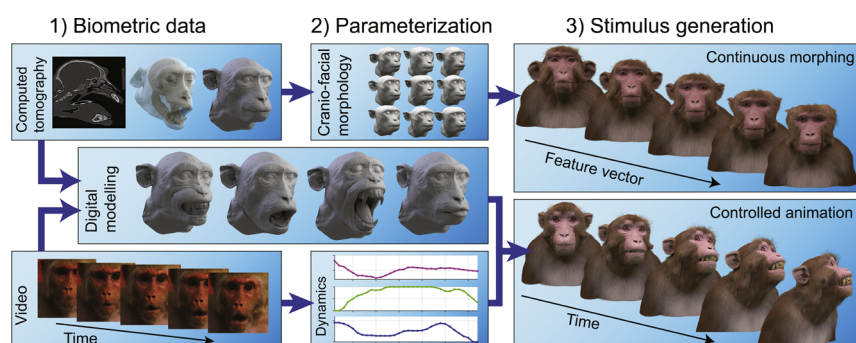# A parameterized digital 3D model of the Rhesus macaque face for investigating the visual processing of social cues

Aidan P. Murphy*, David A. Leopold

Section on Cognitive Neurophysiology and Imaging, NIMH, Bethesda, MD, USA

GRAPHICAL ABSTRACT

ARTICLE INFO

ABSTRACT

Background: Rhesus macaques are the most popular model species for studying the neural basis of visual face processing and social interaction using intracranial methods. However, the challenge of creating realistic, dynamic, and parametric macaque face stimuli has limited the experimental control and ethological validity of existing approaches.

New method: We performed statistical analyses of *in vivo* computed tomography data to generate an anatomically accurate, three-dimensional representation of Rhesus macaque cranio-facial morphology. The surface structures were further edited, rigged and textured by a professional digital artist with careful reference to photographs of macaque facial expression, colouration and pelage.

Results: The model offers precise, continuous, parametric control of craniofacial shape, emotional expression, head orientation, eye gaze direction, and many other parameters that can be adjusted to render either static or dynamic high-resolution faces. Example single-unit responses to such stimuli in macaque inferotemporal cortex demonstrate the value of parametric control over facial appearance and behaviours.

Comparison with existing method(s): The generation of such a high-dimensionality and systematically controlled stimulus set of conspecific faces, with accurate craniofacial modelling and professional finalization of facial details, is currently not achievable using existing methods.

Conclusions: The results herald a new set of possibilities in adaptive sampling of a high-dimensional and socially meaningful feature space, thus opening the door to systematic testing of hypotheses about the abundant neural specialization for faces found in the primate.

1. Introduction

The primate visual system has been shaped through evolution and development to extract behaviourally relevant information from the complex and cluttered sensory inputs that it encounters in the natural environment. This fundamental truth poses a challenge to the reductionist scientific approach that has dominated vision research, and which shapes current thinking about primate brain function. On the one hand, simplified, abstract or artificial stimuli allow for systematic, parameterized testing along stimulus dimensions of interest, whilst keeping other variables constant. On the other hand, their artificial nature raises questions of their relevance to the types of visual experiences that the brain naturally encounters. This tension between rigorous control and ethological validity is perhaps most conspicuous in trying to understand the visual processing of social cues, to which much of the primate brain appears dedicated (Stanley and Adolphs, 2013).

Thanks to continuing technological developments, it is increasingly possible to combine rigorous experimental control with ethological validity (Bohil et al., 2011; Parsons, 2015). State of the art computer generated images (CGI) can accurately model stimuli as complex and dynamic as faces in a manner that approaches true photo-realism. Generating photo-realistic CGI no longer requires a studio budget thanks to open-source ray-tracing engines (such as Blender Cycles) and cheaper access to more powerful graphics processing units (GPUs), either as local hardware or via cloud computing. CGI offers vision researchers the ability to create highly complex and realistic visual stimuli whilst still retaining tight experimental control over variables of interest. Indeed, CGI can allow for increased control over a broad range stimulus dimensions, since the virtual environment is already optimized to generate myriad images based on combinations of parameters.

Faces are an important class of visual objects for all primates, including humans. Owing to their highly complex nature and our exquisite sensitivity to facial appearance, CGI has long struggled to create photorealistic faces based on parameterized models. However, with improvements in high-resolution 3D scanning and capture technologies, it is now becoming easier and cheaper to generate virtual 3D models of real-world objects. Numerous 3D and 4D (dynamic 3D) databases of human faces are currently available for biological and computer vision research (Sandbach et al., 2012; see Supplementary Table S1). The models consist of 3D surface meshes and corresponding texture maps, which are constructed by capturing images of a subject from multiple viewpoints simultaneously. For models of other animals, such as non-human primates, this approach is less practical and often ill-suited to deal with the reflectance properties of hair or fur.

Given the importance of the macaque as an animal model for studying the neural basis of vision in general and face processing in particular (Leopold and Rhodes, 2010), it would be advantageous to have a parametrically controllable face model of this species. Macaques and other primates excel in processing facial information for members of their own species (Dufour et al., 2006; Gothard et al., 2009; Pascalis and Bachevalier, 1998; Pascalis et al., 2002; Sigala et al., 2011). This species bias is reflected in their behavioural responses, such as looking behaviour (Dahl et al., 2009; Sugita, 2008), gaze-following (Emery et al., 1997), as well as the response selectivity of neurons in the macaque brain (Kiani et al., 2005; Minxha et al., 2017; Sigala et al., 2011). Several groups have previously used animated 3D models of the macaque face in their research (Ferrari et al., 2009; Steckenfinger and Ghazanfar, 2009). These models were created manually and were thus unable to capture critical features required to maximize the realism of the images. In contrast, Sigala and colleagues used 3D photogrammetry to capture the surface structure and texture of the faces of real macaque monkeys (Sigala et al., 2011), as has been done with human models (Kustár et al., 2013; Tilotta et al., 2009). As mentioned above, this approach is limited in capturing fine details such as hair or fur, making it problematic for capturing the full face in species like the macaque.

Here we take a hybrid approach, to create an anatomically accurate, fully animatable, 3D model of the Rhesus macaque head and face. We begin with computed tomography (CT) data acquired *in vivo* from anesthetized Rhesus macaques, allowing for accurate modeling of craniofacial morphology and construction of a statistical model of individual variations in 3D face shape. We describe in detail how this digital model was then edited, rigged and textured by a professional CGI artist and applied to models of both facial expression and individual identities. The product is an actualizable, realistic face-space for the Rhesus monkey, in which the individual identity, non-rigid facial expression, and a range of other natural and unnatural transformations can be systematically parameterized. To demonstrate the utility of this parametrically controllable macaque face model, we performed single-unit neurophysiological recordings in three of the functionally defined 'face-selective' regions of the macaque inferotemporal cortex (ML, AF, and AM) while animals were presented with visual stimuli generated using the model.
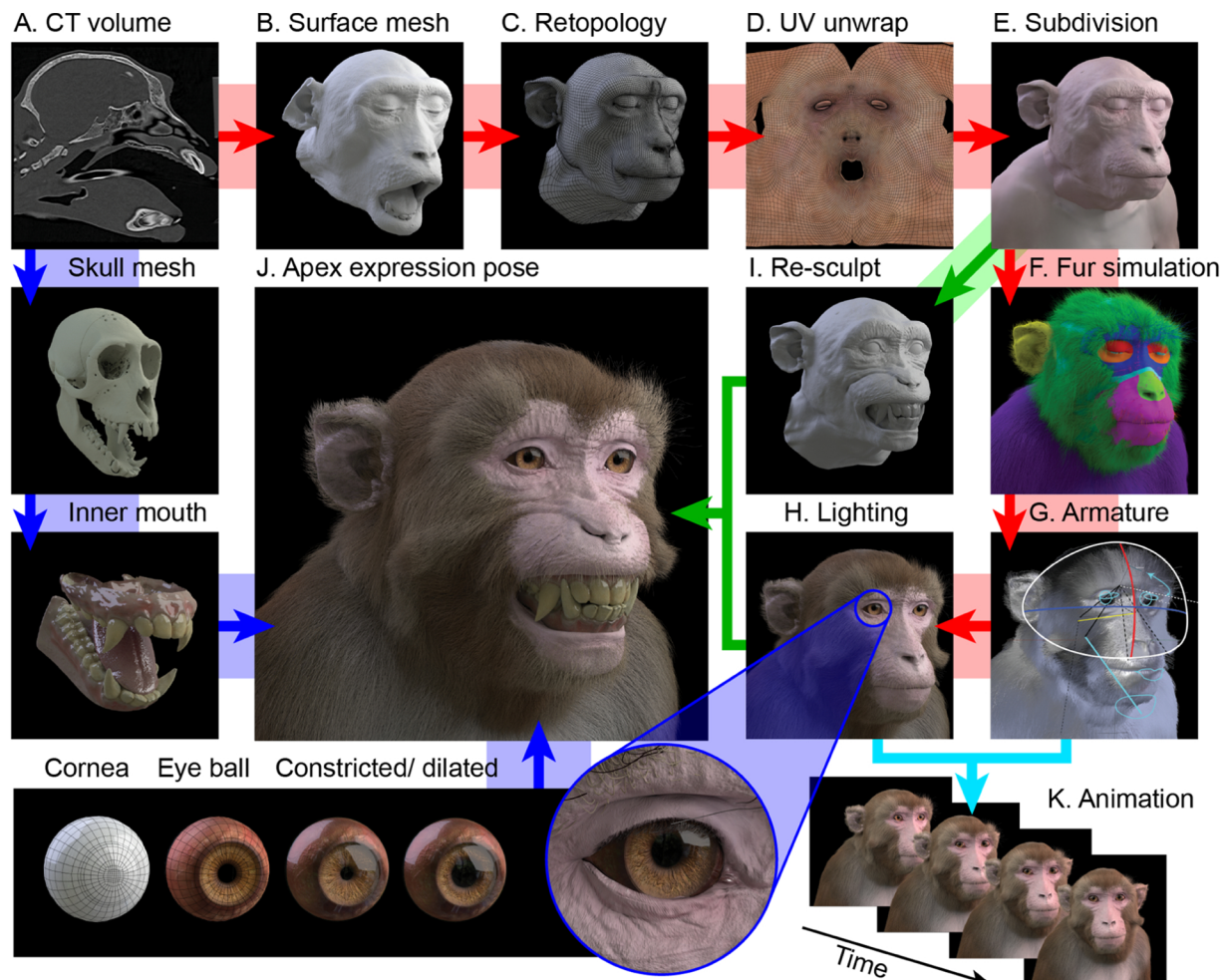
2. Materials and methods

2.1. 3D expression model generation

2.1.1. Computed tomography data

A high-resolution (0.25 x 0.25 mm in-plane) *in vivo* computed tomography (CT) scan of the head and neck of an anesthetized 6-year-old adult male Rhesus macaque (7.1 kg) was acquired from the Digital Morphology Museum (DMM) of the Kyoto University Primate Research Institute (http://dmm.pri.kyoto-u.ac.jp/dmm/WebGallery/index.html). During the scan, the animal lay prone and intubated with the head held in a stereotaxic device. The CT data was processed in 3D Slicer (Fedorov et al., 2012) by finding the threshold intensity of the air and bone in order to segment soft-tissue and skull surfaces respectively. Volumes were then manually edited to remove the intubation tube, stereotactic frame, and any image artifacts before generating surface models which were smoothed and decimated and exported as stereolithography (.stl) files (Fig. 1B).

2.1.2. Surface mesh editing and expressions

Surface meshes were imported into ZBrush (Pixologic Inc., Los Angeles, CA) - a digital sculpting and painting software – for editing by a professional CGI character artist. Soft-tissue deformations (e.g. of the upper lip against the stereotax palette bar) and anatomy cropped by the CT volume (e.g. the edges of the ears) were manually corrected or reconstructed from reference photos of animals of the same age and gender, before a mirror-symmetric, low polygon-count topology was created (Fig. 1C). This mesh was manually edited in order to close the mouth and open the eye lids, which produced the appearance of an awake animal with a neutral expression. A higher subdivision version of this base topology was then created and iteratively edited to enhance fine detail (e.g. pores and wrinkles, Fig. 1E). Using copies of this base mesh as a starting point, new meshes were sculpted to portray eye lid closing, mouth opening, ear flapping, eye brow raising, and five apex facial expressions: 'coo' vocalization, bared-teeth 'fear grimace' display, open-mouthed threat, yawn, and tongue protrusion (Fig. 2). All meshes therefore retained the same number and order of faces and vertices, and differed only in vertex locations. The various facial expression meshes were linked to the neutral (closed mouth and open eyes) base mesh as shape keys, allowing continuous parametric blending of facial appearance between expressive states. The mesh was rigged for animation using bone armatures (Fig. 1G) for head orientation, jaw and eye movements, and shape keys for movements of soft tissue (eye lids, eye brow, ears, lips). The skull and mandible meshes were edited separately to create upper and lower teeth and palette, and the tongue was modelled based on the soft tissue surface (Fig. 1). Additionally, an upper torso was manually created based on photographic references and scaled to fit, but was not included in the renders used for the experiments described here.

Fig. 1. Construction pipeline of the virtual model. Clockwise from top left: A. The CT volume is segmented to obtain soft tissue label. B. Surface reconstruction of the segmented volume yields a dense voxelized mesh. C. The raw mesh is decimated and re-topologized as a cleaner topology with a lower polygon count, to facilitate later editing. D. The retopologized mesh is UV-unwrapped and 2D skin and fur texture images are generated for UV (colour - shown here), displacement (texture), surface normals and specularity (gloss). E. The retopologized mesh surface is subdivided so that fine details (many of which are present in the raw mesh, such as wrinkles and pores) can be sculpted back into the subdivided surface. F. Particle simulation is applied to multiple vertex groups (indicated by colours) of the underlying mesh to grow varied and realistic looking fur and facial hair. G. The model is rigged for movement by linking an armature to certain vertex groups, allowing head movement at the neck, jaw and various other facial movements. H. The environmental lighting of the scene is set up to produce realistic renders of the static model. I. To generate new facial expressions, copies of the neutral surface mesh are edited into new poses with reference to photographic images and video, representing the apex frame of a given expression. *J*. The bared-teeth 'fear grimace' apex expression is rendered with texture, fur and lighting (as in H). Since the teeth and gums are exposed in this expression, these structures were modelled from segmentation of the bone surface and details from the original open-mouth pose. Bottom row: Construction of the eyes. The eye balls are composed of an inner sphere with a concave surface for the iris and a variable aperture of pupil. From left to right: 1) cornea topology; 2) sclera and iris topology; 3) pupil constriction; 4) pupil dilation; 5) eye socket fit. K. Varying parameters of the model over time (e.g. facial expression, eye gaze direction, head direction, etc.) allows rendering of animated sequences.

2.1.3. Anatomical constraints

All models were then imported into the open-source 3D graphics software Blender 2.79 (www.blender.org), where rigged eye balls were added (Fig. 1). The geometry of the eyes was set based on previously published measurements. Specifically, the inter-pupillary distance was set at 35 mm (Bush and Miles, 1996), the axial length of the eye ball was set at 20 mm (Augusteyn et al., 2016; Hughes, 1977; Qiao-Grider et al., 2007; Vakkur, 1967). Each eye ball consisted of an inner textured surface with adjustable pupil diameter, constrained to a maximum and minimum of 7 and 3.5 mm respectively based on published measurements (Gamlin et al., 2007; Ostrin and Glasser, 2004). An outer corneal surface used a mix of transparent and glossy bidirectional scattering distribution function (BSDF) shaders (refractive index = 1.376 based on Hughes, 1977; Vakkur, 1967). Rotation of each eye about its center was linked to track an invisible target object that could be moved in space, thus ensuring realistic conjugate eye movements including vergence (but excluding cyclovergence), which were constrained so as

never to exceed 30°. This enables simple control of the avatar's gaze pattern, including ocular following of virtual objects added to the scene (Supplementary Movie S1). To prevent the eye balls from intersecting with the eyelid surfaces during rotation, shape keys were generated and linked so as to adjust eyelid position during eye rotations, as occurs in real animals (Becker and Fuchs, 1988).

2.1.4. Texture and appearance

Images for use as UV textures (RGB colour images), surface normal, and displacement maps were manually created for the eyes, skin, and fur based on reference images of real Rhesus macaques. High-resolution colour photographs of Rhesus macaques with neutral expressions were acquired from the PrimFace database (http://visiome.neuroinf.jp/primface, funded by Grant-in-Aid for Scientific research on Innovative Areas, "Face Perception and Recognition" from Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan). Realistic fur was generated using particle simulation, and its distribution,

Fig. 2. Modelling of facial expressions. A. Examples of reference photographs of Rhesus macaque species-typical facial expressions (images courtesy of Katalin Gothard and Liz Romanski). From left to right: 1) neutral; 2) bared-teeth 'fear grimace' display; 3) 'coo' vocalization; 4) open-mouthed 'threat'; 5) yawn; 6) tongue protrusion. B. Each apex expression is a copy of the neutral base mesh with a subset of vertex coordinates altered to produce different geometries. C. Expression models rendered with colour textures and fur. D. Expression models are linked to the neutral mesh as "shape keys", which allows for linear interpolation between apex expression shapes (shown here for yawning) for continuous parametric control and enables animation of non-linear temporal dynamics through the use of key frames.

colouration, and overall appearance was adjusted based on analysis of high-resolution photographs of real macaques, using different vertex groups of the underlying mesh surface as seed regions (Fig. 1F). Blender's probabilistic ray-tracing engine 'Cycles' was used for all rendering. Environmental lighting was set up using a high dynamic range (HDR) image and two emission surfaces were positioned symmetrically either side of the avatar, above and in front to simulate indoor lighting conditions. Skin surfaces were rendered using sub-surface scattering shaders, which simulate the way in which light entering the skin is scattered and then emitted.
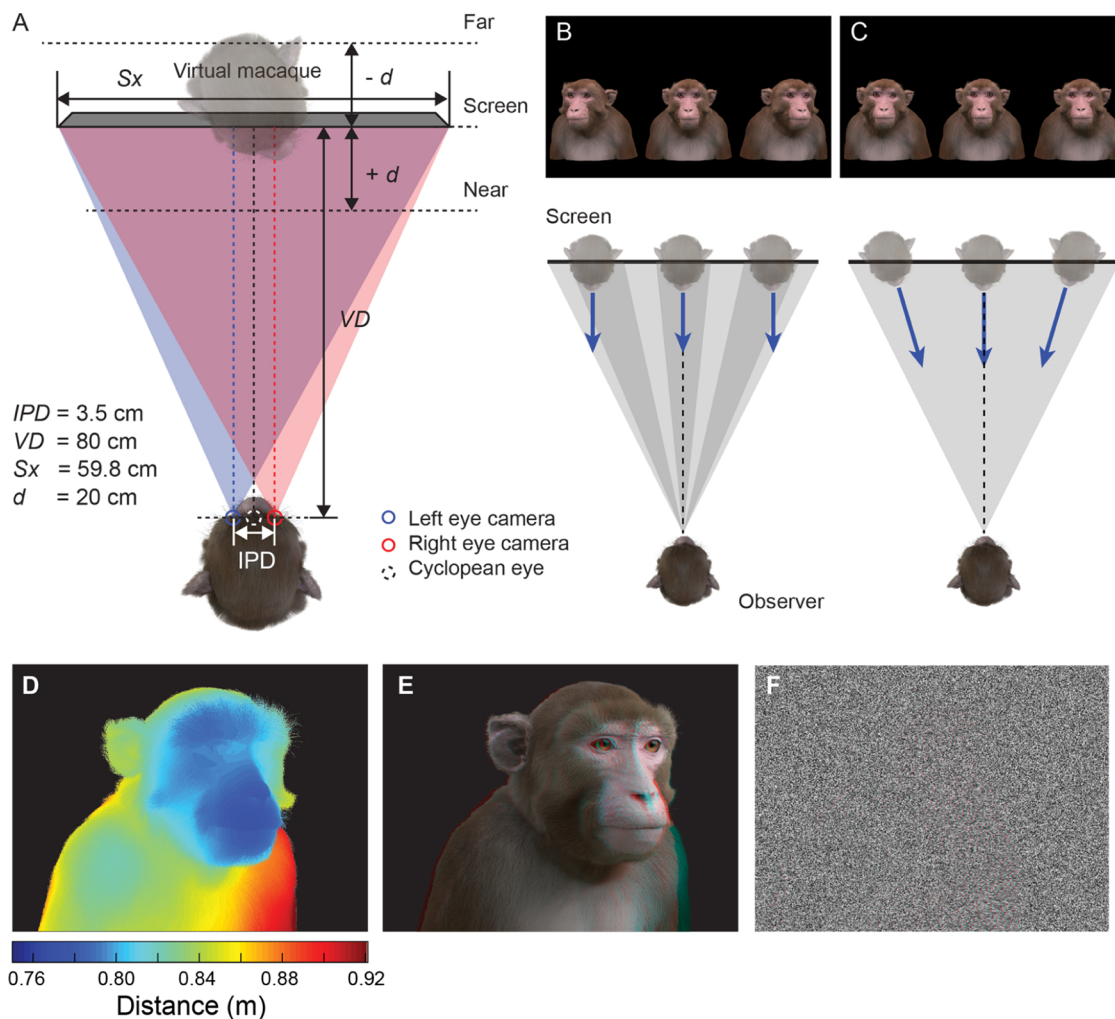
2.1.5. Image rendering

Python scripts were written to automate Blender control of the shape keys, armatures, lighting, material properties and camera position and for rendering high-resolution images and video. Geometrically veridical stereoscopic rendering was achieved by precisely replicating the physical geometry of the experimental presentation setup *in silica*, such that the virtual cameras were positioned the same distance from the 3D model as the real subject's viewing distance from the display screen, using off-axis frusta and an inter-axial distance equal to the subjects' average interocular distance (35 mm) (Fig. 3A). The field of view of the virtual camera was set to match the horizontal angular subtense of the display screen from the subject's perspective. Rendering was performed on a Linux PC running Ubuntu 16.04 OS, with Nvidia Tesla K40 and Nvidia GeForce GTX 1080 GPUs, and the number of samples for rendering was set to 200, resulting in approximate render

times of 2 min per frame (depending on the proportion of the image that the projection of the model occupied). Images were saved in full screen side-by-side stereoscopic format (3840 × 1080) as 32-bit RGBA portable network graphic (.png) files.

2.1.6. Animating naturalistic facial dynamics

In the main experiments presented here, we rendered static images that were presented briefly, as is common in vision neurophysiology. However, a major goal is to understand how the brain processes social cues within the more ethologically relevant context of natural vision. In order to generate dynamic animated sequences that appear as natural as possible we developed methods for estimating the dynamics of facial movement from video footage of real macaques and applied these dynamics to the actuators of the 3D face model. We developed a custom graphical user interface (written in Matlab) to load video clips and manually track the relative motion of different parts of the face over time, frame by frame (Fig. 4; Supplementary Movie S2). The motion parameters extracted from each video clip were written to a spreadsheet file that was read into Blender by a Python script and applied as key frames for animating the macaque model's facial movements. The resulting animations contain similar facial movements to the original clips, but allow for parametric control of other variables, such as head orientation, spatial location, and lighting. Using interpolation, the output animation can also be of higher spatial and temporal resolution compared to the original clip it is based on.

Some of the more subtle changes in facial appearance occur at scales

Fig. 3. Viewing geometry. A. Schematic illustration (view from above) of the virtual replication of real-world viewing geometry. The observer was always located at a fixed viewing distance (*VD*) from the plane of the screen, with the center of the screen aligned to the observer's cyclopean eye. A pair of virtual cameras are spaced the equivalent of the subject's inter-pupillary distance (IPD) apart, with off-axis frusta (blue and red triangles). The position of the virtual macaque in 3D space is restricted by the viewing frusta, which are determined by the size of the screen (*Sx*) and the viewing distance (*VD*). B. When a 2D image is presented peripherally on a flat display, the projection of that image onto the subjects retina becomes distorted due to parallax error. By moving the virtual 3D model to a peripheral location and then rendering the image it is possible to produce retinal stimulation consistent with real-world geometry. C. When the avatar is positioned peripherally but oriented toward the observer, the resulting 2D image is more perceptually similar to the same object at the central location. D. Depth map generated from Z-buffer rendering to high-dynamic range format (Open-EXR). E. Stereoscopic 3D rendering (requires red-cyan anaglyph glasses for viewing). F. Random-dot stereogram (red-cyan anaglyph) generated using the depth map in D, which contains the same binocular disparity content as E, but no other visual cues to depth or object form.

that are difficult to extract from low-resolution reference videos, yet are known to be behaviourally relevant (Kret et al., 2014) and encoded by neurons in the macaque brain (Freiwald et al., 2009). We therefore used published measurements of pupillary constriction and dilation (Pong and Fuchs, 2000), blinks (Guipponi et al., 2014; Schultz et al., 2010), and microsaccades (Hafed et al., 2009) in Rhesus macaques to model the dynamics and probability distributions of these subtle facial parameters (Fig. 4F–G). Random sampling of these distributions allowed us to automate animation of the corresponding parameters to achieve a naturalistic resting state appearance.
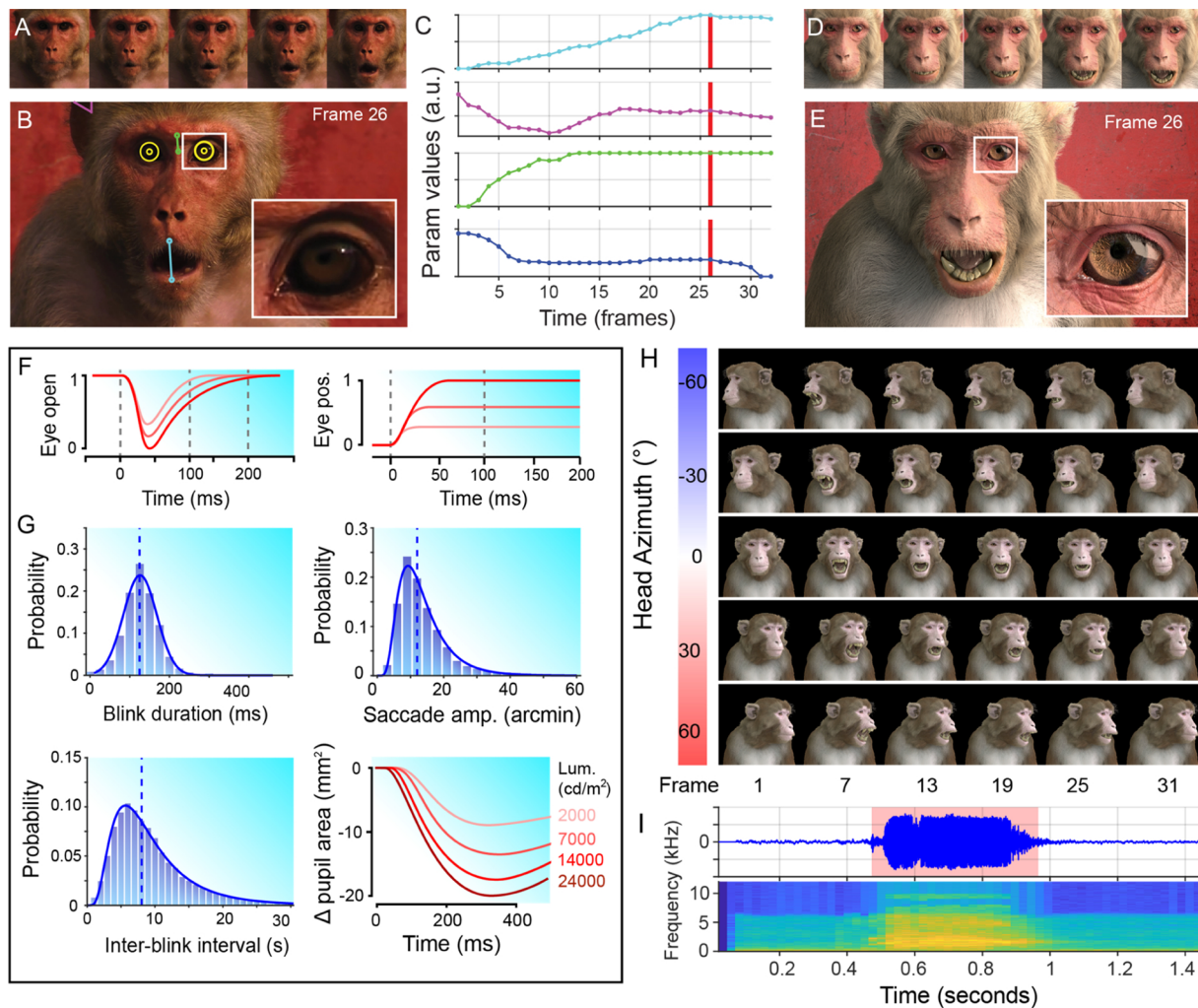
2.2. Morphable identity model generation

2.2.1. Computed tomography data

High-resolution *in vivo* CT scans of the head and neck of 34 anesthetized Rhesus macaques were acquired from multiple institutions (see Table 1). The CT scanners used at each location were: i) KUPRI – Toshiba Asteion 4; ii) NIH – Siemens Force Scanner; iii) UC Davis - GE LightSpeed16. All scans were processed as described above for the

initial individual animal, using 3D Slicer (Fedorov et al., 2012) to segment the volume and construct a surface mesh of the soft-tissue surface.

2.2.2. Surface mesh editing

The raw surface mesh of each individual animal was imported into the commercial software Wrap 3 (R3DS, Voronezh, Russia), and an affine transform was first applied via Procrustes method to bring the data into approximate alignment with the standard base mesh. The base mesh chosen was the low-polygon count (~50,000 faces) re-topologized mesh in neutral expression, with open eye lids and opened mouth that had been manually generated from subject M02 in the earlier stages (see above). Corresponding vertices on the raw mesh and base mesh were identified by manually selecting surface anatomical landmarks such as the inner and outer canthi, nasion, nostrils, tragus, upper and lower lips, etc. (Fig. 5A). The software uses these manually identified correspondences between meshes to perform non-rigid registration of the base mesh to the raw mesh by minimizing residuals for *all* vertices – not just those that were manually selected. Warped meshes were then

Fig. 4. Animating facial dynamics. A. Cropped frames sampled from an input video clip of a real Rhesus macaque, illustrating a change in facial expression over time (reproduced with permission from Off The Fence™). B. The relative spatial configuration of various facial parameters, such as mouth opening, brow raising and ear flapping were estimated in the pixel domain frame-by-frame. C. The extracted time courses for each parameter were then applied to the animation of the corresponding actuator controls of the macaque face model. D. The corresponding frames of the rendered output animation. E. The output animation preserves the natural dynamics and correlated facial movements of the original video clip, but allows greater control and flexibility than would be possible using video footage of real animals. F. Example of simulated dynamics for eye blinks (left), and microsaccades (right) based on measurements from real Rhesus monkeys. G. Probability distributions of blink durations (left), inter-blink intervals (bottom) and microsaccade amplitudes (right) for realistic animation of the avatar in a resting state. Simulation of these parameters was based on normal or lognormal fits to published data (Guipponi et al., 2014; Hafed et al., 2009). H. A subset of frames from an example animation sequence included in the MF3D stimulus set (release 1) is rendered at 5 different head azimuth orientations (rows). I. Accompanying audio waveform and spectrogram for this particular animation, which depicts a 'scream' vocalization.

manually edited to standardize their pose (opening the eye lids to fit the eye balls and closing the mouth) and correct any minor artifacts related to volume acquisition (e.g. cropped ears, deformed lips). All further analysis, stimulus generation and data collection described in the present paper was based on just the first 23 identity meshes that had been edited at the time.

2.2.3. Statistical analysis

The vertex data for each of the cleaned standardized meshes was read into Matlab from. obj files, and averaged together to obtain the sample mean cranio-facial morphology, which was saved as a new surface mesh (Fig. 5B). Principal components analysis (PCA) was performed on the vertex data in Matlab, and a new mesh surface was saved to. obj format for ± 3 standard deviations from the mean for each of the first 10 principal components, which cumulatively explained 89.5% of the sample variance (Fig. 5F). To verify that facial distinctiveness (mathematically defined as the Euclidean distance from the sample mean in the 23-dimensional PC space) did not occur as an artifact of

measurement variability (e.g. CT resolution), we calculated the correlation between CT voxel volumes (Table 1) and distance from the sample mean for all original monkeys in the sample (Fig. 5E), which was not significant ($r = 0.326$, $p = 0.129$). The distribution of facial distinctiveness within the sample approximated a normal distribution, but had a mean of 0.765 (± 0.206 sd) standard deviations from the sample mean (Fig. 5E). The animal whose CT data was selected for construction of the facial expression model (M02, a 6-year-old male) happened to have the most average facial morphology in the sample (i.e. closest to the sample mean at a distance of 0.489σ).

As further analyses of behaviourally relevant variations of cranio-facial morphology in this sample, we utilized the demographic data that was available for the original individuals sampled (Fig. 5G). First, we averaged the surface mesh vertex data for females (n = 5), and a random selection of males (n = 5), to obtain 'average female' and 'average male' meshes. Averaging these two meshes together produced an androgynous average mesh. Displacing each vertex of the average mesh for each sex in the opposite direction to its distance from the

Table 1
Summary of demographic and CT data for Rhesus macaque subjects sampled for the generation of the morphable facial identity model.

ID	Sex	Weight (kg)	Age (years)	Source	Voxel in-plane (μm)	Inter-slice interval (μm)	Distance from mean (σ)
M01	Female	6.7	6.0	KUPRI	225	200	0.525
M02	Male	7.1	6.0	KUPRI	250	200	0.489
M03	Male	5.65	6.8	NIH	300	500	0.646
M04	Male	7.5	7.7	NIH	300	500	0.926
M05	Male	9.45	9.0	NIH	260	300	0.954
M06	Female	9.5	6.0	NIH	290	500	0.552
M07	Female	4.8	6.0	NIH	285	500	0.655
M08	Male	9.5	8.0	NIH	324	500	1.305
M09	Male	12	6.8	NIH	370	300	1.039
M10	Male	7.7	8.4	NIH	350	500	0.796
M11	Female	5.1	7.2	NIH	360	125	0.958
M12	Male	12.5	7.5	NIH	370	250	0.905
M13	Male	10.1	5.2	NIH	370	250	0.689
M14	Male	13.95	10.8	NIH	370	250	0.606
M15	Male	11.25	6.2	NIH	370	250	0.605
M16	Male	11.45	6.2	NIH	370	250	0.827
M17	Male	7.6	5.2	NIH	293	500	0.710
M18	Male	6	5.3	NIH	355	300	0.794
M19	Male	14.1	6.0	NIH	370	125	0.509
M20	Male	6.7	4.8	NIH	370	125	0.652
M21	Male	5.7	4.0	NIH	352	250	0.606
M22	Male	6.8	8.3	NIH	370	250	0.836
M23	Male	8.5	5.6	NIH	293	250	1.022
M24	Male	7.3	4.1	NIH	370	250	–
M25	Male	8.5	6.4	NIH	370	250	–
M26	Male	7.8	4.7	NIH	370	250	–
M27	Male	12.2	24.9	UC Davis	460	625	–
M28	Male	14.2	25.8	UC Davis	310	625	–
M29	Male	11.5	19.0	UC Davis	480	625	–
M30	Male	13.1	14.1	UC Davis	400	625	–
M31	Male	11.1	16.4	UC Davis	360	625	–
M32	–	–	–	UC Davis	340	625	–
M33	Male	16	8.5	UC Davis	390	625	–
M34	Female	7.5	9.7	UC Davis	340	625	–
M35	Female	9.3	7.2	UC Davis	320	625	–
M36	Female	7.8	10.6	UC Davis	290	625	–

androgynous average resulted in 'super-female' and 'super male' meshes, which exaggerated sexually dimorphic variations (Fig. 5H). We applied a similar approach to the demographic variable of age, but limited our sampling to males only since the age distribution for females in our sample was too narrow. This resulted in average old and young male meshes, as well as their combined average and their exaggerated versions (Fig. 5I).

Similarly, by averaging the principal component (PC) values of individuals in each of these demographics, we were able to reconstruct the ten-dimensional hyperplanes through PCA-space that correspond approximately to age and sex. Establishing these trajectories in PCA-space allows one to parametrically shift any individual facial shape identity (described as a combination of PC values) in either direction along the age and sex continua, as has been done previously for human face shape (Blanz and Vetter, 1999).

2.2.4. Rendering

The +1SD meshes generated from the PCA were imported into Blender and transformed to fit the existing rigged eye balls in the mesh eye sockets. Each PC mesh was then linked to the average mesh as a shape key, allowing for the average mesh to be morphed continuously into any identity in the face-space (based on the first 10 PCs). The UV textures, vertex groups and fur simulation of the original base mesh were transferred to the average mesh (Supplementary Movie S3), to allow head and eye movement.

2.3. Neurophysiology

2.3.1. Subjects

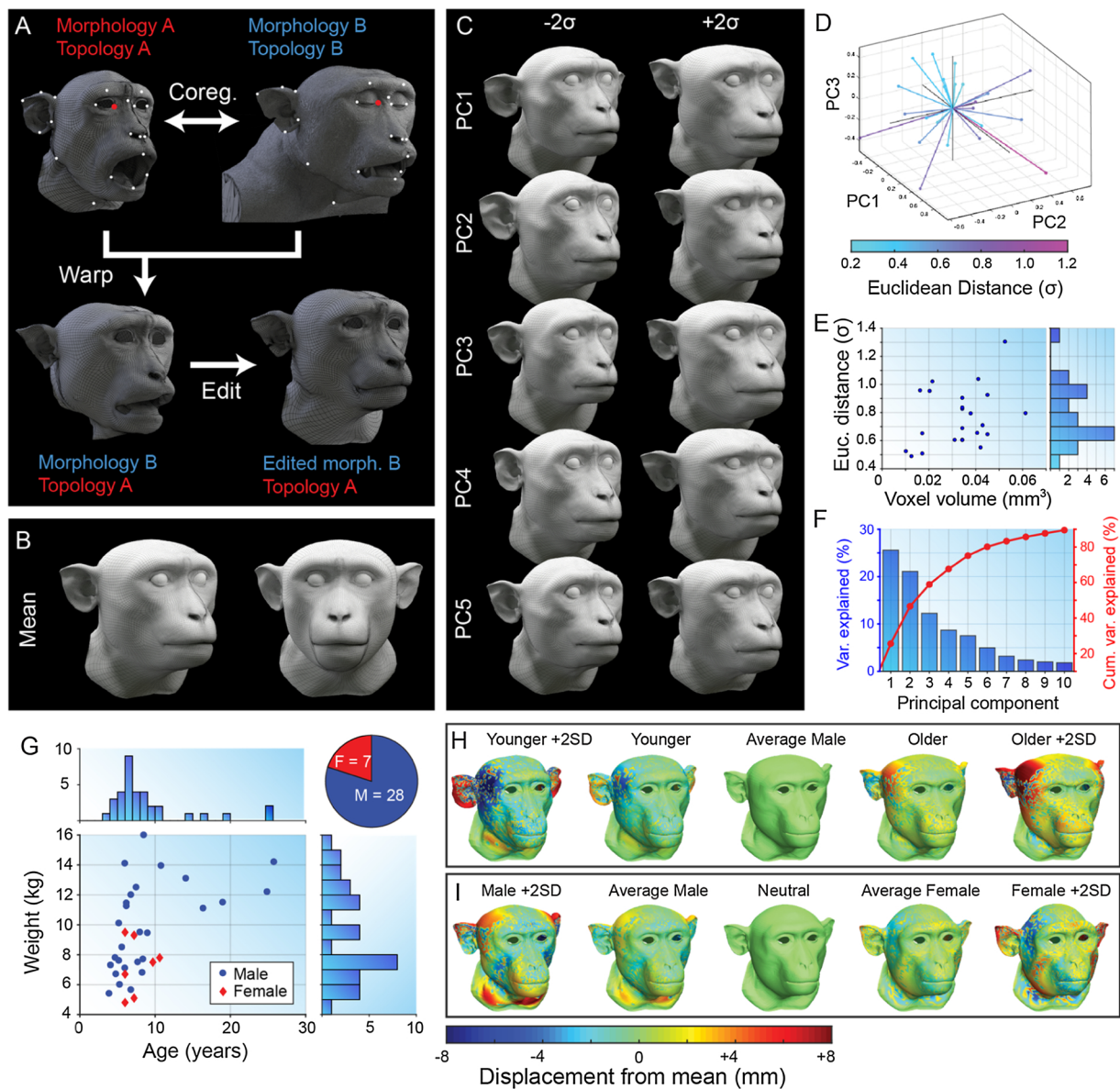Three Rhesus macaques (*Macaca mulatta*) participated in the neurophysiology experiments (2 female, aged 6–13 years, weight 5.6–9.8 kg). All monkeys were implanted with a custom fiberglass headpost used to immobilize the head during experiments, and chronic microwire multielectrodes (MicroProbes, Gaithersburg, MD) targeting functionally-localized 'face-patches' as previously described elsewhere (Bondar et al., 2009; McMahon et al., 2014). All procedures were approved by the Animal Care and Use Committee of the U.S. National Institutes of Health (National Institute of Mental Health) and followed U.S. National Institutes of Health guidelines.

2.3.2. Stimulus presentation

The sixty colour photographs used in the screening experiment were taken from a larger stimulus set (McMahon et al., 2014; Thomaz and Giraldi, 2010), and included 10 exemplars that were personally unfamiliar to the subjects from each of six categories: human faces, Rhesus macaque faces, Rhesus macaque bodies, man-made objects, scenes, and birds. Photographic stimuli were always presented at a diameter of 10°, within a mid-grey circular aperture with cosine edge gradient.

In the main experiment, CGI stimuli were presented on a pair of ASUS VG278H (27″ diagonal, 1920 × 1080 pixels, 120 Hz) arranged in a Wheatstone stereoscope configuration using cold mirrors (Edmund Optics) with eye-tracking cameras (EyeLink II, SR Research, Ontario, Canada) placed behind them. The viewing distance was 80 cm. The luminance outputs of both monitors were measured using a Konica Minolta CA-210 photometer and linearized and matched in range (0.55–376.7 cd/m^2) by calculating gamma look-up tables for each display.

In all experiments, stimuli were presented on a mid-grey background for a 300 ms duration, with an inter-stimulus interval of 300 ms. Each experimental block began with the appearance of a central

Fig. 5. Morphable face model construction. A. Example of manual selection of corresponding vertices on the low-poly base mesh topology (topology A) created from individual M02 (morphology A) and the high-poly raw surface mesh of individual M09 (right). The warping process produces a surface mesh with topology A and morphology B (bottom left), which can then be manually edited (bottom right). B. Sample mean mesh surface. C. First five principal components (mean ± 2σ) of macaque face-space. D. Locations of original sample identities (n = 23) projected into principal component face-space (first 3 PC dimensions only). E. Distribution of CT scan voxel volume for each individual plotted against their Euclidean distance from the sample mean (σ). F. Percentage of variance in sample cranio-facial morphology explained by each principal component. G. Distributions of demographic variables for Rhesus macaque CT data sample (see also Table 1). H. Age trajectory through face-space for males calculated by averaging 5 youngest (2nd column) and 5 oldest (4th column) males, and extrapolating. I. Sexual dimorphism trajectory through face-space calculated by averaging 5 males (2nd column) and 5 females (4th column), and extrapolating. Colour map indicates the displacement of each vertex relative to the mean (middle column) for each mesh. Meshes were aligned via Procrustes method.

fixation cross, which the subject was required to fixate in order for the block to continue. On each trial, between five to ten stimuli were presented sequentially and juice reward was given at the end of the trial provided that the subject's fixation remained within a 1° radius of the center of the fixation marker throughout. Trials on which the subject broke the fixation requirement were aborted without reward and stimulus presentations from aborted trials were not included in the analyses.
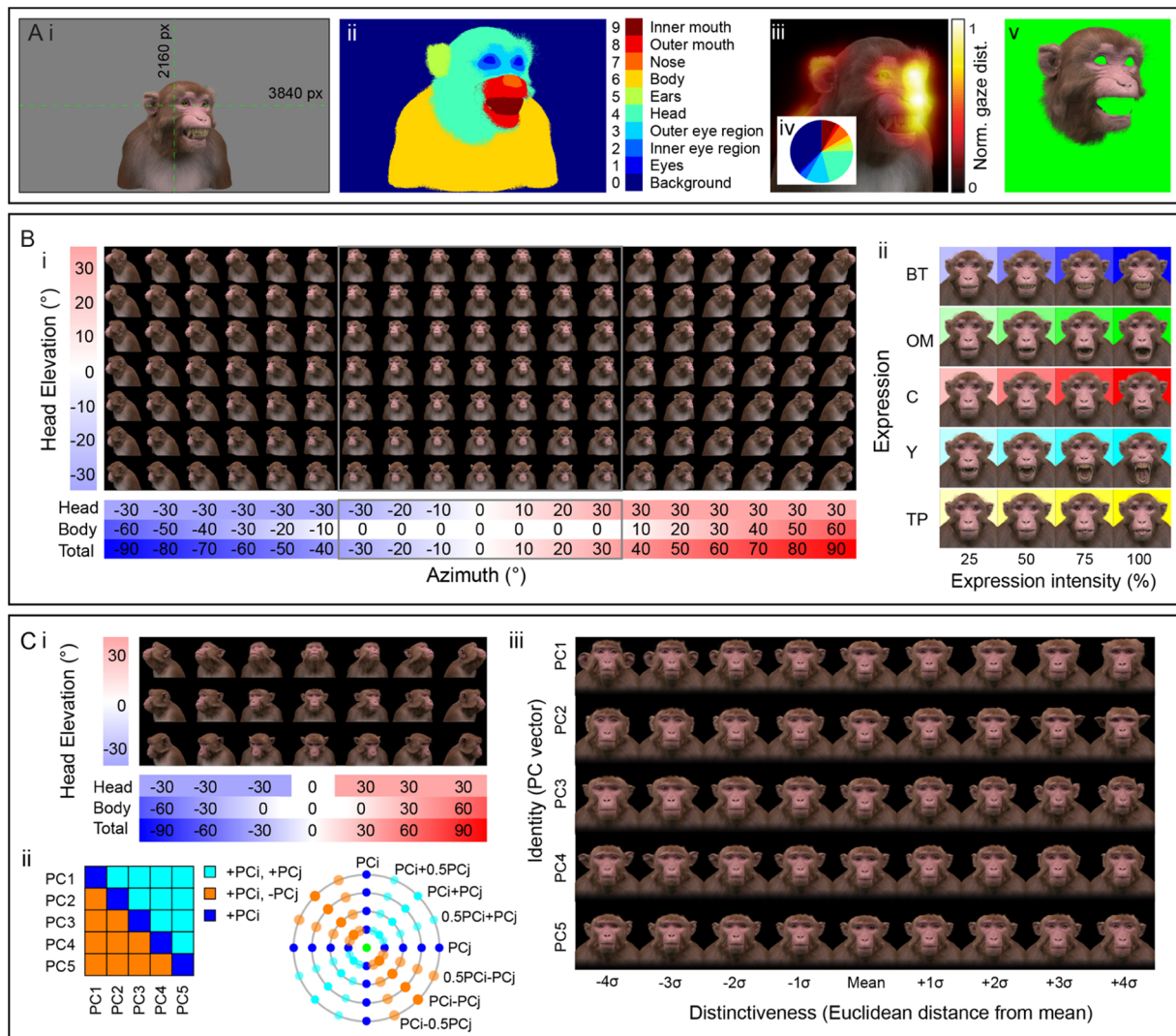
2.3.3. Recording and data analysis

Multi-channel (64–128 channels) data was recorded at 24KHz using a Tucker-Davis Technologies PZ5 preamp and RZ2 bioamp interface (TDT, Alachua, FL). The signal from a photodiode attached to the

corner of the display screen was recorded simultaneously as an analog signal at 1KHz, along with eye position, pupil diameter, and the TTL signal controlling liquid reward delivery. Offline spike sorting was performed using the Matlab-based WaveClus (Quiroga et al., 2004) and utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

2.3.4. Longitudinal tracking of individual neurons

For the electrophysiological data collection in the current study, we utilized the ability to track the same individual neurons across multiple recording sessions. We tracked neuronal identity across sessions based on several lines of evidence (Supplementary Fig. S1). First, we tested neuronal selectivity profiles for an independent stimulus set consisting

Fig. 6. Summary of the publicly released MF3D stimulus set, release 1. A. Stimulus set format and applications. i. Each colour image in the set is rendered as a 3840 × 2160 pixel RGBA image in. png format with 32-bits per pixel. The avatar is positioned such that the center of the screen coincides with the cyclopean eye when the avatar is directly facing the camera. ii. For each colour image, we provide a corresponding label map image (.hdr format) of the same dimensions, where integer pixel values indicate which anatomical structure of the avatar they belong to. iii. An example of a simulated gaze distribution map for the stimulus shown in i. iv. Proportion of fixations on each labelled structure can be easily computed. v. Novel stimuli can be created by using the label map to mask specific parts of the original image. B. Expression stimuli. i. All head orientations rendered for each expression condition (neutral expression shown for illustration): 19 azimuth angles (-90 to +90° in 10° increments) x 7 elevation angles (-30 to +30° in 10° increments) for 133 unique head orientations. ii. Five facial expressions (rows) rendered at four levels of intensity (columns), at each of the head orientations illustrated in F, for a total of 2793 unique colour images. C. Identity stimuli. i. All head orientations rendered for each identity condition (average identity shown for illustration): 7 azimuth angles x 3 elevation angles for 21 head orientations. ii. Identity trajectories through face space were selected through all pairwise combinations of the first 5 principal components from the PCA (which cumulatively account for 75% of the sample variance in facial morphology), at 3 polar angles for a total of 65 unique trajectories. iii. Identities were rendered at eight levels of distinctiveness ($\pm 4\sigma$ from the sample mean in 1σ increments) along each identity trajectory (shown here for the first 5 PCs), plus the sample mean for a total of 10,941 unique colour images.

of 60 photographic images (as described above). While neighbouring IT neurons showed unique selectivity profiles, a given neuron maintained a consistent profile across time (McMahon et al., 2014). Second, we characterized parameters of the spike waveforms for each putative neuron, including peak-to-trough amplitude and duration.

2.4. MF3D stimulus set (release 1)

2.4.1. Static image set

As a first step in the development of a standardized face database for the macaque research community, we have rendered a stimulus set specifically for public release (Murphy & Leopold, 2019a; 2019b). The set consists of 14,000 static rendered conditions, saved as high resolution (3840 × 2160 pixels, 32-bit) RGBA images in. png format

(Fig. 6A, i). The inclusion of the alpha transparency channel allows for compositing of multiple images into a frame, including backgrounds, as well as making it easy to generate control stimuli with identical silhouettes. The high resolution permits down-sampling or cropping as appropriate for the display size being used. The virtual scene was configured such that the avatar will appear at real-world retinal size when images are presented at full-screen on a 27″ monitor with 16:9 aspect ratio, at 57 cm viewing distance from the subject. For each 2D colour image, we additionally provide a label map image, which is an indexed image that assigns each pixel an integer value depending on the anatomical region of the avatar it belongs to (Fig. 6A, ii). Label maps can be used to analyse subjects' gaze in free-viewing paradigms (Fig. 6A, iii–iv), for gaze-contingent reward schedules, or for generating novel stimuli by masking specific structures in the corresponding colour

image (Fig. 6A, v).

The static stimuli of MF3D release 1 are divided into two collections: 1) variable expressions with fixed identity (corresponding to real individual M02; see Table 1); and 2) variable identities with fixed expression (neutral). For the expression set, we varied head orientation (±90° azimuth x ±30° elevation in 10° increments = 133 orientations; Fig. 6B, i), facial expression type (neutral plus bared-teeth 'fear grimace', open-mouthed threat, coo, yawn, and tongue-protrusion = 5) and the intensity of the expression (25, 50, 75 and 100% = 4; Fig. 6B, ii). We additionally include the neutral expression with open and closed eyes, as well as azimuth rotations beyond 90° (100 to 260° in 10° increments) for a total of 2926 colour images. In order to maintain naturalistic poses, head orientation was varied through a combination of neck (±30° azimuth and elevation) and body (±60° azimuth) orientations.

For the identity set, we selected a subset of head orientations (±90° azimuth x ±30° elevation in 30° increments = 21 orientations; Fig. 6C, i), and co-varied facial morphology based on distinct trajectories within PCA-space (n = 65; Fig. 6C, ii), including each of the first five PCs (which together account for 75% of the sample variance in facial morphology; Fig. 5F), with distinctiveness (Euclidean distance from the average face, ±4σ in 1σ increments = 8 levels, excluding the mean; Fig. 6C, iii) for a total of 10,941 identity images.

2.4.2. Dynamic stimulus set

For studies requiring more naturalistic stimuli, we also have the ability to generate a virtually limitless number of animations that, while computationally expensive to generate, promise great flexibility for studying dynamic facial behaviour. Here we have included a small selection of short animations as a proof of concept, which are rendered at 1920 × 1080 pixels and 30 frames per second, encoded with H.264 compression and saved in. mp4 format with a black background. For each action sequence, animations are rendered from 5 different camera azimuth angles (-60, to 60° in 30° increments). All animations feature identical start and end frames, which allows the possibility of stitching multiple clips together using video editing software (such as the video editor included in Blender), to produce longer, seamless movies containing various permutations of action sequences. The animations were produced by manually coding video footage of real Rhesus macaques performing facial expressions and vocalizations, as described above. Over time, we plan to generate and release additional movies animating the face and head.

3. Results

3.1. Single-unit responses to CGI macaque faces in macaque inferotemporal cortex

We recorded single-unit spiking activity of neurons in three face-selective regions of macaque IT, while we briefly presented static visual stimuli. 'Face-selective' neurons were identified using an independent stimulus set consisting of sixty colour photographs of face (macaque and human) and non-face objects, presented in 2D. Within the local populations of face-selective neurons that we recorded, the majority of cells showed a preference for the faces of conspecifics (Rhesus macaques) over those of humans.

We presented static rendered images of the macaque avatar, in which we co-varied multiple parameters. These included head azimuth orientation (7 levels), head elevation orientation (3 levels), facial expression (4 levels), facial colouration and fur length (5 levels), physical size (3 levels), position in depth (3 levels), depth profile (3 levels) and -for human GCI faces- facial identity (2 levels). Neural responses to these parameter variations are illustrated in Fig. 7 for four example neurons.

Many of the neurons recorded exhibited sensitivity to head orientation, with enhanced responses for particular regions of the 2-dimensional head orientation space (i.e. head orientation tuning). For many cells, this tuning to head orientation was robust across variations in the other parameters (e.g. Fig. 7, cells 1–3). Despite the general preference for conspecific faces in the majority of neurons, we also observed that neurons tuned for specific head orientations retained their tuning even when the species of the head changed (Fig. 7, ML cell 2). In contrast, some face-selective neurons appeared to be relatively insensitive to any of the variables tested. Surprisingly, these cells could be strongly modulated by highly specific combinations of variables, such as AM cell 2 shown in Fig. 7, which responded more than twice as strongly to an upward oriented face with bared teeth 'fear grimace' expression compared to any other condition presented.
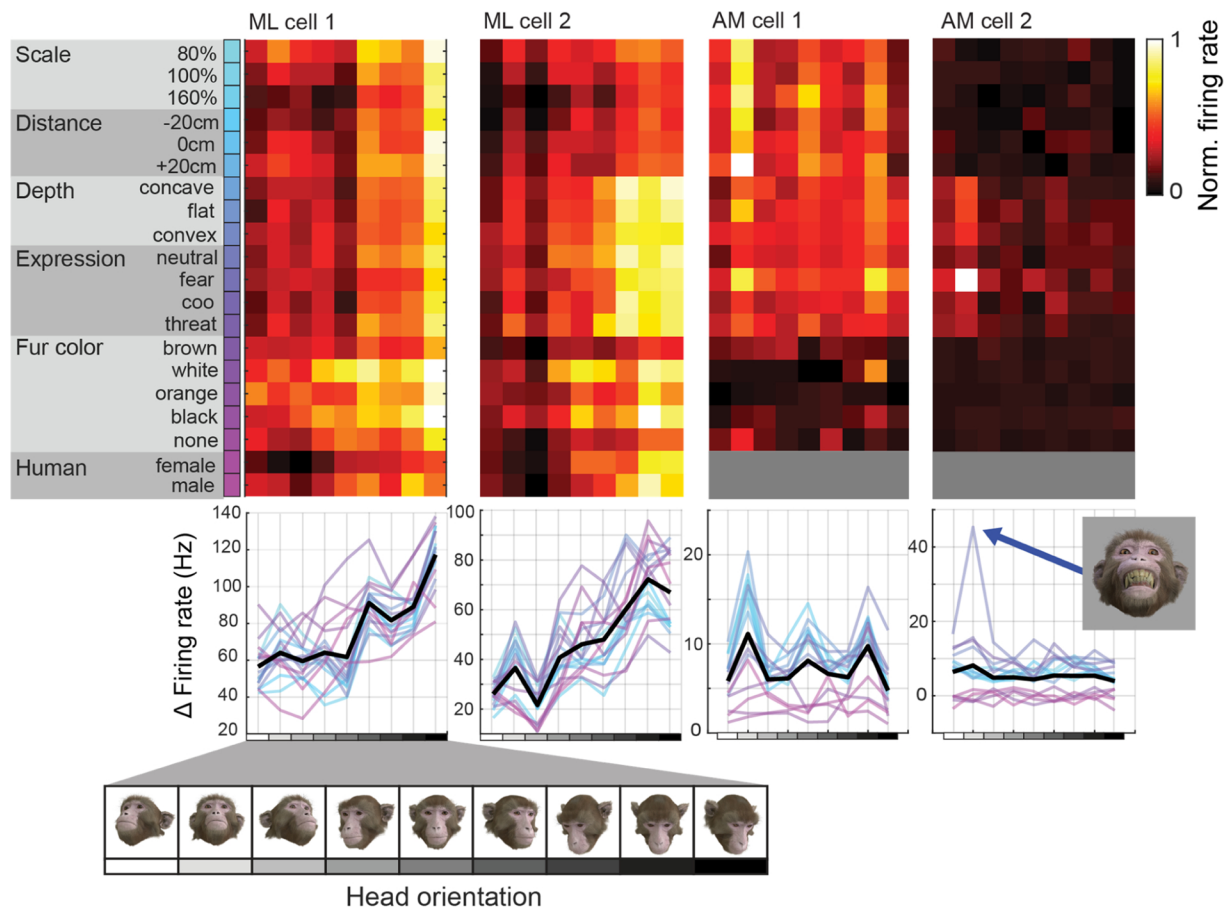
4. Discussion

4.1. Summary

We used CT, photographic and video data collected from Rhesus macaques to develop a realistic virtual 3D model of the Rhesus macaque head and face. The model features various parameters that can be continuously manipulated to vary facial appearance, including facial expression, cranio-facial morphology, head orientation, gaze direction, pupil dilation, fur length and colouration, skin colouration, as well as spatial variables such as physical scale, depth profile and position in 3D space. The model can be used to render images and animations for stereoscopic 3D presentation, improving the ethological validity of experimental visual stimuli. We presented several thousand such visual stimuli to Rhesus macaques while we recorded single-unit responses from three face-selective regions of IT cortex. We observed the tuning of neuronal responses to highly specific parameter combinations of the virtual macaque face model.

The use of a virtual avatar allows for precise control of various behaviourally relevant, continuous variables that various regions the macaque brain appear to be sensitive to, including head orientation (De Souza et al., 2005; Freiwald and Tsao, 2010; Perrett et al., 1985; Tazumi et al., 2010), gaze direction (Marciniak et al., 2014; Perrett et al., 1985; Tazumi et al., 2010), viewing distance (Ashbridge et al., 2000), and lighting direction (Braje et al., 1998; Hietanen et al., 1992) among others.

4.2. Neuronal responses to CGI macaque faces

Given the independent control over multiple variables afforded by the CGI macaque avatar, it is possible to design experiments testing a large number of feature combinations. Together with the need for multiple presentations of each stimulus, this places experimenters at great advantage if they can accumulate data across multiple sessions. Thanks to the development of chronically implantable microelectrodes capable of stably isolating the same population of neurons over weeks and months, this has now become a feasible approach (Bondar et al., 2009; Dotson et al., 2017; Jun et al., 2017; McMahon et al., 2014).

We recorded single-unit activity in three face-selective regions of macaque IT during visual presentation of CGI macaque faces. This approach, combined with chronically implanted electrodes, allowed us to vary a large combination of parameters affecting facial appearance. We observed that the responses of many neurons from all three populations were strong modulated by changes in the avatar's head orientation. This replicates previous findings of head orientation tuning in macaque IT neurons (Freiwald and Tsao, 2010; Oram and Perrett, 1992; Perrett et al., 1991, 1985). However, we were able to go further by showing that head orientation tuning is often robust across variations of other parameters that dramatically transform the retinal image. For a minority of neurons, we were also able to identify unique parameter combinations that strongly increased the cell's firing rate relative to all other stimuli. Such findings would not be possible without the comprehensive mapping of the complex, multidimensional stimulus space

Fig. 7. Neural tuning to head orientation across co-variations of orthogonal variables in four example neurons. Rows correspond to different values of parameters of facial appearance that were systematically varied, while columns correspond to the head orientations that were co-varied. The consistent vertical structure in the data from the first cell (from patch ML) indicates a robust tuning for head orientation that is relatively invariant to changes of other parameters. In contrast, the data from the last neuron shown (from patch AM) shows extremely high specificity for an interaction between facial expression and head orientation, responding only to faces with a fearful expression oriented upward (inset image).

that the macaque avatar allows.

Direct comparison of behavioural and neural data between studies of face processing in the macaque is often hampered by variations in proprietary visual stimuli and presentation paradigms used. By releasing an open-access image repository containing the first standardized macaque face stimulus set with parametric variation of visual features, we hope to provide a common currency to facilitate comparison of future results between laboratories and across measurement modalities.

4.3. Limitations

While our macaque avatar represents a significant step forward in realism compared to previous virtual macaque avatars (Simpson et al., 2016; Steckenfinger and Ghazanfar, 2009), our stimuli are also unlikely to be mistaken for a real animal by either humans or macaques. For certain scientific questions, there is currently no substitute for the authenticity and natural variability of real social interactions, despite the costs to experimental control (Azzi et al., 2012; Ballesta and Duhamel, 2015; Dal Monte et al., 2016). However, the stimuli generated from our model have several important visual features that better approximate reality than 2D photographic images, including their geometrically-accurate stereoscopic depth and their continuously variable parameters.

Similarly, the current model undoubtedly fails to capture all of the subtle facial movements that real macaques are capable of (Burrows et al., 2009). A more complete range of macaque facial expression has

been analysed based on anatomical constraints, and can be described using a facial action coding system for this species (Parr et al., 2010). The 3D dynamics of human facial expressions have previously been successfully modelled by identifying action units of facial expression in 4D data sets (Yu et al., 2012). Utilizing computer vision approaches that fit 3D morphable models to 2D video data might help to overcome some challenges of photogrammetry-based modelling of the macaque. Such data could be used to increase the variety of facial movements in our model whithout the need for time-consuming and subjective modelling by digital artists, whilst also applying spatial and temporal biological constraints.

The generalizability of the 'face-space' model presented here is currently limited by the sample size. Additional data from infant, juvenile and female individuals would improve the representation of these demographics, and improve modelling of these behaviourally important dimensions of facial identity. Adult macaques of various species show a preference for viewing conspecifics, and thus demonstrate the ability to discriminate conspecifics from evolutionarily close heterospecifics of the same genus (Fujita, 1987; Fujita and Watanabe, 1995). Although the majority of neuroscientific research on face processing using non-human primate models is conducted with Rhesus monkeys, it may be useful to produce variants of the model for testing other macaque species, as well as other primate species such as the common marmoset (Mitchell and Leopold, 2015).

The statistical description of a species face-space that has been utilized here and elsewhere (Blanz and Vetter, 1999; Chang and Tsao, 2017; Paysan et al., 2009; Sirovich and Kirby, 1987) is inherently

biased by the sample selection. This includes not only which individuals are sampled from the population, but also the distribution of facial landmarks that are used for alignment. The 3D mesh based approach used here goes some way to minimizing the latter bias, since all ˜25,000 vertices are used as equally-weighted alignment points. However, due to the higher density of vertices around regions of high curvature (such as the pinna), the results may be biased towards certain structures (e.g. some PCs appear to strongly reflect ear position or deformation). Another advantage of our approach over those used in human face analysis is that all of our subjects were anesthetized during scan acquisition, thus removing minor variations in facial muscle tone that inevitably occur when human subjects are instructed to maintain a 'neutral' expression. However, as a consequence of this construction method the model does not currently allow for co-variation of both facial expression and facial identity. Several computational solutions to this problem already exist (Thies et al., 2015; Vlasic et al., 2006), and future developments of the macaque avatar will aim to add expression variation across identities.

A major limitation of our approach to modelling individual variations in facial appearance is the lack of veridical texture and colour information, as well as volumetric information related to fur. Although others have chosen to study shape and texture as independent factors contributing to individual facial identity encoding (Chang and Tsao, 2017; Paysan et al., 2009), these factors are often correlated. One possible approach to overcome this would be to acquire photographic data from the sample subjects in order to generate a texture model, in correspondence with the shape model. While this approach is fairly trivial for hairless human faces, it may be challenging to analyse photographic data in a manner than captures variations in fur length and volume, which may be important cues for visual processing of identity in macaques. The ability to parametrically vary multiple dimensions of fur appearance in the current model makes it possible to test which aspects are most behaviourally relevant to this species.

4.4. Future directions

The stimuli presented here demonstrate how a digital 3D model of the macaque can be used to expand the stimulus parameter space within the context of a traditional vision neurophysiology paradigm – namely, brief presentations of static images during passive fixation. However, a major strength of the digital model lies in the ability to break away from this approach and re-introduce important elements of natural vision to research on object vision and social perception, without sacrificing experimental control and reproducibility.

Motion is an ethologically important visual signal that accounts for the majority of the variance in the primate visual cortex during natural viewing (Park et al., 2017; Russ and Leopold, 2015), and in the case of biological facial motion may be processed by a dedicated network (Fisher and Freiwald, 2015a; Polosecki et al., 2013). Through careful measurement and modelling of the temporal dynamics of facial movement in video footage of real Rhesus macaques, we have been able to animate parameters of the macaque avatar to approximate a naturalistic manner (Movie S2; Fig. 4). The output animations preserve the natural dynamics and correlated facial movements of the original video clips, but allows other parameters of interest to be varied independently. This provides greater control and flexibility than would be possible using video footage or live animals as visual stimuli. Further, the spatial and temporal resolutions of the output video are not constrained by those of the input video, and it is possible to render the output version in stereoscopic 3D format. However, direct measurement of head orientation (e.g. via optical head tracking), eye direction and pupil dilation (via video eye-tracking), and facial muscle activations (via electromyography) in awake macaques during specific natural viewing contexts could facilitate modelling and animation of these parameters in future.

Additionally, colour conveys important social and sexual signals in many animal species, including catarrhine primates who are trichromats (Dominy and Lucas, 2001; Surridge et al., 2003). Females of most macaque species (including Rhesus) show variations in facial colouration over the course of the ovarian cycle (Dubuc et al., 2009; Higham et al., 2010; Pflüger et al., 2014), and animals of both genders can pick up on this signal (Gerald et al., 2007; Waitt et al., 2006). Facial colouration in male Rhesus macaques is also known to influence social interactions (Dubuc et al., 2014; Higham et al., 2013; Waitt et al., 2003), while even individual variations in iris colour may also convey useful information (Zhang and Watanabe, 2007). Powerful node-based image and colour editing tools provided in the Blender software provide a basis for the future addition of statistically constrained colour variations of the virtual macaque model.

In the real world, faces do not appear in isolation but are part of the body. The inclusion of the head, neck, an upper torso in our model is a small step toward increasing ethological validity beyond the traditional use of cropped face images. However, real bodies also provide visual cues about emotional and attentional states and individual identity (Kumar et al., 2017), which are integrated with facial information at higher levels of neural processing (Fisher and Freiwald, 2015b). Expansion of the current model to include the whole macaque body, rigged for species-typical poses, morphable for age, sex and adiposity, and animated for typical locomotor or reaching and grasping behaviours would require a major endeavor, but could be of value to an even broader range of research in this model species. Integration with existing models (Kumar et al., 2017; Putrino et al., 2015) and motion-capture data (Ogihara et al., 2009) could facilitate this.

Another application of the avatar is to insert one or more instances into the context of a virtual 3D environment, allowing for realistic rendering of entire scenes – for presentation on either standard or immersive displays (Supplementary Fig. S2). Virtual reality offers numerous benefits for behavioural and neuroscientific research, and serves to further increase ethological validity whilst retaining experimental control (Bohil et al., 2011; Doucet et al., 2016; Parsons, 2015). A promising avenue for future development will be to adapt the model to operate within a real-time game engine environment that would allow for closed-loop 'social' interactions between subjects and virtual conspecifics using online analyses of subjects' behavioural responses (e.g. eye position, facial movement, vocalizations) and statistical modelling of real social behaviour. Data on how social interactions dynamically modulate gaze (Dal Monte et al., 2016; Ghazanfar et al., 2006), blink rate (Ballesta et al., 2016), pupil dilation (Ebitz et al., 2014), reciprocal facial expression (Mosher et al., 2011) and other parameters of the avatar could be used to model these natural statistics.

4.5. Conclusions

The development of a parametric digital 3D model of the Rhesus macaque face reported here represents an advancement in visual stimulus design for studies of visual object and social processing in this important primate model species. Images and animations generated from the model offer researchers both greater ethological validity and tighter experimental control compared to traditional approaches (e.g. harvesting photographic images from the internet or recording videos of animals whose behaviours cannot be controlled).

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jneumeth.2019.06.001.

References

Ashbridge, E., Perrett, D.I., Oram, M.W., Jellema, T., 2000. Effect of image orientation and size on object recognition: responses of single units in the macaque monkey temporal cortex. Cogn. Neuropsychol. 17, 13–34.

Augusteyn, R.C., Heilman, B.M., Ho, A., Parel, J.-M., 2016. Nonhuman primate ocular biometry. Invest. Ophthalmol. Vis. Sci. 57, 105–114.

Azzi, J.C.B., Sirigu, A., Duhamel, J.-R., 2012. Modulation of value representation by social context in the primate orbitofrontal cortex. Proc. Natl. Acad. Sci. 109, 2126–2131.

Ballesta, S., Duhamel, J.-R., 2015. Rudimentary empathy in macaques' social decision-making. Proc. Natl. Acad. Sci. 112, 15516–15521.

Ballesta, S., Mosher, C.P., Szep, J., Fischl, K.D., Gothard, K.M., 2016. Social determinants of eyeblinks in adult male macaques. Sci. Rep. 6, 38686.

Becker, W., Fuchs, A.F., 1988. Lid-eye coordination during vertical gaze changes in man and monkey. J. Neurophysiol. 60, 1227–1252.

Blanz, V., Vetter, T., 1999. A Morphable Model for the Synthesis of 3D Faces. ACM Press/Addison-Wesley Publishing Co., pp. 187–194.

Bohil, C.J., Alicea, B., Biocca, F.A., 2011. Virtual reality in neuroscience research and therapy. Nat. Rev. Neurosci. 12, 752.

Bondar, I.V., Leopold, D.A., Richmond, B.J., Victor, J.D., Logothetis, N.K., 2009. Long-term stability of visual pattern selective responses of monkey temporal lobe neurons. PLoS One 4, e8222.

Braje, W.L., Kersten, D., Tarr, M.J., Troje, N.F., 1998. Illumination effects in face recognition. Psychobiology 26, 371–380.

Burrows, A.M., Waller, B.M., Parr, L.A., 2009. Facial musculature in the rhesus macaque (Macaca mulatta): evolutionary and functional contexts with comparisons to chimpanzees and humans. J. Anat. 215, 320–334.

Bush, G.A., Miles, F.A., 1996. Short-latency compensatory eye movements associated with a brief period of free fall. Exp. Brain Res. 108, 337–340.

Chang, L., Tsao, D.Y., 2017. The code for facial identity in the primate brain. Cell 169, 1013–1028.

Dahl, C.D., Wallraven, C., Bülthoff, H.H., Logothetis, N.K., 2009. Humans and macaques employ similar face-processing strategies. Curr. Biol. 19, 509–513.

Dal Monte, O., Piva, M., Morris, J.A., Chang, S.W.C., 2016. Live interaction distinctively shapes social gaze dynamics in rhesus macaques. J. Neurophysiol. 116, 1626–1643.

De Souza, W.C., Eifuku, S., Tamura, R., Nishijo, H., Ono, T., 2005. Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. J. Neurophysiol.

Dominy, N.J., Lucas, P.W., 2001. Ecological importance of trichromatic vision to primates. Nature 410, 363–366.

Dotson, N.M., Hoffman, S.J., Goodell, B., Gray, C.M., 2017. A large-scale semi-chronic microdrive recording system for non-human primates. Neuron 96, 769–782.

Doucet, G., Gulli, R.A., Martinez-Trujillo, J.C., 2016. Cross-species 3D virtual reality toolbox for visual and cognitive experiments. J. Neurosci. Methods 266, 84–93.

Dubuc, C., Allen, W.L., Maestripieri, D., Higham, J.P., 2014. Is male rhesus macaque red color ornamentation attractive to females? Behav. Ecol. Sociobiol. (Print) 68, 1215–1224.

Dubuc, C., Brent, L.J.N., Accamando, A.K., Gerald, M.S., MacLarnon, A., Semple, S., Heistermann, M., Engelhardt, A., 2009. Sexual skin color contains information about the timing of the fertile phase in free-ranging Macaca mulatta. Int. J. Primatol. 30, 777–789.

Dufour, V., Pascalis, O., Petit, O., 2006. Face processing limitation to own species in primates: a comparative study in brown capuchins, Tonkean macaques and humans. Behav. Processes 73, 107–113.

Ebitz, R.B., Pearson, J.M., Platt, M.L., 2014. Pupil size and social vigilance in rhesus macaques. Front. Neurosci. 8, 100.

Emery, N.J., Lorincz, E.N., Perrett, D.I., Oram, M.W., Baker, C.I., 1997. Gaze following and joint attention in rhesus monkeys (Macaca mulatta). J. Comp. Psychol. 111, 286.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., 2012. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn. Reson. Imaging 30, 1323–1341.

Ferrari, P.F., Paukner, A., Ionica, C., Suomi, S.J., 2009. Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. Curr. Biol. 19, 1768–1772.

Fisher, C., Freiwald, W.A., 2015a. Contrasting specializations for facial motion within the macaque face-processing system. Curr. Biol.

Fisher, C., Freiwald, W.A., 2015b. Whole-agent selectivity within the macaque face-

processing system. Proc. Natl. Acad. Sci. 112, 14717–14722.

Freiwald, W.A., Tsao, D.Y., 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330, 845–851.

Freiwald, W.A., Tsao, D.Y., Livingstone, M.S., 2009. A face feature space in the macaque temporal lobe. Nat. Neurosci. 12, 1187.

Fujita, K., 1987. Species recognition by five macaque monkeys. Primates 28, 353–366.

Fujita, K., Watanabe, K., 1995. Visual preference for closely related species by Sulawesi macaques. Am. J. Primatol. 37, 253–261.

Gamlin, P.D.R., McDougal, D.H., Pokorny, J., Smith, V.C., Yau, K.-W., Dacey, D.M., 2007. Human and macaque pupil responses driven by melanopsin-containing retinal ganglion cells. Vision Res. 47, 946–954.

Gerald, M.S., Waitt, C., Little, A.C., Kraiselburd, E., 2007. Females pay attention to female secondary sexual color: an experimental study in Macaca mulatta. Int. J. Primatol. 28, 1–7.

Ghazanfar, A.A., Nielsen, K., Logothetis, N.K., 2006. Eye movements of monkey observers viewing vocalizing conspecifics. Cognition 101, 515–529.

Gothard, K.M., Brooks, K.N., Peterson, M.A., 2009. Multiple perceptual strategies used by macaque monkeys for face recognition. Anim. Cogn. 12, 155–167.

Guipponi, O., Odouard, S., Pinède, S., Wardak, C., Ben Hamed, S., 2014. fMRI cortical correlates of spontaneous eye blinks in the nonhuman primate. Cereb. Cortex 25, 2333–2345.

Hafed, Z.M., Goffart, L., Krauzlis, R.J., 2009. A neural mechanism for microsaccade generation in the primate superior colliculus. science 323, 940–943.

Hietanen, J.K., Perrett, D.I., Oram, M.W., Benson, P.J., Dittrich, W.H., 1992. The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. Exp. Brain Res. 89, 157–171.

Higham, J.P., Brent, L.J.N., Dubuc, C., Accamando, A.K., Engelhardt, A., Gerald, M.S., Heistermann, M., Stevens, M., 2010. Color signal information content and the eye of the beholder: a case study in the rhesus macaque. Behav. Ecol. 21, 739–746.

Higham, J.P., Pfefferle, D., Heistermann, M., Maestripieri, D., Stevens, M., 2013. Signaling in multiple modalities in male rhesus macaques: sex skin coloration and barks in relation to androgen levels, social status, and mating behavior. Behav. Ecol. Sociobiol. (Print) 67, 1457–1469.

Hughes, A., 1977. The topography of vision in mammals of contrasting life style: comparative optics and retinal organisation. The Visual System in Vertebrates. Springer, pp. 613–756.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., 2017. Fully integrated silicon probes for high-density recording of neural activity. Nature 551, 232.

Kiani, R., Esteky, H., Tanaka, K., 2005. Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. J. Neurophysiol. 94, 1587–1596.

Kret, M.E., Tomonaga, M., Matsuzawa, T., 2014. Chimpanzees and humans mimic pupil-size of conspecifics. PLoS One 9, e104886.

Kumar, S., Popivanov, I.D., Vogels, R., 2017. Transformation of visual representations across ventral stream body-selective patches. Cereb. Cortex 29, 215–229.

Kustár, A., Forró, L., Kalina, I., Fazekas, F., Honti, S., Makra, S., Friess, M., 2013. FACE-r—a 3D database of 400 living individuals' full head CT-and face scans and preliminary GMM analysis for craniofacial reconstruction. J. Forensic Sci. 58, 1420–1428.

Leopold, D.A., Rhodes, G., 2010. A comparative view of face perception. J. Comp. Psychol. 124, 233.

Marciniak, K., Atabaki, A., Dicke, P.W., Thier, P., 2014. Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus. Elife 3, e03222.

McMahon, D.B.T., Bondar, I.V., Afuwape, O.A.T., Ide, D.C., Leopold, D.A., 2014. One month in the life of a neuron: longitudinal single-unit electrophysiology in the monkey visual system. J. Neurophysiol. 112, 1748–1762.

Minxha, J., Mosher, C., Morrow, J.K., Mamelak, A.N., Adolphs, R., Gothard, K.M., Rutishauser, U., 2017. Fixations gate species-specific responses to free viewing of faces in the human and macaque amygdala. Cell Rep. 18, 878–891.

Mitchell, J.F., Leopold, D.A., 2015. The marmoset monkey as a model for visual neuroscience. Neurosci. Res. 93, 20–46.

Mosher, C.P., Zimmerman, P.E., Gothard, K.M., 2011. Videos of conspecifics elicit interactive looking patterns and facial expressions in monkeys. Behav. Neurosci. 125, 639.

Murphy, A.P. and Leopold DA (2019a). MF3D R1 Expressions. figshare. Dataset. https://doi.org/10.6084/m9.figshare.8226029.v1.

Murphy, A.P. and Leopold DA (2019b). MF3D R1 Expressions. figshare. Dataset. https://doi.org/10.6084/m9.figshare.8226311.v2.

Ogihara, N., Makishima, H., Aoi, S., Sugimoto, Y., Tsuchiya, K., Nakatsukasa, M., 2009. Development of an anatomically based whole-body musculoskeletal model of the Japanese macaque (Macaca fuscata). Am. J. Phys. Anthropol. 139, 323–338.

Oram, M.W., Perrett, D.I., 1992. Time course of neural responses discriminating different views of the face and head. J. Neurophysiol. 68, 70–84.

Ostrin, L.A., Glasser, A., 2004. The effects of phenylephrine on pupil diameter and accommodation in rhesus monkeys. Invest. Ophthalmol. Vis. Sci. 45, 215–221.

Park, S.H., Russ, B.E., McMahon, D.B.T., Koyano, K.W., Berman, R.A., Leopold, D.A., 2017. Functional subpopulations of neurons in a macaque face patch revealed by single-unit fMRI mapping. Neuron 95, 971–981.

Parr, L.A., Waller, B.M., Burrows, A.M., Gothard, K.M., Vick, S.-J., 2010. Brief communication: MaqFACS: a muscle-based facial movement coding system for the rhesus macaque. Am. J. Phys. Anthropol. 143, 625–630.

Parsons, T.D., 2015. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front. Hum. Neurosci. 9, 660.

Pascalis, O., Bachevalier, J., 1998. Face recognition in primates: a cross-species study. Behav. Processes 43, 87–96.

Pascalis, O., de Haan, M., Nelson, C.A., 2002. Is face processing species-specific during the first year of life? Science 296, 1321–1323.

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T., 2009. A 3D Face Model for Pose and Illumination Invariant Face Recognition. IEEE, pp. 296–301.

Perrett, D.I., Oram, M.W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, P.J., Thomas, S., 1991. Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Exp. Brain Res. 86, 159–173.

Perrett, D.I., Smith, P.A.J., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., Jeeves, M.A., 1985. Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc. R. Soc. Lond. B 223, 293–317.

Pflüger, L.S., Valuch, C., Gutleb, D.R., Ansorge, U., Wallner, B., 2014. Colour and contrast of female faces: attraction of attention and its dependence on male hormone status in Macaca fuscata. Anim. Behav. 94, 61–71.

Polosecki, P., Moeller, S., Schweers, N., Romanski, L.M., Tsao, D.Y., Freiwald, W.A., 2013. Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. J. Neurosci. 33, 11768–11773.

Pong, M., Fuchs, A.F., 2000. Characteristics of the pupillary light reflex in the macaque monkey: metrics. J. Neurophysiol. 84, 953–963.

Putrino, D., Wong, Y.T., Weiss, A., Pesaran, B., 2015. A training platform for many-dimensional prosthetic devices using a virtual reality environment. J. Neurosci. Methods 244, 68–77.

Qiao-Grider, Y., Hung, L.-F., C-s, Kee, Ramamirtham, R., Smith Iii, E.L., 2007. Normal ocular development in young rhesus monkeys (Macaca mulatta). Vision Res. 47, 1424–1444.

Quiroga, R.Q., Nadasdy, Z., Ben-Shaul, Y., 2004. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. 16, 1661–1687.

Russ, B.E., Leopold, D.A., 2015. Functional MRI mapping of dynamic visual features during natural viewing in the macaque. Neuroimage 109, 84–94.

Sandbach, G., Zafeiriou, S., Pantic, M., Yin, L., 2012. Static and dynamic 3D facial expression recognition: a comprehensive survey. Image Vis. Comput. 30, 683–697.

Schultz, K.P., Williams, C.R., Busettini, C., 2010. Macaque pontine omnipause neurons play no direct role in the generation of eye blinks. J. Neurophysiol.

Sigala, R., Logothetis, N.K., Rainer, G., 2011. Own-species bias in the representations of monkey and human face categories in the primate temporal lobe. J. Neurophysiol. 105, 2740–2752.

Simpson, E.A., Nicolini, Y., Shetler, M., Suomi, S.J., Ferrari, P.F., Paukner, A., 2016. Experience-independent sex differences in newborn macaques: females are more social than males. Sci. Rep. 6, 19669.

Sirovich, L., Kirby, M., 1987. Low-dimensional procedure for the characterization of human faces. JOSA A 4, 519–524.

Stanley, D.A., Adolphs, R., 2013. Toward a neural basis for social behavior. Neuron 80, 816–826.

Steckenfinger, S.A., Ghazanfar, A.A., 2009. Monkey visual behavior falls into the uncanny valley. Proc. Natl. Acad. Sci pnas-0910063106.

Sugita, Y., 2008. Face perception in monkeys reared with no exposure to faces. Proc. Natl. Acad. Sci. 105, 394–398.

Surridge, A.K., Osorio, D., Mundy, N.I., 2003. Evolution and selection of trichromatic vision in primates. Trends Ecol. Evol. (Amst.) 18, 198–205.

Tazumi, T., Hori, E., Maior, R.S., Ono, T., Nishijo, H., 2010. Neural correlates to seen gaze-direction and head orientation in the macaque monkey amygdala. Neuroscience 169, 287–301.

Thies, J., Zollhöfer, M., Nießner, M., Valgaerts, L., Stamminger, M., Theobalt, C., 2015. Real-time expression transfer for facial reenactment. ACM Trans. Graph. 34 183-1.

Thomaz, C.E., Giraldi, G.A., 2010. A new ranking method for principal components analysis and its application to face image analysis. Image Vis. Comput. 28, 902–913.

Tilotta, F., Richard, F., Glaunès, J., Berar, M., Gey, S., Verdeille, S., Rozenholc, Y., Gaudy, J.-F., 2009. Construction and analysis of a head CT-scan database for craniofacial reconstruction. Forensic Sci. Int. 191 112-e1.

Vakkur, G.J., 1967. Studies on Optics and Neurophysiology of Vision. University of Sydney.

Vlasic D., Brand M., Pfister H., and Popovic J., Face transfer with multilinear models. In *ACM SIGGRAPH 2006 Courses (SIGGRAPH '06)*. ACM, New York, NY, USA, Article 24.

Waitt, C., Gerald, M.S., Little, A.C., Kraiselburd, E., 2006. Selective attention toward female secondary sexual color in male rhesus macaques. Am. J. Primatol. 68, 738–744.

Waitt, C., Little, A.C., Wolfensohn, S., Honess, P., Brown, A.P., Buchanan-Smith, H.M., Perrett, D.I., 2003. Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. Proc. R. Soc. Lond., B, Biol. Sci. 270 S144-S6.

Yu, H., Garrod, O.G., Schyns, P.G., 2012. Perception-driven facial expression synthesis. Comput. Graph. 36, 152–162.

Zhang, P., Watanabe, K., 2007. Preliminary study on eye colour in Japanese macaques (Macaca fuscata) in their natural habitat. Primates 48, 122–129.