”. There was a short break of 20 seconds between each set and each trial. The order of the sets were different between the two rendering face methods but consistent across all participants.

IV. Results

The main contribution of this paper is the design approach of the MorphFace. Fig. 8 shows the final outcome of MorphFace with geometrical differences of the face across 6 gender-ethnicity groups implemented by controlling 4 physical DoFs and details of skin color and facial activation units projected onto it. For simplicity, coefficients of the linear combination of AUs in each row are kept constant. The second row of Fig. 8 expresses the same level of pain (100% pain) from different gender-ethnicity backgrounds. 64% of the participants ”strongly agree” that MorphFace looks realistic and resembles a real person, 45% of the participants rated MorphFace to be more realistic than the 2D virtual faces, and the other 55% rated them equally realistic.

Next, we investigate the question as to whether images of the same level of pain but with different gender and ethnic backgrounds would lead to any differences in perception when they are projected onto the 3D MorphFace vs a 2D surface. The top plot in Fig. 9 shows difference in variances across participants in perceiving the same incremental change of pain levels for the same faces. The y-axis is the Numerical Rating Scale (NPRS-11) [32] of the difference between perceptions of a painful face and the corresponding neutral face. The bottom plot shows larger percentage differences in perceived pain intensities between Asian and White participants for Black Male ($p = 0.023$, Kruskal-Wallis test), White Male ($p = 0.051$, Kruskal-Wallis test), and Asian Female ($p = 0.014$, Kruskal-Wallis test) when using MorphFace. The corresponding percentage differences was also significant for Asian Female ($p = 0.032$, Kruskal-Wallis test) when observed as a 2D virtual face. Gender did not show to have any effect in how pain intensity was perceived from this experiment. The effect of rendering methods between MorphFace and the 2D virtual faces were also evaluated, and 4 participants had statistically significant difference ($p < 0.05$, Mann-Whitney U-test). And across all 6 faces, the White male and Asian male faces

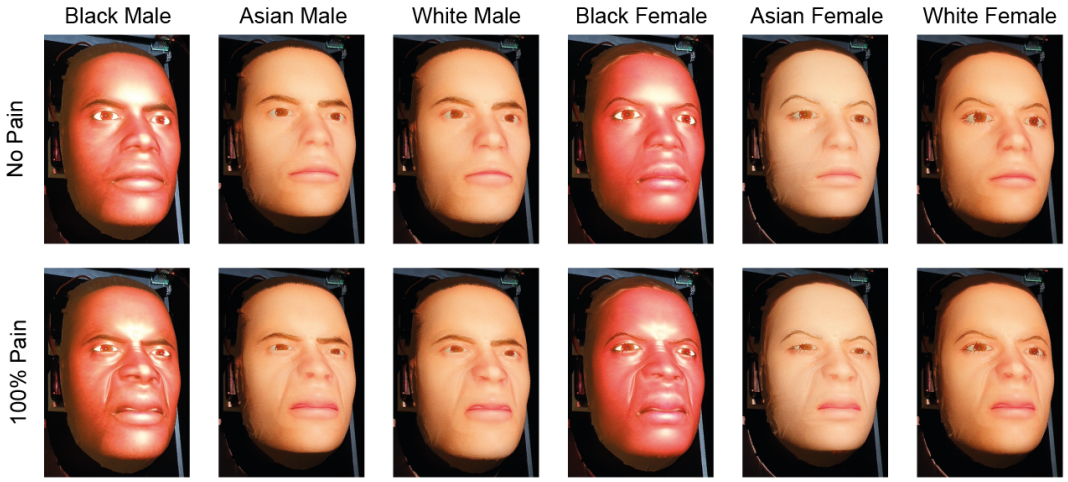

Fig. 8. MorphFace with matching physical and virtual faces. Top panel shows the faces with neutral expressions and bottom panel shows faces with the facial expressions of 100% pain. Ambient lighting condition was maintained consistent throughout the experiment.

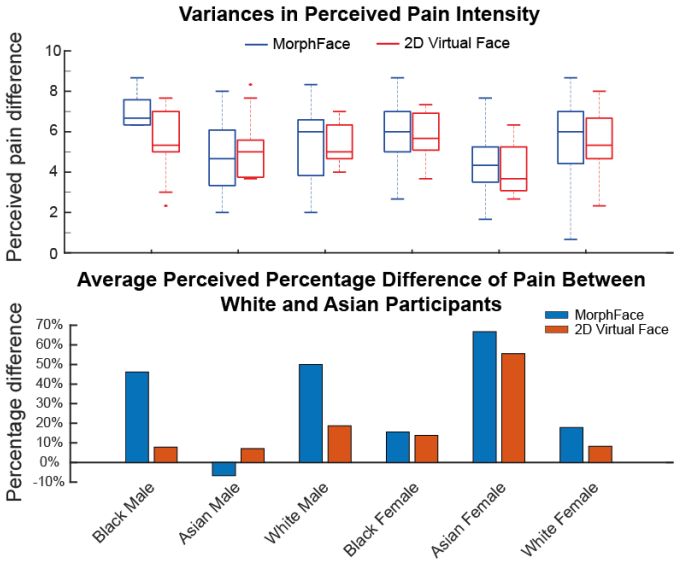

Fig. 9. Top: Participant's perceptual variance of incremental pain when the same pain level is projected to the MorphFace and on a 2D surface representing different gender and ethnicity backgrounds. Bottom: Percentage difference in perceived pain intensity between Asian and White participants on different faces ($(Avg_{Asian} - Avg_{White})/Avg_{White}$).

showed to have the lowest statistical power ($p_{WM} = 0.14$, $p_{AM} = 0.82$). This may be because the participants are all White or Asian, and are more used to seeing these two faces from past experiences.

Overall, these results show that there are differences in the process of information acquisition when the same pain level with different gender-ethnicity backgrounds rendered using MorphFace vs projected onto a 2D surface. This is important because the purpose of training is to develop internal models that are relevant to real patient examination. When we take the two perception results in Fig. 8 and Fig. 9 together, this paper highlights the potential usefulness of MorphFace as a controllable robotic simulator to deliver a quantifiable training to physicians on gender and ethnicity backgrounds of patients during physical examination.

V. Conclusions and Future Directions

In this paper, we proposed a novel hybrid morphable face capable of providing pain expressions from faces representing different genders and ethnicities. An experiment with 11 naive participants showed that their perception of fixed incremental pain had different variances across the 6 gender-ethnic groups of MorphFace. Moreover, MorphFace showed to amplify perceptual differences compared to an equivalent 2D virtual face. The statistical power of faces with the same ethnicity as the participants also showed to be the lowest, and 4 participants had a statistically significant difference in their perception of the same incremental pain presented from MorphFace vs. a 2D virtual face.

Hybrid systems allow for the study of emotion perception via *synthetic in vivo experimentation*, providing rich virtual contexts for participants [33], as well as high levels of accuracy due to the controllability of emotional expressions [34]. Work using Virtual Human (VH) interaction shows that VH is capable of influencing humans in such interactions, but to be perceived as engaging and believable, VH agents need to have movable faces [35]. VH is also capable of eliciting positive and negative emotional states in human conversational partners during both speaking and listening phases [36]. Nevertheless, VH-interactions represent 2D rather than 3D interactions. However, the process of information interaction from a 2D surface and a 3D landscape such as a human face are different. Therefore, hybrid systems such as MorphFace have the potential to build perception skills [37] that are more relevant to real patient examinations. To the best of our knowledge, no prior studies have used hybrid systems that produce variations in ethnicity and/or gender to further such training objectives.

A future study will involve conducting a detailed analysis of the interaction effect among the participants' gender-ethnicity

background and that of the robotic face presenting different pain levels, and the question as to how participants will change perception after a training session where they are repeatedly given feedback of the actual pain level being projected.

In addition, we will integrate the MorphFace with a robotic patient with sensorised organs [38], [39] to test the system in a closed-loop manner. The efficacy of this platform to carry forward the experience of pain for future training of general practitioners also needs to be tested. Our future developments will also start to include other facial features that might be part of abdominal disease. For example, the white of the eye may be an easy place to see that the patient is jaundiced, or the conjunctiva suggest that the patient is anaemic. We envision that this work will open up a new dimension of research related to robotic patients with facial expression rendering.

REFERENCES

[1] H. L. Wagner, C. J. MacDonald, and A. Manstead, "Communication of individual emotions by spontaneous facial expressions." *Journal of personality and social psychology*, vol. 50, no. 4, p. 737, 1986.

[2] J. Endres and A. Laidlaw, "Micro-expression recognition training in medical students: a pilot study," *BMC medical education*, vol. 9, no. 1, pp. 1–6, 2009.

[3] J. B. Engelmann and M. Pogosyan, "Emotion perception across cultures: the role of cognitive mechanisms," *Frontiers in psychology*, vol. 4, p. 118, 2013.

[4] T. Koda and T. Ishida, "Cross-cultural study of avatar expression interpretations," in *International Symposium on Applications and the Internet (SAINT'06)*. IEEE, 2006, pp. 7–pp.

[5] T. Koda, T. Ishida, M. Rehm, and E. André, "Avatar culture: cross-cultural evaluations of avatar facial expressions," *AI & society*, vol. 24, no. 3, pp. 237–250, 2009.

[6] M. Fölster, U. Hess, and K. Werheid, "Facial age affects emotional expression decoding," *Frontiers in psychology*, vol. 5, p. 30, 2014.

[7] A. R. Gonçalves, C. Fernandes, R. Pasion, F. Ferreira-Santos, F. Barbosa, and J. Marques-Teixeira, "Effects of age on the identification of emotions in facial expressions: a meta-analysis," *PeerJ*, vol. 6, p. e5278, 2018.

[8] U.-S. Donges, A. Kersting, and T. Suslow, "Women's greater ability to perceive happy facial emotion automatically: gender differences in affective priming," *PloS one*, vol. 7, no. 7, p. e41745, 2012.

[9] W. C. McGaghie, S. B. Issenberg, M. E. R. Cohen, J. H. Barsuk, and D. B. Wayne, "Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? a meta-analytic comparative review of the evidence," *Academic medicine: journal of the Association of American Medical Colleges*, vol. 86, no. 6, p. 706, 2011.

[10] A. Ziv, P. R. Wolpe, S. D. Small, and S. Glick, "Simulation-based medical education: an ethical imperative," *Simulation in Healthcare*, vol. 1, no. 4, pp. 252–256, 2006.

[11] J. M. Weller, D. Nestel, S. D. Marshall, P. M. Brooks, and J. J. Conn, "Simulation in clinical teaching and learning," *Medical Journal of Australia*, vol. 196, no. 9, pp. 594–594, 2012.

[12] "SimMan 3G," accessed 17-05-2020. [Online]. Available: https://www.laerdal.com/gb/products/simulation-training/emergency-care-trauma/simman-3g/

[13] "Pediatric HAL," accessed 17-05-2020. [Online]. Available: https://www.gaumard.com/s2225

[14] "SIMROID," accessed 17-05-2020. [Online]. Available: https://www.morita.com/group/en/products/educational-and-training-systems/training-simulation-system/simroid/

[15] D. F. Glas, T. Minato, C. T. Ishi, T. Kawahara, and H. Ishiguro, "Erica: The erato intelligent conversational android," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 22–29.

[16] "Sophia," accessed 17-05-2020. [Online]. Available: https://www.hansonrobotics.com/sophia/

[17] M. Gilbert, S. Demarchi, and I. Urdapilleta, "Facshuman a software to create experimental material by modeling 3d facial expression," in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 2018, pp. 333–334.

[18] T. D. Lalitharatne, Y. Tan, F. Leong, L. He, N. Van Zalk, S. De Lusignan, F. Iida, and T. Nanayakkara, "Facial expression rendering in medical training simulators: Current status and future directions," *IEEE Access*, vol. 8, pp. 215 874–215 891, 2020.

[19] A. Bermano, P. Brüschweiler, A. Grundhöfer, D. Iwai, B. Bickel, and M. Gross, "Augmenting physical avatars using projector-based illumination," *ACM Trans. Graph.*, vol. 32, no. 6, Nov. 2013.

[20] "Furhat Robotics," accessed 17-05-2020. [Online]. Available: https://furhatrobotics.com/

[21] "socibot," accessed 17-05-2020. [Online]. Available: https://www.engineeredarts.co.uk/socibot/

[22] 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 *IEEE International Conference on Humanoid Robots*, 2011, pp. 99–104.

[23] S. Daher, J. Hochreiter, R. Schubert, L. Gonzalez, J. Cendan, M. Anderson, D. A. Diaz, and G. F. Welch, "The physical-virtual patient simulator," *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, vol. 15, no. 2, pp. 115–121, 2020.

[24] K. Hayashi, Y. Onishi, K. Itoh, H. Miwa, and A. Takanishi, "Development and evaluation of face robot to express various face shape," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, 2006, pp. 481–486.

[25] R. Schubert, G. Bruder, and G. Welch, "Adaptive Filtering of Physical-Virtual Artifacts for Synthetic Animatronics," in *ICAT-EGVE 2018 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, G. Bruder, S. Yoshimoto, and S. Cobb, Eds. The Eurographics Association, 2018.

[26] D. Ma, J. Correll, and B. Wittenbrink, "The chicago face database: A free stimulus set of faces and norming data," *Behavior research methods*, vol. 47, 01 2015.

[27] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[28] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[29] B. Amos, J. Harkes, P. Pillai, K. Elgazzar, and M. Satyanarayanan, "OpenFace: Face Recognition with Deep Neural Networks," http://github.com/cmusatyalab/openface, accessed: 2015-11-11.

[30] Y. F. Wen, H. M. Wong, R. Lin, G. Yin, and C. McGrath, "Inter-ethnic/racial facial variations: a systematic review and bayesian meta-analysis of photogrammetric studies," *PloS one*, vol. 10, no. 8, p. e0134525, 2015.

[31] "makehumancommunity." [Online]. Available: http://www.makehumancommunity.org/

[32] R. H. Dworkin, D. C. Turk, J. T. Farrar, J. A. Haythornthwaite, M. P. Jensen, N. P. Katz, R. D. Kerns, G. Stucki, R. R. Allen, N. Bellamy *et al.*, "Core outcome measures for chronic pain clinical trials: Immpact recommendations," *Pain*, vol. 113, no. 1, pp. 9–19, 2005.

[33] S. Marsella and J. Gratch, "Computational models of emotion as psychological tools," *Handbook of emotions*, pp. 113–132, 2016.

[34] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology," *Psychological inquiry*, vol. 13, no. 2, pp. 103–124, 2002.

[35] C. Qu, W.-P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx, "Human perception of a conversational virtual human: an empirical study on the effect of emotion and culture," *Virtual Reality*, vol. 17, no. 4, pp. 307–321, 2013.

[36] C. Qu, W. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx, "Conversations with a virtual human : synthetic emotions and human responses," *Computers in Human Behavior*, vol. 34, pp. 58–68, 2014.

[37] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological science in the public interest*, vol. 20, no. 1, pp. 1–68, 2019.

[38] L. He, N. Herzig, S. de Lusignan, and T. Nanayakkara, "Granular jamming based controllable organ design for abdominal palpation," *IEEE Engineering in Medicine and Biology Society*, vol. 2018, 2018.

[39] N. Herzig, L. He, P. Maiolino, S. Guamán, and T. Nanayakkara, "Conditioned haptic perception for 3d localization of nodules in soft tissue palpation with a variable stiffness probe," *PLOS ONE*, vol. 15, 07 2020.