

2016

A Retro-Projected Robotic Head for Social Human-Robot Interaction

Delaunay, Frederic C.

<http://hdl.handle.net/10026.1/4871>

<http://dx.doi.org/10.24382/1436>

Plymouth University

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

A RETRO-PROJECTED ROBOTIC HEAD FOR
SOCIAL HUMAN-ROBOT INTERACTION

by
Frédéric Delaunay

Centre for Robotics and Neural Systems

University of Plymouth

frederic.delaunay@plymouth.ac.uk

in partial fulfillment for the degree of

Doctor Of Philosophy

Friday, the 30th of October 2015

Frédéric Claude Delaunay,

A Rear-Projected Robotic Head for Social Human-Robot Interaction

—

Abstract

As people respond strongly to faces and facial features, both consciously and subconsciously, faces are an essential aspect of social robots. Robotic faces and heads until recently belonged to one of the following categories: virtual, mechatronic or animatronic. As an original contribution to the field of human-robot interaction, I present the R-PAF technology (Retro-Projected Animated Faces): a novel robotic head displaying a real-time, computer-rendered face, retro-projected from within the head volume onto a mask, as well as its driving software designed with openness and portability to other hybrid robotic platforms in mind.

The work constitutes the first implementation of a non-planar mask suitable for social human-robot interaction, comprising key elements of social interaction such as precise gaze direction control, facial expressions and blushing, and the first demonstration of an interactive video-animated facial mask mounted on a 5-axis robotic arm. The LightHead robot, a R-PAF demonstrator and experimental platform, has demonstrated robustness both in extended controlled and uncontrolled settings. The iterative hardware and facial design, details of the three-layered software architecture and tools, the implementation of life-like facial behaviours, as well as improvements in social-emotional robotic communication are reported. Furthermore, a series of evaluations present the first study on human performance in reading robotic gaze and another first on user's ethnic preference towards a robot face.

Contents

0.1	Acknowledgements	15
0.2	Declaration	17
1	Introduction	20
1.1	Open Issues and Limitations	20
1.1.1	User Expectations and the Uncanny Valley	21
1.1.2	Practical Concerns in Mechatronics	23
1.1.3	Other Limitations	27
1.2	Challenges Addressed in This Work	29
1.2.1	Objectives	29
1.2.2	Contributions to Knowledge	29
1.2.3	Plan for Work	30
2	Background	32
2.1	Being Social	38
2.1.1	Evolutionary Perspective	39
2.1.2	Social Robots for a Social Species	43
2.2	Towards Natural Human-Robot Interaction	46
2.2.1	Non-Verbal Communication	49
2.2.2	Reliance on Faces	51
2.2.3	Robotic Facial Guidance	53
2.2.4	Robotic Head Technologies	57

3	LightHead, a Social-Emotional Robot	63
3.1	Motivations for Innovation	63
3.1.1	Socially Guided Machine Learning and Non-verbal HRI	65
3.1.2	Overcoming Other Technologies' Limitations	66
3.2	Retro-Projected Robotic Faces	67
3.2.1	Background	67
3.2.2	The Mask	68
3.2.3	The Projection	71
3.2.4	Benefits of Computer Generated Imagery	74
3.3	Expectation-Driven Embodiment	81
3.3.1	The Head	81
3.3.2	The Spine	84
3.4	Control	87
3.4.1	Existing Systems	87
3.4.2	Overview	89
3.4.3	ARAS	90
3.4.4	CHLAS:	98
3.4.5	HMS	102
4	Measuring Eye Gaze Readability	108
4.1	Experimental Protocol	111
4.2	Results	114
4.3	Discussion	117
5	Robotic Social Influence in Human Tutelage	120
5.1	Learning Mechanism	121
5.2	Experimental Protocol	122
5.2.1	Simulated Experiment	122
5.2.2	Robotic Experiment	124
5.3	Results	127

5.3.1	Impact of Active Learning	129
5.3.2	Other Aspects	131
5.4	Conclusion	133
6	Influence of Robot Ethnicity	135
6.1	Experimental Protocol	137
6.1.1	Targeted Participants	137
6.1.2	Stimuli	137
6.1.3	Survey Platform	138
6.1.4	Questionnaire	139
6.1.5	Outlier Removal	142
6.2	Results	143
6.2.1	Inter-Ethnicity Analysis	145
6.2.2	Analysis of Interaction Effects	147
6.2.3	Analysis of Personality Test	148
6.2.4	Semantic Differential Analysis	149
6.3	Discussion	153
6.3.1	Online Survey Platform Issues	153
6.3.2	Conclusion	154
7	Discussion and Conclusion	156
7.1	Renewed State of the Art	156
7.1.1	Improvements over Mechatronics	156
7.1.2	Refined Human-Robot Interaction	158
7.1.3	Limitations of Retro-Projected Animated Faces	160
7.1.4	Summary	161
7.2	Opportunities for a New Technology	162
7.2.1	Industrial Aspects	162
7.2.2	Research Aspects	165
7.3	Impact and Follow-up Studies	167

7.3.1	Collaboration within the University	167
7.3.2	Collaboration with Externals	168
7.3.3	Related Subsequent Works	169
7.3.4	Spin-off and Patent	172
7.3.5	Insights Gained from Outreach Events	173
7.4	Future Work	177
7.4.1	Long Term Interaction	177
7.4.2	Holistic Affective Models	177
7.4.3	Delineating Models' Transferability	178
A	Schematics of LightHead Version 4	181
B	CHLAS Documentation	183
C	Active Learning Experiment	204
D	Ethnic Preferences Experiment	209
	References	217

List of Tables

3.1	Design iterations of the LightHead.	77
3.2	A vector of the internal matrix constituting the pool of 63 Action Units. Allows for proprioceptive information through polling the current value.	93
4.1	Example of an experimental sequence for a pair of participants. P1 and P2 swapped their seats and repeated the sequence.	112
4.2	Statistical difference in mean performance of different display types	115
4.3	Statistical difference tests for the four different displays between the two different angles. The difference between viewing angles for dome, mask and human is significant, and for flat it is not.	116
4.4	Participants' mean euclidian distance and angular errors in gaze reading from 1.5m, for both viewpoints and each condition. N=12, eyes-to-object=0.5m.	119
5.1	Overview of the occurrences of ambiguities and subsequent confusions over all experimental sessions (50 rounds, 39 participants).	128
6.1	Distribution of respondents across all ethnic groups (N=87).	145

6.2	Counts of favourite robot version against respondents' ethnic group (N=87).	146
6.3	One-way ANOVA for each ethnic version of the robot stimuli (N=78).	147
6.4	Two-way ANOVA (interaction between country and participant's ethnicity) for each ethnic version of the robot's monologue (N=78).	148
6.5	Three-way ANOVA (interaction between gender, age group and participant's ethnicity) for each ethnic version of the robot's monologue (N=78).	149
6.6	Questionnaire items and their contribution to 1st factor extracted with PCA for exploratory factor analysis (N=78). . .	152
6.7	Most contributing questionnaire's items to factors 2 to 5, extracted with PCA for exploratory factor analysis (N=78). . .	153
7.1	Comparative overview of established robotic head technologies against R-PAF heads.	163
7.2	Distribution of the 111 collected open comments collected from the museum's visitors over 4 days (N=230). Some comments belong to more than 1 category.	174
C.1	LightHead's utterances in active learning condition.	204
C.2	Detail of the participants' game success and alignment for both active learning and baseline conditions.	205

List of Figures

1.1	The Uncanny Valley, illustrated with examples and taking into account the effects of movements on perception of familiarity (from (MacDorman, 2005)).	22
2.1	Some examples of contemporary robot heads (from left to right): an avatar displayed on screen <i>GRACE</i> (Gockley, Simmons, Wang, Busquets, & DiSalvo, 2004), the <i>iCub</i> mechatronic and LED head (Beira et al., 2006), the <i>Nexi MDS</i> mechatronic robot head (Breazeal et al., 2008), the <i>Robothespian</i> head with mobile phone screens and animated jaw by Engineered Arts Ltd. and the <i>Actroid DER3</i> android robot head by Osaka University and Kokoro Company Ltd.	58
3.1	The LightHead robot, the fourth and last version. See all versions in table 3.1	64
3.2	a & b) front projection, one of the Disney’s Haunted Mansion singing busts – from (Mine, van Baar, Grundhofer, Rose, & Yang, 2012). c) retro-projection, Hashimoto’s Kamin-FA1 robot – from (M. Hashimoto & Morooka, 2005).	68

3.3	Left: mould and mask. The mould requires sanding to smooth the layers still visible and drilling in the ridged areas (e.g. eye sockets); this prevents trapping air pockets so the vacuum process correctly shapes these areas. Right: foldings can appear if the temperature for vacuum forming is too high or the plastic too thin.	70
3.4	The LightHead’s parts: (a) left view, (b) top view, (c) perspective. For clarity, some parts are only outlined in (a) and (b). Parts 1 to 11: laser-cut PETG (see appendix A), 12: moulded HIPS mask, 14: moulded HIPS cover, 14: tip of KatanaHD400s-6M robot arm, A: Microsoft Lifecam Cinema, B: fisheye lens Nikon FC-E8, C: electret microphones, D: Optoma PK301.	73
3.5	An attempt at simulating crying with LightHead’s virtual face. This effect was not exploited in experiments.	78
3.6	Augmented expression through textual information with the LightHead’s virtual face. This effect was not exploited in experiments.	81
3.7	The KatanaHD400s-6M kinematic chain mapped to the spine of LightHead (angle ranges in degrees).	84
3.8	LightHead’s software architecture, splitting animation, reactive and affective control, and cognition of the robot. From bottom to top: Abstract Robotic Animation System (ARAS), Character Hi-Level Animation System (CHLAS), High-level Management System (HMS).	89
3.9	Summary of the software designed and implemented during this thesis.	91

3.10	Five of Ekman’s Six Basic Facial expressions with the 1st design of LightHead: happy, disgusted, surprised, frightened and angry.	94
3.11	Computation of eye (center in E) orientation (Θ_z) from focal point F and eyes mid-distance M.	100
3.12	A behaviour implemented with two event-based parallel state machines: cognition in blue, face-following in red.	104
4.1	The four facial interfaces providing gaze sequences	109
4.2	Wollaston’s effect applied to the Mona-Lisa: the faces appear to gaze at different location although the pairs of eyes are identical.	111
4.3	Experimental setup: pairs of participants seated viewing the four displays straight and at 45°.	113
4.4	Mean Error (cm) in gaze reading for each display for a distance of 2m with front and 45° seating positions (N=12) . . .	114
4.5	Mean of user preferences for each display (N=12).	116
5.1	Comparative overview of the success in communication of the baseline and active learning strategies (AL) of 50 language games. Values are averaged over 50 complete simulations. . .	123
5.2	Experimental setup: the participant faced the LightHead robot; both shared the context procured by the touch-screen.	125
5.3	The GUI which presented the exemplars and words to the participants.	126
5.4	Comparative overview of the success in communication of 50 language games. Values are averaged over 50 complete simulations (N=39).	128
5.5	Distribution of tutors’ alignment with LightHead’s cues against game success for AL and non-AL conditions.	130

5.6	Influence of LightHead’s learning behaviour on guessing game success, split by gender and learning condition.	132
6.1	The various skin designs used for the ethnic preference study. Each stereotyped ethnic group was implemented as a “skin” overlay for the robot. From left to right: White Caucasian, Black-African, Middle-Eastern, North East Asian and Alien (control).	138
6.2	Mean ranking and SD for each ethnic version of the robotic monologue (N=87).	144
6.3	Distribution of rankings for the Alien design (N=87).	144
6.4	Mean Big5 profiles. Top: by ethnic group, bottom: by country (N=78).	150
6.5	Scree plot of the exploratory factor analysis (N=78). Inflection point appears at the fifth factor.	151
7.1	Top-left: Mask-bot (adapted from (Pierce et al., 2012) and (Kuratate, Matsusaka, Pierce, & Cheng, 2011)), top-right: Furhat (permission from Al Moubayed), bottom-left: Hoque’s mask (adapted from (Hoque, Onuki, Kobayashi, & Kuno, 2011)), bottom-center: a reduced scale face by Misawa (adapted from (Misawa, Ishiguro, & Rekimoto, 2012b)), bottom-right: HALA (adapted from (Fanaswala, Browning, & Sakr, 2011)).	170
7.2	The Lighty prototype as commercialized by the spin-off Synthelligence until 2015 ⁷ . Projected face, form-factor and some materials have been updated compared to LightHead v4.	172

A.1 Laser-cut parts of the LightHead's chassis as in version 4 (*l* and *r* suffixes refer to left or right editions of a part); all parts are 3mm thick PETG except # 6 which is 6mm thick. 1: front frame, 2: side panels also housing microphones, 3: lens side-grippers, 4: main lens holder, 5: grippers and PK301 bridge, 6: main base, 7: KatanaHD400s-6M adapter, 8 & 9: cables holder, 10: back frame, 11: frames bridges. 182

0.1 Acknowledgements

Before diving in the depth of the most challenging essay I've come to write, I wish to express personal as well as less directed gratitude.

With my personal award of the most social and stimulating supervisor, I wish to thank you Tony for letting me nurture my curiosity. These four years witnessed your rise from lecturer to professorship, and from supervisor to respected friend.

Then of course, Joachim de Greeff, accomplice and saviour of critical situations, talented sophist, mediator, and social gatherer. I owe you much, thanks for your friendship. To you I wish to find the position that could provide as much long-term satisfaction as possible, as I believe the rest is already covered.

To Vadim Tikhanov and Zoran Macura, I can't imagine who could provide better introduction to the life of a research student. Greatest feelings to the Schindlers for being the most friendly British family I have come to know (special touring award for Robert). Thanks to Martin Peniak for our stimulating philosophical discussions, intense parties or both simultaneously. Similar thoughts to Alex Smith for the various entertaining ideas on robotics and beyond! Indeed the Alien House, whose dean Musaab Garghouthi has become worldwide famous for his impromptu barbecues. Also thanks to Thomas Roc and Frédéric Verret, two remarkable and unexpected comrades...

Already omitting too many others, and to prevent unforeseen frustrations, I wish to express many thanks to my reviewers, other uncited friends and acquaintances part of my life in the South West. You know why we enjoyed spending time with each other, so come back to me to carry on if life allows us to.

Acknowledgements to the CRNS fellows and Plymouth University staff,

Davide Marocco & Elena Dell' Aquila, Guido Bugmann, your support has been deeply appreciated. Carole Watson and Lucy Cheetham, your care to minimise the impact of academic bureaucracy on my life has literally been legendary. Also, a mention specially dedicated to the supporting people: Martin Woolner, David Scott, Gregory Nash. Your attitude just stands out.

Then, within the key academics who opened to me the gates of research, I particularly wish to thank in chronological order Jean-Luc Zarader and Pierre-Yves Oudeyer. I think it's fair to say I would never have enjoyed the pleasure of indulging passionately in the field of robotics without your trust.

Also responsible for my success, my remote friends and family. As being part of my private life, my acknowledgements to you all are more personal and go beyond this thesis.

Finally, I feel the need to thank all of those who foolishly tried to hinder my efforts in reaching the goals I have come to achieve: you have only confirmed abnegation is no match to self-determination.

0.2 Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This work was supported by the CONCEPT project (EPSRC EP/G008353/1) and the EU FP7 ITALK project.

Publications and conference presentations :

F. Delaunay and T. Belpaeme (2012). Refined human-robot interaction through retro-projected robotic heads. In *IEEE Workshop on Advanced Robotics and its Social Impacts*. Munich, Germany.

J. de Greeff, F. Delaunay, and T. Belpaeme (2012). Active robot learning with human tutelage. In *Proceedings of the joint International Conference on Developmental Learning (ICDL) & Epigenetic Robotics*. San Diego, USA.

F. Delaunay, J. de Greeff, and T. Belpaeme (2010). A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *HRI 2010*. Osaka, Japan.

J. de Greeff, F. Delaunay, and T. Belpaeme (2010). Socially guided machine learning for conceptual knowledge on robots: an overview of the CONCEPT project. In *The Postgraduate Conference for Computing: Applications and Theory*. Exeter, UK.

F. Delaunay, J. De Greeff, and T. Belpaeme (2009). Towards Retro-

projected Robot Faces: an Alternative to Mechatronic and Android Faces. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Toyama, Japan.

J. de Greeff, F. Delaunay, and T. Belpaeme (2009). Concept acquisition through linguistic human-robot interaction. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Toyama, Japan.

J. de Greeff, F. Delaunay, and T. Belpaeme (2009). Human-Robot Interaction in Concept Acquisition: a computational model. In *IEEE 8th International Conference on Development and Learning* (pp. 1-6). Shanghai, China.

P.Y. Oudeyer, and F. Delaunay (2008). Developmental exploration in the cultural evolution of lexical conventions. In *8th International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. Falmer, UK.

Posters :

Ninth International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Retro-projected faces effectiveness on gaze reading, Venice, Italy, November 12-14 2009.

CNRS Inauguration, A Bright Cost-Effective Head for Robots: Improving HRI with Back-Projected Facial Masks, Plymouth, UK, 2009.

Word count for the main body of this thesis: 41622

Date: *Friday, the 30th of October 2015*

Signed:

A handwritten signature in blue ink, consisting of several loops and a horizontal line, positioned to the right of the word "Signed:".

Chapter 1

Introduction

As people respond strongly to faces and facial features, both consciously and subconsciously, faces are an essential aspect of social robots. Robotic faces and heads until recently belonged to one of the following categories: virtual, mechatronic or animatronic. Natural human communication is necessary to the diffusion of social humanoid robots, however it appears the current state of mechatronics suffers from limitations that limit efforts in this area.

1.1 Open Issues and Limitations

Despite the solutions available to social robot designers, no particular robot technology can currently claim full user satisfaction, and perhaps this may never happen. People are notoriously difficult to please, and not only are aesthetic preferences towards robots comparable with other consumer oriented products, tastes also differ with robotic technologies. This section summarizes the open challenges currently faced by social robotics researchers and companies designing robot heads.

1.1.1 User Expectations and the Uncanny Valley

Robot head and face design has a profound effect on our relation with respect to a robot. The high-dimensional design space, with a wide choice of techniques, materials and aesthetics, makes it difficult to evaluate the impact of every aspect during lab evaluations and eventually in real-life scenarios. Extensively exploring the influences at play in robotic designs, MacDorman gathered in (MacDorman, 2005) a number of important insights. Noticeably, his work encompasses the effects of mixing aesthetic designs leading to the notion of character coherence. How many trial and error cycles are needed to yield a full understanding of what makes a successful design is probably not a relevant question to ask: in effect every new technology brings new possibilities. Moreover, each new generation of users is exposed to an ever increasing amount of technology, which in turn changes expectations and designs.

However focusing on a robot's affordances, attempts to define a principled understanding of designing social humanoids emphasizes the importance of facial features whether they are explicit ('designed') or merely suggested. In a study by DiSalvo et al. (DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002) participants shown pictures of robots rated the eyes and mouth as most significant for social interaction. Users expect humanoid robots to interact primarily through vision and auditory channels, without which interaction would be drastically impoverished.

Less obvious is the presence of non-explicitly intended features and how typically human they look. In this same study, DiSalvo et al. theorize that the closer the face and its animation resembles that of a human face, the closer the interaction edges towards uncanniness. However, it is important to note that users' expectations do not necessarily coincide with a robot's features degree of realism. Robot design should advertise particular functionalities, but strictly copying a human appearance only advertises the

potential for the robot to have human-level performance. As such, it is often better to steer clear of constructing a human simulacrum.

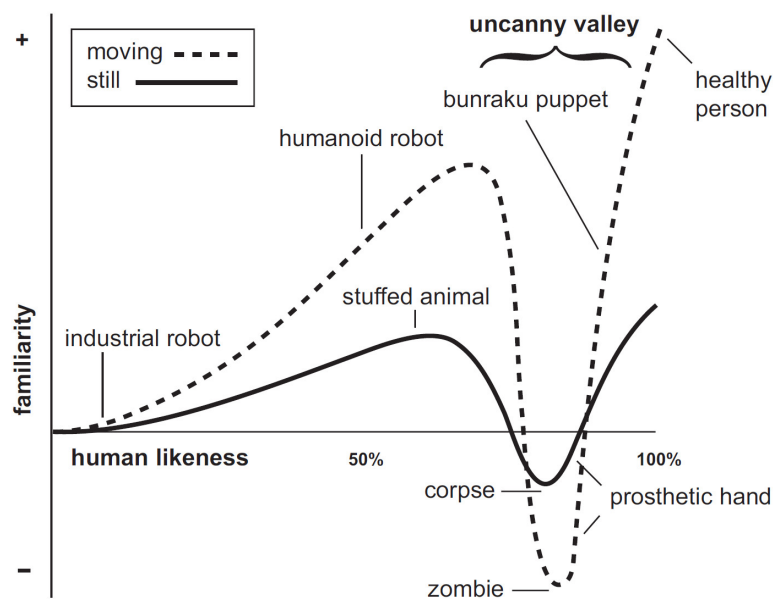


Figure 1.1: The Uncanny Valley, illustrated with examples and taking into account the effects of movements on perception of familiarity (from (MacDorman, 2005)).

The notion of artificial character uncanniness that Mori coined as the *uncanny valley* (Mori, 1970), persists as a contentious topic for human robot interaction. In short, Mori argued that as the design of robots gradually progresses towards robots which more closely resemble humans, at a certain point an “uncanny valley“ is encountered in which humans perceive the robot as not familiar at all. Conducting many studies on the topic, McDorman notes how controversial this question actually is and in a later study with Matsui (Matsui, Minato, MacDorman, & Ishiguro, 2005), applies human movement to an android in an effort to delineate the issue. Perhaps our innate cognitive processes involved in detection of faces and various forms of

sensory prediction, eventually eliciting empathy, are disturbed by incongruent signals, leading to the unease reported as the uncanny valley. Androids, and especially their movement, have not yet reached a level of performance which is perceived as convincing by humans, and thus feelings of uncanniness can halt the progress made on that front. In (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2011), functional resonance imaging used on the brain area responsible for action perception and prediction shown stronger suppression effects on the twenty participants watching a video of an android. Maybe an inventory of the neurological processes at play will eventually be constituted as new studies exploiting brain imaging continue to contribute to the understanding of perception of uncanniness in humans. Whereas the uncanny valley embodies the recognition of the problem, McDorman and other researchers challenge its simplistic definition (see Bartneck et al in (Bartneck, Kanda, Ishiguro, & Hagita, 2007) for a dedicated paper) and this thesis would also promote a critical perspective on this issue.

In general, it can be argued that for a robot to be successful in its interaction with humans, its mode of presentation needs to be clear: either it is presented as a non-human character which makes no pretence of being human-like, or the robot does resemble humans closely enough (in all relevant aspects) to overcome the uncanny valley. As the latter type has not yet been achieved, and indeed may take a while longer, it can be argued that an approach in which a design stays coherent, with a robot not trying to be as human as possible (such as the Nao for instance), is more feasible to achieve effective human-robot interactions. This is of course not to say that the type of interactions should be limited because of this design choice.

1.1.2 Practical Concerns in Mechatronics

Building a mechanically animated face requires a considerable amount of resources and the technologies involved have a number of drawbacks which

are discussed below along with possible alternatives.

Mechanical Complexity

Currently, even if novel technologies such as dielectric polymer-based muscles or memory shape alloys reached niche applications, servos and pneumatics still hold robot designers' preference for actuation; alas, these cannot yet match the power yield or density of their biological counterparts. By requiring extra pumps and lacking precision, pneumatics are unfit for facial animation, hence servos actuate conventional robot facial features as well as android's skin underlying structures. Many times the size of its relative facial muscle, a servo can only occupy space behind the facial bones and transmit the mechanical force to a specific facial feature. Therefore only a certain number of servos can be housed in the head volume, each one adding weight and noise, reducing space, complicating the mechanical design and eventually requiring adequate cooling.

Power Requirements

Not all actuation techniques drain the same amount of power. As faster or stronger servos drain more, they also release more heat, up to the point where active cooling is required, adding to the power consumption. Control boards, active sensors and lights also require their share of electric current. Adding to the weight of servos, the chassis, sensors, electronics, cabling and plastics stress the smallest – but also weakest – servos, and often the solution is to opt for more robust ones that drain more power. In that regard, virtual characters do not suffer these problems: the power consumption of their displays stays almost constant and heat can dissipate freely without a cover.

Costs

The higher price of mechatronic designs finds many justifications. Although robots are often publicized, the social robotics market has yet to mature and become mainstream enough for the industry to mass produce cheaper robotic components, and hopefully compatible whole robotic parts; hence most robots continue to bear custom designs, with each innovation raising costs. Fortunately, with the 3D printer market taking-off, a vision where even the most complicated mechanical design becomes affordable to small budgets is becoming a reality (see for instance (Griffey, 2012)), however 3D printing is no comprehensive fix. One can reasonably think servos and components would eventually become cheaper as automation becomes ubiquitous, however it seems unlikely that a specific android face would become standard: we easily accept having the same car as our neighbour but it may be confusing to see our personal android's face anywhere else but home. Today though, ultimately the fewer the number of sensors, actuators and other components that enter the conception of a robotic face, the cheaper it becomes to produce and maintain.

Motion

Typical actuator motion is not convincing: slow, linear and/or jerky actuations poorly simulate human facial muscles which fully stretch in a only a few hundred milliseconds displaying smooth, non-linear dynamics. Trying to achieve these sort of motions with conventional actuators is an interesting mechanical and control challenge. As human facial expressions change in a very short amount of time, video is more suited than mechanical servo mechanisms, with the smoothness of the animation only constrained by the video frame rate.

Expressiveness

Non-android faces are often composed of solid parts for each feature (e.g.: eyebrows and eyelids) and sometimes a flexible mouth such as FloBi's. These visible features essentially follow a linear movement in contrast to androids' flexible skin, and combining multiple actuators to address this issue is rarely practical. A limited number of actuators restricts the number of facial expressions available in a robot's repertoire, and to our knowledge, only the Nexi MDS houses as many as 15 DOF. As android faces emphasize realism, one would expect their faces to be very expressive, however attempts to provide android faces with more DOF (for instance Albert HUBO (Oh et al., 2006) has 28 facial DOF) have yielded mixed results. In fact, an android's face is subject to the same servo housing issues as other robots, consequently latest androids manage facial expression with roughly the same number of DOF as most expressive non-android faces (for instance Actroid F has 12 DOF). However, this implies compromises and such androids fail to create realistic mouth deformations, a visual effect most noticeable while they talk.

On the other hand, virtual faces experience none of these issues: their expressiveness is only limited by the quality of the 3D model and its animation, and the resolution of the device displaying them.

Uncanniness

Androids are more prone to uncanniness. As their physical appearance is closest to humans, actuator capabilities and skin interaction is still an issue with current technology. This is most obvious when androids engage in fast gestures, as precise control of momentum remains problematic in robotics. Also the jerkiness some androids experience may unconsciously remind us of motor diseases not uncommon in humans (such as Parkinson's), typically unmasking the failed imitation. Mouth animation, and lip-synchronization in particular, fail to look realistic and contrast with the other achievements.

Moreover, noise originating from actuators participate in the general uncanniness, even more present during face to face interaction. This effect seems less important in non-realistic robots, and perhaps this carries a cultural bias: films are notorious for associating non-android robots with servo noises, while no such sound effects were present with Terminator *T-800*, StarTrek's *Data*, etcetera, until their robotic nature needs to be emphasised. Other auditory effects take part in the uncanniness experienced with robots. While constant, fan noises may be filtered out by the brain in the long run, but a fan changing speed unexpectedly can also generate distraction. Adding to the uncanny effect, the problem with loudspeaker placement emerges either when the speech source is distinctively not localised in the mouth, or when placement in the robot head modifies speech spectral profile (e.g: loss of high frequencies or resonance).

Despite the progress made, shortcomings in the illusion of life in virtual characters still attracts the attention of non-expert people. For the illusion to work best, many aspects of physics must be reproduced: simulating all light behaviour passing through skin and other materials is notoriously difficult to achieve in real-time, believable collision management require complex engines, and hair, fluids, clothing and skin foldings also rely on specific algorithms to appear realistic. Nonetheless, CGI steadily pursues its way towards photo-realistic quality, and we may witness photo-realistic real-time character animation in the very near future.

1.1.3 Other Limitations

Avatar Flatness

While an avatar displayed on a screen satisfies some robotic projects, their strongest limitation comes from the Mona Lisa effect investigated by Rogers et al. (Rogers, Lunsford, Strother, & Kubovy, 2003) or Maurer (Maurer, 1985). Essentially, portrayed eyes always appear with a gaze direction rel-

ative to the viewer himself, regardless of his viewpoint. Consequently two viewers do not see the same object being gazed at, and more critically, a moving viewer does not perceive a fixated gaze. In addition, the lack of facial three-dimensional geometry to advertise social and natural interaction abilities often negatively impacts on the interaction, and while one could argue that 3D monitors exist, their price remains a prohibitive constraint.

Behavioural Limitations

Even if speech synthesis has known widespread adoption over the recent couple of years – most notably on “smart phones” – these speech models typically sound monotonic without proper prosody, although interesting proposals exist (see (Scherer, 2003) for a review). Alternatively, better human speech production is possible and attractive low-dimensional models have been proposed by Nicolao and Moore in (Nicolao & Moore, 2012). The most convenient method to tackle multi-modal congruence of emotional content may be through analysis of human behaviour, however the very definition of emotion and its principled model have not yet reached definite consensus amongst researchers. Therefore, avoiding patterned interactions and achieving believable robotic behaviour on the long term may elude the community for a while.

Some of the facial expressions used to convey social signals are culturally universal: they are produced and understood by all. Ekman and Friesen (Ekman & Friesen, 1969) accumulated evidence showing that six basic facial expressions are inter-cultural: joy, sadness, fear, disgust, anger and surprise; however many others are culture-specific. Indeed, this is also the case for head movements (e.g. the signalling of agreement or attention employs nodding in Western cultures, instead many South Asian cultures as well as Bulgarians use head wobbles). As such, it would be desirable for a robotic face and its behaviour to adapt to cultural settings.

1.2 Challenges Addressed in This Work

Consequently, a novel approach to naturally interactive robotic faces is much needed and that is what the work reported in this thesis addresses.

1.2.1 Objectives

The objectives of this research program were twofold: create a robotic platform exploiting non-verbal HRI to elicit natural social-emotional communication, and employ this robot for socially guided acquisition and teaching of knowledge (de Greeff, Delaunay, & Belpaeme, 2009; de Greeff & Belpaeme, 2011).

To this end, a major milestone was the creation of an independent non-mobile anthropomorphic robotic agent, comprised of a iCub-based robotic head displaying a computer-animated face, and a standalone industrial robotic arm offering less than 500g of payload. This required the design and manufacturing of parts, writing controlling software and interactive behaviours such as joint attention and turn taking so as to demonstrate capability to engage in human-robot interaction. The robot's knowledge and active learning system were to be evaluated in a robot-tutelage scenario in which the robot could crane over objects laid out before it.

1.2.2 Contributions to Knowledge

Therefore, I present the R-PAF technology (Retro-Projected Animated Faces, also known as RAF in my previous publications): a robotic head displaying a real-time, computer-rendered face, retro-projected from within the head volume onto a mask, as well as its driving software designed with openness and portability to other hybrid robotic platforms in mind.

Contributions to knowledge exposed in this thesis include the design of

the R-PAF technology through the creation of the LightHead interactive robot, the study of non-planar robotic facial displays, their measured eye gaze readability by human viewers, and the demonstrated effectiveness of a R-PAF robot to engage in and exploit social human-robot interaction. Overcoming important limitations of mechatronic heads, the R-PAF technology represents an alternative development to mechatronic and flat-screen based robotic heads, allowing a wide range of customisations and multiple cost savings.

Retro-projected faces not only offer a refreshing take on robotic head technologies, but also provide a great potential as a research platform for HRI and human focused fields. As reported in section 7.3.3, several scholars followed the path laid out in this research with R-PAF heads, exploring realistic faces, new social capabilities with conversational agents, social cueing and small form factor faces.

1.2.3 Plan for Work

This chapter discussed current issues with approaches to facial animation for social robotics and framed the work reported in this thesis. Next, chapter 2 brings forward the rationale behind the need for social interaction with current and future humanoid robots, and reviews the related state of the art in the field Human-Robot Interaction. Chapter 3 explores the design challenges emerging in social robotics and details the solutions developed in this work. Chapter 4 looks at the question of robotic eye gaze readability through a first novel experiment and provides measured results of human performance with the LightHead robot supporting the robot's ability to provide precise gaze from both front and side-facing users. Chapter 5 investigates the effectiveness of robotic social cueing in a robot-tutelage experiment demonstrating benefits in active learning. Chapter 6 addresses the issue of user-robot ethnic alignment in a second novel experiment using

the crowd-sourcing platform CrowdFlower and provides insights over local versus individual ethnic alignment of a robot's facial appearance. Finally, chapter 7 summarises the main results of the research, provides a comparative analysis with anthropomorphic social robots equipped with either a flat-screen or mechatronic head, itemises the original scientific and engineering contributions in this work, and mentions industrial impact.

Chapter 2

Background

Robotics has a rich and dynamic legacy rooted in the belief that machines can relieve humans from labour, and to a greater extent, be endowed with enough intelligence to take a crucial role in human societies. While the term and depiction of a robot in R.U.R by Karel Čapek dates back to 1920, the field of robotics really started with the development of cybernetics (Norbert Wiener, 1948): the study of the structure of regulatory systems. Although laying the foundations for machine control and automation, it was necessary for research in cybernetics to acknowledge the difficulty of general purpose problem solving and to focus on domain-specific issues more suited for classical analytical methods. With subsequent improvements in control and automation, a number of precisely framed problems found robotic solutions. Consequently, the industry mainly driven by the need for more robust and faster production in factories, successfully introduced programmable robots – typically a manipulating arm – solely relying on position to accomplish their task, and thus with far less capabilities than Čapek’s robots.

Towards the end of the 20th century, availability of multiple types of sensors and advances in their integration allowed relaxing strong structural constraints on robotic operating environments. Thanks to combined scientific and technological advancements, along with progress in computing, a

new breed of machines (dubbed *service robots*) could then sense their environment, react to changes and gain mobility: key properties to initiate the market of robotics we know today. Each generation of service robots came with the ability to handle a wider range of practical problems such as quality control, optimised logistics, continuous automated surveillance and many other tasks in hazardous or sterile environments.

This trend is indeed still ongoing and nowadays one objective of research is to relax service robots' environmental constraints: ideally, they should adapt to their environment, in terms of mobility and manipulation, but also in terms of appropriate decision making. Landmark achievements are usually demonstrated by the industry releasing technological products with significant impact on our societies (e.g: telecommunications and mobile phones). Similarly, efforts to open service robots' operating environments recently yielded exciting applications such as autonomous vehicles (cars, UAVs, submarines, planet rovers, etc.), robotised experimental research (micro-organism selection for bio-genetics) and assisted surgery (compensation of patient pulse and surgeon's movement control).

Although service robotics is spurring tremendous changes in human activities, the nascent market of domestic robots is promised to have a stronger influence on people, as these robots' interaction with humans is more direct and potentially longer. Public service robots remain specialized to a specific task, thus their method of interaction can be kept minimal. On the other hand, domestic robots in general are designed to serve humans best, and as such seek natural integration into the family.

Motivations to pursue research and industrial efforts in the area of domestic robots are numerous, and backed up by many surveys. With a rare large number of participants, Arras & Cerqui (Arras & Cerqui, 2005) asked 14 questions on various aspects of robotics and their usefulness to 2042 visitors attending a Swiss exhibition held in 2002. When questioned: "Could

you imagine to live on a daily basis with robots which relieve you from certain tasks that are too laborous for you?”, 71% replied positively. Also, 83% of the polled reported they would accept a robot to help partially regain independence; a claim fitting a trend of ageing population and longer life expectancy in most developed countries.

A general public interest in domestic robots in particular is indeed undeniable, and unsurprisingly a few high-technological companies endeavoured to create the first general public domestic robot. The robot dog AIBO – the brainchild of Sony released in 1999 – is described as the first commercially successful domestic robot from the entertainment industry. Although this effort certainly satisfied its audience despite the robot’s expensive price, AIBO’s limited usefulness restricted adoption by a broader range of customers and eventually led to the discontinuation of its commercialisation. On the other hand, the Roomba is an autonomous vacuum cleaner (introduced in 2002 by iRobot) and was designed with a clearly defined purpose. With more than 8 million units reported sold in 2012, it is considered a sound commercial success. Nonetheless, a study of users’ feedback by Forlizzi & DiSalvo (Forlizzi & DiSalvo, 2006) suggests that even if this domestic robot satisfies customers, people expect more intelligence from it.

A robot with predefined and static knowledge is not pragmatic for unconstrained environments: it is hard to delineate all possible use-cases and environmental conditions, and the usefulness of a robot unable to adapt to new environments is very limited. However, one can assume that domestic robots will get more intelligent and expand the number of simple domestic tasks users would delegate them, fulfilling their purpose of freeing people from some simpler house-keeping chores. It is reasonable to predict this market will only grow and we can expect the presence of autonomous robots to be more common in homes. But then, projecting this trend in the future, two main problems arise: integration of domestic robots into human activ-

ities, and management of several task-specific robots. Currently domestic robots run their program without accounting for human presence, however blending autonomous behaviour with human activities raises a whole new set of challenges (collaboration and disturbance avoidance is a contextual problem). Management of these robots, if handled by users themselves, can be daunting for most, let alone the industrial challenge of robot interoperability. Opinions on the future of domestic robots diverge, however a popular idea also constantly promoted by the media and arts is that robots should be intelligent, and certainly much more than the current domestic robots.

The personal robot stands as an alternative to multiple task-specific domestic robots. This general purpose, anthropomorphic robot would also be capable of holding conversations, and finally match current users' expectations.

Integration of multi-purpose robots in human society has been a long standing goal and it is no surprise that research progressed towards a type of robot shaped after human physiognomy: the humanoid. For instance ASIMO (Honda Corporation, 2000) and QRIO (Sony, 2003), while not being of the same size, have bipedal locomotion, arms and hands, as well as vision and audition sensors fitted in the head. By contrast, current domestic robots are limited by their embodiment: a wheeled robot can hardly handle steps or use stairs, and the lack of an arm or hand prohibits any form of grasping. Humanoid robots can fully exploit our human environment and thus blend in at no cost, which is certainly why personal robots are physically designed in this manner. Eventually, these robots should eliminate specific modes of interaction, in direct opposition to the best computer interfaces where humans still have to adapt to machines; with personal robots, machines adapt to humans through their embodiment and cognitive abilities.

The push and desire for personal robots mainly comes from consumers'

expectations of domestic robots, which have been sampled by several surveys across diverse populations. Copleston and Bugmann (Bugmann, 2011) polled 442 subjects from five age groups with an open questionnaire (no predefined answer) on their hypothetical use of a personal domestic humanoid. The questionnaire proposed seven questions such as “You get up and get ready for your day, what will you ask your robot to do today?” and “You have booked two weeks holiday and plan to go away. What will you ask your robot to do while you are gone?”. From this study, the first three most popular categories extracted were “Housework”, “Food Preparation” and “Personal Service”. These categories encompass diverse tasks reported by the participants: for instance housework relates to tidying, cleaning (vacuuming, washing dishes, cleaning floors, baths, sinks and windows..), water plants, make beds and laundry. Indeed, no current domestic service robot can handle all these tasks (an obvious issue is the amount of tools required) and even if a very smart design would allow overcoming that limitation, the real problem is how such a robot could be tuned to fit the various specifics of each household. For this, a robot needs not only to learn from its environment but also from its owners (preferences, house policies and lifestyle), which cannot be achieved without a form of intelligence.

The need for smarter domestic robots is also supported by another study by Ray (Ray, Mondada, & Siegwart, 2008) which concluded people would prefer to interact with robots using speech. Most entertainment robots are provided with some degree of speech recognition mainly used as a way to instruct them with predefined commands. After exploring these robots’ capabilities, such a strongly framed, one-way communication loses the entertainment value of the interaction. Although experienced users can extend a robot’s behaviour through computer interfaces (ie. programming tools), ideally no programming would be required for that matter. Instead users would use natural speech and gestures with the robot to carry their in-

tention with the same level of understanding commonly present between humans. Considering that natural speech processing is well established in domain-specific applications such as call transfer, booking and GPS systems, one would expect their usage in robotics to be more commonplace. Unfortunately natural speech processing remains problematic, requiring a great deal of domain-specific knowledge and computation. Using natural speech in robotics and more so in open-domain personal robotics, requires symbol grounding: a comprehensive bi-directional relation of symbols (words) with an embodied experience that connect them with the environment (see Harnad (Harnad, 1990) for further reading on symbol grounding).

The field of robotics regroups many aspects (design, control, intelligence) in diverse sub-fields and school of thoughts: classical machine learning, bio-inspired and developmental robotics to name a few. Arguably, all these specialisations exist because there is yet no general purpose theory which covers all aspects of robotics. Although the current state-of-the-art in artificial intelligence prevents natural language interactions with robots, alternative forms of communication and interaction are possible. Inspired by human-to-human interactions, a growing body of studies from the field of Human-Robot Interaction (HRI) are revealing the effectiveness of non-verbal communication applied to robotics.

Human-Robot Interaction covers various aspects of robots interacting with humans. HRI studies interactions socially, psychologically, personally, ergonomically, etc – often considering several aspects at the same time. Such a broad definition is typical to a young, complex and dynamic field of study, with many interdisciplinary connections. Once again, deep philosophical questions about HRI – e.g. Can machines be considered conscious? Should they have rights? How such a powerful kind of agents would impact mankind? – were exposed almost a century ago in arts and literature (Čapek K., Asimov I.), and obviously we're not yet able to answer them.

However, these questions were formulated assuming humans and robots interact similarly and naturally. In fact, it took researchers to explore the characteristics and define the problems of natural embodied robotic interaction to shed light on the crucial importance of paralinguistic communication at the heart of our social and emotional interactions.

2.1 Being Social

Working in robotics presents many opportunities to discuss the fascination robots have on the public. Often, questions such as “Will robots replace humans?” or “Is it possible for robots to take over the world?” find their origin in popular culture, and however tempting it is to disregard them (considering how simple current robots are) these questions are too fundamental to be discredited without debate. In fact, this questioning is about ourselves and how we relate to technology in general, reformulating them: Can technology ever challenge our own complexity? How can we always be sure of a robot’s intentions? Evaluating human complexity is usually a matter of observable behaviour and maybe popular interest comes from the increasing ability of robots to imitate us. Asking whether robots would ever have a motive to overthrow humans reveals how acute our reliance is on interpreting each other’s intentions. Fundamentally, it is because we define ourselves as social beings that these questions matter to us; because we mutually relate to peers to evaluate ourselves, seek wisdom for decision making, and at a basic level, ensure survival. Integrating intelligent robots in our societies may have profound consequences over the social structure we live in, and regardless of the changes, this process will ultimately expand the understanding of our own kind.

2.1.1 Evolutionary Perspective

Usually, the term “social” brings to mind how humans connect to others, and only occasionally do we leave this human-centric conception. As a matter of fact, humans are not the only species that exhibit a social structure; many other mammals such as deers and meerkats, birds such as the great tit, insects such as ants and bees (and even micro organisms) establish strong bonds with other members of the community they belong to. The meaning, expression and context of these bonds constitutes a communication essential to the formation of a social group. Exploring these aspects from a human-human interaction standpoint unravels innate and acquired means of tacit mutual understanding which robots must tap into for genuine integration in any of our social environments.

Arguably, the mere act of sensing others is a form of passive one-way communication. A lot of information is passively disclosed: age, gender, lineage and overall health are given away by the body. However behaviour carries more: social status can be inferred from the number of followers and degree of dominant behaviour, an evasive movement reveals the presence of a threat, a gathering helps spotting a large resource. When environmental pressure is low (for instance the Amazonian forest has a stable climate with uninterrupted essential resources), evolutionary processes generally lead to more bio-diversification and specialised species. On such principle, it is conceivable that this one-way communication underwent an evolutionary shaping to become two-way, specialising individuals to analyse and react to each other’s behaviour. Assuming genetic transmission of fundamental behaviour (environmental and community fitness) to all members of a group, behaviours become the de-facto communication protocol. For instance, eye gaze, and the ability to use eye gaze to transmit intent and other forms of information, is essential for primates (Emery, 2000). Behavioural, non-verbal, communication is indeed an essential aspect of human-human communica-

tion, so much so that providing robots with the means to analyse behaviour promises abilities such as robust verbal communication, social context inference, and cultural awareness.

Occurring in most social species as a means to improve chances of survival, cooperation strongly motivates the consensual establishment of overt intentional behaviours. To shed light on the result of this epigenetic process, Bratman (Bratman, 1992) explored the preliminary mechanism of mutual responsiveness of intention. Cooperative behaviours can be innate, but for humans and some primates, cooperation is immediate and localised, relative to the completion of a task and reliant on the establishment of joint attention (see (Tomasello, 1999)). Cooperative species may have evolved intentional behaviour with co-occurring subtle non-verbal signals mediated by gaze, or facial expressions to optimise completion of tasks and quickly engage in key survival behaviours such as repelling a predator. No matter how sophisticated these conventions are, it is very likely that they would eventually constitute the basis of a social language. Luc Steels's language games (Steels, 1997) present a general theory on the emergence of a communicative consensus in a population of agents able to interact and adapt according to the success of communication; which in our case is the completion of the cooperative task itself. Oudeyer and Kaplan (P.-Y. Oudeyer & Kaplan, 2007) revealed how agents playing the language games maximise communicative robustness by exploiting and adapting their communication channels, further supporting the idea of evolution towards multi-modal social communication.

The field of evolutionary social psychology offers an insightful and generally appealing approach to explaining various aspects of our social behaviour. Covering this field's axes of study and all of the related key contributions in support of this thesis is obviously beyond the scope of this document; however a few particularly relevant ideas are mentioned in this section amongst

others from related fields. The curious reader wishing to familiarise himself with evolutionary social psychology can refer to Neuberg (Neuberg, Kenrick, & Schaller, 1998) and acquire more in-depth knowledge with the book by Barkow et al. (Barkow, Cosmides, & Tooby, 1992).

The Expression of the Emotions in Man and Animals (Darwin, 1872) brings a general understanding of the phylogeny of our facial features and limbs as a means of non-verbal communication. A striking example comes with the evolution of eyebrows: “the eyebrows are continually lowered and contracted to serve as a shade against a too strong light; and this is effected partly by the corrugators”. The activation of these muscles is linked to other stimuli potentially damaging for the eyes: wind, carried particles of dust or sand, and liquid projections. Perhaps our feeling of disgust and its associated facial expression are a legacy of this general reaction to disagreeable stimuli. The facial expression of anger also uses the corrugators to lower the eyebrows and reduce exposure of the eyes; arguably a mechanism to protect our most precious sense in the case of a fight. Regardless of the origin of these facial expressions, it is clear that our species evolved to expand the primary usefulness of the muscles controlling our two main facial organs’ area: eyes and mouth.

Similarly, free hands created a new social communication channel, most likely comprised of rough arm movements at first, and extended to subtler hand and finger movements over evolutionary processes. Moreover, many primitive gestures and postures (such as hugging, finger pointing, shrugging) acquired different meanings over the geographical expansion of our species, and along with other factors of differentiation, these localised consensus spurred culture. For instance, hugging in public, and more generally public physical contact, is considered rude by Asian Indians but understood as a display of friendship by Westerners. Trying to be thorough, a presentation of cultural diversity found in embodied social communication would mention

full body poses, interpersonal distance, signing, and more, but these fall beyond the scope of this work.

Presumably, the evolutionary approach supports the inception of a basic model of social behaviour common to all cultures, which appears most relevant for a first generation of culture-agnostic social robots; but at a later stage, cultural and cross-cultural awareness should match a robot's perceived intelligence. However, if robots could reach the next step – by fully adopting a culture and behaving accordingly – challenging societal considerations could arise. Culture may remain such a human-specific trait that it may become questionable for a robot to actively engage in all behaviours of a particular culture, potentially raising identity issues similar to the many inter-cultural community clashes in history.

Across all cultures though, it stands that our shared embodiment and shared cognitive mechanisms support the detection of intent and empathy. Social cognitive neuroscience (Ochsner, 2004) and the theory of mind (Meltzoff & Decety, 2003) converge towards the recognition of our innate ability to transpose a person's feelings and intentions to ours. The co-evolution of brain and facial features eventually allowed the detection of internal states in others; imitation of a facial expression easily induces in others the same feeling at its cause. Quoting Darwin from (Darwin, 1872): “The force of language is much aided by the expressive movements of the face and body”, a statement definitely backed-up a century later by studies from Ekman and Friesen (Ekman & Friesen, 1969) on detecting deception and micro facial expressions.

Indeed evolution shaped many other aspects of human behaviour. Emotional displays given off by the head go beyond facial expressions: e.g. pupil dilation, gaze, prosody of speech or even skin conductance also reveal internal states. Conversely, specific stimuli bias our behaviour and reasoning, an effect referred to as priming. Despite advances in fields such as neuro-

biology and neuro-psychology, comprehensive understanding of human non-verbal behaviour is not yet complete. even as quickly as science progresses, it sounds fair to say quite some time and research effort is needed to catch up with 5 million years of evolution, and to endow robots with the ability to naturally interact with our human nature.

2.1.2 Social Robots for a Social Species

The need for social robots is supported by a deeply rooted human trait revealed by several studies: humans behave socially with robots.

A study by Tanaka et al. (Tanaka, Cicourel, & Movellan, 2007) placed a non-social QRIO robot in a class of Japanese toddlers. Over the course of 5 weeks, behavioural reports described how the pupils treated the robot as a member of their group, showing care and integrating it in their daily activities. It may be that QRIO sharing the toddlers' size and overall aspect helped them to relate to it, nonetheless, they spontaneously treated the QRIO as a social partner.

In Andrea Thomaz' experiment *Sophie's kitchen*(Thomaz, 2006), participants taught the agent Sophie how to bake a cake with objects available in a virtual kitchen; the agent acted visually and could only display object-oriented attention with head gaze. At any time of the baking process, Sophie could receive participants' positive or negative feedback through a computer mouse. Thomaz theorises participants used social reasoning as they reportedly interpreted the agent's gaze as intentional. More importantly, they "felt positive feedback would be better for learning", which correlates with 69.8% of positive feedback given by the participants. Thomaz underlines that machines need to be aware of human teaching biases: actions are interpreted as goal-oriented (ruling out random exploration) and guidance is mostly given with positive feedback. Arguably, positive feedback isn't most people's primary teaching method, a fact that suggests psychological and

cultural traits, but the surprise in this study is how participants anthropomorphised the agent, and appeared to care for its “feelings” in its attempts to learn cooking skills. Maybe the task to learn and the representation of the agent promote a specific social mindset in people, however, exploration of these dimensions ask for additional studies.

Cynthia Breazeal pioneered extensive research on the social aspects of human-robot interaction with the Kismet (Breazeal, 2002) and Leonardo (A. G. Brooks et al., 2004) robots. The latter capitalises on the experience with Kismet, and features 65 degrees of freedom necessary to display a rich set of social behaviours. Thomaz – née Lockerd – and Breazeal (Lockerd & Breazeal, 2004) stress the importance of transparent states to establish collaborative learning as implemented on Leonardo. A continuous provision of robotic social cues guarantees the teacher’s ability to infer the robot’s learning confidence, allowing the teacher to regulate complexity by guiding the acquisition of the most relevant examples for a task. Additionally, they mention robot commitment: social readability supports overt robot intentional behaviour, dedication to acquiring a specific skill, and may also be key to maintaining a teacher’s engagement. Continuous robotic social cues also benefit learning with “Just-in-time Correction”, the possibility for the teacher to provide timely feedback on the robot’s actions.

Arguably, it takes more time to formulate and verbalise an accurate description of a learner’s error than it takes to provide a social cue. Perhaps, in these interactions, people usually build short abstract sentences such as “not like that”, or “put that stuff back in there” because they believe the established social context carries enough information. Consequently, this underlines how a speech-capable robot can overcome difficult utterances if it adequately monitors and interprets social cues.

Establishing and maintaining recurrent robotic social interactions in the long term presents additional issues: the curiosity associated with the nov-

elty of the robot fades away and machine driven interactions tend to be repetitive, which may eventually shadow efforts spent on the robotic social behaviour. Literature on long term human-machine interaction includes works by Bickmore and Picard (T. W. Bickmore & Picard, 2005) in which a virtual relational agent interacts on a daily basis with a hundred of participants over the course of 4 weeks. The agent with social-emotional and relationship building skills, was “respected more, liked more, and trusted more[..], additionally, users expressed a significantly greater desire to continue working with the relational agent after the termination of the study”. These findings not only introduce the idea that long-term relationships with agents are possible, but also that people welcome these abilities and look forward to these kind of interactions. More recently, Bickmore focused on the means to maintain long-term engagement (T. Bickmore, Schulman, & Yin, 2010), and found that “increased variability in agent behaviour leads to increased engagement and self-reported desire to continue interacting with the agent”.

The result obtained with subtle variability in the agent’s visual and syntactical changes may also impact the future of robot design, aesthetically and functionally. It may be that people would feel more engaged with robots capable of changing subtle visual aspects such as a clothing, accessories or even finer facial details. It may also be that beyond verbal communication, variability in robot behaviour would have similar effects. However, personal robots have yet to catch up with the degree of subtlety virtual agents are capable of, alas, current hardware limitations profoundly hinder efforts in this area.

Whether considered for long-term or short-term interaction, robot personality touches a sensitive part of our human nature and brings societal reflections. Currently, this topic falls short of complexity in existing robots, certainly because most have a functionally oriented programmatic nature,

however, they are not devoid of personality either. Syrdal, Dautenhahn, Woods et al. in (Syrdal, Dautenhahn, Woods, Walters, & Koay, 2007) consider the influence of robot anthropomorphism on perceived personality, and their findings suggest people assign personality traits to robots in much the same way they do between themselves. In light of anthropomorphism, the interactive nature of robots, their observable behaviour and physical appearance endow them with a form of personality admittedly limited by a lack of coherent self-awareness. Slow, precise and rigid methods of problem solving (such as pouring and bringing a cup of tea in the same predictable way) would probably let them seem careful and perhaps stupid; traits that might irritate people on a long-term basis. Such a scenario points to the many reasons why social robots, and more so personal robots, should develop and adapt user-compatible personalities, a view also supported in (Dautenhahn, 2004). Psychological studies such as (Terveen & McDonald, 2005) report people seek other with similar personalities, consequently incompatible human-robot personalities might degrade interaction quality. Therefore, it appears rather reasonable to promote robots in social environments where careful movements are commonplace and where users do not expect tasks to be completed in record time.

2.2 Towards Natural Human-Robot Interaction

Advanced robots available for the mass market – beyond being much too expensive – are so limited in their interactive skills that they are restricted to simple tasks or specific niche environments. Retrospectively, these robots have proved to be more of a futuristic promotion than anything useful. However things are changing as various kinds of robots start to meet the needs of a diverse audience, and progress on personal robots brings about compelling advancements.

In healthcare, robot assisted therapy is already more than a decade old

(Schraft, Schaeffer, & May, 1998) but is now emerging as a promising way of helping patients or the elderly. Paro (Wada & Shibata, 2007) is a contemporary example used in elderly and autistic child care; RI-MAN (Onishi et al., 2007) is a more advanced and potentially helpful robot providing assistance such as lifting and holding a human. Even if their primary objective is to support the therapy, these robots are still limited by design to a specific interaction: Paro is a pet robot producing seal sounds and responsive to touch, and RI-MAN will just locate a human through face detection and sound localization. Surely one would wish for entertainment or conversation during a stay in hospital, but none of these robots have the ability to provide interaction beyond what they are programmed for.

The toy industry also has the potential to generate a strong appeal for robotics. Compared to legacy robotic toys, of which, Furby (Hampton & Chung, 2003) is an example, latest robots feature improved processing power, more degrees of freedom, more sensors and indeed, more skills (eg. locomotion, face recognition, etc.). Unfortunately they still remain limited in their intelligence and ability to communicate with their owners. Aldebaran's Nao robot (Monceaux, Becker, Boudier, & Mazel, 2009) current design mainly features limited speech recognition that triggers specific programmed behaviours and limited walking capabilities that seem very robotic. While this kind of robot is legitimately targeted at children, the restricted interaction limits the possible extension to robotic enthusiasts of all ages, amongst whom, researchers are currently the main users.

Finally, robots are expected to fill many public areas. Some are already serving as guides in museums like the Rackham (Clodic et al., 2006), as performers like the RoboThespianTM(Engineered Arts Ltd, 2006) or in shopping malls like the Wakamaru (Kanda, Shiomi, Miyashita, Ishiguro, & Hagita, 2009). Also, robots are replacing interactive kiosks: see (Lee, Kiesler, & Forlizzi, 2010) for a study. Operating around the clock, they

would provide information and services through speech and gestures; tasks that humanoids or androids (like the Actroid SAYA (T. Hashimoto, Hiramatsu, Tsuji, & Kobayashi, 2007b)) could certainly handle. However, even if thanks to a dialogue system these robots appear to be smarter in order to work with crowds, the interaction they offer is definitely short-term, often utilitarian and directed to specific scenarios.

On the other hand, the appeal for personal robots pushes robotic skills boundaries towards natural interaction. Even if current research is focused on obstacles to their industrialisation (power consumption, materials, autonomous reasoning, safety, etc.), their inter-personal purpose place them at the forefront of natural interaction research. Naturally, copying human physiognomy is a key advantage for personal robots as a similar embodiment primes mirroring and learning the specifics of inter-personal behaviour.

Ideally, humanoids and androids would offer a natural form of interaction, using their human-like embodiment as humans do, not only with gestures, poses, facial expressions and believable human ocular and mouth movements, but also respecting implicit conventions such as interpersonal distance. A naturally interacting robot is a socially and emotionally aware robot, manipulating non-verbal behavioural cues to the level any human possesses. With such skills, robust, unconstrained and rich communication would be possible, the net effect being less emphasis on verbal interaction, shorter dialogues, less need for assessment of understanding and improved reliance on robots, in short: trustworthy robots.

Beyond trust (and novelty), the motivational argument behind public and personal robots industrial efforts is the engagement they elicit from people, especially grabbing people's attention and keeping them connected so that the robot can successfully deliver its message. Public robots in shopping malls should engage their audience to make purchases, others should engage people to use them as a primary means to get information and ser-

vices, and personal robots failing to engage their users would risk disuse in the long term. However, engagement remains hard to measure objectively, and usually bound to a specific goal-oriented context. On the topic of human-robot engagement, extensive research by Sidner must be mentioned, such as Sidner et al. (Sidner, Lee, Kidd, Lesh, & Rich, 2005) in which the metric relies on mutual gaze and gestures, but most research in this regard is conducted with avatars: Yukiko and Nakano emphasise the importance of timing, Bickmore et al. (T. Bickmore et al., 2010) explore long-term engagement. Measures of engagement in a disembodied agent have also been carried out, as described by Yu, Aoki and Woodruff who propose a method based on speech (Yu, Aoki, & Woodruff, 2004).

2.2.1 Non-Verbal Communication

Robot feedback has been so far fairly limited. Aibo (Fujita, 2001) features mainly blinking LEDs, and latest versions provide additional text-to-speech capabilities, as does Nao. However, for both, a *single* acknowledgement of command is given through light, sound samples, or simply engaging in a new activity. Although such feedback can be sufficient if given immediately, processing verbal commands (e.g: speech recognition and action planning) on computationally limited platforms always creates a significant delay between robot perception and action. For non-expert users, this can be enough time to suppose a communication issue, and often makes matters worse by entering a vicious loop when users repeat their command or utter new ones; this usually overflows the robotic system and eventually the loss of control elicits frustration. We're all familiar with a similar effect computers suffer while under heavy computational load, and struggle to update visual information. To keep the interaction engaging, a robot should actively and *continuously* display that internal processes are at work, and certainly not stand-still. Copying human verbal interactions, a robot should exploit par-

alinguistic communicative channels to fill the gap created while processing verbal commands or other computationally intensive operations, or at least allow interruption.

Despite identification of several human communication channels, communicating in a human manner still represents a real robotic challenge: natural actuation remains difficult probably because we lack hardware and control software able to match human characteristics, and more importantly, research has yet to determine all of the *detailed* aspects of human-to-human communication. Analysis of recorded human-human conversations presents ways to empirically extract non-verbal content (for head and facial study see Ford et al. (C. C. Ford, Bugmann, & Culverhouse, 2010)), but this approach comes with considerable effort since human labelling is required as minute and brief details elude current software, which prevents automation of the task. Nonetheless, there exists a large quantity of literature on the exploitation of less detailed social and emotional content supported by non-verbal channels. Most focus on gestures, body stances, head movements, blinks and gaze shifts with avatars (early key book can be found in (Cassell et al., 1994)) however transferability of avatar research to robots needs to be evaluated in regard to robots' physicality.

Paralinguistic communication also lies at the heart of dialogue management, especially in a multi-party context. While Duncan (Duncan, 1972) identified non-verbal modalities and inferred usage models, Kendon (Kendon, 1967) and later, American sociologist Goffman (Goffman, 1981) analysed human group behaviour analysis, describing conditions of turn taking emergence and how this takes place through gazing. Later, many others tested refined behavioural models on avatars (Bohus & Horvitz, 2011) and replicated turn taking behaviour on robots such as (Bennewitz, Faber, Joho, Schreiber, & Behnke, 2005). Turn taking also exploits hand and head gestures, and of course, verbal language.

Often disregarded in HRI, timing also conveys non-verbal content. While bio-mechanics and most task-driven behaviours require timed coordination (e.g: gait, focus, speech), psychological studies reveal how it is an essential aspect of paralinguistic and interactive behaviours. Interactional synchrony is explored by Kendon (Kendon, 1970), by Bernieri and Rosenthal (Bernieri & Rosenthal, 1991), while Cassell et al. place gesture/speech synchrony at the base of a virtual character animation system in (Cassell et al., 1994). As for HRI, only a few studies specifically target timing and synchrony; for instance in (Yamazaki et al., 2008), Yamazaki et al. consider the ways non-verbal actions should be timed at specific points in their robot guide’s talk, and found the audience’s non-verbal responses increased, suggesting improved engagement. On the other hand, non-congruent synchronised movements can reveal deceptive behaviours (see Ekman (Ekman & O’Sullivan, 2006) for detecting genuine facial expressions); improper timing or conversational delays generate discomfort especially in a dyad.

2.2.2 Reliance on Faces

Social robots are particular in their need to interact in a natural manner with people. Usually, service robots bear a pragmatic design limited to an utilitarian appearance fit to their specific tasks, such as surveillance or vacuum cleaning. Alternatively, a social robot’s design must allow tapping into the human propensity for social interaction to effectively engage in activities as diverse as providing information, entertainment, education, support and encouragement. While social interaction involves multiple modalities, the most important “interface” for human-to-human interaction is the face. The face contains most of our socially relevant senses, and is the source for several highly salient social channels such as facial expressions, eye gaze and verbal communication.

Faces are important to people. Newborns rely on an innate ability to

detect faces to establish social bonds (Brå ten, 1998). Recently evidence from neuro-imagery revealed how this process lies at the base of empathy and theory of mind in later stages of development (Meltzoff & Decety, 2003). Joint attention, a dyadic interaction where the face and the eyes play a crucial role, is at the root of social cognition (Tomasello, 1995).

As faces are supportive of primate and human social cognition and interaction, it seems only natural that machines with faces will stand out of the ordinary. Robot faces can foster social interaction and rarely leave observers unmoved: often the robot's face facilitates social interaction, sometimes it disrupts the interaction, but never does it not have an impact on the relation between user and robot. Faces, including robot faces, are so important to us that often the presence of faces influences us subconsciously: we do not consciously *read* a face, but rather *experience* a face and its actions at a more basal level (Hadjikhani, Kveraga, Naik, & Ahlfors, 2009). Experiencing faces cannot be turned on or off: the brain is continuously processing visual input for faces (which is very poignant in pareidolia: the seeing of faces in random patterns) and is trying to work out their significance. As such, robots with faces will always be treated as special, the question however is how to implement a robotic face and how to bring this face to life to achieve desired effects; either effects desired by the user or effects desired by the robot designer.

In the last two decades, affective and social communication has received increased attention from the Human Computer Interaction (HCI) and Human Robot Interaction (HRI) communities. Although affective computation in HCI can improve the user experience (Picard, 1997), the experience and interaction do not require an anthropomorphic device. In contrast, in HRI affective computation is necessarily two-way: not only does a social-emotional system need to monitor a user's facial behaviour to extract a social-emotional state, but social robots also need to display interpretable

affective states. Consequently, the HRI community working on social robots relies on anthropomorphic devices and faces to support affective human-robot interaction.

2.2.3 Robotic Facial Guidance

At its most basic level, robotic faces serve to support communication and convey information. The face can acknowledge understanding, display unavailability, signal the intention to reply, channel the focus of attention (important in joint attention) and display internal state changes. Moreover, a robot face can be persuasive; Kidd (Kidd, 2008) for example studied different support devices for weight loss programmes and showed how a robotic weight loss assistant with a simple face persuaded participants to stick with the programme for longer. Also the type and implementation of the robot face is relevant; Fischer et al. (Fischer, Lohan, & Foth, 2012) for example show how the appearance and responsiveness of a robot face has an influence on the complexity of language used by users when giving instructions to the robot.

Often a robot face does not have to look natural: a subset of features can readily produce desired effects. Blow et al. (Blow, Dautenhahn, Appleby, Nehaniv, & Lee, 2006) for example present the KASPAR robot with emphasis on dimensions of face design for “minimal expressive features to create the impression of sociability as well as autonomy”. KASPAR uses skin-coloured rubber and displays “fairly natural-looking” facial expressions with only 6 degrees of freedom (contrasting with 47 degrees of freedom in the human face (Ekman & Friesen, 1969)). It has been successfully used with children diagnosed with autistic spectrum disorders to engage them in social communication.

Beyond KASPAR though, a range of androids have been demonstrated

as well, which have a larger number of mechatronic actuators controlling a flexible synthetic skin. In this category, the Hanson robot faces gained popularity through widespread diffusion on the web of one of the first video recordings of the Albert Hubo and Joey Chaos robot heads (Hanson, 2005). Of course Ishiguro's androids (Sakamoto, Kanda, Ono, Ishiguro, & Hagita, 2007) also generated significant attraction.

Nonetheless, expressive facial animation in robots has been traditionally implemented using mechatronic devices. Kismet is one of the earliest and most classic mechatronic expressive robot, with all features –such as eye lids, eye brows, lips and ears– being physically implemented and controlled by electric motors. Other examples are the Philips iCat (van Breemen, Yan, & Meerbeek, 2005) which has a cat-like head and torso with motorised lips, eye lids and eye brows, and the MDS (mobile, dexterous and sociable) robots, which have motorised eyes, eye lids and a mouth (*MDS project at the Personal Robots Group, MIT Media Lab*, 2008).

Before robots featured facial expressions, research in this area gathered the video game and movie industry. Both developed techniques to record human performances and smoothly play these animated facial expressions on different virtual characters. Incidentally, level of detail in films pushed Computer Graphics Imagery (CGI) to constantly raise the quality of facial animation, and the first popular and widely recognized achievements towards realism came with the well received *Final Fantasy: The Spirits Within* in 2001. However, believable synthesis brings a tougher challenge, and research continues to yield many techniques for facial animation (for a comprehensive survey see (Noh & Neumann, 1998) or (Schroeder, 2008)). Facial expression models also differ in their space of representation: the component and intensity approach described by (Smith & Scott, 1997) is one of the many schools of thought on the topic in psychology, and similar concepts are applied in robotics. Bartneck et al. (Bartneck, Reichenbach, &

Breemen, 2004) conducted an experiment modifying geometric intensity of facial features on the iCat robot and showed a linear relationship between geometrical intensity and perceived intensity of expressions.

In light of existing efforts in facial expression modelling, it is not surprising some robotic systems instead carry a flat screen monitor to display a synthetic face. Whilst the hardware cost of these robots is considerably less due to the use of off-the-shelf components, it is often felt that these attempts to endow the robot with an affective character are not as successful as the previously mentioned mechatronic solutions.

Humans are exceptionally good at inferring where others are looking. This ability highly facilitates the establishment of joint attention, deemed to be very important for a wide variety of interaction schemes, both between human-human (Deboer & Boxer, 1979; Langton, Watt, & Bruce, 2000), robot-robot and human-robot (Nagai, Asada, & Hosoda, 2006) interaction. This has been acknowledged in the HRI field for quite some time and several studies have proposed algorithms for gaze direction detection, both in humans and other robots (Atienza & Zelinsky, 2002; Yoo & Chung, 2005; Ruiz-Del-Solar & Loncomilla, 2009) (see (Hansen & Ji, 2010) for a survey of eye and gaze detection).

Appropriate eye gaze behaviour facilitates interaction: for instance in (Yoshikawa, Shinozawa, Ishiguro, Hagita, & Miyamoto, 2006), a responsive robotic gazing system increases the feelings of people being looked at, thus enhancing the interaction experience. Related to this, in (Miyachi, Nakamura, & Kuno, 2005) and (Miyachi, Sakurai, Nakamura, & Kuno, 2004) a bidirectional eye contact method was described that facilitates the communication between a robot and a human. In (Picot, Bailly, Elisei, & Raidt, 2007), a virtual agent displayed on a flat-screen monitor was able to interpret scenes and direct its gaze in a lifelike manner.

In human-robot interaction, the detection of gaze direction can be considered from both perspectives: the robot detects gaze direction in the human partner and vice-versa. The ability to detect the direction of some agent's gaze needs to be present for both interacting partners, hence it is very important a human can easily perceive where his/her robotic partner is looking. This is of significant interest in developmental robotics where the robot-human dyad supports mental development. In young children, for example, cyclical changes in gaze to and from the adult serves as a signal function of the infant's affect, which in turn modulates the adult's behaviour towards the infants (Deboer & Boxer, 1979). Cognitive psychology shows how gaze direction reading is essential in joint visual attention (Langton et al., 2000) or how object permanency can be read from the gaze being fixed on the expected location of an occluded object. In adults gaze is a powerful signal; gaze aversion, for example, is used to signal thinking such as in the consideration of a question (McCarthy, Lee, Itakura, & Muir, 2006).

If, however, a robot's design includes neither facial expression nor eye gaze, head gaze can provide a fallback mechanism for the provision of robotic social guidance. Many salient cues such as direction of attention or conversational management nods can be expressed with a pan & tilt neck, and the same actuation mechanism can also provide emotional cues in the limit of cultural conventions. Z6PO, the popular science-fiction robot has no animated face but its behaviour convinced a large audience; also famous, ASIMO is designed with a simplistic black visor as a face, and PR2 from Willow Garage has nothing close to a humanoid face either, yet their head movements manage to convey overt social cues. This type of gaze is interpreted and impacts human behaviour: in (Mumm & Mutlu, 2011), Mumm and Mutlu manipulated Wakamaru's gaze and found mutual gazes increased the overall distance male participants maintained with the robot, but no such effect was found with females.

Finally, the freedom of robot design allows the exploration of other forms of facial guidance. Ears are popular with Leonardo, Simon and Nabastag. The rationale behind these designs relies upon our ability to interpret pet behaviour, especially so considering our significant co-existence with dogs and horses in which ear position is salient and congruent with distinctive behaviours interpretable socially. With robots using color signals like the Nao, social interpretation becomes difficult and leaves us with feelings which may be shared and described in similar terms by others, a priming effect investigated by psychologists (for further description see (Maljkovic & Nakayama, 1994)).

2.2.4 Robotic Head Technologies

Arguably, a robot's face defines its identity, and in this regard, many designs and technologies are available. However, to support facial animation, all robots require a head, movable only through a mechanical neck. For head gazes, the most basic robot neck features 2 degrees of freedom (DOF) with a pan and tilt unit, a mechanical design selected for the QRIO and Nao. Nevertheless, a 3rd DOF enables head movements to appear more natural and increases the number of social gazes robots such as the Wakamaru or the latest version of ASIMO can perform.

Mechatronic Heads

Mechatronics groups all disciplines involved in digitally controlled mechanical actuation and remains the primary method of humanoid robotic movement; unsurprisingly, this is also the case for robotic head animation. For rich interactive scenarios, a robot should at least display its eye-gaze by orienting his eyeballs – a minimal solution adopted for the Robovie (Mutlu,

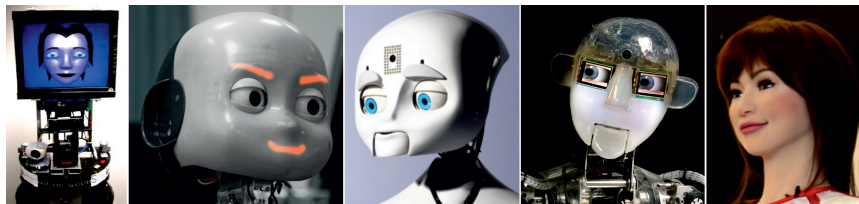


Figure 2.1: Some examples of contemporary robot heads (from left to right): an avatar displayed on screen *GRACE* (Gockley et al., 2004), the *iCub* mechatronic and LED head (Beira et al., 2006), the *Nexi MDS* mechatronic robot head (Breazeal et al., 2008), the *Robothespian* head with mobile phone screens and animated jaw by Engineered Arts Ltd. and the *Actroid DER3* android robot head by Osaka University and Kokoro Company Ltd.

Shiwa, Kanda, Ishiguro, & Hagita, 2009) – and for emotional interaction, movable brows, eyelids, and lips are necessary to endow the robot with facial expressions. This is the approach taken for the robot *KOBIAN* (Zecca, Endo, Momoki, Itoh, & Takanishi, 2008) which emphasises displaying natural behaviour and facial expressions with 7 DOF (three for the eyes, and one each for upper eyelids, eyebrows, jaw and lip), all managed by actuators embedded in the head. In addition, *KOBIAN*'s arms and legs enable the performance of congruent gestures. Aesthetically, these often conservative designs offer simple shapes and replace skin with a hard cover, usually opting for smoothed plastic surfaces. Movable facial features are typically made to look salient, hence facial design do not need to be particularly realistic for users to pick up social signals. Belonging to this category were previously mentioned *Kismet*, *Simon*, and *Nabastag* which have most facial features, *Nexi* (pictured above) however does not feature lips even if it has a mouth.

Recently, the robot head *FloBi* (Hegel, Eyssel, & Wrede, 2010) introduced a modular aesthetic design to mechatronic faces with magnetic facial features. Users of this 15 DOF robotic face can physically change facial features such as hair, eyebrows or lips, all coming in various colors and shapes

so the robot's gender becomes controllable, and theoretically even all skin color could be changed. Many more mechatronically driven social robots could be mentioned, but their variety makes it impractical to detail the subtleties of each and every one of them.

Android Heads

The key characteristic of androids is their intentional high similarity to human appearance and exceptionally realistic skin deformation. Due to the high number of actuators and non-linear interaction with the synthetic skin (a technology evolved from animatronics), android faces are typically more expressive than the above mentioned mechatronic heads. Previously cited robots in this category are SAYA and KASPAR, but most popular examples continue to be the Ishiguro's Geminoids (Sakamoto et al., 2007) and androids heads from Hanson Robotics (Hanson, 2005). Androids seem to exert a particular attraction on the general public as they can be perceived as more sophisticated, and that may be true for the efforts deployed in their underlying actuation. The better the quality of actuation, the closer androids get to human physical capabilities: smaller actuators mean more degrees of freedom for facial expression, quicker movements and better control help behavioural realism. However facial expression still uses the same kind of servos found in mechatronic robots, and even if electro-mechanical motors and compressed-air muscles compete for actuation of the neck and other limbs, the general consensus is that serious contender technology can be expected such as electro-active polymers (Bar-Cohen, 2006) or their graphene-enhanced version (Liang et al., 2012).

Aesthetically, the flexibility, colour and texture of the synthetic skin raises the realism of android faces to new heights with each new generation of materials. Skin comes in localised variable thickness to account for the different deformation of fat tissues and foldings (see (Bickel et al., 2012) for

improvements), and veins, bulges or beauty spots can be reproduced, even make-up can be applied easily. Although only seldom explored (see a survey in (Argall & Billard, 2010)), touching a android’s synthetic skin promises a more natural feeling, also paving the way for – ethically debatable – intimate HRI, mostly disregarded in research but demanded nonetheless. Finally, synthetic hairs also contribute to the overall impression of a human, at least from the distance. Even though only a few companies in the world have acquired the extensive experience needed to build state-of-the-art androids, these robots carry a great potential for HRI once reservations against their appearance and use will fade.

Virtual Characters

A convenient and economic way of implementing a robotic head is to mount a computer screen on a robot and use it to display an animated avatar’s head. Virtual heads are getting more attraction in robotics as often this option aims towards mechanical simplicity and lowest maintenance use. Depending on the weight and size of monitors, mounting these displays on mechatronically articulated necks may require dedicated capable hardware, thus a fixed neck design serves as a maintenance-free solution. Valerie, Grace and George (Gockley et al., 2004) are virtual robots heads using this technique, while Baxter (Guizzo & Ackerman, 2012) features a smaller screen for eyes, only actuated by a pan and tilt neck.

A computer rendered virtual head – also the first occurrence of a “talking head” – has a wide range of freedom in terms of aesthetics and functional design and allows extending exciting areas already explored with avatars such as lifelikeness and human-like behaviours, see (Prendinger & Ishizuka, 2005) and more recently (Pelachaud, 2005) and (Peters & Qureshi, 2010). For HRI, the aesthetic freedom creates opportunities to propose alternative designs in real-time or on a particular occasion (e.g: new year’s eve,

venue of a special guest); moreover visual issues originating from a defective 3D modelling can be fixed without any hardware modification. A monitor in place of a head fosters key advantages compared to the aforementioned robot head technologies as the screen estate permits the display of other forms of information along with the virtual face. For instance, at Carnegie Mellon University, speech bubbles augment Valerie's utterances. Arguably the possibilities are boundless, for example pictures or animations can enhance visual feedback acting for thoughts or emotional status, maps can help a robot receptionist's direction, and so on.

Early on, CGI researchers (see (Wojdel & Rothkrantz, 2005) for modelling) and vision researchers (Pantic & Rothkrantz, 2000) have based their work on Ekman's facial action coding system (FACS) which has been refined over the years and yielded the newest version in 2002. Briefly, FACS divides the face in 44 basic Action Units (AU) that are involved in facial expressions. Each AU stands for a muscle or set of muscles visually modifying a specific facial feature and the coding system precisely describes all these modifications per AU.

Mixed Technologies

Semiconductor light sources technologies can be used to implement faces as well, and a range of innovative designs adapt and/or mix these technologies to overcome limitations of the aforementioned robot heads. Some designs successfully merge various technologies: *Robothespian* (Engineered Arts Ltd, 2006) uses mobile-phone displays to animate the eyes, mechatronics for animating the chin and Light Emitting Diodes (LED) to control the colour of the face. Another approach is possible as demonstrated by the *iCub* robot (Beira et al., 2006). *iCub*'s plastic head has a volume similar to that of young child but inspiration stops there: behind the smooth semi-transparent plastic face, three sets of LEDs implement the eye brows along

with the mouth, mechatronic eyes (a camera in each eye socket) and eyelids. The resulting facial expressions however, are necessarily stateful and consequently appear far less realistic. Robots Simon and FloBi also rely on LEDs, although in these cases, light only provides a means to colourize the face.

Finally, Hanson's Zeno (Hanson et al., 2009) mixes an android deformable skin with mechatronic eyelids and eyes in a non-realistic child robot about 50cm tall. The result amongst users have yet to be thoroughly evaluated.

Chapter 3

LightHead, a Social-Emotional Robot

3.1 Motivations for Innovation

The motivation to innovate came from the realisation that most robot heads have a restricted ability to explore the limits of non-verbal Human-Robot Interaction whereas interaction remains essential to the progress of robotics and exchange of knowledge with connected research domains such as artificial intelligence.

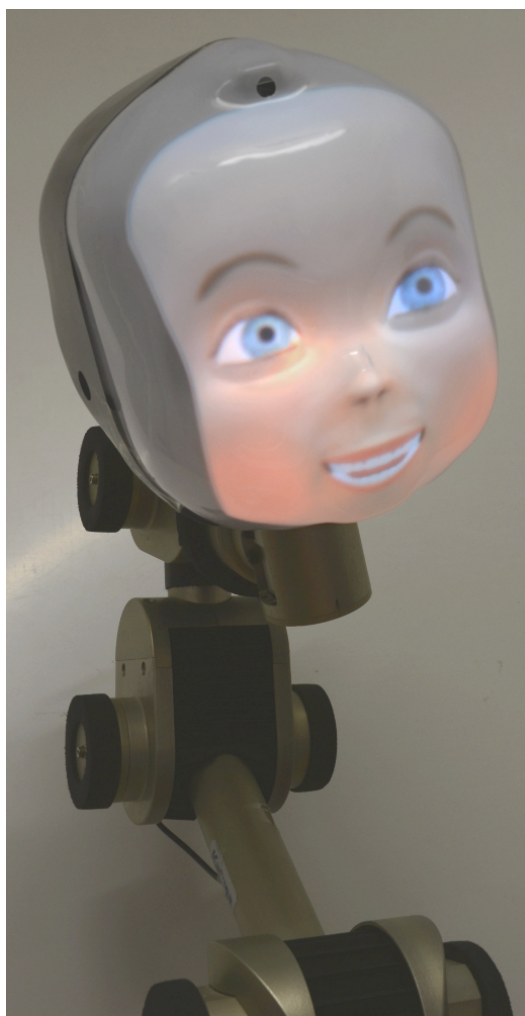


Figure 3.1: The LightHead robot, the fourth and last version. See all versions in table 3.1

Research and development of the robot commenced in 2008, supported by the EPSRC funded CONCEPT project (EPSRC EP/G008353/1). Originally, the project aimed at achieving human-to-robot and robot-to-robot tutelage, unfortunately budgetary constraints prohibited the purchase of a second robotic arm for robot-to-robot tutelage. Besides, eye-tracking as well as gesture interpretation represented significant endeavours within the allowed time frame and resources, thus efforts shifted towards endowing the

robot platform with social signalling capabilities.

3.1.1 Socially Guided Machine Learning and Non-verbal HRI

A robotic platform immediately advertises its social status when having a face, preferably regardless of the angle they are looked at. Appearance also focuses users' expectations and promotes specific interactions, hence providing a robot with a child face entices engagement in non-verbal interaction. It also fosters tolerance, in particular regarding non conformance to cultural standards, such as a lack of manners. Moreover, for human tutelage scenarios, a robot should naturally establish and sustain user engagement through emotional displays as this complements socially guided machine learning. Failure to do so means users may have neither enjoyed the interaction or felt they overcame the system's limitations, and consequently their desire to engage in further tutelage may fade. Particularly fitted for emotional communication, facial expression on many robot heads suffers from the shortcomings mentioned in the previous chapter.

Additionally, to support the interaction, the robot head is mounted on a robot arm, with the arm acting as a spine and neck. Thus, the robot can dispense social signals from head movements and scan the environment, but also to crane over a table, to for example inspect objects presented to the robot. A robot arm needs to be safe: if non-compliant, the only arms deemed safe for close interaction with people remain those with payload restriction. This limitation excludes many technologies currently used for implementing social robot heads and faces. As such, a custom projection-based system is designed, which not only addresses the weight issue, but at the same time improves over many aspects of existing robotic heads.

3.1.2 Overcoming Other Technologies' Limitations

R-PAF technology (Delaunay, de Greeff, & Belpaeme, 2009, 2010), also known as retro-projected faces or RAF, was proposed to address physical limitations of existing robot head technologies. R-PAF relies on the retro-projection of an animated image of a face onto a semi-transparent surface, moulded to match the geometry of a face. Both the projector and the semi-transparent mask are mounted on a chassis, which can be attached to a robot body, such as a robot arm or a mobile platform. The face animation projected onto the mask is generated in real-time by a computer, also used to control the robot arm.

The robotic prototype presented here is dubbed LightHead and has the appearance of a young child (see figure 3.1). However, the design of the mask is flexible and can be adapted to a more adult physiognomy. This is the approach taken for example in (Al Moubayed, Beskow, Skantze, & Granström, 2012; Kuratate et al., 2011) and the recent Socibot by Engineered Arts whom received a demonstration of the technology in 2009. Moreover, after the shaping of the face as a mask, the projection allows further modifications of the aesthetic design. In fact, two main faces were deployed over the iterations of the prototype (see figure 3.1) and four alternatives of the latest iteration have been customised for an experiment (see figure 6.1).

Current projectors are hard to mount inside a constrained space (i.e: AlMoubayed and Kuratate used a mirror to project the image of medium-sized video projectors, which sits outside the head volume), however specific revision of the design enabled fitting all equipment within the head volume. Thus, the back of the head has a cover completing the skull, while respecting the proportion and dimensions of a young child's head. As the projector and optical equipment is confined in the head volume, both are invisible, inaudible and do not create distraction.

Although the face contains visual and auditive sensors, the projection apparatus represents most of the head’s weight. Without any mechatronic element and using only plastic for the face and internal structure, not only weight stays at a minimum, but also maintenance and power consumption.

Consequently a retro-projected head is relatively affordable due to the use of off-the-shelf components and the low-cost of materials. While some elements of the head need bespoke manufacturing, the materials are readily available, and as such have little impact on the total cost. In addition, as retro-projected robot heads have no moving parts, the mean time between failure only comes from the projector (i.e. at least 10,000 hours). This contrasts with mechatronic and android technologies, where due to wear and tear, the face needs regular maintenance and sometimes costly repairs.

3.2 Retro-Projected Robotic Faces

3.2.1 Background

Exploring the modification of human perception with the projection of an image (or video stream) onto objects may have started with the first projector technology and still continues today. Often set up in augmented/mixed reality research, this process usually bears the name of *shader lamps*. In the area of facial projection, although Naimark et al. proposed the *Talking Head Projection* (Michael Naimark, Nicholas Negroponte, & Chris Schmandt, 1980) as early 1980, their inspiration – the *Singing Busts* – appeared in the haunted mansion of the Walt Disney’s amusement park (see (Mine et al., 2012)) and are probably the first popular occurrences of shader lamp faces. Of course, the projected material could only rely on film and interactive robotic applications were not possible.

The R-PAF technology described in the rest of this document is based on a retro-projection version of shader lamps. Most likely, previous ventures

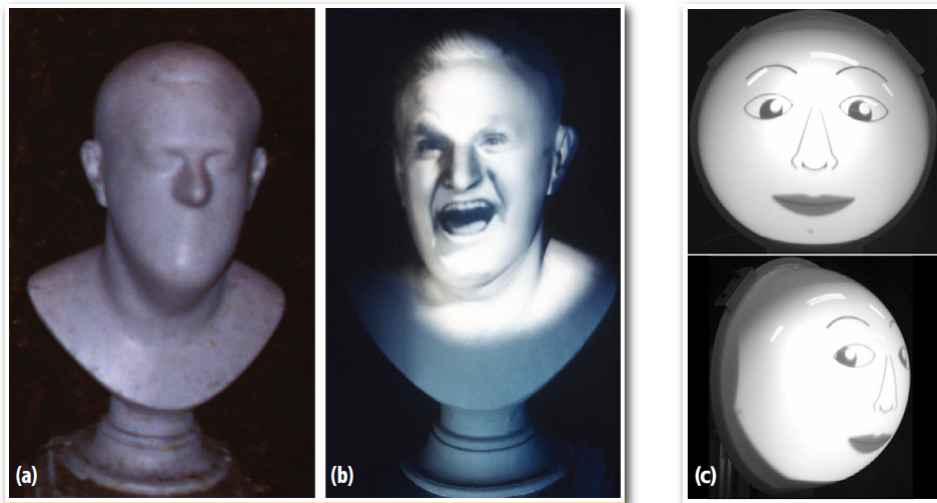


Figure 3.2: a & b) front projection, one of the Disney’s Haunted Mansion singing busts – from (Mine et al., 2012). c) retro-projection, Hashimoto’s Kamin-FA1 robot – from (M. Hashimoto & Morooka, 2005).

in this direction were motivated by avoiding shadows and hiding projection equipment of shader lamps. The earliest publication may come from Hashimoto (M. Hashimoto & Morooka, 2005) in 2005, who partially realised such a setup with the Kamin-FA1 robot, but limited the shape of the facial display to a hemisphere. Unsurprisingly, other scholars such as (Karahalios & Dobson, 2005), or later (Lincoln et al., 2009) also reported the potential of animatronic shader lamps for robotics, but to my knowledge, LightHead represents the first case of the method successfully applied to a robotic head.

3.2.2 The Mask

Invariant Features

Similar to the shader lamps, the geometry of the mask only expresses the invariant facial features found on a face (ovoid shape, nose, eyeballs) while others features like mouth and eyebrows are moving most of the time, and

thus are not geometrically expressed by the mask. The smooth geometry balances between aesthetic freedom (e.g. eyes can take various sizes and shapes) and ease of identifying the display as a face, suggesting the social abilities of the robot. This improves on the Kamin-FA1 robot face (M. Hashimoto & Morooka, 2005), in which a system is described where a line drawing is projected into a semi-sphere, which appears to be the frosted shell of a light fixture.

Often, HRI research and experimental studies target dyadic interactions, yet social interaction calls for other scenarios where a robot has to interact with more than one person. In those contexts, being able to read the robot's gaze is a requirement so that each participant in the interaction can monitor the robot's gaze direction, thus supporting turn-taking in multi-party conversations. By geometrically expressing the nose, head gaze can be picked up immediately, moreover social robots' design should include eyes and eye control that supports natural eye communication, which permits gaze direction following and most importantly joint attention.

A flat display of the eyes does not allow gaze to be directed: the so-called *Mona Lisa effect*. Instead, mimicking mammalian eyes by using a convex surface enables directed gaze (see also (Moore & Series, 2002)). Consequently, LightHead's mask has two curved areas for the display of eyelids and eyeballs (full sclera, iris and pupil), which is key to reaching a satisfying level of social interaction. In (Delaunay et al., 2010), the effectiveness of convex eyes for reading gaze direction was evaluated, confirming that 3-dimensional eyes provide a substantial advantage over flat eyes when reading eye gaze (see further 4). Interactants were able to read LightHead's eye gaze just as they did with people, both when facing the robot and when viewing the robot from a 45° angle.



Figure 3.3: Left: mould and mask. The mould requires sanding to smooth the layers still visible and drilling in the ridged areas (e.g. eye sockets); this prevents trapping air pockets so the vacuum process correctly shapes these areas. Right: foldings can appear if the temperature for vacuum forming is too high or the plastic too thin.

Material and Process

For the video stream to be seen through the facial mask, a light-permissive material is required that does not restrict shaping freedom. Vacuum forming presents the best option for creating such a mask: these well-know process and plastics are cost-effective solutions and the end result offers a pleasant smooth finish.

The mould sets the geometry of the face and fixes its overall aspect. For its creation, the original iCub face model was taken as an inspiration and has been reworked in a CAD suite to regroup all parts into one single model and adapt it to a solid virtual mould. This required capping holes to create inner volume for the material, shaping eyes spherically, re-expressing the chin and creating the housing for the forehead's camera. Then, the mould was generated with a rapid prototyping machine (a ZPrinter 310 by Z Corp)

printing a high-performance composite, sufficiently resistant to heat (over 150°C) as required for the next step.

To shape the mask, a sheet of thermoplastic was vacuum formed over the mould. For the material, neutral, white-tint, opalescent, High Impact Polystyrene (HIPS), 1.5 mm thick appeared the best choice. HIPS comes in a variety of transparencies and thicknesses, which facilitates experimentation with the level of detail captured from the mould and the level of image sharpness: a thinner plastic results in brighter and sharper images, as opal HIPS has a tendency to diffuse light. During forming, although thicker layers ensure the smoothest shapes, those thinner than 1.5 mm tend to create foldings at sharper angles as seen in figure 3.3. HIPS also allows further tooling to, for example, smooth edges, precisely fit unworkable connected elements, drill venting holes in the back cover or simply glue sensors and accessories.

3.2.3 The Projection

For optimal display, the projected beam should be evenly distributed and cover the widest facial area possible. Thus the projector's normal ray should meet the centre of the mask, assuming mask and face are aligned. However, most off-the-shelf projectors do not include documentation with schematics of the normal ray, and many efforts are spent in inferring them, integrating and designing appropriate housing for the device as well as securing their position. As most projectors have a 16:9 aspect ratio, they must be set in a vertical orientation. This better fits the roughly oval shape of human faces, although depending on the particular projector, that might further complicate their integration¹.

¹over the multiple projectors tried, often this implied removal of the projector's case and adaptation of the chassis to hold reassembled components.

Throw Distance

Any projection system has a distance issue: projectors are designed with an optimum distance range for the picture to be viewed at, determined by the view angle. The mask surface in LightHead is about 15×18 cm, requiring at least 40 cm of projection distance for commercially available projectors. In addition, for the projector to be contained within the head volume, a very small projector is used — a so called pocket or pico projector. Fitting a small and light projector inside the head contributes to the aesthetic quality (and by extension the interactive quality) of the head, a heavier projector potentially complicates the mechanical design by adding weight to the support, limiting the head's motion range and —as the projector would sit outside the head's volume— would add inertia and potentially image instability during quicker head movements.

Fisheye Lens

The most convenient solution to shorten the projection distance is to use an ultrawide-angle lens or fisheye lens, as mentioned in (M. Hashimoto & Morooka, 2005). However fisheye lenses have their own issues. The projection becomes non-linear as the image is compressed near the center and stretched outwards near the corners. The many lenses that comprise a good quality fisheye lens reduce the amount of light passing through, a serious limitation for the weakest portable projectors. Smaller fisheye lenses suffer from chromatic aberration which splits and shifts the original colours near the edges of the image, perceived as a mono-chromatic ghost image. Finally, a good fisheye lens for photography can be costly, taking up space and adding significant weight to the head. For the LightHead robot, this last point took special importance considering the requirement for the head to be light (a design requirement was to keep the arm payload under 400 g). A Nikon FC-E8 lens was used for LightHead; the FC-E8 has close to no

chromatic aberration, a field of view slightly over 180° and measures 74mm in diameter by 50mm long and weights 205g. Also, the radial projection fits the facial volume of the mask and small image distortions are fixed by software.

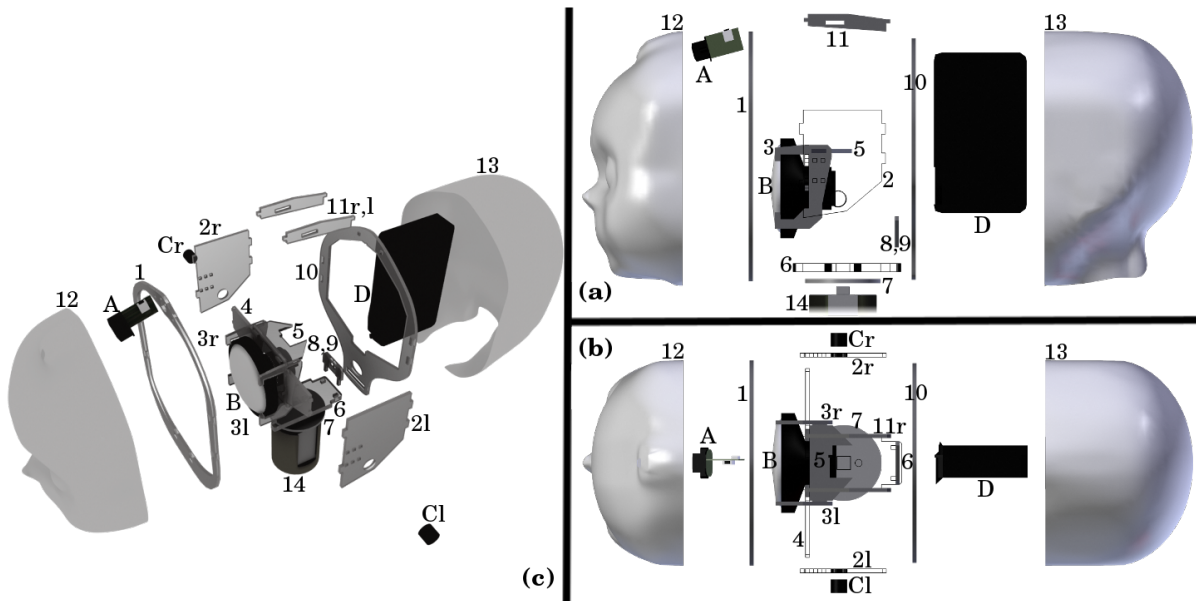


Figure 3.4: The LightHead's parts: (a) left view, (b) top view, (c) perspective. For clarity, some parts are only outlined in (a) and (b). Parts 1 to 11: laser-cut PETG (see appendix A), 12: moulded HIPS mask, 14: moulded HIPS cover, 14: tip of KatanaHD400s-6M robot arm, A: Microsoft Lifecam Cinema, B: fisheye lens Nikon FC-E8, C: electret microphones, D: Optoma PK301.

Projectors Types

Unfortunately the radial projection of the lens also alters the focus of the image, in proportion to the size and planar aspect of the face. While Digital Light Processing (DLP) technology, which is typically used in video projectors, is bound to have this problem, laser projection technology does not. Laser based projection has the advantage of always displaying pic-

tures in focus since spatially and temporally coherent light does not require focal adjustment; also, passing through the fish-eye lens, laser beams are deflected but remain coherent. These projectors have become affordable and portable, due to the use of micro-electromechanical systems technology. Version 3 of the LightHead used the Microvision ShowWX+, providing an 4:3 image, with a brightness of 15 LED Lumens – although looking much brighter, probably over 50 lumen – while matching the size constraints. Unfortunately, this device tends to overheat when operating continuously in a confined area such as the robot skull, thus requires a form of cooling . The inclusion of a small and silent 5V DC fan providing a thin constant stream of air permits several hours of uninterrupted operation.

Nevertheless, a brightness of 15 Lumens cannot match most common indoor lighting conditions, thus, it appeared necessary to replace this model with a brighter projector able to address this limitation. Equipped in version 4 of the LightHead, the Optoma PK301+ reaches 75 Lumens, creating a much improved facial image. However, the powering light source of this projector falls into the DLP category and needs precise focusing. Additional constraints re-emerged: more heat is produced, space is taken by the heat sinks which are actively dissipated by a fan, more space is used by extra digital components, and the case must be modified to properly fit the lens.

Finally, no off-the-shelf device brings a perfect solution (i.e: lightweight, laser-based, small size, silent, passively cooled) but technological improvements keep a steady pace, suggesting these drawbacks would be solved eventually.

3.2.4 Benefits of Computer Generated Imagery

Perhaps the most interesting aspect of retro-projected faces lies in the possibilities offered by using a generated video feed. Undeniably, computer graphics being so ubiquitous, this field of research draws worldwide atten-

tion and sustains a prolific community of researchers – see (Parent, 2012) for a comprehensive survey of algorithms and techniques. Moreover, virtual character animation remains an area of intense research on which retro-projected robot faces capitalise.

With CGI comes a great deal of tools, talented animators, and a wide range of real-time visual effects, all readily available. In terms of computing capabilities – even considering the cheapest models – current 3D graphics chips are powerful enough for elaborated real-time animation and picture effects. For instance, a pixel shader² can compute the projection matrix adapting the generated video to the mask’s geometry (or implement the fitting method described in (Lincoln et al., 2009)). Also, CGI libraries such as OpenGL provide a single operation for specifying the position of a light-source to automatically compute shadow effects beyond those described by Hashimoto in (M. Hashimoto & Morooka, 2005).

Although the prototype is controlled from a standard PC, the computational power of contemporary embedded devices can handle animation of the projected face. Current trends in consumer electronics and embedded computing points to SoCs³ that could not only run facial animation, but whole robotic systems.

Aesthetic Freedom

As opposed to other robotic heads, retro-projection offers a great deal of aesthetic freedom limited only by the level of geometric detail given by the facial display. The face animation is entirely implemented in software and this creates design opportunities of which a number have been explored. In essence, the face can range from a simple cartoon-like animation (perhaps

²Also known as a fragment shader. Shaders are small programs specific to the Graphic Processing Unit, as opposed to the Central Processing Unit that runs most of a program.

³for ”System on Chip”: computer chips embedding most computer hardware on a single integrated circuit.

as simple as the line drawings of Kamin-FA1 (M. Hashimoto & Morooka, 2005)) to a playback of video-recorded faces.

Over the four iterations of LightHead, the robot presented two main faces (as seen in figure 3.1) while retaining the same facial aspect ratio. Early prototypes featured a Japanese cartoonish face⁴ (in the style of mangas) with very contrasting facial features: orange eyebrows, dark eyelashes and pupils and pink mouth over an almost white skin. Even if present, the nose was kept discreet. This proved satisfying initially, but the design tended towards a female, ruling out gender based experiments; and since most participants were British, a Japanese designed robot may have raised a culture mismatch. Consequently, the second iteration of the facial design (LightHead version 3) offered a gender neutral, Caucasian child face. With thinner eyebrows, more realistic pupils and eyelashes, reinforced nose presence and mouth, as well as a more natural skin colour, this design created by another artist⁵ met CONCEPT's cultural setting.

In contrast with other robotic head technologies, the same robotic mask can support several facial variations effortlessly and without time consuming manual operations. Moreover, for each version of a face, colours and style are changeable at run-time. Adapting skin, eye and facial hair color, or even age (for adult faces) is possible while the robot interacts with a particular user. A discussion of the possibilities can be found in section 7.2 of this document.

Extended Animation

Also interesting is the wide range of conceivable visual effects often overlooked in HRI. There are other effects that are hard to achieve with other robot face technologies: simulating sweat, tears or changing pupil dilation

⁴“maid-san” 3D model from author FEDB <http://fedb.blogzine.jp/BA/body.zip>

⁵Bruno Dorbani: bruno.dorbani@gmail.com





Version	Hardware	Software
1: proof of concept 	<ul style="list-style-type: none"> • Acer H7530D office projector, 1600 lumens • raw vacuum-formed HIPS mask • fixed setup 	<ul style="list-style-type: none"> • manga-style facial design • Blender3D game engine proof of concept: keyboard-based interactive animation. • subset of 32 FACS muscles
2: orientable head 	<ul style="list-style-type: none"> • Aiptek V10 pico-projector, 15 lumens • Nikon FC-E8 fisheye • cut out mask • laser-cut PMMA chassis holding all elements • KatanaHD400s-6M mount 	<ul style="list-style-type: none"> • ARAS first implementation with AU pool • ARAS robotic arm FW kinematics support • ARAS script player • CHLAS first implementation • HMS pyVision + face detection support in helper libraries
3: complete head 	<ul style="list-style-type: none"> • Microvision ShowWX+, 10 lumen (appears > 50 lumen) • Microsoft LifeCam Cinema Webcam HD 720p • electret stereo microphones • skull cover • fan cooling 	<ul style="list-style-type: none"> • cartoon-like facial design • ARAS speed control (robotic arm) • CHLAS v1.0, Acapela TTS support • CHLAS script player • CHLAS instincts: breathing, basic blink, coactuator • HMS basic face recognition based on HSV histogram • HMS face tracking robotic gaze
4: smoother mask 	<ul style="list-style-type: none"> • redesigned mask's forehead • Optoma PK301+, 75 lumen 	<ul style="list-style-type: none"> • ARAS real-time editable movement dynamics • CHLAS instincts: conversational blink model, gaze control • CHLAS eSpeak TTS support • HMS system configuration and facial expression library editor

Table 3.1: Design iterations of the LightHead.

are just a few examples. There are two reasons for LightHead to support the latter. Not only reacting to variable light conditions – as a changeable weather is common in Plymouth – adds to the illusion of life, but pupil dilation also convey emotional cues (for a study of the correlation of pupil dilation with mental activity see (Beatty, 1982)) suitable for tutelage interactions. Blushing adds to the emotional effects and as it is straightforward to implement on a cartoonish face, it was added at no cost.



Figure 3.5: An attempt at simulating crying with LightHead’s virtual face. This effect was not exploited in experiments.

Tears, however, were attempted as seen in figure 3.5 but not fully deployed on the system. It was felt this effect would elicit mixed feelings as successful implementation was not guaranteed. Specifically, the effect would introduce additional complications (such as a generating a sobbing sound) thus potentially disrupting the character coherence. Nonetheless, effects such as tears and sweat call for experiments testing the emotional impact with realistic facial designs, an option not available with physical heads. On the other hand, retro-projected faces open a wider range of interaction, bringing affective displays to a new level by capitalising on possibilities of-

ferred by CGI.

In further push towards realism, lip synchronisation can be made authentic, by reproducing minute physical deformations (e.g. progressive parting of central lips when opening the mouth), or adapting lips reflectivity in relation to their dryness. In the same idea, animation of the tongue can improve readability of speech with better visemes⁶, such as drawing the tongue on the teeth, with /ðə/ (i.e: IPA transcription for "the") for instance.

Refined Interaction

Without mechatronic components in the head, noiseless operation becomes possible, further approaching natural human-robot interaction. Actuator noise brings no interactive improvement and rather conflicts with our acquired concepts of life: no species produce constant noise from actuation of their muscles and limbs (although their effect on the environment usually does), instead most animals employ sound as a means of communication. This is particularly relevant in case of facial movements, as actuation noises prevent the illusion of a robotic autonomous mind. In effect, a noiseless actuation eventually allows the accidental realisation that some subtle robotic behaviours evaded our attention. Such observation helps to consider the fact that the robot may have many more undetected self-motivated behaviours. Moreover, if a robot gives away every gaze or facial expression, constant solicitation of our attention forces filtering actuator noise which may add stress in long-term interactions.

Similarly, a virtual head platform grants tight control of the animation dynamics (see section 3.4.3), a critical aspect when generating believable characters that convey the illusion of life. Ekman and Friesen in (Ekman & Friesen, 1982) reveal key timing differences between fake and spontaneous facial expressions. Besides, some humans also uncover concealed emotions

⁶the visual aspect of the lips, tongue and jaw for a specific phoneme.

with very short timed and small facial expressions (Ekman & Friesen, 1969), that the same authors coined *micro expressions*. Android (and more so mechanical) faces still cannot display these minute details: skin is not thin enough, actuators would reach uncannily high pitched sounds to achieve speeds required. In contrast, retro-projected faces open the way to experimentation with robots adopting specific human behaviours through subtle robotic expressions.

Amplified and Augmented Expression

Virtual characters have no physical restrictions, hence facial animation parameters may be modified, the CGI rendering technique can be adapted, and alternative animation or visual effects explored.

For facial animation systems that tolerate out of bounds parameters, over-expression does not damage the hardware. To represent a sensible approach, an over-expression needs to keep a form of visual coherence, or recall established stereotypical cultural expressions, likely borrowing a repertoire from cartoons. Moreover, their expression requires a relevant social context, such as acting or storytelling with children. This *amplified* expression can support comical or horror effects, inaccessible to androids that may likely tear apart the flexible skin.

Finally, *augmented* facial expression can take place displaying an overlay of text as in figure 3.6, icons or videos over the less animated parts of the face, most likely over the forehead. This technique creates a means for robust or explicit expressions: a handshake over a smiling face drawn upon reaching a conversational agreement, a textual information over a sad face upon critical error, etc. Undoubtedly, as a clear facial area is a prerequisite to augmented expression, this option contributes better to non-realistic faces where textural detail stays low and with which users better tolerate unexpected elements.



Figure 3.6: Augmented expression through textual information with the LightHead’s virtual face. This effect was not exploited in experiments.

3.3 Expectation-Driven Embodiment

3.3.1 The Head

Perhaps of all parts of a robot, the aesthetic appearance of the head forges the greatest expectations upon first contact. For the CONCEPT project, emphasis was on soliciting simplicity to lower users’ expectations previous to any interaction.

In terms of design, LightHead’s mask is an adaptation of the iCub face cover (see fig. 2.1)⁷ for its rather simple and elegant aesthetics, resembling a young infant through the large size of its eyes and its high forehead. The child-like design immediately comforts the user with a non-threatening character, also suggesting fragility and in all likelihood, innocence. Even when switched off, the robot’s smooth mask identifies the primary communicative interface provided by the system; the salient rounded eyes also inform on vision capabilities, while the nose completes the face and serves as a head gaze hint.

⁷see also www.robotcub.org

Despite the eye shapes of the mask, vision is localised elsewhere and as no sensor equipment can be mounted between the projector and mask, sensors need to be positioned outside the projection. In the case of non-mobile robots, a camera can be fixed somewhere in their close environment providing them with a third-person view, however some tutoring interactions require close interpersonal distance which may obstruct robotic vision in such setups. Consequently, LightHead has a front-facing camera (Microsoft LifeCam Cinema 720p) located in the forehead, whose housing has been moulded into the semi-transparent mask and as such does not distract. Nonetheless, the small hole stays visible, providing insight for the enquiring user. In fact this detail eludes the youngest interactants who tend to attract the robot's attention by presenting stimuli right in front of the eyes. For this non-experimental case, forging user expectations must be addressed by other means.

Even if the head cover lacks ear-like shapes, obvious holes drilled in respect to human proportions gives away the robot's hearing capabilities. A microphone is set in the head at the location of each ear, put to use with a simple auditive attention system. Accessories like auricles can enhance aesthetic design and from a practical point of view, also narrows the range of directionality of sound detection, reducing acoustic input from the back and focusing the robot's auditive attention to the front and sides.

Further enhancing projected robotic heads, the back cover is created with the same process and material as the mask but serves a triple purpose.

Essentially protecting internal components such as electronics and cabling, the robot's skull affords *hazard-free* interactions: children can safely touch the head, and head movements will not hook users' clothing or jewellery.

Aesthetically, a hard shell makes it possible to further match the robot's degree of realism in realistic anthropomorphism settings (e.g. photo-realistic

virtual face) by attaching features such as hair, auricles or wear on accessories such as a hat.

A cover completes the head so the robot better meets users' *expectations* of a human shape. This not only reinforces its social abilities, but also prevents users from being distracted with very robotic features such as apparent machinery (e.g. projector and cables) or the uncanniness of a face without any head volume. It is crucial to limit to the minimum the novelty effect experienced by participants when introduced to this novel technology so experiments can be kept reasonably short, and collected data captures the essence of the interaction rather than people's curiosity.

For the facial animation, representing a life-like human face may yield to some of the uncanny valley effect. In order to avoid this issue, a non-realistic face was specifically modelled for this robot following a few key requirements: aged around four or five years old, genderless, Caucasian to match experimental demographics and neither realistic nor too cartoonish. As mentioned previously, the first facial model was deemed overly simplistic and motivation came from the perspective of endowing the robot with finer facial expressions.

It was felt the presence of a speaker in the head was necessary to meet users' expectations as human hearing is very sensitive to sound location. Simple informal tests helped realise using external loudspeakers –even if sufficient for providing the robot with a voice– can be slightly unsettling when the head gaze is significantly shifted from its neutral position. Hence, a standard loud-speaker was placed behind the mask to improve speech directionality, however accounting for resonance and voice distortion presented its own challenge and halted efforts in this direction.

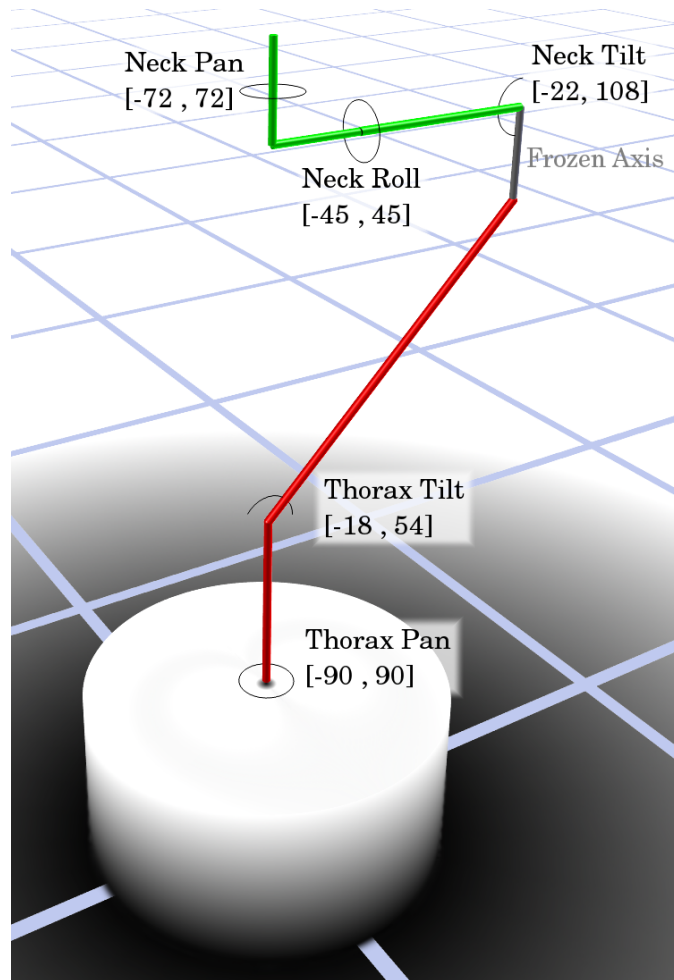


Figure 3.7: The KatanaHD400s-6M kinematic chain mapped to the spine of LightHead (angle ranges in degrees).

3.3.2 The Spine

The face mask is mounted onto a custom chassis holding the projection system, camera, microphones and head back cover. The complete head is attached to the tool plate of an anthropomorphic robot arm, a Neuronics Katana HD 400s 6M, which procures six degrees of freedom. Considering actuator noise though, the actuation of LightHead's robotic arm has been kept deliberately within speeds offering the best balance between acceptable

speed and most quiet movements possible.

Spinal Sections

From the head attachment, the first three joints form a neck, one remains static and the next two are being used as the spine (see figure 3.7).

Because a neck adds complexity to a robotic design, it usually serves more than a social function, and maybe this is why most robots equipped with a neck have sensory input in the head as opposed to a less confined area. In effect, the bigger the robot and the more complex the mechanical design, the quicker and more energy efficient it becomes to direct head-mounted sensory inputs. Otherwise the robot would need to move more body parts – hence more weight – activating for instance the hips or at worst, the locomotion system.

Often less actuated than a neck, a fully orientable and articulated torso is sometimes used for anthropomorphic robots (see (Ly, Lapeyre, & Oudeyer, 2011; Potkonjak, Svetozarevic, Jovanovic, & Holland, 2011)), although rarely can the whole spine bend like a human's. In order to get closer to this capability and mimic human poses, two joints of the robotic arm (from base to neck) enable pan and tilt movements of the thorax while the next axis (joining neck and thorax) did not represent significant additional freedom and was left frozen for simplicity. In some tutelage sessions, the robot spine bent forward to alternatively crane over several objects, and along with constant visibility of the rounded eyes of mask, the illusion of inspection was successful. Even if the Katana 400s would implement rolling of the thorax, few scenarios would actually exploit this axis as humans rarely make such movements.

Both neck and thorax serve the primary purpose of endowing the robot with the capacity to focus its attention within a surrounding world. That is: scan its environment, orient the most effective sensor in response to a

stimulus, lock onto a face or salient object and follow it. On the other hand, static robots equipped with fixed sensors indeed lose track of their target as soon as it goes out of frame. Although these active behaviours carry a social meaning as well, neck and thorax also support signals that are exclusively social.

Social and Aesthetic Aspects

Usually, HRI studies apply eye and head gaze in a congruent manner –with LightHead not being an exception– but head gestures convey additional non-verbal cues. Acknowledgement (nodding), disapproval (shaking), questioning (tilted head) are powerful social signals for Western cultures that are produced with the neck. As mentioned previously, many other cultures communicate with head gestures, and it stands to reason that culture-aware robots necessarily need to make use of a 3 DOF capable neck.

Socially, the thorax becomes important to regulate and respect interpersonal distance. Experimental results from Walters et al. (Walters et al., 2005) not only illustrate people are closer to humanoid robots, but also that their personality can help estimate the distance at which they would likely approach a humanoid. Although this particular arm limits proxemics (see (E. T. Hall, 1966)) to personal space, no experiments on the impact of LightHead on interpersonal distance were conducted.

Aesthetically, for LightHead the spine also acts as a support to tie data and power wires, preventing them from hanging out of the head’s cover, which would certainly detract from the clean design of the robot. However, such a bare chest maintains a pronounced robotic appearance in line with the obvious platform limitation: LightHead does not feature arms. Of course this choice corresponds to the CONCEPT project’s objectives (cf. section 3.1), and even if gestures would fit non-verbal communication, time, human and financial resources imposed prohibitive constraints. In any case

a surprising fact marks the low impact on the robotic character: over the many people introduced to the robot, few of them raised this topic.

3.4 Control

Efforts in designing not only a functional robot with virtually unlimited facial expressions, vision, hearing, speech but also head movements could only bare fruit with a system able to exploit these capabilities to create a naturally expressive robot. As such, software able to react in a timely manner, robust to load and allow integration of sensor data is fundamental to retro-projected robot faces.

3.4.1 Existing Systems

An ideal robot control system should be robust, portable, versatile and accessible. Because it is notoriously difficult to achieve, such a goal is still the topic of several projects either from academia (e.g *Robot Operating System* ROS (Quigley et al., 2009)) or industry (e.g URBI (Baillie, 2005)). Highly constrained environments (e.g. factory lines) significantly moderate the complexity of the task, however social robots are envisioned to interact with humans in a dynamic and complex world, creating an entirely different challenge. Dedicated HRI operating systems have been proposed. For instance, Fong et al. (Fong, Kunz, Hiatt, & Bugajska, 2006) provide a structured software framework for coordinating human-robot teams through different user interfaces and using a variety of robots, although support of social dyadic interaction seems to have room for improvement. Breazeal (Breazeal, 2002) uses a reactive system, based on a subsumption architecture, to regulate behaviours of social robots, and Kuratate et al. (Kuratate et al., 2011) integrated OpenHRI (Matsusaka, 2008) to the Mask-bot. While this serves reactive social robots well, interaction scenarios delineated for the CON-

CEPT project need both the reactive element and extension of behaviours over time. Behaviours that, for example, are needed to let the robot act naturally as an engaging receptionist, museum guide or tutor. Nonetheless these dedicated systems have still failed to reach mainstream use, perhaps from of lack of contributors, and unfortunately they do not indicate they would fit LightHead’s needs either. On the other hand, URBI and ROS – even if each does not aim exactly at the same use case – are likely to fit most robotic problems insofar as control is a complex issue. In the case of *Willow Garage*, openness of the software, best programming practices, and definite established popularity⁸ has the advantage of recruiting contributors. For instance, Rich et al. packaged as a ROS node means to recognize engagement between a human and a humanoid (see (Rich, Ponsler, Holroyd, & Sidner, 2010)). Later, another contributor shared a ROS driver for the Katana arm, but development of LightHead’s system was well underway and integration into ROS was impractical. Nevertheless, LightHead’s purpose is not task-centred but rather focused on interaction and lifelike behaviours, and as such relies on an alternative custom software solution.

⁸Willow Garage supports OpenCV (Open Source Computer Vision library) used in countless projects.

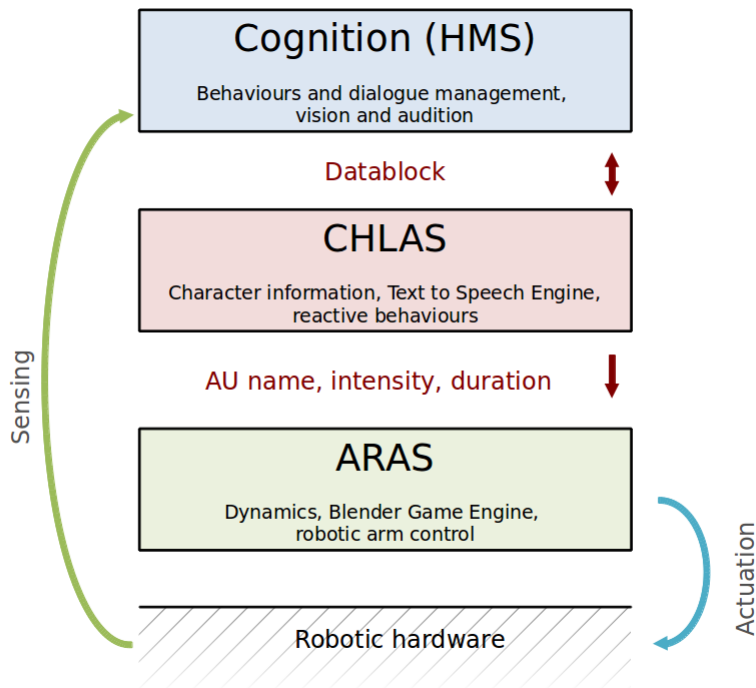


Figure 3.8: LightHead’s software architecture, splitting animation, reactive and affective control, and cognition of the robot. From bottom to top: Abstract Robotic Animation System (ARAS), Character Hi-Level Animation System (CHLAS), High-level Management System (HMS).

3.4.2 Overview

Aesthetics and functional design are often at the heart of arguments about the uncanny valley, however behaviour is at least equally, if not more, important. In order to separate actuation from behaviour a layered system was designed as in figure 3.8:

1. a low-level animation system is responsible for managing and abstracting hardware actuation (Abstract Robotic Animation System or *ARAS*),
2. a mid-level system (Character High-level Animation System or *CHLAS*)

merges reactive behaviours with commands from the next level, transmitting animation info to ARAS in a timely manner,

3. a top-level (High-level Management System or *HMS*) that most often implements cognition, and having direct access to sensors.

In this section the benefits of this architecture is discussed and how behaviour supporting social interaction can be simply implemented.

For the reader also involved in programming, it may be worth knowing that all systems' source code are written in the high-level scripting python language, which contributes to the simple migration of the system. Also, the software has been released under the GNU Public License on the popular GitHub platform⁹ so that other researchers can freely use, modify and distribute the system or its parts for integration into their own work as long as they keep referencing LightHead's. Clarity of documentation and source code are key for broader dissemination of a software, and in the same idea, the designed software interface was kept simple and consistent. Consequently, the systems communicate through a human-readable, clear text protocol, a decision that grants ease of script writing, reuse of common tools and simple integration with third-party software although at the expense of optimization.

3.4.3 ARAS

Abstraction of hardware is limited to actuators and mechanical design: from a user's perspective, focus stays on moving essential body parts regardless of those parts' design details. Sensors, however, are directly handled by higher levels in the software architecture (figure 3.8): this data remains free of any bias. For instance, cognition can directly poll sensory input (e.g. camera, microphones, raw proprioceptive actuator values) if needed,

⁹<http://github.com/Dfred/concept-robot>

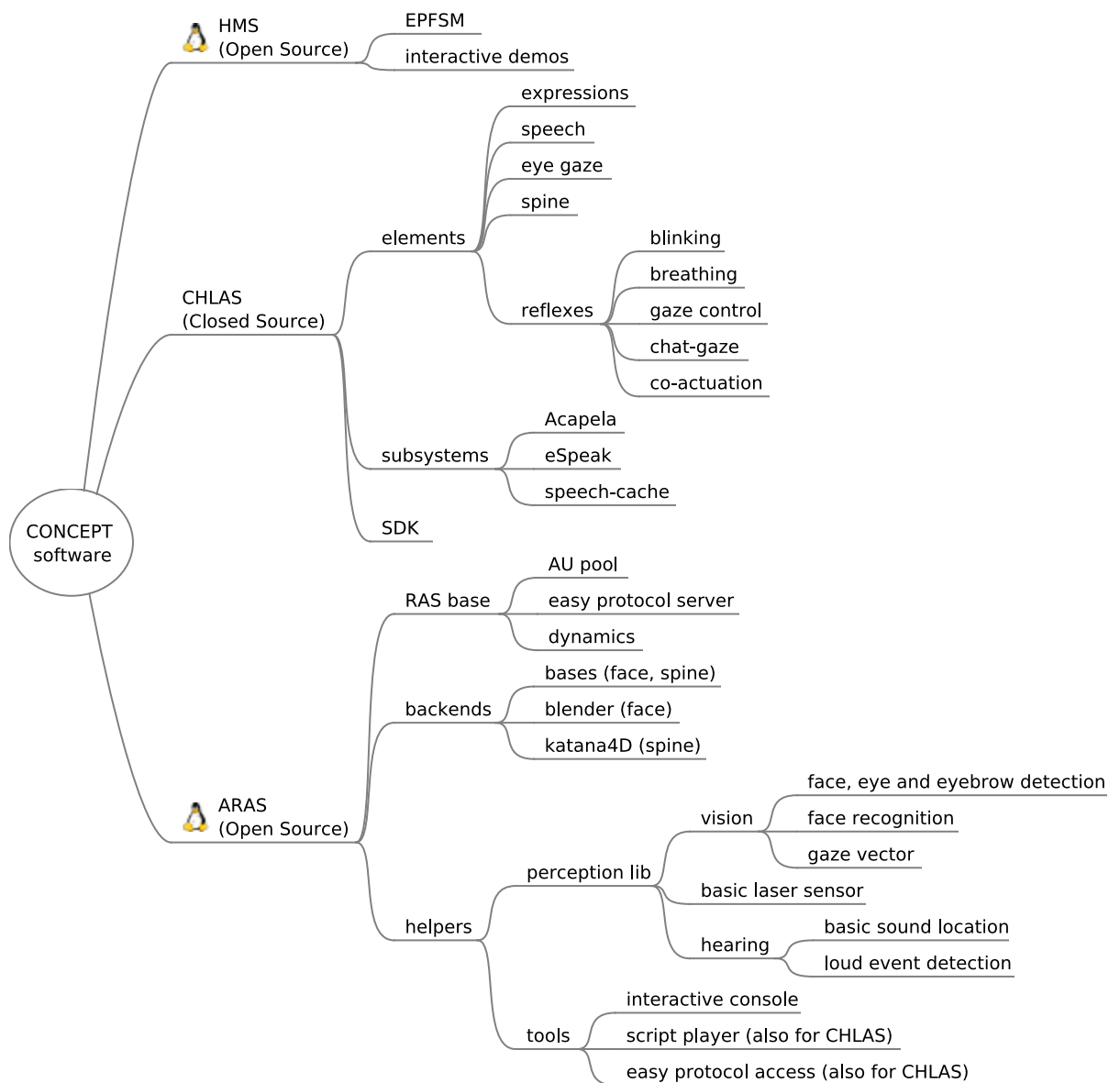


Figure 3.9: Summary of the software designed and implemented during this thesis.

or use available helper functions (see section 3.4.5) such as captured user face position. Real-time movement information for each available – virtual or physical – actuator of the system is unified and shared by the Action Unit pool (described further) hence maintaining proprioceptive information

at no computing cost.

FACS Baseline

Abstraction guarantees variations of the hardware platform have no impact on the algorithms and functionalities available to users of the system. As facial expressions are at the core of the system, inspiration is taken from the Facial Action Coding System by Ekman and Friesen (Ekman & Friesen, 1969). In short, FACS splits the face in Action Units (AU): a single facial muscle or a group of muscles responsible for a localised visual modification of the face. For instance, each eyebrow can be modified by three AUs: inner and outer brow raisers, and a brow lowerer.

Originally, FACS lays out a five degree discreet valuation of AU intensity unsuitable for animation, whereas normalized values lays the mathematical foundations required for computation of the finest animations. Other expressive heads use a normalized and contemporary version of FACS (for an example in robotics see (T. Hashimoto, Hiramatsu, Tsuji, & Kobayashi, 2007a)), which is also the basis of the LightHead system. Normalization fits muscles very well because their activation is bounded, but normalization also applies to angles of AUs managing the orientation of the eyes, tongue, and generally for each element of the skeleton. As mentioned previously, ARAS defines extra AUs for animation of the tongue as well as affective effects such as level of blushing and pupil dilation, other scholars also extended FACS for mechatronic robots (see (Kühnlitz, Sosnowski, & Buss, 2010) for instance). Moreover, FACS defines too many AUs. Multiple instances occur where individual AU define each opposed muscles involved in a specific linear movement, for instance AU61&62 bound to the horizontal orientation of eyes. In these cases, an 'average' AU name convention was used (e.g. AU61.5) to represent the same movement. A comprehensive list of modifications can be found in annexes, page 200. With this baseline,

ARAS ensures a common framework for *backends*, each of which implements operations specific to a piece of hardware.

Action Unit Pool

ARAS maintains a centralized pool of all Action Units in black-board fashion, so that any software component can read data whereas only backends can update contents. Hence, such an architecture focuses optimization to a single critical part of the software. The AU pool receives constant iterative updates until each AUs has reached its target value, and backends remain free to pick the values relevant to them at the hardware's poll rate capabilities.

Action Unit
Base Value
Delta Value
Target Duration
Delta Duration
Derivative Value
Current Value

Table 3.2: A vector of the internal matrix constituting the pool of 63 Action Units. Allows for proprioceptive information through polling the current value.

One extra benefit of abstraction through the AU pool lies in the transparency offered up to machine learning: algorithms stay unmodified whether they deal with virtual animation or hardware actuation. This makes possible the mapping of facial expression or poses to any other sort of input; for instance, one can imagine the robot learning natural facial behaviour – or to a greater extent natural motion – from data the facial vision systems recorded after human performance.

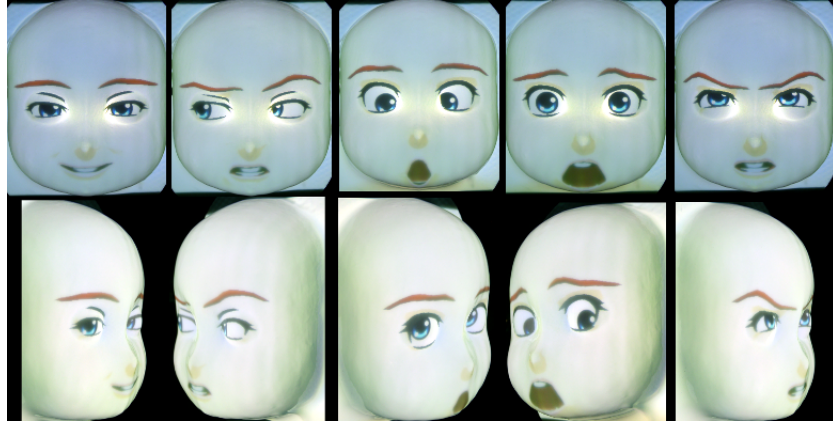


Figure 3.10: Five of Ekman's Six Basic Facial expressions with the 1st design of LightHead: happy, disgusted, surprised, frightened and angry.

Animation and Dynamics

Keeping a simple mathematical formalism of animation ensures minimal efforts to transfer animation states (animators refer to these as animation keys) to third-party tools. The resulting environment stays close to that described in (Saerbeck & Breemen, 2007). An animation state AS (also defined as an expression or pose) can be expressed as a vector:

$$\vec{AS} = \begin{bmatrix} au_0 \\ \cdot \\ \cdot \\ au_N \end{bmatrix}$$

where N is the cardinality of the supported AUs set.

Then, the transition from \vec{AS}_t to \vec{AS}_{t+1} generates an animation \vec{A} such that:

$$\vec{A}_i = \vec{AS}_t + T\left(\frac{\vec{AS}_{t+1} - \vec{AS}_t}{i}\right) \forall i \in [0, d]$$

with d ensuring replay of predefined animations at various speeds.

In effect, each animation state is defined by a set of triplets (AU identifier, target value, duration). Complex animations then consist of a sequence of

animation states bound to a specific duration. Since an animation stops after reaching its last state, this method follows the principles described by Thomas and Johnston (Thomas & Johnston, 1995). To summarize, onset, sustain and decay of animated facial expressions or gestures consist of simple transitions between animation states.

Transitions leave room for dynamics: T , a function monotonically increasing on d , defines the shape of transitions. To keep dynamic functions simple, T only needs to be defined on the range $[0, 1]$ and a factor accounts for each AU distance in $\vec{AS}_{t+1} - \vec{AS}_t$ to compute the amount of movement. Backends extract discreet values whenever possible, adapting the transition to the available computing power - or frame rate. Efficiency of backends is crucial since precise control over the dynamics of facial expression adds more realism to animation: for a study on humans see (Pantic & Patras, 2006), while in (Oda & Isono, 2008) experiments reveal how typical onset profiles do not apply to all facial expressions. In light of these observations, and to display changes of affect, dynamic functions can be redefined during system operation to impact on the whole robot behaviour.

Finally, ARAS's external interface enforces issue of instructions (AU, target value and transition time) in a transactional manner, necessary to start multiple animations at a given time.

Virtual Face

Beyond meeting the needs for controlling the robotic setup, the animation system has been designed for shared and long-term use, allowing scalable realism and exchangeable rendering subsystems.

Independence of the rendering subsystem remains possible because the facial animation does not employ a specific technique. In essence, two main approaches exist: a first family of methods relies on morphing often based on photographs (such as (Pighin, Hecker, Lischinski, Szeliski, & Salesin,

1998)), and a second, more popular technique employs 3D rendering (for a seminal publication, see (Waters, 1987)). Since the CONCEPT project states no intentions to pursue a photo-realistic design, and with the wide choice of open-source 3D modellers available, the latter option was chosen.

A particular face can implement any subset of the FACS' Action Units without modification of the system. To set up a 3D face two methods were tested: using a template model featuring all AU effects on which a texture is applied (more likely to be used for mapping real faces) or an original 3D model scaled to fit the proportions of the template model upon which AUs effects are modelled.

To display the face, a 3D model created with the Blender3D modeller is rendered by its own game engine using an orthographic camera so no perspective distortion occurs, keeping distances constant. As a baseline, the face is modelled without muscular activity (which is equivalent to all AUs set at 0 intensity). Next, the visual effect of each AU is defined by the linear translation of vertices, rotation of objects or hierarchical geometric modifiers (also known as “rigging” for animators). It is possible (and often the case) that some vertices belong to more than one AU and conflicts can arise. However AU normalization allows precise blending of AUs together – additionally, rules can be applied similar to Wojdel (Wojdel & Rothkrantz, 2005) – keeping facial expressions consistent and scalable across 3D models. The method and end result stays close to recent works also based upon FACS, such as (Krumhuber, Tamarit, Roesch, & Scherer, 2012). However one further step remains for the robotic setup: the software compensates for fisheye visual distortion through Blender3D's projection matrix.

As light is projected from within the head volume, it prevents fitting sensors in the mask volume. Hence eye gaze representation is indirect: eyes are displayed as they *should* be, not as they *have to be* when using actual cameras in place of eyes. Mapping eyes' surface to the mask's can use

a polynomial method such as (M. Hashimoto & Morooka, 2005) although other vector-based methods calibration exist, such as (Lincoln et al., 2009). ARAS does not enforce any limitation on eye orientation so the system may be able to support non-anthropomorphic robots. Responsibility for such a work is incumbent upon the CHLAS.

Spine

The spine backend – beyond ensuring safe operation – drives the robot arm from the set of spine AUs, for which AU values represent angles. Although this particular LightHead’s backend directly maps the arm’s joint space, the AU method nonetheless abstracts the kinematic chain of the robot arm. Thus, a different mapping could associate several connected sections of the arm to a single AU and account for another spinal design, such as the long neck of a dragon.

In line with such mapping method, the implemented spine backend relies on forward kinematics (FK). Incidentally, tests of the inverse kinematics (IK) solver shipped with the KatanaHD400s-6M revealed this solution takes too much time to process (sometimes more than 800ms), preventing reactions in a timely manner. Therefore, no particular IK method accompanies ARAS, and if needed, IK would rather fit higher level software such as the HMS.

Since the AU pool provides proprioceptive information, discrepancies between target and actual arm angles arise upon issue of new spine instructions because of the arm’s inherent mechanical and communicative latency. In order to minimize the communication problem, the spine backend source embeds an updated version of the Katana open-source drivers: latency for setting all arm axes is reduced to a minimum.

Finally, on the Katana arm, timed iterations of motor position cannot achieve uninterrupted animation as each new position abruptly stops the arm in a very jerky movement. Instead velocity control achieves smooth

motions through a software PID controller that takes into account the arm's communicative latency, as well as current movement dynamics function.

3.4.4 CHLAS:

The CHLAS sits between the top-level system (the robot's cognitive system) and ARAS.

While ARAS embeds the – robotic or virtual – character's aesthetic design, CHLAS defines aspects of the character's personality. The configuration includes predefined static or animated facial expressions and poses, voice settings and reactive behaviours. Thanks to abstraction, character personality stays transferable to a totally different character managed by ARAS. Currently two characters have been created with this system: LightHead and HALA2 the robot receptionist in Carnegie Mellon University Qatar: an Arabic female, about thirty years of age (see figure 7.1), obviously different from LightHead's childish design. Bringing ARAS to the HALA robot (for details refer to 7.3.2) and meeting requirements called for a new software architecture I designed, and later updated as the CHLAS. Ultimately subsequent iterations initiated by other collaborations opened the range of scenarios this system has to offer.

During human-human interaction, actions and behaviours may be interrupted by a change of thought or an instinctive reaction to an external stimulus. CHLAS is equipped with a way to gracefully interrupt an animation, which accommodates for the character's reactive behaviours. This effectively supports the impression by users that the robot is aware of its physical and social environment, which enhances HRI scenarios.

Similar design principles applied to ARAS guided the development of this software component. Since a thorough documentation can be found in the annex of this thesis, page 183, only outstanding aspects of the CHLAS are presented.

Fusion of Channels

Human behaviour, whether considering a task, communication or interaction, appears synchronised. In this regard, the CHLAS enforces synchrony with its high-level clear-text command interface (a *datablock*) with each one allowing specification of actions on all robotic channels. Therefore, facial expression, vocal utterances, eye gaze, head gaze and generally spine configuration, and finally instinctive behaviour parametrisation (e.g. breathing rate) are guaranteed to be synchronized if sent together in the same command.

For an interactive system to avoid the uncanny valley, time is of essence, and although relaxed requirements favoured non-realtime operating systems, proper scheduling stands as a critical asset. Therefore the core machinery of the CHLAS mainly hosts a software scheduler and the rest of its components adhere to time constraints. Without a reliable system, experimental conditions vary too much over the participants and intermittent delays spoil overt communication channels such as facial animations.

Endowing the robot with speech expanded interactive scenarios beyond non-verbal communication, and later proved an essential asset in collaborations with other researchers. However, poor lip synchronization at best distracts, at worst confuses, depending on a participant's reliance on this cue, therefore particular graphical and computational attention was given to this modality. Lip-sync however remains the task of the text-to-speech (TTS) engine plugged into the system, which should produce both speech samples and phoneme information, to be translated into visemes. In that regard, support to the open-source *espeak*¹⁰ has been implemented, mapping TTS' phonemes to visemes. However the disappointing voice synthesis

¹⁰see <http://espeak.sourceforge.net>

quality called for another implementation. In contrast, Acapela¹¹ proved very satisfactory and unexpectedly close to natural reading speech. Technically, this TTS feeds the system with visemes in the form of lip parting and tension, mouth width and curvature, top and low teeth visibility, jaw opening, and vertical tongue position; respectively mapped to Action Units 25, 24, 20 & 18, 13 & 15, 10, 16, 26 and 93Y & 94.

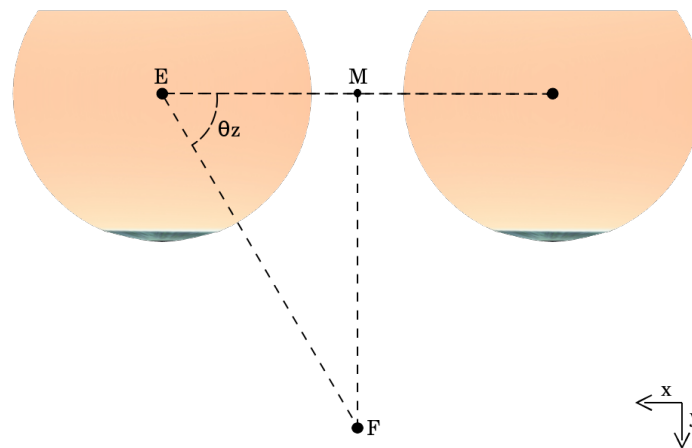


Figure 3.11: Computation of eye (center in E) orientation (Θ_z) from focal point F and eyes mid-distance M.

For the eye-gaze channel, the CHLAS primarily converts the focal vector to vergence. Using a right-hand oriented system (positive Y forward), indirect gaze representation computes each eye orientation from a reference point (i.e: the middle of both eyes) and a relative 3D vector indicating the focal point. Hence, eye orientation is computed using the arc tangent:

$$\Theta_z = -\arctan\left(\frac{F_x \pm E_x}{F_y}\right)$$

with Θ_z the eye orientation on Z axis, F the focal point vector and E the

¹¹see <http://www.acapela-group.com>

eye coordinates as in figure 3.11. Orientation on the X axis consists in a trivial variation.

The system also takes into account human eye-gaze motion based upon publication by Baloh et al. (Baloh, Sills, Kumley, & Honrubia, 1975) whom observed and charted performance of the eyes under various conditions. Despite availability of many other resources on the topic, the implementation simply computes ocular rotation velocity from angular distance and allows room for a more complex simulation.

In a very straightforward manner, part of the datablock specifying spine configuration just requires identification of a spinal section, its desired orientation and time of movement. Spinal section identifiers regroup multiple AUs, abstracting details of the kinematic chain: the cervicals on a humanoid robot could range from a single panning movement to a fully featured anthropomorphic neck with multiple actuators.

Natural and Instinctive Behaviours

Part of human behaviour that conveys the illusion of life serves a biological function: breathing, blinking (as opposed to winking), or saccades for instance, are in fact often unconscious and reveal emotional states. In this manner, CHLAS splits conscious operation of a character and “instinctive” autonomously generated behaviours implemented by several modules. Co-actuation of eyelids with eye-gaze and a natural blinking model (C. C. Ford et al., 2010) are amongst available instinctive behaviours. When conflicting actions are requested by the conscious and the instinctive behaviours for the same robot part, priority is given to the former.

Co-actuation ensures that even cartoonish faces appear natural. For instance, eye orientation has an impact on facial features: when gazing up, eyelids and eyebrows lift to free the field of view. Such detail carries special

importance since the eyes are the primary point of focus during interaction. In effect, the co-actuator instinct recreates this effect for vertical gaze and some visemes.

Gaze control was added as another component amongst instincts. Essentially, gaze control reads the gaze vector to compute orientation of the spine while meeting weighted tolerance constraints of each section. Such a behaviour, rather than being totally instinctive in humans, participates in the illusion of a natural embodiment linking conscious attention and unconscious movement.

Such routines of the CHLAS constitute a repertoire of natural reactions in line with their human counterparts. Humans though enjoy a great control over their embodiment, reflexes and innate behaviours, which they can consciously suppress in favour of other actions. The system offers means to such a mechanism through the deactivation of any routine, at any time, allowing the higher-level system to take over those aspects entirely.

All along the development of the robotic platform, apparent improvements towards natural interaction resulted from each aforementioned modules. Perhaps those components coincide to meet users' expectations and hopefully help in reversing negative first opinions expressed on initial encounter with LightHead. However such informal surveys took place only during public settings and could not justify a specific publication.

3.4.5 HMS

The High-level Management System conceptually holds the place of any software communicating with the CHLAS such as de Greeff's active learning system, which is covered in his own thesis (Joachim de Greeff, 2012). Consequently this package currently stands as a collection of tools to ease endowing LightHead with intentional behaviours, such as motivated interactive learning.

Helper Libraries

To facilitate setting up experiments, three main helpers come along the open-source software: a library of perception algorithms, an advanced state machine and a script player. These benefit the integration, development, and testing of third-party software to be connected to the CHLAS or ARAS. For instance, this approach helped Joachim de Greeff in the development of the graphical user interface (GUI) (see figure 5.3) deployed for his experiment. Alternatively, for non-interactive scenarios the script player stages LightHead’s performances, a method which supported the recordings featured in experiments described in chapters 6 and section 7.3.5. The perception library encompasses both vision and audition, although eventually audition was not deployed over the course of the experiments.

Based upon `pyvision` (David S. Bolme, 2008), the vision helpers exploits functions for camera access, facial and eye detection while hiding their specifics through a configuration file. Built over `pyvision` are the vision routines which feature the generation of focal vector from detected face, and histogram-based facial recognition in the HSV color space. In turn, these primitives stem vision-based behaviours ranging from illumination-independent color perception of objects to histogram-based facial tracking. Also available is the real-time display of the robot’s camera including textual information such as frame rate.

The audition helper provides functions to retrieve and monitor acoustic pressure in decibels for each available channel, thus making it possible to enrich LightHead’s natural behaviour with an auditive reflex. With the stereo microphones embedded in LightHead’s skull, constant monitoring of each channel’s signal power raises events whenever statistical difference occurs or levels reach a specific threshold. Consequently, the robot can appear aware

of its surrounding environment, orienting itself towards the loudest source and reacting to a door violently opened or a person's sneeze for instance.

Event-based Parallel Finite State Machines

Finite State Machines (FSM) formalise state transitions of systems and present the benefit of clear visual representation even for programming-illiterate users. While FSMs can model a variety of logical systems – beyond software and electronics – they fit particularly well the control of automata. However interactive robots and especially research on novelty and curiosity driven behaviours hinge better on event-based finite state machines (EFSM): control depends on events usually generated from the environment. Although directly applicable in this work, the simplicity principle called for the breakdown of a complex EFSM behaviour into simpler models.

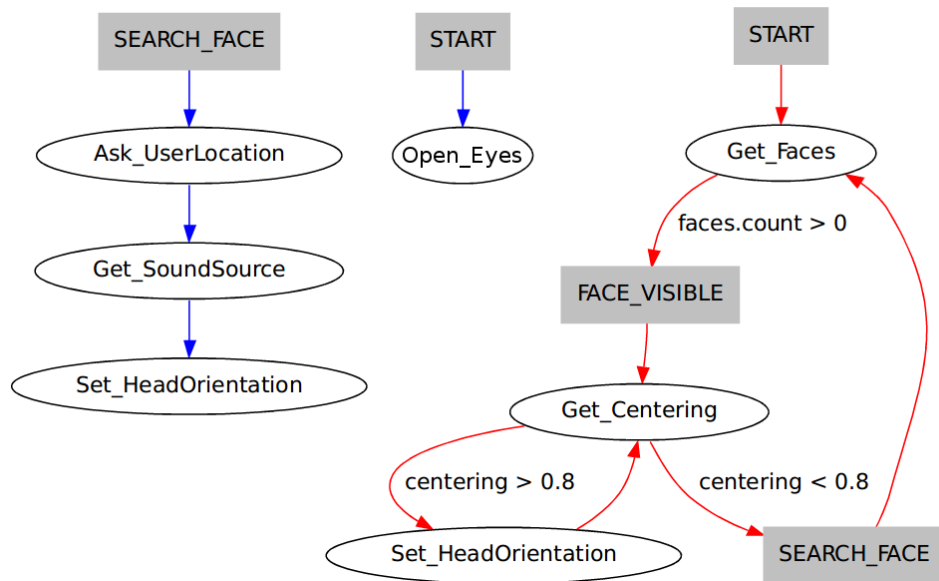


Figure 3.12: A behaviour implemented with two event-based parallel state machines: cognition in blue, face-following in red.

However for these behavioural models to interact, they would need to work in parallel and share states through a *blackboard system*: so is the rationale behind the inception of EPFSM. This approach remains very similar to (R. A. Brooks, 1986) or to the more recent ROS nodes, and certainly shares a common objective with behaviour-based robotics: breaking down complex behaviours into simpler and reusable ones. However, a key aspect of this effort was to create a shareable library of behaviours, which requires portability of logic. Hence key aspects behind the design of EPFSM must be mentioned:

- animation logic shall remain at CHLAS protocol only and remain free of direct/raw actuation, so to allow portability to other supported robotic platforms (e.g. by ARAS);
- states shall be shared between all state-machines and allow their synchronisation;
- events shall be shared states originating from sensing routines;
- logic building shall rely mostly on events;
- basic events shall be provided by the framework (e.g. START, STOP);
- advanced events shall be provided by helper libraries (from sensing routines);
- custom shared states shall rely on basic events or other custom states to create behaviours;
- behaviours shall avoid state name collisions and target self-containment for stackability and modularity reasons;

In particular, the behaviour stackability design requirement allows presenting the user a layered approach: each layer can describe a modal logic (e.g. auditive attention) so that a global (stacked) view displays the full

complexity of the behaviour. In turn, simple behaviours can be layered to finally coalesce in a complex behaviour.

Technically, the EPFSM-based behaviour builder takes for input a definition of shared-memory machines to run in parallel, each in the form of trigger states, boolean function to run, and new state to transit to upon success. Combining helpers and EPFSM, LightHead can adopt a visual search behaviour as well as react to sound as represented in figure 3.12; synchronization of routines depends on the broadcast of the new state to concurrent machines thanks to the blackboard system. Although state machines remain a convenience, they benefit integration of machine learning and behaviour programming by grouping reusable routines in a single logical block, which ultimately eases understanding and updating of algorithms.

Next Reading The following chapters describe a series of experiments evaluating the effectiveness in non-verbal communication of both Light-Head’s hardware and software. The CONCEPT project consists of two complementary topics: human-robot learning and human-robot interaction. The former is covered in De Greeff’s thesis (Joachim de Greeff, 2012) and the latter in this thesis. Eventually though, both domains were merged for the experiment covered in Chapter 5.

The first experiment details investigation of the readability of robotic eye-gaze across non-mechatronic facial displays; while the second experiment assembled a typical scenario of the CONCEPT’s project in which a human teaches a socially guiding robot learner. Further, are presented findings from the crowd-sourced exploration of users’ preferences to robots in relation to their own ethnicity. Finally insights from public displays are reported and commented in section 7.3.5.

Chapter 4

Measuring Eye Gaze

Readability

Reading eye gaze direction is crucial in proto-communicative child-caretaker interactions as it supports, among others, joint attention and non-linguistic interaction. It has been argued (Scassellati, 1998) that reading gaze direction, and by extension, joint attention is important for developmental robotics. While most work has focused on implementing gaze direction reading on the robot, little is known about how the human partner in a human-robot interaction is able to read gaze direction. To the best of my knowledge, no such experiment has been reported previous to the publication of this study (see (Delaunay et al., 2010)), however follow-up works such as (Beskow & Al Moubayed, 2010) endorsed this research direction.

This first experiment addresses the following two questions: (1) What factors influence the ability of people to infer where another (artificial) agent is looking? (2) What is the influence of the physiognomy of an agent's face and eyes on the user's ability to infer where it is looking?

To gain insights, an experiment was devised asking human subjects to judge the gaze direction of four different types of facial interface: (1) a real human face, (2) a human face displayed on a flat-screen monitor, (3) an animated

face projected on a semi-sphere and (4) and an animated face projected on the 3D mask (figure 4.1).

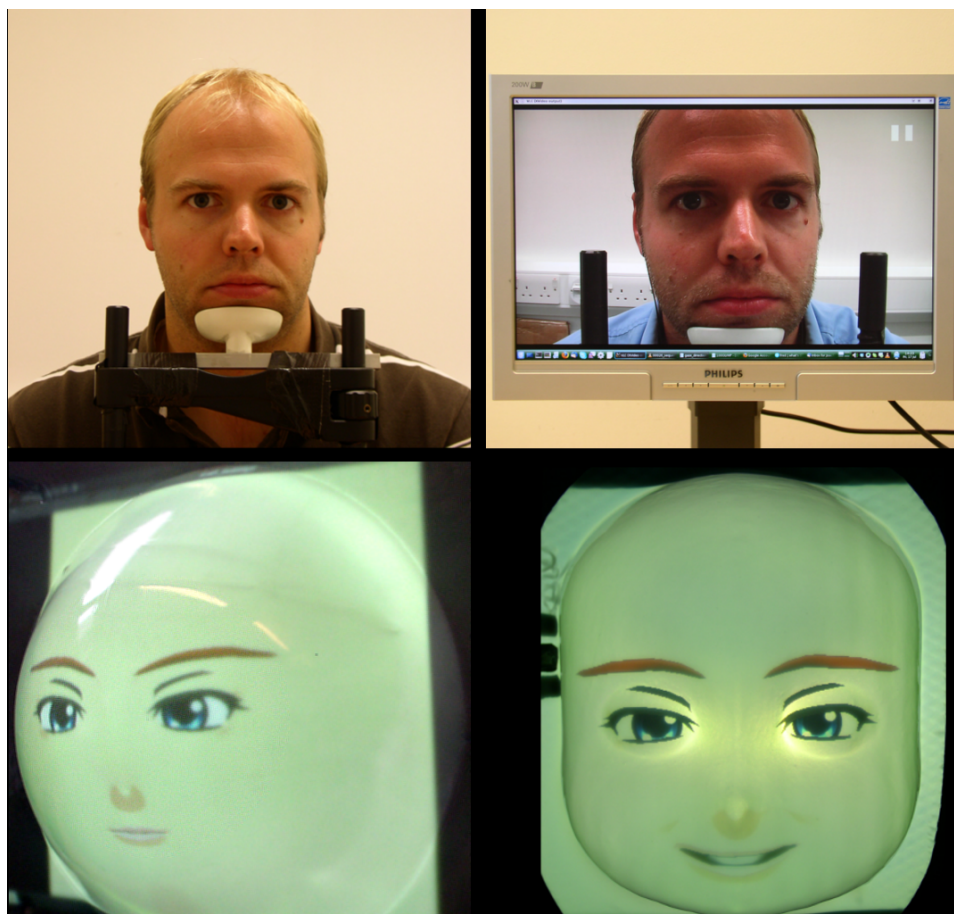


Figure 4.1: The four facial interfaces providing gaze sequences

The rationale behind the different face types is as follows:

1. the real human face will serve as the null-hypothesis, it is assumed that a real human face will work best for assessing gaze direction.
2. the recording of the human face displayed on a monitor serves as a baseline to assess how much the lack of 3D structure influences gaze reading.
3. the LightHead, providing a 3D structure and rounded eyes,

4. the same robotic face, this time projected into a hemisphere.

The last condition serves to evaluate the technology of Hashimoto (M. Hashimoto & Kondo, 2008), who evaluated a similar robotic setup. No android robot face was introduced, such as the Albert Hubo or the Ishiguro’s androids due to budgetary constraints.

Human eyes are unique: no other other animal—including primates and apes— have such a large visible sclera to iris ratio (Kobayashi & Kohshima, 1997). The spherical shape of the eye also facilitates reading gaze direction, which allows one to infer the position of the iris not only when facing a person head on, but also when seeing someone from viewpoints other than frontal.

The information gleaned from the spherical shape of the eyes is distorted when a face is displayed on a 2D surface. However, under certain conditions the distortion is minimal: if a video of a person is shown on a screen with the viewer sitting at the relative location where the camera was, the distortion is minimal. The reason for this is that the 3D to 2D transformation is consistent: we have no problems reading gaze direction from the video because we are *aligned* with the camera’s line of sight. However, by moving away from this ideal position, the visual transformation is no longer relevant to the viewer’s perspective and other cues are needed to read gaze direction.

Figure 4.2 illustrates the 2D gaze interpretation problem, described in two ways: the Mona Lisa effect occurs with a face represented as *looking at us*, which persists regardless of the viewer’s perspective; while the Wollaston’s effect (Wollaston, 1824) occurs with eye manipulation of *head gazes* not directed to the viewer. In real environments, we remain free of this effect as depth of vision and other visual clues (such as reflections) help constructing geometric information. As such, multiple observers at different viewpoints can read a person’s gaze, although with respect to the observer-subject dis-

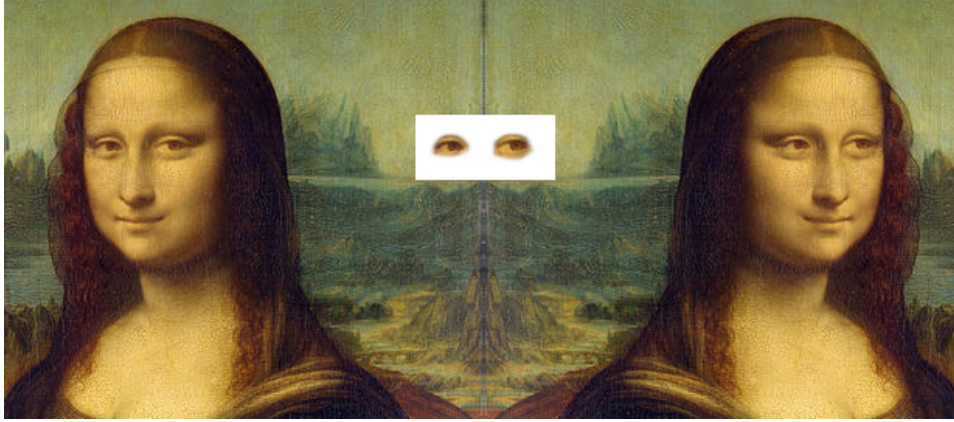


Figure 4.2: Wollaston's effect applied to the Mona-Lisa: the faces appear to gaze at different location although the pairs of eyes are identical.

tance. Similarly in multi-party scenarios, ensuring readability of robotic eye gaze reading benefits the transmission speed of non-verbal messages such as turn taking, as opposed to slower head gazes.

4.1 Experimental Protocol

Two different viewpoints for observing four facial displays were evaluated. With only a single experimental sequence for all participants, a participant's performance could increase over the four sessions, and such a training effect might have tainted the results. Therefore it was ensured the face presentation sequence varied over participants, shuffling the order of sessions for each pair of participants. Under such a principle, 24 (4!) unique sequences of facial display presentation are possible, each of which could be tried twice by the same pair of alternating participants. Hence 24 participants were arranged following 2 different sequences from either one of the viewpoints, which generated a total of 96 records, that is 12 records for each condition.

Participants sat 1.5 meters in front of a transparent screen, behind which they could see the facial display at an additional 0.5 meters (see figure 4.3).

Display	Straight Viewing (0°)	Side Viewing (45°)
Natural Human face	P1	P2
Human face on flat monitor	P2	P1
Animated face on semi-sphere	P1	P2
Animated face on 3D mask (LightHead v1)	P2	P1

Table 4.1: Example of an experimental sequence for a pair of participants. P1 and P2 swapped their seats and repeated the sequence.

One participant faced the display straight (aligned with its normal) and the other at a 45 degrees angle from the facial display's normal. To obtain a metric, the transparent screen appeared as a grid, divided in 10 rows and 10 columns, which bore the numbers 0 to 99 from top left to bottom right, each cell measuring 5x5 cm. The grid stood upright between the participants and the facial display so that the distance from eyes of the face to the numbers of the grid would increase evenly from the center of the grid; this would not be the case if the numbered grid was laid flat in front of the face. The position and size of the grid also ensured downward facing eyelids could not hinder the interpretation of gaze direction when gazing at the bottom of the grid.

A single session consisted of the face looking at a sequence of 50 randomly generated numbers, switching to the next one after a fixed delay of 5 seconds. As numbers are pseudo-randomly generated, the participants were instructed that the same number can appear multiple times in a number sequence.

Once a number was gazed at, an auditory signal was given indicating to the participants that they could perform their observation. A delay of 5 seconds was long enough to give the (human) face enough time to find

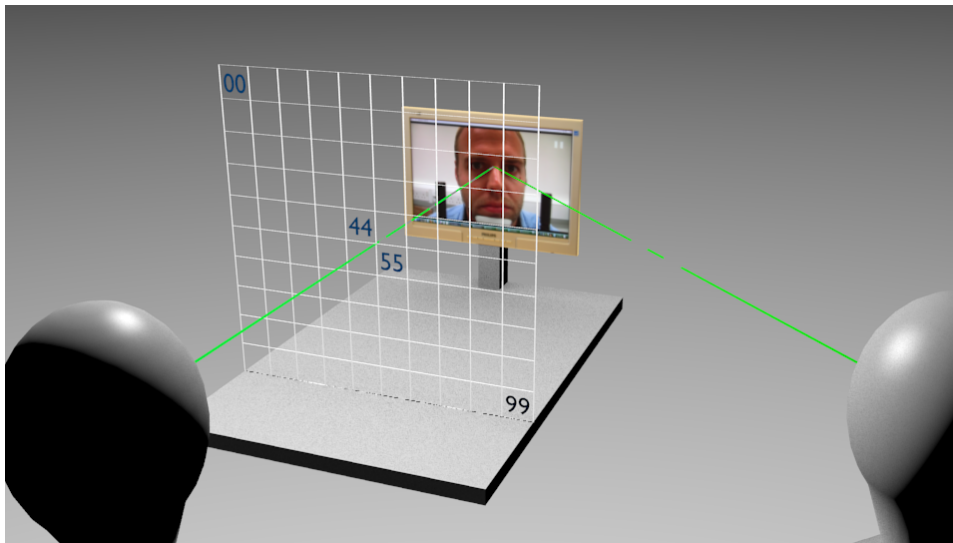


Figure 4.3: Experimental setup: pairs of participants seated viewing the four displays straight and at 45° .

the proper number and for the participants to write down their observations afterwards. When the face was a human (one of the experimenters), the number sequence was played over earphones worn by the experimenter so it could not be heard by the participants. In the case of the video, the face consisted of a pre-recorded sequence of the same experimenter looking at a number sequence. In the two animated faces cases, the number sequence was generated on the fly and fed into the animated face control module. The same auditory signal was played when the face was looking at the next number to ensure consistency among sessions.

The participants were asked to write down the number they thought the face was gazing at on a paper sheet. Handwriting allows participants to quietly report – or correct – their results in a very natural way, and only requires basic equipment. Participants were also asked to perform as best as they could and not to cheat by looking at each others' notepads. Observing the participants while they performed the experiment ensured they were obeying these rules.

The sequence of numbers written by each pair of participants was compared to the actual sequence and the difference was calculated using the euclidean distance between the cell on the grid the participant reported and the cell the robot gazed at. In this way, the difference between a participant's sequence and the real sequence is expressed as a mean error distance.

4.2 Results

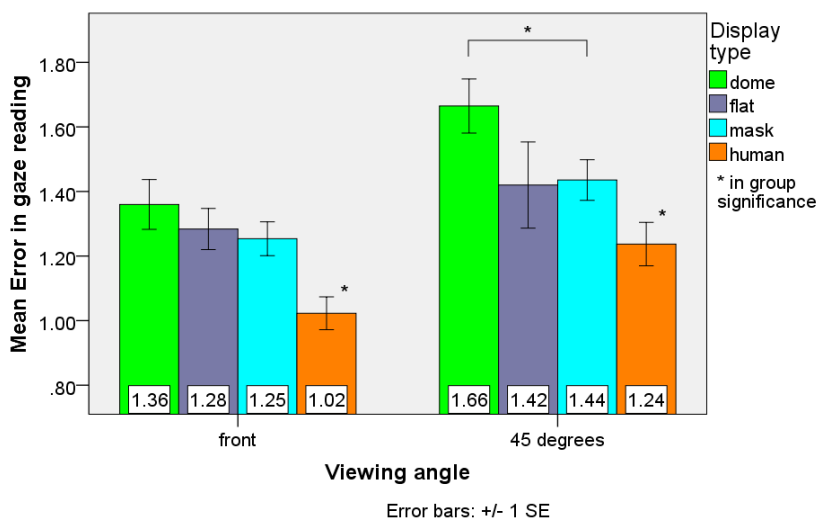


Figure 4.4: Mean Error (cm) in gaze reading for each display for a distance of 2m with front and 45° seating positions (N=12) .

Not rejecting the null hypothesis, performance for the human face was best (lowest mean error): when having to guess at which number the human face was looking, the participants had an average error of 1.13 when combining data from both viewing angles. As the numbers were 5cm apart, this means that participants were in average about 5.65cm (5cm x 1.13) off in guessing where the real human gaze rested. The mask and flat faces appear to be next in accuracy for guessing where the eyes are looking, followed by the dome face; however results are not statistically significant.

A 4 x 2 analysis of variance (ANOVA) on gaze interpretation error showed main effects of both display type, $F(3, 88) = 8.121$, $p < .01$, and looking angle, $F(1, 88) = 14.438$, $p < .01$. However, no interaction effects were observed, $F(3, 88) = 0.419$, $p = .740$.

Post-hoc comparison of the ANOVA using a Tukey test shows that the participants' performance between the human condition and all other conditions was significant, while this was not the case for any other comparison (see table 4.2).

condition	versus	<i>p</i>
human	dome	0.000
	flat	0.027
	mask	0.035
mask	dome	0.146
	flat	1.000
dome	flat	0.176

Table 4.2: Statistical difference in mean performance of different display types

A first observation is that participants for all conditions performed above chance, and that the error distance is less than expected by chance for all displays. The difference between faces was tested for significance using an unpaired two-tailed t-test. The difference in performance between human versus all other faces turned out to be significant, as was the difference between mask versus dome, while the difference between mask versus flat and flat versus dome was not.

Unsurprisingly, the difference in performance between the two different viewpoints revealed it is much easier for participants to determine the gaze direction when viewing the facial display straight, as opposed to a side view (see table 4.3). An ANOVA for the independent variable “display type”

	dome	flat	mask	human
<i>p</i> value	.014	.367	.037	.018

Table 4.3: Statistical difference tests for the four different displays between the two different angles. The difference between viewing angles for dome, mask and human is significant, and for flat it is not.

showed statistical significant difference at 0°: $F(3, 44) = 5.992, p = 0.003$; and at 45°: $F(3, 44) = 3.690, p = 0.019$. This difference between viewpoints was significant for the human, mask and dome, but not for the flat screen, due to the large variance in performance.

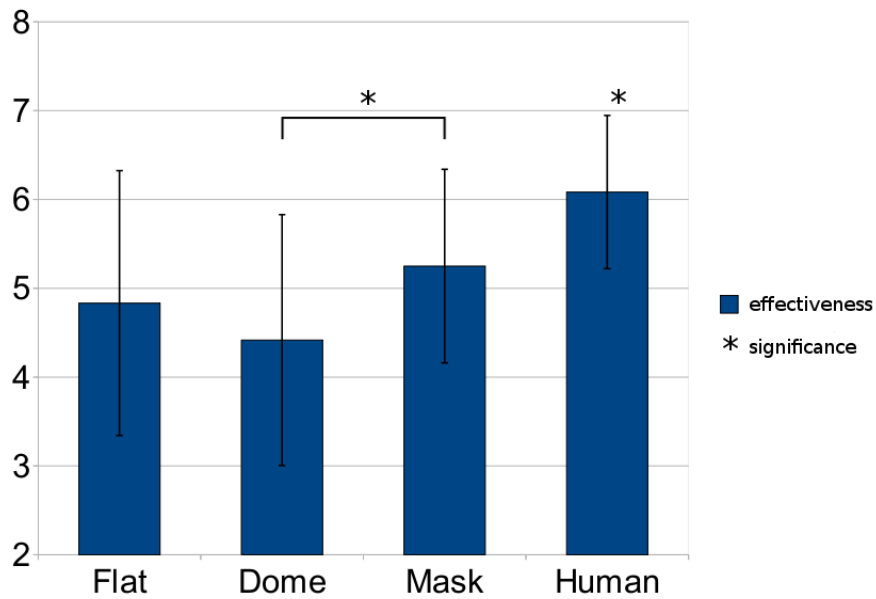


Figure 4.5: Mean of user preferences for each display (N=12).

After performing the experiments, participants were asked to subjectively rate their experience describing how effective they found each of the four different faces at conveying information about gaze direction. This was rated on a seven-point Likert scale, with the range: 1-very ineffective,

2- ineffective, 3-somewhat ineffective, 4-undecided, 5-somewhat effective, 6-effective, 7-very effective. Results show that participants find the human the most effective in terms of gaze information, followed by mask, flat and dome (see figure 4.5). The difference between human and all other faces is significant, as is the difference between mask and dome, while the difference between flat and mask and flat and dome is not.

4.3 Discussion

With this first study, such inquiry into the influences on eye gaze reading was inherently limited: for instance conditions where a human face is projected into LightHead's mask were not tested as this proved technically challenging and highly uncanny, nor was tested the CGI face on a flat screen. Nonetheless the most representative conditions were selected and this helped unravel the behavioural aspect merged in aforementioned geometric and aesthetic concerns; a process detailed hereafter.

The results indicated a 3D mask with a projected animated face is a valid setup for which participants are still able to infer the gaze direction. Participants' increase of error when shifting to side view of the mask remained similar as with the human. Assuming participants managed to read the human's eye gaze instead of guessing it, the similarity of the standard deviation between mask and human conditions suggests that participants manage to read it too, regardless of the viewer's position. This was reflected in the fact that standard deviation for the flat screen is more than twice the mask's on the side view condition, where it seems the gaze is at some vague distant point in space, rather than at the number grid. Besides, this correlates with the perception that the face is looking at the corners of the screen when it is supposed to look at a number in the grid corner.

However, the relatively good performance of the flat screen was interesting, especially considering the side view condition. Comparing results

of the front view between the mask and flat-screen video, participants performed more or less equally well (difference in performance is not significant). Upon further investigation, participants occasionally reported trying to *reason* about the gaze location of the human video rather than naturally detecting it, finding it helpful to see the eyes of the human face employing recognisable search strategies when looking for the next number. For instance, when switching from number 12 to 86, typical human search behaviour would be to drop the eyes first from the 2nd line (20-29) to the 8th line (80-89), and then move along the horizontal axis from 82 to 86. It might be the case that other participants did not consciously observe this but were nevertheless sensitive to this information, although it remains evident participants engaged in gaze guessing in the side view condition. In contrast, the projected animated face (being computer controlled), would drop its gaze directly from one number to the next. This suggests that animated faces missing visual search behaviours might impair the interactants' ability to infer robotic gazing direction, and that their presence may be also beneficial in other scenarios.

Geometrically, the participants' high error with the dome came less as a surprise. Moreover, no dome compensation was in place, although even if present, the low curvature might not have proven visible to the naked eye; however this condition exploited a normal vector calibration. The reasoning on the high eye gaze reading error is that it represents the condition with the less visual cues: there is no nose geometry to help assessing the head gaze (Wollaston effect), aesthetics are minimal since no realistic cues on the face suggest geometry, and no search behaviour was implemented.

Finally, measured errors are provided to help a robot designer select a facial display type, table 4.4 summarises average angular and distance errors for each viewpoint and condition. Those values only reflect results obtained

condition	viewpoint	distance error (in cm)	angular error (in degrees)
human	0	5.1	5.86
human	45	6.2	7.14
mask	0	6.25	7.2
mask	45	7.2	8.31
flat	0	6.4	7.37
flat	45	7.1	8.19
dome	0	6.8	7.84
dome	45	8.3	9.6

Table 4.4: Participants’ mean euclidian distance and angular errors in gaze reading from 1.5m, for both viewpoints and each condition. N=12, eyes-to-object=0.5m.

with our particular set up. It worth noting that the gazer-cell distance is greater for cells further away from the grid’s center, hence it is likely that those values would slightly change with another run of the experiment, however current results suggest a sub-centimetre discrepancy.

Chapter 5

Robotic Social Influence in Human Tutelage

The experiment described in this section embodied a key objective of the CONCEPT project by framing a controlled evaluation of de Greeff's and Delaunay's work (Delaunay et al., 2009; de Greeff et al., 2009). Consequently, this occasion tested the viability of the integration as a proper social robotic system.

Advances in machine intelligence and in the concept of information (for a seminal book, see (Floridi, 2004)) initiated a paradigm shift, departing from good old fashioned artificial intelligence to account for the unpredictability of unstructured environments. As emphasized earlier in this document (see section 2.1.2 in particular), social and personal robots are to evolve in such environments and need to specialise in natural interaction, a crucial aspect of their ability to expand their knowledge through human tutelage.

The following study aims to investigate how the LightHead acting as a learning robotic agent can acquire categories through interaction with a human tutor, acknowledging previous studies demonstrating the improved effectiveness of social communication within learning mechanisms (Thomaz,

2006; Cakmak, Chao, & Thomaz, 2010). Exchange of non-verbal signals serves as a regulatory system between learner and tutor to assess and adapt the transmission of essential constituents of a particular idea or skill. The hypothesis is that LightHead’s provision of social cues can influence the teaching strategy of participants.

5.1 Learning Mechanism

The robot’s underlying machine learning system employed in this experiment has been the sole work of Joachim de Greeff which links grounded knowledge¹ with a mechanism for the transfer of linguistic symbols.

For the learning mechanism, Steels’ *Language Games* (Steels, 1997) are employed as a means for an agent to learn a new vocabulary from another tutor agent. Modelling the dynamics of linguistic interactions between agents, several variations exist as the general mechanism allows for the manipulation of a range of intrinsic parameters such as the number of agents simulated, the communication properties or even the agents’ learning strategies (for further exploration, see (Steels & Kaplan, 2000; Belpaeme & Bleys, 2005; P.-Y. Oudeyer & Kaplan, 2007; P. Oudeyer & Delaunay, 2008)).

A *context* and *topic* represent respectively the environment and one of its elements. Each topic is directly accessible to all agents which describes them to others through a *word*, and each agent maintains its own word/topic matrix. In this experiment only two agents play the game. To be more precise, an interaction of the game unfolds according to the following steps:

1. the speaker agent selects a topic from the environment
2. from the topic, the speaker looks up a word in its word/topic matrix and communicates it to the other listening agent

¹Details of the conceptual space used to represent knowledge can be found in (de Greeff, Delaunay, & Belpaeme, 2012)

3. from the word (likely to be unknown to the listener at the beginning of the game), the listener looks up a topic in its own word/topic matrix
4. a comparison of the communicated and interpreted topic (which relies on an alternative method such as pointing to an object) determines the success of communication
5. a strategy decides which agent updates its word/topic matrix.

The mechanism guarantees that an iteration of games between the agents eventually leads to a shared lexicon, provided the modification of the matrix avoids confusion.

5.2 Experimental Protocol

The grounded elements at the base of the concepts learned by the agent, are exemplars of the Zoo Data Set from the UCI Machine Learning Repository (Frank & Asuncion, 2010), a simple database constituting of 7 different categories: MAMMAL, FISH, BIRD, INVERTEBRATE, AMPHIBIAN, INSECT and REPTILE. As MAMMAL included “girl”, this exemplar was removed to avoid confusion as pilot studies revealed some participants had no idea humans were mammals, which left 100 exemplar animals with 16 different properties such as “airborne” or “predator”.

5.2.1 Simulated Experiment

A pilot test of the experiment simulated the baseline version of the game² (non-AL) against an active learning version(AL) in which the learner agent

²To be exhaustive, the form of active learning in use for this experiment tightly matches (de Greeff et al., 2009), except for a specific modification of the topic selection which proves “marginally more effective” (see (Joachim de Greeff, 2012)). However this doesn’t carry particular relevance to the HRI aspect of this experiment.

actively influences the tutelage session by deciding on the topic which it knows least.

Classic language games require about 10^4 iterations to reach a consensual vocabulary with only few confusions within the agent population. Although studies with a high number of agents might use a greater order of magnitude, even 100 iterations could not be practical with humans. Therefore, a simulation was limited to 50 interactions – in line with the planned human-robot version – and 50 simulations were conducted to obtain an average measure. Each context consisted of 3 randomly selected animal exemplars.

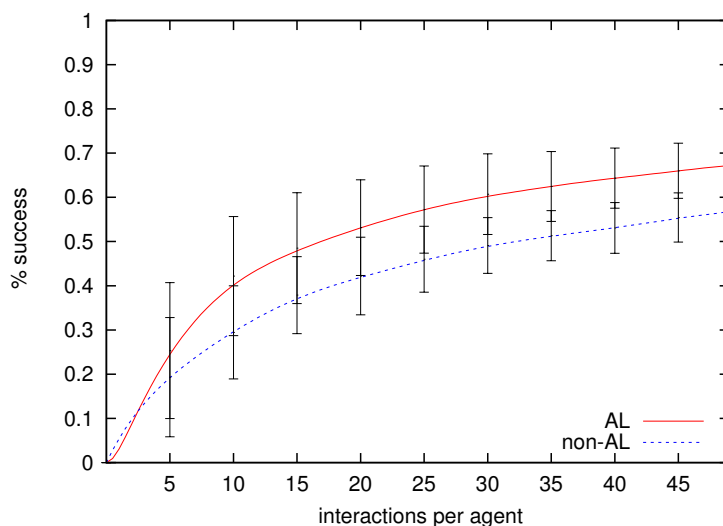


Figure 5.1: Comparative overview of the success in communication of the baseline and active learning strategies (AL) of 50 language games. Values are averaged over 50 complete simulations.

Figure 5.1 illustrates the better performance of the AL condition: only after 5 iterations, AL achieves and keeps a higher game success rate. In addition, a two-sample t-test indicates the difference in overall performance for the two conditions is significant ($p < 0.001$).

5.2.2 Robotic Experiment

For the robot learner and human teacher version of the experiment, 20 females and 21 males (mean age = 24.8) were recruited around the university’s campus and paid £7.50 for their participation. Each of them were then randomly assigned to interact with either the AL or non-AL version of the robot (LightHead v3). Of course, the robotic version of the experiment was set to reproduce the experimental protocol of the simulation: each session consisting of 50 rounds, with a 3-exemplar context.

For the robot to convey social signals, LightHead’s face had to remain visible at all times for the sitting participant. More importantly, for the AL condition, the eye gaze direction of the robot also needed to effectively cue the participant, and allow the latter to identify the exemplar corresponding to the robot’s learning preference.

In between them, a touch-screen mediated the context and allowed the participant to select any of the displayed elements. To guarantee a robust protocol, the touch events generated by the participant were directly channelled to the robot, although LightHead adopted behaviours providing the illusion of natural perception.

Preliminary to the participant’s session, the experimenter gave a brief explanation of the guessing game, followed by practice rounds on colour categories, in order to habituate the participant to both the robot and game mechanics. Upon expressing satisfaction with the tutorial, the experiment started and the participant proceeded with the teaching of animal categories.

Present during the experiment, the experimenter sat a couple of meters behind the participant but deliberately appeared working on something else. After the experiment had begun, whenever a participant interrupted the interaction to ask a question, the experimenter would reply with an evasive answer as to not give away any clues. His presence ensured consistency of



Figure 5.2: Experimental setup: the participant faced the LightHead robot; both shared the context procured by the touch-screen.

the experimental protocol, and allowed him to restart the projector³ on two occasions.

To present the categories (words) and exemplars (domains), a graphical user interface displayed on the touch-screen (figure 5.3). For every round, the touch-screen displayed 3 random pictures of animals along with 7 buttons each labelled with an animal category.

The interaction mechanics of a game followed this sequence:

1. LightHead examines the pictures and asks the participant to think of one of the available animal and its category
2. the human teacher silently decides on the animal picture then he/she touches the corresponding category button on the GUI
3. LightHead acknowledges the category and guesses the animal selected by the participant

³This experiment was conducted using the version 3 of the LightHead which, at the time, relied on a ShowWX+ and did not include the cooling method mentioned in section 3.2.3. Overheating caused the shut-down of the projector, hence a blank robot face.

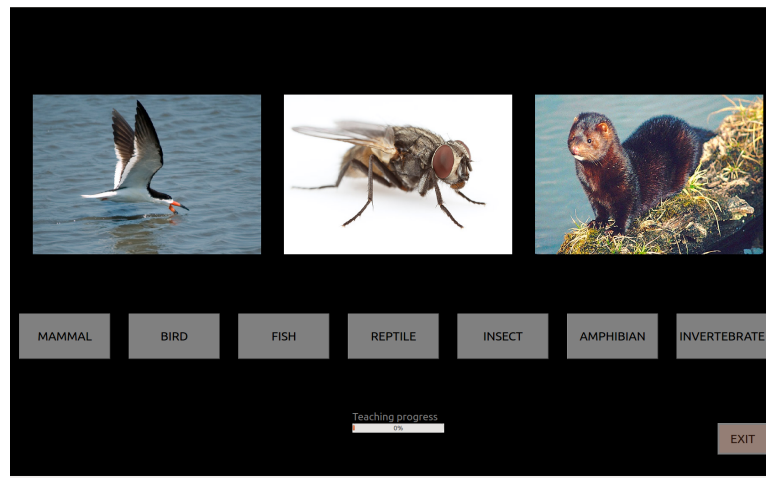


Figure 5.3: The GUI which presented the exemplars and words to the participants.

4. the teacher communicates his initial choice to the robot by touching the corresponding GUI picture
5. the robot confirms the outcome of the game with either a positive or negative facial expression and verbal statement, and the next round starts.

Production of synthetic speech supported game transitions from one step to another. Non-verbal language alone might have introduced uncertainties regarding the state of the robot, moreover verbal statements reiterated some instructions of the game in a less formal manner. Overall, it stands to reason verbal communication would reinforce the teacher-learner engagement. For instance, at step 3, after LightHead gazed at the guessed animal, the robot uttered a sentence such as “is this the topic?” or “is this the animal you were thinking of?” in an attempt to appear involved in the game. A full list of statements is available in the annex p.204.

Even if LightHead’s facial expressions and nods were congruent to uttered speech, significant attention was given to the non-verbal behaviour,

also seeking to elicit engagement. At step 1, such efforts supported an essential cueing difference: in the non-AL condition, the robot either moved back a bit and gazed at the participant; while in the AL condition, it alternated gazing at a particular exemplar and to the participant, as well as making a verbal statement along the lines of “what about this one?” or “I would like to learn this”. Additionally, LightHead exploited its camera to perform face tracking and visually gaze at the participant, participating in the illusion of life. The gaze was interrupted by blinks, occasional gaze shifts and indeed, attention to the pictures.

At the end of the experiment, the participant was asked to fill in a questionnaire (reproduced in the annex p.209) about their experience with the robot and their topic choice strategy. Then, they were given a short debrief which was also the opportunity to ask questions.

5.3 Results

With two discarded sessions due to the overheated projector (participants #6 and #7), 19 AL and 20 non-AL entries constitute the experimental data.

As a global result, all participants succeeded in teaching the animal categories. Figure 5.4 illustrates the similarity with simulation (cf. figure 5.1): after 5 games, AL achieves and keeps a higher game success rate. On the 10th game played, a two-sample t-test indicates the difference in performance for the two conditions is significant ($p < 0.001$). Moreover, the average success for AL was 0.626 ($SD = 0.077$), and 0.566 ($SD = 0.087$) for non-AL; this difference is significant (two-sample t-test, $p = 0.028$).

Once all experiments were completed, it emerged that participants might have been confused by proposed contexts with two (or three) animals belonging to the same category. For instance with 2 mammals (e.g: an otter and a squirrel), the teacher might have selected the otter as a topic hence

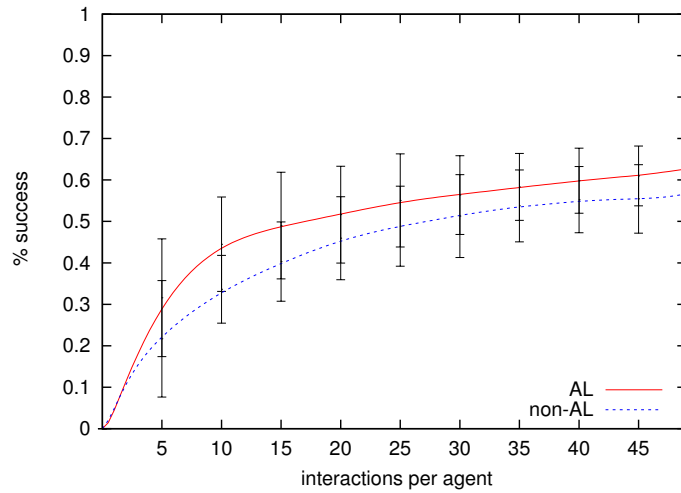


Figure 5.4: Comparative overview of the success in communication of 50 language games. Values are averaged over 50 complete simulations (N=39).

communicate the category MAMMAL, while the learner might have guessed the other MAMMAL, i.e. the squirrel. Table 5.1 provides a general overview of the rate of proposed contexts with multiple animals belonging to the same category (ambiguities) and the rate of participants selecting an animal in such category (confusion⁴).

Experimental Condition	Ambiguities		Confusions	
	Rate	SD	Rate	SD
without Active Learning	32%	0.07	44%	0.20
with Active Learning	30%	0.06	56%	0.10

Table 5.1: Overview of the occurrences of ambiguities and subsequent confusions over all experimental sessions (50 rounds, 39 participants).

From a +12% increase of confusion compared to non-AL, it appears participants in the AL condition did not specifically try to avoid confusions. As

⁴The participant confusion is not established in this experiment, so the wording expresses a potential participant confusion instead.

such potential misunderstanding of the game was unexpected, the experiment did not include tracking the participant's initial choice. Interviews of few participants revealed some of them concurred with the robot on proper classification whether or not the animal was their own choice, therefore leading to a successful guess. Unfortunately it is not known if most others considered these occurrences to be a failure.

5.3.1 Impact of Active Learning

Because this experiment represents a conclusive aspect of the CONCEPT project, the impact of the AL condition bears a significant importance and is likely to influence the nature of the future work.

To obtain a measure of teacher and learner *alignment*, the occurrences of the same topic preferred by the learner and selected by the tutor were counted over the total number of rounds in the game. For instance, if the learner's preferred topic was chosen by the tutor only half of the time, alignment is .5. In the case of the non-AL condition, topic alignment was .32 (SD=0.08), that is, almost 1 in 3 times did preferred⁵ and selected topics were similar, which corresponds to a random choice since the robot did not provide any cue. On the other hand, in the AL condition, it was .56 (SD=0.18), signifying alignment of topics occurred above chance ($p < 0.001$) and confirming that, on average, half of the participants' rounds were influenced by LightHead's cueing behaviour. However, there is important variation in how sensitive participants are to the robot's guidance: participant #16 obtained .38, while participant #8 obtained .94.

From an observation of figure 5.5, the spread of results within the AL condition is clear, thus no correlation has been found between alignment

⁵Indeed in the non-AL condition there is no agent preference *per se*. The word is kept for ease of comparison with the AL condition.

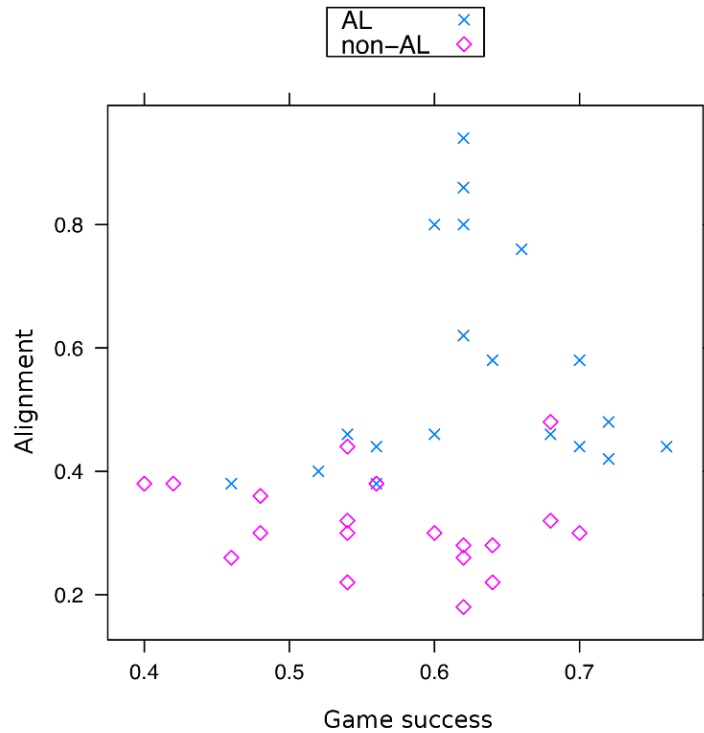


Figure 5.5: Distribution of tutors’ alignment with LightHead’s cues against game success for AL and non-AL conditions.

of interactants and guessing game success (Pearson’s $r = 0.09$). Indeed this discrepancy is further matched with participant’s answers to this questionnaire’s question “On what basis did you choose the animal examples as topic?”: only four participants acknowledged following the robot’s cues. Most of the other participants reported instead having employed their own strategy, driven by their knowledge or affinity towards the animal exemplars. Some others answered their choice was random, however their alignment value suggests they were enticed to follow the learning path suggested by the robot.

5.3.2 Other Aspects

A general gender analysis of the game in active learning condition reveals alignment scores of 0.59 for female participants and 0.54 for males. However this is not significant, and again, this is due to the large differences between individuals.

Further investigation with 2-factor ANOVA on gender and active learning condition yields significant interaction effect on game success ($p = 0.04$). Figure 5.6 illustrates how effective active learning was for female participants, as opposed to the lack of effect with males, suggesting female participants are more sensitive to social cues offered by the robot. Such results present more contrasted differences than previous psychological studies on female's ability to decode non-verbal behaviour (J. Hall, 1978), and that observation might be related to the fact that a tutelage scenario is a goal-directed interaction.

More interaction effects appeared with analysis of the 7 Likert scale questions of the debriefing survey (see annex C). For each of the seven questions, the null-hypothesis is the absence of gender influence on the various questions asked (two-sample t-test). Moreover, the interaction effect across gender and active learning condition on the participant's rating was performed with a 2-factor ANOVA. Answers of interest are reported hereafter and commented.

Question 2 "How do you rate the robot's behaviour?", presents a significant interaction effect between gender and active learning since female participants rated the actively learning robot to be more natural: $F(1, 35) = 8.517, p = 0.006$. The perception of the robot as a natural partner suggests females of this study felt the robot's behaviour appropriately matched the interactive scenario, a result to compare with their better game performance in the AL condition. For female participants, perhaps the AL condition elicits better incentive to adapt to the robot because it matches their expectation

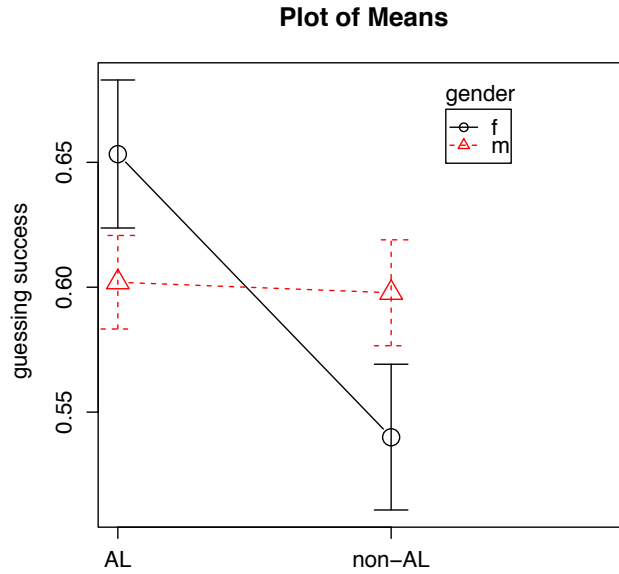


Figure 5.6: Influence of LightHead’s learning behaviour on guessing game success, split by gender and learning condition.

of a learning child partner.

Question 4 “Who was in control of the teaching sessions?”, female participants judged the robot to be more in control than male participants. The gender factor is in this case very significant (two-sample t test: $p = 0.002$). However no interaction effect was found: $F(1, 35) = 1.814, p = 0.187$. This result rejects the null hypothesis for this question and supports the idea that females of this study considered more important to let the robot take control. Besides, one might propose the judgements is closer to their expectations than their experience.

Question 8 “How smart do you think the robot is?”, presents a significant interaction effect⁶ since female participants of the AL condition deemed the

⁶In this study α was set to 0.05, however in stricter studies $\alpha = 0.01$ or below thus this result would not be considered significant.

robot to be smarter: $F(1, 35) = 6.229, p = 0.017$. A tentative interpretation of this result might be that females of this study consider the robot's behaviour to match those of curious children, although a more probable explanation might be that the robot utterances in the AL condition were supportive of this perception.

As general observation, on average female and male participants present an opposite rating trend to questions 2, 4 and 8 depending on the robot's learning condition: male participants judged the robot less natural in AL compared to those in non-AL; they also perceived being more in control in the non-AL than in the AL conditions; and slightly smarter in non-AL than in AL condition.

Finally, an analysis of the personality questionnaire – of the Big-5/OCEAN dimensions (see (Groom, Bailenson, & Nass, 2009)) – only revealed the conscientiousness dimension to be most correlated with the alignment score, but this was not significant (Pearson's $r = 0.424, p = 0.071$). Personality traits might help recognise trends in participant's ability to interact with a social robot, although much more evidence is needed to draw sound conclusions. Discussions with psychology researchers hinted towards greater openness with the participants scoring better in game success and alignment, however no significant correlation stands out in that regard.

5.4 Conclusion

This study demonstrated the active machine learning system provides consistent improvement over non-AL across both simulated (teaching agent always followed learner topic) and embodied interactions (participant had no particular directive). In general, a congruent deployment of social cues (i.e: head gaze, alternated eye gaze / eye-contact) and enticing speech proved a valid approach to the implementation of robotic active learning, which was

more effective on the female participants of this study. Acknowledged gender differences suggests a need for further studies on AL strategies matching the tutor's gender, and additionally adapting the robot's social cueing strategy in regard to the tutor's teaching method (authoritative or socially guided).

Follow-up studies might replicate the experiment testing robustness to different embodiments with a socially-diminished robot (e.g: only capable of head gaze) and fully fledged humanoid such as ASIMO, or even an android. Finally, follow-up studies should employ a dataset containing simpler categories to ensure full participant confidence in their ability to classify exemplars, as well as addressing confusions by instructing teachers to accept valid guessed classifications from the robot.

Chapter 6

Influence of Robot Ethnicity

Across the globe there is a great diversity of cultures and phenotypes, a diversity to which people are particularly sensitive. Given that people feel more comfortable when operating in an in-group (i.e. people with a similar cultural and educational background, language and ethnicity), a social robot designer could hypothesise that robot-user cultural and ethnic alignment would be desirable, or just shun this aspect altogether by endowing the robot with a non-humanlike face, as no ethnicity-aware design principles exist yet.

While there are several psychological studies which investigate inter-cultural and inter-group preferences (Hewstone, Rubin, & Willis, 2002), most of these only cover North American culture. This is also the case for studies employing avatars: in (Groom et al., 2009) an immersive virtual environment reflects the user's avatar with a Black-American ethnicity, while in (Gong, 2008) users expressed ethnic preferences through the selection of the avatars' facial properties to constitute teams of various purposes. However, there are no studies on the transferability of these insights to robot embodiments. In addition, a world-wide collection of data, instead of a study limited to one geographical region, would be preferable to reveal

common influential factors. The approach reported here focuses on users' preferences rather than discrimination: within a global world, social robots could and should offer a choice of ethnic appearances or could provide a means to display the face most likely to please a user.

The motivation behind this experiment was to question whether people have preferences towards social robots overtly displaying a specific ethnicity, and to investigate the nature of such an effect if present. To restrict the study somewhat, a decision was made to limit the expression of ethnicity to the face, and capitalise on R-PAF's ability to digitally update the facial texture: colour of skin and eyes, size of features and other minor – yet perceptible – modifications.

Equipped with a R-PAF, the LightHead robot supports such an experiment without requiring new hardware, whereas androids and their modification (i.e. creating a new facial skin) still requires prohibitively high costs. Although all participants in the study reported being adults, the LightHead robot's child-like design brings this study close to (Mahan, 1976), which addresses identification and preference of Black or White children, and (Jordan & Hernandez-Reif, 2009) on skin tone preferences using cartoon characters. Additionally, a further study of inversion effects, first reported in (Mahan, 1976), is desirable; some cultures present this trait as a cosmetic preference. For instance, a proportion of white Westerners prefer tanned skins, while a proportion of South East Asians prefer a whiter skin.

Therefore, two null-hypotheses were formulated:

1. misalignment between the robot's and user's ethnic facial features does not play a particular role in people's preferences for the robot's design;
2. gender, age group, or culture do not influence people's preferences for the robot's design.

6.1 Experimental Protocol

Compared to earlier experiments, testing these hypotheses called for a large number of participants with a view to gathering sufficient data to have enough statistical power.

6.1.1 Targeted Participants

To test the first hypothesis requires respondents from all ethnicities. As such, a global population sample –probably encompassing all cultures– was targeted. Furthermore, the second hypothesis also requires a balanced gender and age across the conditions. To this one, a decision was made to run the study as a cloud sourcing experiment, giving access to both a large pool of participants, with potentially varied ethnic background, age and gender.

6.1.2 Stimuli

To present a sufficiently complex stimulus, we opted for an animated video of the robot, rather than a still image. Consequently, all ethnic variations shared a purpose-built single scenario in which the LightHead robot (version 4) performs a 55s monologue describing an imaginary tour of a robotic museum. The script player (see 3.4.5) and robotic system ensured the performance was similar across the different ethnic designs.

As the LightHead robot displays a White Caucasian face across all versions, three stereotyped facial variations were created (Black-African, Middle-Eastern, North East Asian) with an extra control condition in the form of an Alien face, each differing in skin tone and facial features as seen on figure 6.1. Eventually five videos were recorded, each from the same front-view of the robot’s monologue, placing the participants in the visitor’s viewpoint.

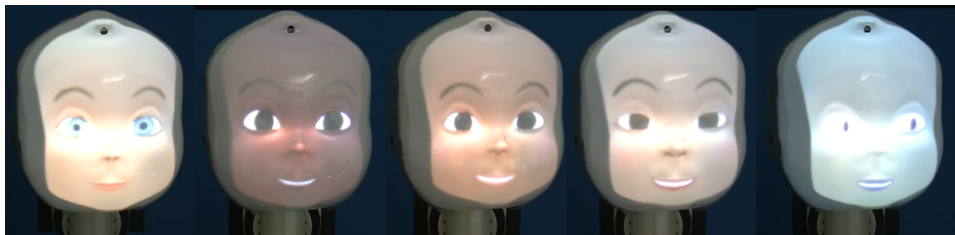


Figure 6.1: The various skin designs used for the ethnic preference study. Each stereotyped ethnic group was implemented as a “skin” overlay for the robot. From left to right: White Caucasian, Black-African, Middle-Eastern, North East Asian and Alien (control).

The rationale to the facial designs was to offer only four easily identifiable, *stereotyped*, ethnicities participants could relate to, and one that did not suggest an existing ethnicity. In addition to the skin tone, a few facial features differ: the White Caucasian face has blue eyes, the Black-African face and N.E. Asian face have larger nostrils, and the N.E. Asian eyes also display an epicanthic fold. In contrast, the Alien face shows a blueish skin colour, and no iris nor eyebrows.

6.1.3 Survey Platform

Crowd sourcing is the distribution of a task to a large number of contributing internet users, which Wikipedia being a well known example. Some crowd sourcing platforms specialise in survey taking and attract participants with a small remuneration, and allow access to a global pool of respondents, thereby moving away from local studies. In recent years, crowd sourcing gained more attention amongst scholars conducting experiments (Kittur, Chi, & Suh, 2008). Most popular remains the Amazon Mechanical Turk (AMT) platform (Chen, Menezes, Bradley, & North, 2011; Mason & Suri, 2011), but its restriction to only allows survey to be set up by US-based residents ruled

out this solution. Crowd Flower¹ does not have this restriction, and acts as a proxy to other crowd-sourcing platforms, including AMT. Crowd Flower, at the time of writing, offers respondents a wage of \$7 per 30min. In addition Crowd Flower, with some small modification in the study implementation, also offers useful benefits:

- participants can be selected across all continents and languages;
- it can balance number of female and male participants;
- it prevents surveying the same participant more than once;
- it has a trust indicator for each survey taken;
- the completion of a current group of questions is required before being shown the next;
- it can reject results based on checker questions.

At the time of this experiment's inception, the *trust* feature was considered an desirable as it allowed the removal of outliers, but later proved to be unusable for this study. Crowd Flower also allows checker questions – known as *gold* questions – which sound like an attractive feature of the platform, unfortunately the intricacies of implementing gold questions limited their deployment and effectiveness.

6.1.4 Questionnaire

The online survey is based on a form reproduced in Appendix C. Eventually, 89 seven-interval Likert-format items were used in the questionnaire. All questions were reviewed for clarity and effectiveness by trained psychologists.

The questionnaire contained the following groups of questions:

¹see <http://crowdflower.com>

1. personality test (46 items);
2. ranking of the five robot ethnic versions (5 items);
3. affinity test for the favourite robot (35 items);
4. participant's ethnic group (1 item);
5. level of experience with technology as well as explicit checker questions (5 items);
6. three optional free text questions.

A personality test known as the Big Five (John, Donahue, & Kentle, 1991) — or OCEAN — was used. It relies on a 44 questions, and was chosen to allow to study of potential correlations between personality and robot preferences. Other personality tests exist — such as the Myers-Briggs Type Indicator (Myers, McCaulley, Quenk, & Hammer, 1998) — but the Big Five is more established. The scoring procedure of the Big Five test consists of computing the mean value for all items falling into one of 5 categories: openness, conscientiousness, extroversion, agreeableness and neuroticism. Each dimension carries several reverse coded items (not to be confused with reversed items) which are used to validate the responses.

Participants were then asked to rank the five robotic guides so that a finer metric on the average preference could be computed.

Next, participants were asked to rate their favourite robot guide along a semantic differential scale (Snider & Osgood, 1969) specifically created to measure the connotative meaning of cultural objects (34 questions) which is also employed in the evaluation of a participant's attitude towards objects or concepts, and includes a likeability factor. In this case the investigation was intended to narrow the nature of the affinity towards the participants' favourite robot guide with an exploratory factor analysis.

Participants were asked to classify one of the 8 pre-defined racial groups or

meta-groups² to which one of the robot’s ethnic versions might correspond. An optional second racial group was also available to moderate the initial answer, and specify a mixed ethnicity.

Finally, participants were asked to report their experience with computer and robotic technology in 2 separate questions. Finally, three optional open questions were asked about overall feeling, possible improvements and other opinions, in the hope these would offer supplementary insights.

Early Tests and Checker Questions

In crowd sourced studies, it essential to include checker questions to spot unreliable respondents. Initially, only the following items were used:

- A question on the favourite robot being “honest or dishonest” appears twice (7-Likert format, reversed 6 items later), the same response is required on both occasions;
- “What could you say influenced your rating?” (10 multiple-choice question, only “facial appearance” effectively valid);
- If the last open questions contained nonsense text, the response was rejected.

However a first pilot test revealed 25% of the participants completed the questionnaire in less than 10 minutes whereas a in-house pilot showed that at least 15 minutes were required to take the survey. Indeed, such data correlated with missed checker questions, indicating a significant part of the respondents were answering hastily, thereby affecting the results. Therefore the following items were added for the final version, with the aim of capturing more robust data:

²As there is no consensus on the classification of the human phenotype, the classification proposed in <http://www.racialcompact.com/racesofhumanity.html> was adopted. This particular classification has a relative small number of classes, which suits the purpose of the experiment well.

1. In the personality test, a new 7-Likert format question: “[I see myself as someone who] Can reply honestly to a questionnaire”, with respondents to indicate they are honest
2. In the semantic differential, a new 7-Likert format question “The robot is more...” [a guide - a visitor], with guide being the only correct answer
3. In the knowledge assessment, 7-Likert format question “How familiar are you with subspace quantum robot technology?”, with respondents required to indicate they are unfamiliar with this non-existing technology.

Ultimately, filtering responses from the two less obvious checkers (on rating influences and the free-text questions) presented a real challenge: the former is strongly biased by the respondent’s perceptive capabilities, while no clear interpretation can be made from the latter. Thus, only the 7-Likert format mentioned above were used as checker questions.

6.1.5 Outlier Removal

Three surveying sessions collected data from a total of 225 respondents. Before performing data analysis to extract meaningful values, a first pass to remove answers from unreliable respondents was necessary.

Straightforward Filtering

Crowd Flower assigns an exclusive identifier to each of its members, therefore 3 respondents taking the survey twice were removed.

Despite technical efforts to enforce responding to all questions, another 3 members managed to leave some items unanswered and thus were removed. Despite instructions on how to rank the preference of robots, the platform did not automatically prevent invalid answers. Another 19 respondents were

dropped (four without a first choice, fifteen with more than one). Seven respondents duplicated rankings beyond the 2nd favourite robot but those were kept for analyses. Respondents picking the same answer over all items amounted to 4 additional cases. No duration limits were technically in place, thus another 35 cases were removed as those respondents took less than 7'51" to complete the survey (five 55s. videos + ninety-eight items requiring about 2s. to read and answer).

In total, this method filtered out 64 cases, leaving 161 to further filtering.

Filtering with Checker Questions

Using the 4 obvious checker questions, 20 participants failed to state they could reply honestly to the questionnaire (scoring under 6), 34 members failed to acknowledge the robot was a guide although this was explicitly written and clear from the robot's verbal story (scoring under 6), in 8 cases respondents failed to provide a consistent answer to the semantic differential checker (scoring difference over 2), and 12 members declared being at least somewhat knowledgeable in the non-existent subspace quantum robot technology (scoring above 2). In total, 77 more cases were deemed unreliable with this method (about 62.67% of the total cases collected).

6.2 Results

In the following analyses, the dependent variable is the ranking value (a continuous integer in the range [1-5]) for each ethnic version. Indeed, no participant belongs to multiple groups as only the main ethnicity is considered in this study, and for statistical significance the alpha value is fixed at 0.05.

Mean of rankings for each robot's ethnic version permits us to arrange

a global order of preference as shown in Figure 6.2. The North-East Asian version was favoured most, followed by the White Caucasian, Black African and Middle-Eastern. Unsurprisingly, the control version (Alien) ends up least favourite, indicating that globally participants prefer human facial designs over non-human designs.

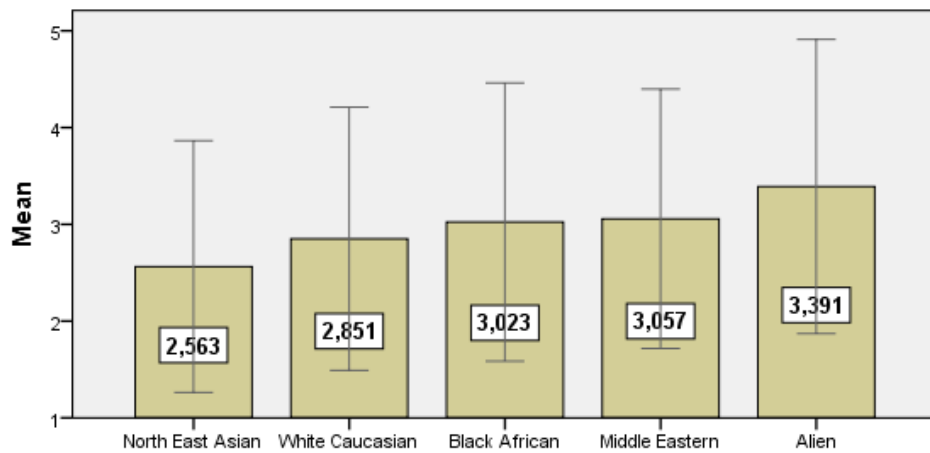


Figure 6.2: Mean ranking and SD for each ethnic version of the robotic monologue (N=87).

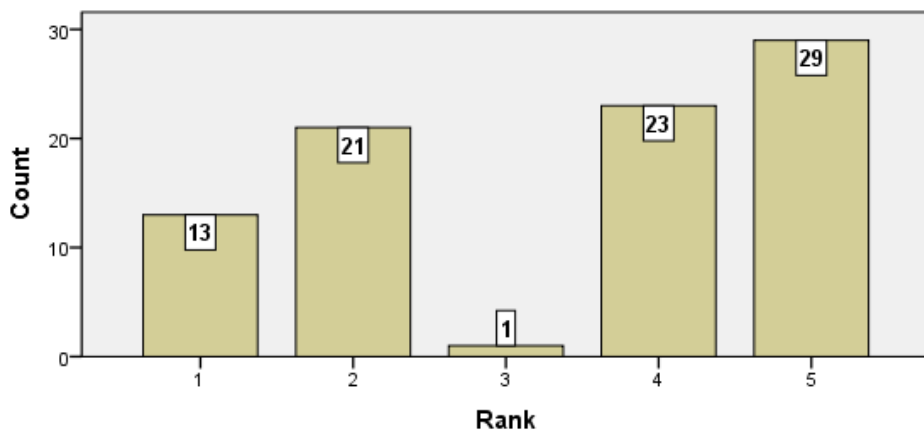


Figure 6.3: Distribution of rankings for the Alien design (N=87).

A more detailed global ranking of the Alien version is reported in Fig-

	Count	Percent
American Indian Australian Aborigine or Melanesian	1	1,1
Caribbean	1	1,1
North African, Arabic, Persian...	3	3,4
North East Asian (Japanese, Korean, North Chinese, ...)	3	3,4
Central African or black	4	4,6
South East Asian (Chinese, Vietnamese, ...)	13	14,9
Indian or Bangladeshi	15	17,2
White Caucasian	47	54,0

Table 6.1: Distribution of respondents across all ethnic groups (N=87).

ure 6.3, showing a contrasted preference for this condition: a single case expresses no preference or rejection, but 59% rank it last or next to last, and 39% rank it first or second.

6.2.1 Inter-Ethnicity Analysis

The distribution of respondents' ethnic groups (Table 6.1) indicates the White Caucasian group accounts for 54% of the sampled population, while the Indian or Bangladeshi group accounts for 17.2% and the South East Asian group for 14.9%.

Participants' inter-ethnicity preferences for their favourite robot are reported in Table 6.2 and show no specific trend.

Independent variables represented by a single case prevent running Tukey post-hoc tests for analysis of variance for interaction effects, hence the removal of the 3 cases representing a single ethnic group (“American Indian, Australian[...]”, “Central African or Black” and “Caribbean”); and 9 cases representing a single country (Argentina, Algeria, Egypt, Great Britain, Jamaica, Pakistan, Russia, Sweden and Turkey) for two-factor analysis. With

	Alien	Black African	Middle-Eastern	North-East Asian	White Caucasian	Total
American Indian, Australian Aborigine, or Melanesian	0	0	0	0	1	1
Caribbean	0	0	0	1	0	1
Central African or black	0	3	0	1	0	4
Indian or Bangladeshi	0	6	3	5	1	15
North African, Arabic, Persian...	1	0	0	1	1	3
North East Asian (Japanese, Korean, North Chinese, ...)	0	0	2	0	1	3
South East Asian (Chinese, Vietnamese, ...)	3	3	1	4	2	13
White Caucasian	9	8	7	10	13	47
Total	13	20	13	22	19	87

Table 6.2: Counts of favourite robot version against respondents' ethnic group (N=87).

N=78 and considering the independent variable *participant's ethnic group* only the mean ratings of the Alien design was found to have a statistically significant difference between the participants' ethnic groups as determined by one-way ANOVA: $F(4, 73) = 3.03, p = 0.023$ (see Table 6.3). Yet, a Tukey post-hoc test revealed no particularly statistical significant difference between groups.

Robot ethnic version	<i>p</i>-value	F statistic F(4,73)
Alien	0.023	3.03
Black African	0.163	1.683
Middle-Eastern	0.201	1.536
North-East Asian	0.348	1.132
White Caucasian	0.837	0.359

Table 6.3: One-way ANOVA for each ethnic version of the robot stimuli (N=78).

Thus, these results suggest no particular correlation between a participant's ethnicity and their favourite robot, confirming the first null-hypothesis.

6.2.2 Analysis of Interaction Effects

Results reported in this section also disregard single-case ethnic groups or countries (N=78). Upon inspecting the sampled population's properties, the following can be noted:

- genders are no longer balanced: 45 Females (57.7%), 33 Males;
- cases are biased towards USA (59%) with 46 cases, and in descending order: Canada (12 cases, 15.4%), India (8 cases), Philippines (7 cases), Germany (3 cases) and Malaysia (2 cases);
- cases represent most the 26-35 years old age group (41 cases, 52.6%), and in descending order: 18-25 years old (18 cases), 36-50 (12 cases)

and over 50 (7 cases).

Hence a strong bias exists in the demographics towards residents of North America, females and between 26 to 35 years old.

To detect a possible cultural/inter-ethnic influence, a two-way ANOVA was conducted that examined the effect of the participant’s ethnic group and country of residence on all 5 versions of the monologue (Table 6.4). Results indicate no statistically significant interaction effects between ethnicity and country.

Robot ethnic version	<i>p</i>-value	F statistic F(3,65)
Alien	0.963	0.094
Black African	0.903	0.189
Middle-Eastern	0.154	1.812
North-East Asian	1.000	0.002
White Caucasian	0.238	1.443

Table 6.4: Two-way ANOVA (interaction between country and participant’s ethnicity) for each ethnic version of the robot’s monologue (N=78).

To investigate a possible gender/maturity/inter-ethnic influence, a three-way ANOVA was also conducted, examining the effect of the participant’s ethnic group, age and gender on all 5 versions of the monologue (Table 6.5). Results indicate no statistically significant interaction effects between ethnicity, age and gender on robot preference.

Thus, these results suggest no particular interaction effect between the participants’ ethnicity, age and gender on robot preference, holding true the second null-hypothesis.

6.2.3 Analysis of Personality Test

Figure 6.4 summarises personality profiles by ethnic group and country. A one-way ANOVA revealed a strong statistically significant difference ($F(4, 73) =$

Robot ethnic version	<i>p</i>-value	F statistic F(3,54)
Alien	0.613	0.608
Black African	0.327	1.177
Middle-Eastern	0.077	2.405
North-East Asian	0.334	1.160
White Caucasian	0.615	0.604

Table 6.5: Three-way ANOVA (interaction between gender, age group and participant’s ethnicity) for each ethnic version of the robot’s monologue (N=78).

4.765, $p = 0.002$) between ethnic groups for the Openness dimension. However no statistically significant difference was found in other dimensions. A post hoc Tukey test precises significant differences in openness between South-East Asian and Black African ethnicities ($p = 0.022$) by about 1.44 points, and between White Caucasian and Black African ethnicities ($p = 0.005$) by about 1.56 points. In any case, those results have to be put in perspective: only three cases are representative of the Black African ethnicity.

Also, a one-way ANOVA yielded no statistically significant difference between countries for each of the Big5 dimensions. Consequently, it appears no personality bias on the sampled population could have influenced the results in the two previous sections as determined with the Big5 personality test.

6.2.4 Semantic Differential Analysis

Exploratory factor analysis relies on principal component analysis (PCA) and is well suited to survey research to determine the underlying dimensional structure of a questionnaire. A scree plot (figure 6.5) visually repre-

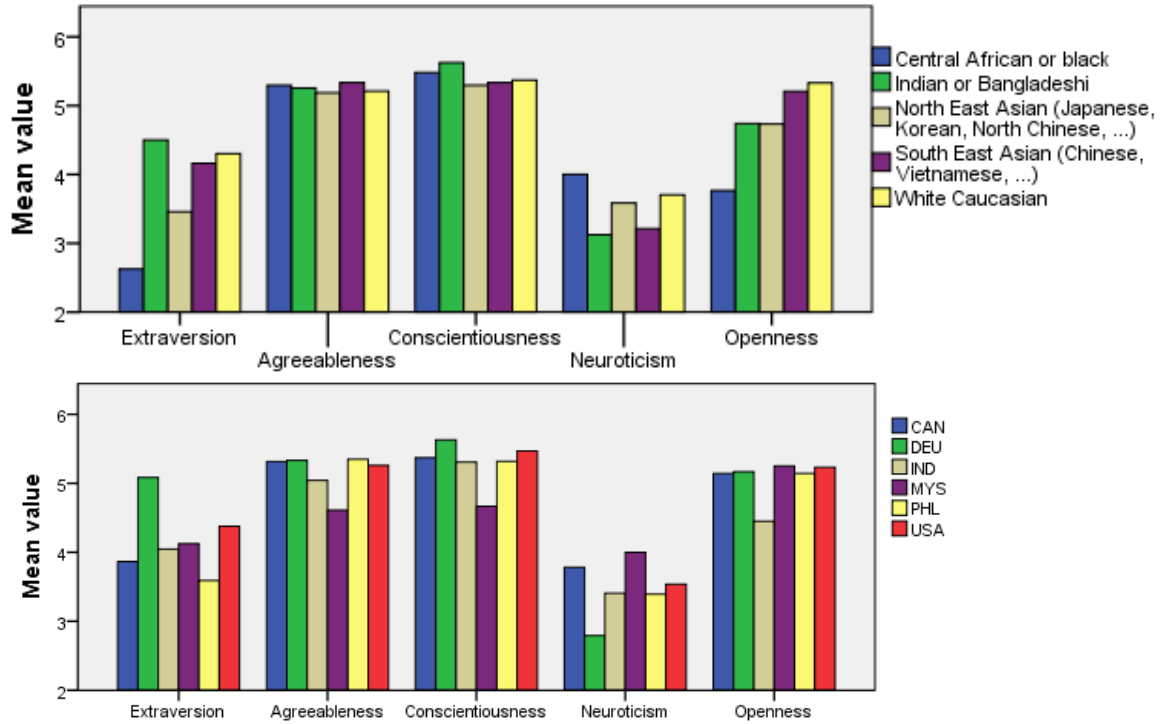


Figure 6.4: Mean Big5 profiles. Top: by ethnic group, bottom: by country (N=78).

sents eigenvalues obtained with the principal axis factoring method (a PCA) against the component rank, and helps the researcher select the first meaningful components. A first factor (12.6) explaining 37% of the total variance was revealed, followed by four smaller factors explaining 23.34% (61.44% with 1st factor) of the total variance of the semantic differential data.

Correlation scores — constructed from an extracted factor and item’s scores — for the first factor are sorted in table 6.6. Highest scores for each top items correspond to “personal”, “engaging”, “kind”, “friend” and “warm”, thus likely describing the participants’ affinity towards their favourite robot. Such results seem to indicate respondents rather considered the robot as a person than a product, further indicating appropriateness of the robot’s

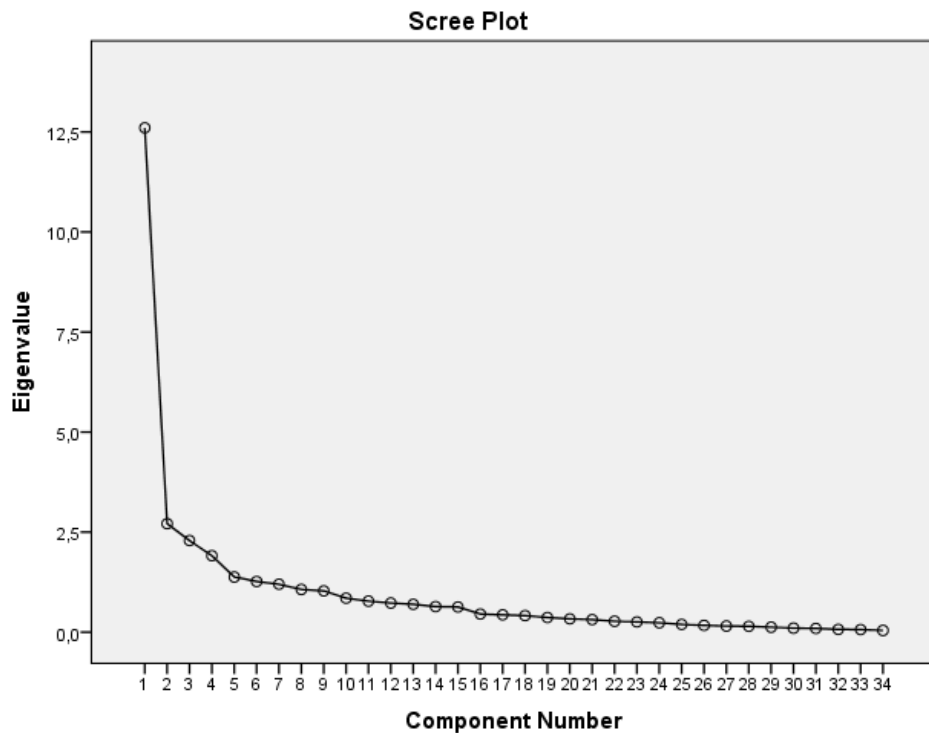


Figure 6.5: Scree plot of the exploratory factor analysis (N=78). Inflection point appears at the fifth factor.

behaviour in a museum guide scenario. This should however be confirmed with a more tailored instrument, such as the Godspeech Questionnaire.

The items presented here, although having the highest correlations available, do not reach strong scores (1 being a perfect correlation with the factor, and 0 representing no correlation at all).

Next, four factors' correlation values were sorted, the item correlating most is reported for each factor in table 6.7. According to these factors, it seems the participants' attitude towards their favourite robot hinged on notions of trust ("honest") and attention ("interested", "distracted"), potentially those looked for in a museum guide.

Certainly less specific, "humanlike" offers at least aesthetic and be-

Item	Correlation	Item	Correlation
impersonal - personal	0.778	indifferent - interested	-0.417
unengaging - engaging	0.757	decisive - indecisive	-0.525
unkind - kind	0.748	honest - dishonest	-0.564
not as a friend - as a friend	0.741	exciting - boring	-0.576
cold - warm	0.739	diligent - lazy	-0.619
abnormal - normal	0.727	active - passive	-0.632
low quality - high quality	0.699	friendly - unfriendly	-0.641
stupid - intelligent	0.681	responsible - irresponsible	-0.678
impolite - polite	0.640	I liked - I disliked	-0.693
insensitive - sensitive	0.639	good - bad	-0.727
weak - strong	0.627	trustworthy - untrustworthy	-0.771
non humanlike - humanlike	0.598	lively - deadpan	-0.780
unbalanced - balanced	0.587		
slow - fast	0.540		
dishonest - honest	0.537		
standard - unique	0.437		
engaged - distracted	0.400		
child - adult	0.399		
masculine - feminine	0.372		
serious - Fun	0.364		
traditional - contemporary	0.217		
affordable - expensive	0.088		

Table 6.6: Questionnaire items and their contribution to 1st factor extracted with PCA for exploratory factor analysis (N=78).

havioural interpretations, but this lack of specific interpretation corroborates the Alien version being ranked least favourite: participants did prefer a human-like guide.

Factor rank	% of variance	Most participating item
2	7.97	indifferent - interested
3	6.73	dishonest - honest
4	5.62	engaged - distracted
5	4.06	non humanlike - humanlike

Table 6.7: Most contributing questionnaire’s items to factors 2 to 5, extracted with PCA for exploratory factor analysis (N=78).

6.3 Discussion

It is fair to say this experiment not only tested the potential familiarity effects elicited through user-robot appearance alignment, but also the effectiveness of a crowd sourced approach to conducting survey-based experiments.

6.3.1 Online Survey Platform Issues

It has to be noted this experiment was conducted on Crowd Flower in 2012, so that the platform must have evolved towards addressing at least part of the limitations reported in this document. More to the point, crowd sourcing remains a recent technology and our understanding of its effectiveness and limitations is still evolving. Payments of wages to “crowd workers” certainly appeals to respondents, but without adequate measure, it also creates an incentive for quick and low quality responses.

To summarise, exploring the data sheds light on a few issues:

- respondents tend to rush through the survey (or son occasion let a survey sit for an unreasonable amount of time);
- random answering, setting up and testing checker (“gold”) questions was also a financial challenge;

- multiple participation of the same worker to updated versions of the survey as no list of undesired workers could be specified;
- lack of balanced age groups and country of residence as immediate availability of surveys biases results towards the population active at publication time;
- limited number of countries (20 included at maximum)
- lack of language control as many respondents apparently could not understand english, provided untranslated localised labels in their data, or wrote plain foreign language comments.

Moreover, the anonymity of the participants leaves a lot of data out of the experimenter’s control, such as age, gender or language. Crowd Flower’s *trust* value of certain obvious outliers can be surprisingly high (a participant’s trust was 0.8 but he completed the survey in 3 min) although most respondents’ trust was below 0.5.

A combination of programming and custom checker questions (obvious or reversed items) might have reduced the amount of work need to filter outliers from the raw data. To this end, assumptions on taking the survey should remain an exception, opting for strong checks in the questionnaire(s) instead. In effect, detecting all cases of untrustworthy respondents in data can lead to increasing biases and raises the risk of erroneous conclusions.

6.3.2 Conclusion

In this crowd-sourced experiment, participants ranked 4 ethnic versions of a robot. The extra, non-ethnic version (an alien version used as control) is least favoured, and no correlation with the participant’s ethnicity nor country of residence was found. Further investigation might reveal why the alien face (which displayed an unnatural skin colour as well as unnatural eyes) was not the least favourite in 66% of the cases. It might be that, in

this scenario, the participants do assume that an alien face might have some appeal.

Because of the many issues revealed after adopting this early version of this online platform to conduct the study, the sampled population is not representative: it is too small and mostly representative of the White Caucasian group from North America. Hence, this study might result in more contrasting responses if deployed on a global scale. Alternatively employing very realistic faces might bring to light results closer to earlier work in psychology. Currently no clear conclusion from this study can be drawn, nor guidance provided on Human-Robot ethnic appearance alignment.

Chapter 7

Discussion and Conclusion

At a crossroad between mechatronic and virtual faces, R-PAF represent a promising alternative to established robotic facial technologies. In this thesis, the motivation, conception and evaluation of the R-PAF solution have been detailed through the robotic platform *LightHead* that I realised during four years of studies. Simultaneously, key topics potentially disruptive of the current robotic landscape were introduced, which this chapter summarises, laying exciting perspectives for the future of robotic facial displays.

7.1 Renewed State of the Art

At the heart of the innovations brought by retro-projected animated faces technology lies the replacement of actuators with video: new capabilities gained by departing from current robot head technologies.

7.1.1 Improvements over Mechatronics

Cost – Perhaps the most attractive strength of R-PAF technology is the cost. Although designing a head, parts and moulds requires multiple skills and experience, rapid prototyping and vacuum forming techniques employed in the production keep getting cheaper. Essentially, these belong to long es-

tablished industry standards and do not require purpose-built techniques or materials. In effect cheap plastic materials constitute the translucent mask, cover and chassis, while off-the-shelf components such as the pico-projector and fisheye lens remain the most expensive parts. Current trend indicates pico-projectors will continue towards affordability, as opposed to the electromechanical components used in the mechatronic and android faces, which seem to have stagnated technologically and economically.

Moreover, compared to other actuator-based technologies, the use of a projector considerably reduces the maintenance costs of R-PAF heads. The mean time between failure for LED projectors can reach 20,000 hours and as no mechanical components are involved in the face animation, there are virtually no other parts prone to failure. Hybrid Laser/LED projective solutions exist as well¹, sharing the same robustness level. Moreover, these keep a sharp image at any distance within the projection range (0.2 to 2 meters) as opposed to manual focus imposed by standard projectors.

Yield – The liberation from mechanical actuation and its complexity relieves a retro-projected head from most of the issues exposed in detail in chapter 1 (section 1.1) and improves a robot overall yield. Such robotic heads are lightweight – potentially less than 300g – leaving behind other technologies, thus requiring less demanding actuated necks. Additionally

Animation – Capitalizing on avatar technology and unbounded by electromechanical constraints, R-PAF technology grants designers with the freedom to implement an unlimited range of facial animations and enables realistic state-of-the-art lip-sync. Video animation also allows unlimited facial expressiveness and reactivity. With software actuation range and speed, caricatural expressions come as an extra benefit and actuation dynamics remain devoid of constraints.

¹such as Explay's Colibri compact mobile module

Appearance – Whereas the aesthetic freedom of R-PAF robots’ non-actuated features such as ears or hair and generally facial geometry remains unchanged compared to existing humanoids, a R-PAF virtual face grants total aesthetic freedom over all facial features. Also, in contrast to android heads, a R-PAF head clearly displays its robotic identity with no possible confusion for the user, which results in better acceptance. Thanks to the projected computer-generated video, any facial design in the realistic-simplistic spectrum remains a matter of texture update. A R-PAF mask can accommodate exaggerated features: e.g. larger eyes with which humans readily sympathise (DiSalvo et al., 2002).

Equally compelling, facial design can change on demand, for example presenting a female character to male users and vice versa. This morphing ability authorises on the fly evaluations of facial designs, as opposed to the hardware change required by other humanoids.

7.1.2 Refined Human-Robot Interaction

Retro-projected faces can display a number of social and emotional communicative cues, which are hard or impossible to display with traditional mechatronic robot faces.

Computer graphics are conducive to the introduction of visual effects: as with LightHead, sweating and blushing add a noticeable amount of emotional information to the user, to convey particular emotions such as excitement or embarrassment. Whilst these artefacts are rarely used in HRI, they may be well suited for long processing times (sweating) or task failure (blushing), in conjunction with non-conversational vocal fillers. Additional emotional effects are possible, such as tears of sadness or sometimes happiness, pupil dilation (for a study of the correlation of pupil dilatation with mental activity see (Beatty, 1982)), eye saccades and micro facial expressions (see Ekman (Ekman & Friesen, 1969)), particular lip movements (puckering,

biting, pressing, etc.) to express many culturally-specific facial expressions such as doubt, stress, disagreement, etc. Indeed, simpler ones, such as facial colour change, are straightforward and carry potential².

Also authorizing more than the traditional gamut of facial features, the generated face grants other visual signals rooted in our cultural background. For instance, cartoon-styled characters might exploit exaggerated expressions,

Actuation noise distracts, and increases proportionally with the speed of motion, making matters worse. During interaction, humans faces are never totally at rest, therefore constant facial animation is required to elicit the illusion of life in believable characters. For a robotic head with noticeable operating noise this can hinder interaction or worse, become an annoyance and risk breaking interaction altogether. Retro-projected faces do not suffer this issue, letting users – as long as facial behaviour is natural – experience faces naturally rather than trying to interpret them.

Another strong advantage to the absence of actuators is silent actuation: electric or pneumatic actuators housed in mechatronic and android faces make a very noticeable – and too often distracting – acoustic noise which R-PAF heads are devoid of.

The speed at which the face can respond is only limited by the refresh rate of the projector, a crucial aspect of HRI applications where responsiveness is key to achieving successful interaction. Projection escapes all forms of jerkiness, and enables the reproduction of very fast human movements such as blinks, although the Geminoid DK³ demonstrates recent progress in this specific matter.

²For an example of effective use of facial colour - and other facial features - see the RoboThespian from EngineeredArts, see <http://www.engineeredarts.co.uk>

³visit <http://geminoid.dk> for videos

7.1.3 Limitations of Retro-Projected Animated Faces

R-PAF robots do not necessarily compete with established robotic head technologies, rather they offer an alternative solution to the provision of a social user interface capable of emotional signals. Like any other solution, it also comes with shortcomings that help delineate its areas of deployment.

As retro-projected faces typically employ pico projector technology, light intensity currently generates around 100 ANSI Lumens. Such brightness provides perfect visibility for standard indoor lighting, but prohibits placement in brightly lit and daylight environments. Arguably, this does not represent an important limitation as currently, social robots are mostly confined to indoors settings. For instance, CMUQ's robot receptionist HALA2, despite being placed in the middle of wide and fully lit hall, sits in a dimmed booth in which her face appears sufficiently bright. Considering the current pace of technological progress in pico projection, one can only foresee newer models gradually improving in brightness and resolution, relaxing the limitation to indoor environments. On the other hand, the radiance of these faces fits darker conditions by providing a stronger sense of presence, which usually catches the attention of users not yet familiarized with the robot.

The volume between the projector and mask must be kept free to permit the projection of the face, which imposes a restriction on the mounting of sensors. As such, cameras cannot be set in the eyes, where they would be usually located in mechatronic heads. However, alternative camera placements are possible: in the forehead similar to LightHead, or away from the face such as on the shoulders, or directly in the surrounding environment (e.g on a desk). For HALA2, active accessories such as sensors mounted in jewellery were also considered.

Although the aesthetics of retro-projected faces enjoy total freedom, the moulded mask sets a definite facial shape. Consequently, the biometrics of

a particular mask do not fit all faces, for instance the eye-to-eye distance – which varies noticeably amongst individuals – must match the eyes’ spherical shape of the mask. Although facial aspect ratio and features such as nose shape can be adapted, these modifications might hinder people’s ability to identify the related individuals. Finally, the rigid mask cannot follow natural geometric deformations occurring with large facial movements such as a wide open mouth.

In that regard, most realistic – but quite uncanny – are androids such as the Geminoid-F with a human-like flexible skin, although this comes at the cost of several hours of work in case of replacement.

7.1.4 Summary

Balancing these advantages, the major drawback to retro-projected animated faces become visible with external lighting significantly brighter than the head’s projector.

To summarise section 7.1, Table 7.1 provides an overview comparing retro-projected faces with mechatronic faces and android faces, and the following list summarises the main contributions of retro-projected faces.

- Face actuation no longer suffers from physical, mechanical and actuation limits, such as inertia, acoustic noise or mechanical complexity.
- Retro-projected faces allow the display of additional communicative signals, including an animated tongue, iris dilation, blushing and other socially salient cues in a straightforward manner.
- The aesthetic design is no longer fixed, but can be changed on the fly during operation, for the robot to adapt its appearance and suit the preference of the user.
- Retro-projected robot heads remain light, they require minimal maintenance and are very affordable compared to alternative technologies.

Because flat-screen heads suffer from the mona-lisa effect, it appears more appropriate to compare physically actuated heads with R-PAF heads as the latter share essential aspects with virtual heads whilst improving on visibility.

In short, retro-projected faces can not only overcome some limitations of HRI imposed by current technologies, but also provide opportunities to endow robots with more subtle physical, behavioural, and dynamic aspects of a human robot interaction, along with the various insights gathered by related research fields.

7.2 Opportunities for a New Technology

While traditionally robot faces are implemented using mechatronically actuated heads, retro-projected face technology improves on a number of properties that have been obstacles to making commercially viable robotic faces. This benefits both scholars and the industry.

7.2.1 Industrial Aspects

The simple construction and design freedom of retro-projected animated faces endows this technology with a significant potential for the mass market, bringing personal robotic costs down enough to broaden the deployment of social robots. Although many potential applications could emerge from R-PAF technology – with telepresence as an obvious starting point – shedding light on novel concepts appears more interesting than building an inventory of specific applications.

Character coherence underlines the feeling that a robot’s body and head need to match each other in appearance and ability: if the body suggests certain physical and social affordances, they need to be matched by the head and face. On many levels, robotics have not yet reached human level perfor-

	retro-projected	Mechatronic	Android
Development cost	Relatively low	High	Very high
Maintenance cost	Very low	Medium or High (mechanical parts)	High (idem + wear on flexible skin)
Aesthetic freedom	High (software)	Usually fixed	Low and time consuming
Expressive range	High (Software)	Limited	Limited
Realism	Medium or Low	Low	High
Texture	Unnatural	Unnatural	Closest to human skin
Uncanniness	Limited	Low	High
User acceptance	High	Relatively high	Relatively high (but uncanny)
Power drain	Low	Medium or high	High
Acoustic noise	None	Present	Present
Weight	Low	Average	Relatively high
Reactivity	Fast	Medium	Medium
Lighting constraints	Indoor only	None	None

Table 7.1: Comparative overview of established robotic head technologies against R-PAF heads.

mance and have only begun their integration into society. As such, R-PAF might be particularly well suited to fit the current state of development in humanoid robotics, avoiding unmatchable expectations from users.

Sitting in between the realism of android faces and the mechanical ap-

pearance of mechatronic faces, R-PAF robotic heads do not impose a strict specific choice towards either boundary and allows further refinements of social details. Moreover, software updates could help aesthetics tally with enhancement of a robot's social and cultural capabilities, by increasing facial realism for instance. A wide range of aesthetic freedom permits adjusting such elements so that hardware production is left unmodified whilst social specialisation belongs to a third party.

In effect, a modular approach to robotics emerges in the field thanks to the Robotic Operating System ⁴. This effort should spur actors of specific domains of expertise to develop ROS-compatible, state-of-the-art modules, to be connected with other robotic solutions; the more modular, the wider the range of possible applications. A ROS compatibility layer is available for the ARAS software, but more generally R-PAF embraces such principle thanks to its inherent capability for visual adjustments and absence of actuators.

Early adoption of the technology might occur in public environments where social robots offer value by eliciting natural interaction supported rich expressiveness. In these scenarios, R-PAF heads also facilitate *personalised* HRI, where a social robot can offer a more individualised interface, adapting to the user's preferences and interaction style. Public robots could provide a personalised service, such as in care giving (e.g. in hospitals and nursing homes) or guiding (e.g. in museums, shopping malls and airports).

Furthermore, a robotic face could contribute to the overall performance of other robotic applications. In effect, the Baxter robot (Guizzo & Ackerman, 2012) paves the way to cooperative factory robots: working with industrial robots remained until now a potentially hazardous activity, only possible with the adoption of strict safety protocols. Along with actuator compliance, integration of social cues and predictable behaviour now guaran-

⁴ROS – see (Quigley et al., 2009)

tees mutual awareness and renders robotic cooperation safer. The peripheral visibility of R-PAF heads and their readable directed gaze improves on flat facial displays integrated to factory robots such as Baxter.

Finally, the reduced power consumption might benefit mobile social robots as well. Honda's ASIMO does not yet feature a social face, however this humanoid embodies a vision shared by other industrial actors, and one may wonder why ASIMO's design still excludes a face. The autonomy of a mobile social robot relies on energy savings achieved over multiple design levels and R-PAF heads most definitely contribute to this objective.

Arguably, R-PAF technology only carries minor limitations, and trends confirm the progressive disappearance of limited brightness, certainly a temporary drawback of R-PAF as brighter LED projectors appear on a yearly basis. One can only speculate on this research's effective impact on the industry, nevertheless an application with those compelling benefits would definitely participate in raising public awareness towards social robotics and in motivating further research in this area.

7.2.2 Research Aspects

Exploring the uncanny valley – Closer to android research, robots equipped with retro-projected faces represent an ideal platform to refine the definition of the Uncanny Valley. This can be directed by studying human facial behaviour in minute detail, and applying extracted principles to R-PAF robots through comprehensive implementations. Subtle social-emotional signals serve to make the robot appear more natural and replicated facial behaviours open further investigation of synchrony, dynamics and contextual-awareness. For instance, the LightHead allows a controlled study of the effects of the cues referred to in psychological studies (eye saccades and micro expressions (Ekman & Friesen, 1969)). Related to the Uncanny Valley, matching users expectations and investigating the properties

of character coherence particularly fit such robotic embodiments.

New aspects of HRI – The ethnic influence experiment (chapter 6) introduced how R-PAF technology offers a unique potential to explore new aspects of HRI. In particular, dynamic adaptation of facial appearance facilitates the study of the following topics:

- robotic facial individualisation
- human to robot facial identity transfer
- remote presence
- user-robot ethnicity alignment
- inter-cultural facial behaviours
- embodied amplified or exaggerated facial expressions

Ultimately of course, the goal dwells on a principled theory of robotic facial design in which R-PAF heads might support ground work.

Mixed displays of explicit information on a robotic face, such as text and/or icons have so far been only technically possible with avatars on a flat-screen. This is unexplored in these studies and it remains unknown how users would experience this and how they might benefit from augmented facial expression.

The LighHead platform also calls for exploring the role of physicality. The virtual world in which avatars reside shapes the nature of their possible user interactions, preventing the establishment of naturally shared references and limiting exploitation of the sense of touch. With the provision of directed gaze and a touchable mask, R-PAF robots support blending virtual and physical boundaries: several existing avatar projects could bridge the gap of both worlds using R-PAF heads as a surrogate.

Benefit to other fields – In addition, as retro-projection is flexible and fast, it opens up possibilities as a tool for experimental psychology. The controlled manipulation of social cues, such as the rate of eye blinking during dyadic interactions (C. C. Ford et al., 2010), has up to now been limited by hardware, while manipulation of pupil dilation remained impracticable. Retro-projected faces do offer the potential for experimental psychologists to carefully tailor experimental conditions to lay bare the various impacts of facial responses in social interactions.

Finally, as mentioned in the introduction, robots assist in autistic child therapy (Robins, Dautenhahn, & Dickerson, 2009), an ideal application for the technology: not only do R-PAF robots present a facial area that remains robust to manipulation and safe to interact with, but their level of social affordances can be adapted to the patient’s progress.

7.3 Impact and Follow-up Studies

The novelty of the R-PAF technology and potential to enhance robot social communication has created several opportunities for collaboration over the four years of this work.

7.3.1 Collaboration within the University

Joachim De Greeff As mentioned in the last paragraph of chapter 3, the work-packages of the CONCEPT project were distributed to De Greeff and myself. Hence, our collaboration spanned over all of the project’s duration – including outreach events – and is explicitly labelled in this document. Refer to De Greeff’s publications for further reading about his work on active learning.

Christopher Ford – Ford, as research student from the University of Plymouth, focused on gaze behaviour during human to human conversa-

tions. In his first experiment, Ford recorded his conversations with several participants through a bidirectional camera-to-monitor system, and later annotated the participants' actions in terms of facial expression, blinks, gaze direction, head movements and speech. This symbolic data composed a rich body of sequenced behaviours which I transferred onto the LightHead robot as an early informal evaluation of the impact of human behaviours on LightHead's lifelikeness. To that end I created a performance player, procuring a simple, clear-text script utility to be reused in other scenarios. Replayed behaviours helped realize that human performance elicits a much more natural experience than randomly generated ones.

Ford's follow-up work (see 7.3.5) resulted in the creation of a more believable blink model as my implementation following ARAS' dynamics model enforced several refined formalisations and acted as a comprehensive validation.

7.3.2 Collaboration with Externals

Majd F. Sakr – Majd F. Sakr is the coordinator of the Computer Science Program at the Carnegie Mellon University in Qatar (CMUQ) and associate teaching professor in the Computer Science department at Carnegie Mellon University. Both CMU Pittsburgh and CMU Qatar are involved in a robot receptionist project based on the GRACE (Gockley et al., 2004), initiated and mainly authored by Professor Reid Simmons. HALA, CMUQ's robot receptionist (Fanaswala et al., 2011), features a flat screen to display a non-realistic virtual face. In 2010, Sakr expressed interest in modernising the robot using an Arabic R-PAF head – albeit more realistic than LightHead's – to study the influence of socio-cultural norms and the nature of interactions during human-robot interaction within a multicultural setting, yet primarily Arabic. The subsequent effort included an approach departing from Sim-

mon’s work to meet LightHead’s requirements, now dubbed CHLAS (see section 3.4.4) to reflect the significant differences from HALA’s.

Additionally, because HALA2 required a complete new facial design, the specificities of qataris was investigated: morphology, facial expressions and head movements were recorded and analysed to extract salient features. These activities and the recurrent interactions with a virtual head robot were an excellent opportunity to approach the Uncanny Valley conjecture from a different perspective, as well as evaluating it against those of the animator responsible for the 3D modelling. Unfortunately, no hardware implementation could be done in time due to complications with the Qatari customs.

Nonetheless, modernizing the HALA robot resulted in a very positive impact: a culturally-fitting, coherent character driven by a more reactive, extensible and portable avatar solution, integrated with a conversational agent. Such a robust solution allowed for further development of the robotic receptionist, and new research questions to be explored.

7.3.3 Related Subsequent Works

The advent of portable projectors instilled desire to explore projection-based animated faces, and undoubtedly, early demonstrations confirmed that trend. Over the last three years, other scholars also reported comparable studies exploring different dimensions of the design space.

From the Technical University Munich, Kuratate’s Mask-bot (Kuratate et al., 2011) opens exploration of the use of photo-realistic facial designs. Although replicating a person’s face on a robot can suffer from an aberration with wide mouth openings, this modus operandi directly tackles the uncanny valley problem which was avoided not to diverge from CONCEPT’s objectives.



Figure 7.1: Top-left: Mask-bot (adapted from (Pierce et al., 2012) and (Kuratate et al., 2011)), top-right: Furhat (permission from Al Moubayed), bottom-left: Hoque’s mask (adapted from (Hoque et al., 2011)), bottom-center: a reduced scale face by Misawa (adapted from (Misawa et al., 2012b)), bottom-right: HALA (adapted from (Fanaswala et al., 2011)).

Additionally, Kuratate’s fabrication method improves on image sharpness by spraying the transparent plastic mask with a thin layer of projection-specific paint. Alas, reproducing the method proved overly difficult.

Pierce and Kuratate (Pierce et al., 2012) also depart from the traditional robotic head volume format: Mask-bot differs significantly from LightHead as the robot just presents a face with little dissimulation of the projection system. Mask-bot does not rely on a fully fledged robot arm but instead

mounts the mask on a 3-DOF neck in which the projector might be housed in the future.

From the KTH Royal Institute of Technology, the FurHat robot (Al Moubayed et al., 2012) is part of the IURO project⁵. Furhat displays a non-realistic avatar face projected onto a translucent mask realised with a 3D printer. Although this process has the advantage of removing the moulding phase, the printing process creates ridges on the surface of the mask, even if only perceptible within the intimate interpersonal distance. Furhat’s original design includes a furry hat as a replacement to a full skull as well as a partial concealment of the retro-projection system.

Also using a retro-projected animated face, Hoque et al. (Hoque et al., 2011) investigated the effectiveness of gaze behaviours for attracting and controlling human attention. A key difference with previously mentioned designs appears upon examination of the mask’s facial features: their expression is much more detailed, restricting the areas and freedom of animation. This might be the rationale behind containment of the projection to the eye region only. Hoque reported the blinks were effectively conveyed, along with head cues comprised in the robot’s repertoire of social actions.

Even though in (Misawa, Ishiguro, & Rekimoto, 2012a) Misawa also implemented a R-PAF telepresence surrogate system, a more imaginative take on retro-projection can be found in (Misawa et al., 2012b) which describes a scaled down projected face in order to explore the effects of intimate communication. Remarkably, both of Hoque’s and Misawa’s designs exploit the down scaling issue caused by short projection distance with two perpendicular approaches: Hoque’s setting appears⁶ to scale up projected items to

⁵Interactive Urban Robot, see <http://www.iuro-project.eu/>

⁶publication’s pictures make the use of a fisheye lens very unlikely.

obtain regular sized eyes, while Misawa conserves the small factor intentionally.

7.3.4 Spin-off and Patent



Figure 7.2: The Lighty prototype as commercialized by the spin-off Synthelligence until 2015⁷. Projected face, form-factor and some materials have been updated compared to LightHead v4.

Considering LightHead a solid proof of concept as well as the short time-

⁷latest LightHead version available at <http://www.manymakers.fr/LHx>

to-market of the R-PAF technology, my research activities let room to the creation of a start-up in social robotics. Founded in May 2013 and based in Paris area, Synthelligence SAS® (SIRET #792872012, Créteil, France) brought to market an adult-looking version of a R-PAF head known as Lighty, which its early prototype is pictured in figure 7.2. Unfortunately Synthelligence folded in 2015 and I am since then carrying efforts to propose new versions of the product with optimized and certified re-implementations of the software created during the thesis.

In 2009 Plymouth University had evaluated patentability of my design of the LightHead as an original invention, however a start-up calls for patents as means to protect and develop its business. Hence, I applied for a very similar patent in October 2013: *AVATAR ROBOTIQUE DE TÊTE À VIDÉOPROJECTION* (demande INPI #1360230) which can be translated as “robotic avatar head with video projection”. The patent describes integration of all necessary electronics (sensors, projector and computation) for a retro-projected face into a fully functional standalone human-sized robotic head.

7.3.5 Insights Gained from Outreach Events

This section groups less structured evaluations of the LightHead robot in non-controlled environments. Nevertheless, these experiments do provide relevant insights into how retro-projected robot heads are perceived and might be used.

Arguably, controlled studies authorize framing the evaluation of an interactive robot system in tightly controlled conditions, and a tacit element of such experimental protocols is the nature of the participants. Usually, participants are sympathetic to robots: for obvious reasons, researchers use financial rewards to attract participants, lure curious people with capti-

vating descriptions, or recruit colleagues and students. On the other hand, exposing a robot to the general public elicits various emotions and reactions, some of which particularly unpredictable, nonetheless very insightful.

London Science Museum

The presence of LightHead at RobotVille (London Science Museum, 1 - 4 December 2011) was the opportunity to introduce the robot to the public in a less formal manner, and receive comments from a wider range of interactants than those typically recruited for lab-based experiments.

Representing the University of Plymouth’s CONCEPT project, Joachim de Greeff and myself ran an autonomous version of LightHead v3, driven by face detection and tracking, also displaying the status of the facial detector to the visitors. Although not our initial intention, visitors were enticed to express a bipolar opinion about the robot: either “cute” or “creepy”. Over the 4 days of public display, a total of 230 interviewees (88 males and 143 females, 54 children and 172 adults) reported their opinion. Overall, 120 participants considered the robot “cute” versus 73 for “creepy”. Additionally, our interactions with the public allowed us to collect 111 open comments, further labelled with four classes, as seen on table 7.2.

Aesthetics	Functional	Reasoning	Emotional	Cultural references
62.2% (69)	18.9% (21)	25.2% (28)	14.4% (16)	7.2% (8)

Table 7.2: Distribution of the 111 collected open comments collected from the museum’s visitors over 4 days (N=230). Some comments belong to more than 1 category.

As expected the aesthetics of the robot are first to attract people’s attention and elicit sharing their opinion. However, the number of participants using cultural references was expected to be much higher as the design bares – at least in principle – a resemblance with the “Sony NS-5” robot from the

film *I-Robot*. As only 3 participants made references to this film, it might be that the design does not necessarily entice a connection with that fictional robot, and that the freedom of design allowed by technology would not be limited by such cultural references.

Another unexpected outcome of this venue is the positive effect of interaction over the a priori feeling towards the LightHead robot: over multiple occasions visitors who initially considered the robot as “creepy” came back to report having changed their mind and leaned towards cute. Although it is possible their discovery of the other displayed robots participated in changing their mind, this insight might help us refine the ways to investigate the boundaries of uncanniness in retro-projected animated faces supported by an articulated neck.

Crowd-sourced Evaluations of Social Blinking

Human blinking not only moistens the cornea, but also takes part in non-verbal communication. In (C. C. Ford et al., 2010), Ford investigated blinking behaviour and later observed most blinks occurring during face to face conversations do not appear to have a biological origin. The simplistic blinking model of the LightHead’s system was initially designed to provide a basic sense of lifelikeness, and called for improvement through a collaboration with Ford. Therefore, the integration of Ford’s basic blink model into the LightHead’s system served a dual purpose: as a improvement of the life-like autonomous behaviour of the robot, and as a experimental platform to further refine the model. The latter has been published in (C. Ford, Bugmann, & Culverhouse, 2013).

For this experiment, crowd-sourcing evaluations of LightHead’s performance using video records presented the same advantages as cited with the ethnic preference experiment in chapter 6. Thus, an annotated participant recording was used as a baseline for LightHead’s performance which included

head gestures, both head and eye gaze, facial expressions, and speech. Since in (C. C. Ford et al., 2010) the participant was recorded during a dialogue with the experimenter, the participant’s speech was reproduced using the TTS embedded in LightHead’s system – thus with different voice characteristics – while keeping the experimenter’s apart. Eventually the Caucasian version of the LightHead v4 was filmed, then the full dialogue reconstructed with the experimenter’s speech.

Four versions of the dialogue were created, such that the LightHead’s blinking behaviour was manipulated to one of the following conditions:

- LightHead’s blinks are generated every 5s, this served as a control condition (most robotic behaviour);
- the blinks are generated after a delay (within a 0.1 to 4.9s range, using a uniform distribution), every 5s;
- the blinks are played from the analysed human performance;
- the blinks are generated by Ford’s refined model.

Ford first evaluated the last 2 conditions in an uncontrolled environment, during a public presentation of the videos at a Science festival. 84 participants were asked their preference between the human-based blinks and those generated by the model. Preferences figures are even, which suggest the LightHead running the blinking model appears as believable as the one with human blinks. Splitting results by gender, 54% of polled males preferred the human blinks whereas 53% of females favoured the model blinks. Even though the human performance transferred on the robot was recorded from a male, these figures do not suggest a particular gender effect.

For the crowd-sourced version of the experiment, a reduced version of the questionnaire used for the ethnic preference experiment only included the

semantic differential, the videos and an open question for the participants to leave comments. At the time of this writing, the experiment on CrowdFlower has just completed with 262 unfiltered participants, hence no result can be reported yet.

7.4 Future Work

Throughout this chapter, profoundly diverging follow-up works to R-PAF heads have been introduced or envisioned, and the amount of topics to choose from suggests federating projects together might help modularity and exchange of methods. However, before such a network emerges, the LightHead platform could benefit from the research topics mentioned next.

7.4.1 Long Term Interaction

In order to investigate mid-term interaction issues and multi-user interaction, relocating the robot in a public area such as a mall could effortlessly familiarize visitors to the presence of a robot. A small windowed booth could constitute an ideal robotic shelter, letting unconditioned participants interact naturally with the robot, without causing disturbances. With these relaxed conditions, establishment of engagement and strategies for long term support would gain strong experimental credibility albeit directed to the English culture. Also, analysis of user behaviour should help selecting most robust robotic behaviours for stationary public service robots like museum guides, receptionists, etc.

7.4.2 Holistic Affective Models

As robotics continue to deploy and strengthen bridges with various facets of human behaviour, our tendency for empathy contributes to the need for a sense of – emergent or forged – coherent robotic personality if we are to

accept robots as real social actors. Because the nature of a person's emotional interactions reveals aspects of her personality, an investigation of the principles of emotional congruence is needed. Advancements in affective computing could result from the joint effort of scholars from the University's Psychology department. Aiming at a holistic approach to emotional influence, the LightHead robot could support further studies in facial expressions, motor and timing dynamics, head poses and gestures, gaze and saccades as well as utterances. These objectives imply only little software updates to the CHLAS in order to offer a single parameter for the emotional value, and join together psychology projects and results, some of which are already available in publications.

At Plymouth University, groups such as the CRNS and the Cognitive Institute are initiating a tradition of modelling and replicating human behaviours identified in psychology through cognitive science. In the case of the CRNS, robotic evaluations of these models are eventually carried out on the iCub. However, this scheme could not yet comprise facial behaviours on the grounds that iCub can only accommodate a limited number of static facial expressions.

7.4.3 Delineating Models' Transferability

Arguably, potential limits in transferring human behavioural models to robots evoke the Uncanny Valley conjecture, thus identification of these limits may also be bound to a lack of consensual evaluation protocol. However focused experiments, such as those conducted with Ford, suggest a detailed investigation might generate insights in the Uncanny Valley evaluation, insofar as they tightly frame measurements to minute aspects of specific modalities.

It is not yet known if empiric sampling of modalities might result in the identification of all possible social affordances, but such an approach could

be supported by a recent Bayesian model of the Uncanny Valley (Moore, 2012): not only by delineating the pool of social cues, but also smoothing the curve through identification of the elements that should continue to belong to the human realm.



Appendix A

Schematics of LightHead

Version 4

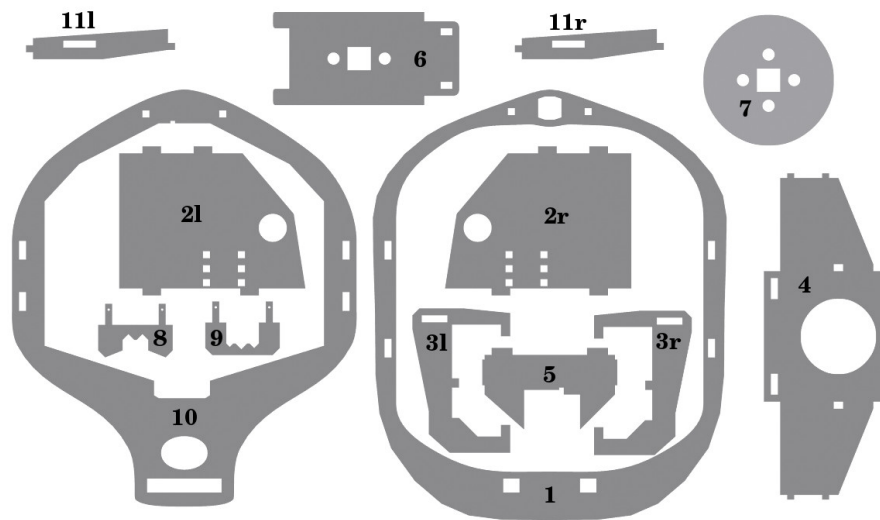


Figure A.1: Laser-cut parts of the LightHead's chassis as in version 4 (*l* and *r* suffixes refer to left or right editions of a part); all parts are 3mm thick PETG except # 6 which is 6mm thick. 1: front frame, 2: side panels also housing microphones, 3: lens side-grippers, 4: main lens holder, 5: grippers and PK301 bridge, 6: main base, 7: KatanaHD400s-6M adapter, 8 & 9: cables holder, 10: back frame, 11: frames bridges.

Appendix B

CHLAS Documentation

CHLAS

Design document
current version: 2.1

author:

Frédéric Delaunay <frederic.delaunay@plymouth.ac.uk>,

original Work

May 2010 - September 2011

with co-author:

Imran Fanaswala,

and supported by

Majd Sakr & Brett Browning (CMUQatar)

Abstract:

This document describes the design of the CHLAS Server working with an ARAS (such as LightHead) server. Design focus is on both external and internal interfaces, modularity of the system and portability of the source code itself.

This system is a fork (from the 30th of August 2011) and extension of "Expression2" which was initially designed for the HALA robot receptionist at CMUQ (LGPL) by Frédéric Delaunay and implemented/tested also with the support of Imran Fanaswala (CMUQ).

TABLE OF CONTENTS

TABLE OF CONTENTS

Foreword

Preliminary design

General description

Character personality aspects

General IO

From the HMS

Datablock syntax

Blank values

Commands

Intensity and Duration

Transforms and Vectors

Vector Space Orientation

Queuing

Interruption

To the HMS

To the ARAS

Sequence diagram with the HMS

Animation

Modules description

Dispatch

EASI

Face

Speech

Gaze

Spine

Reflexes

Conflict Resolver

Detailed Design

Implementation of Reflexes

Internal Format

ARAS' Target Frames

Appendix

List of Implemented FACS Action Units

Modifications from FACS

Animation guidelines

State-based animation

Event-based animation

Attack discrepancies in Target Frames

Foreword

Preliminary Design does not require a specific programming language of implementation. On the opposite, *Detailed Design* should hold comprehensive information for a programmer to implement the software.

A requirement is defined through the use of **shall**.

A recommendation is defined through the use of **should**.

A possibility is defined through the use of **may**.

The rationale behind this formalism is to help the validation and test process.

Preliminary design

General description

The Character High-Level Animation System (CHLAS) lays the foundation of a character's behaviour.

It is the gateway for processing:

- animated facial expressions
- utterances
- eye gaze
- head, neck, shoulder and thorax movements
- other reflex behaviours

Conceptually, CHLAS is driven by a High-level Management System (**HMS**) which handles all cognitive processing (analysis of input, action planning...). CHLAS allows a HMS to animate a (robotic/virtual) character in a timely and consistent manner without knowledge and management of the underlying character itself.

CHLAS itself is abstracted from implementation (physical or virtual robot) by an Abstract Robotic Animation System (**ARAS**). For instance, the *LightHead* server is such a hardware-abstracted robotic management system.

An ARAS abstracts the implementation of facial animation by using an evolution of *FACS*¹ (Facial Action Coding System). These Action Units (**AU**) are normalized and represent intensity of muscle activation or angles. For a list of all modifications see Appendix **Modifications from FACS**.

Consequently, low-level animation (and rendering if applicable) of a character is done by the ARAS, which receives abstracted actuation instructions from CHLAS.

To summarize, CHLAS is an interface between a HMS and an ARAS :



An ARAS may have multiple backends, allowing it to animate robots as well as virtual characters.

Character personality aspects

The Character's observable personality is defined in two ways:

- how the 3D artist creates the 3D model and the muscular deformations (ARAS)
- how Action Units are combined to create a specific expression (CHLAS).

¹see Ekman, P., & Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Considering a *receptionist*, she might engage in various activities:

- Idle: nothing particular to do
- Typing: on her keyboard
- Phoning: talking on the phone
- Inviting: when a new person appears within her area of interest
- Greeting: when first engaging in a dialogue
- Talking: when having a discussion

Note: These could be defined in a finer set of activities.

Each of these activities can use a similar set of facial expressions and gestures, but a specific activity may require specific animations. Emotional states also influence these activities.

The CHLAS itself provides two ways to provide a behavioural sense of personality:

- through expressive animations
- through reflexes

However, the CHLAS doesn't provide a fully-featured personality mostly because it has no access to sensors, and let the management of contextual high-level animation to the HMS. These concepts are developed further in the document.

General IO

All IO shall be in clear text, UTF-8.

In the rest of the document, **EOL** stands for End Of Line and embodies both `\n` or `\r\n` standards.

A valid set of elements' instructions is a ***datablock***.

Disconnection from the HMS can interrupt the connection with the ARAS.
Disconnection from the ARAS shall report a DSC status to the HMS.

From the HMS

The HMS is responsible for interaction and task management, hence the CHLAS shall receive high level information:

Element	Instruction Description
expression	the predefined (facial and/or gestural) animation to display
speech text	text delimited by double quotes (i.e.: ") to be uttered by the Character. Text is UTF-8 and thus can be of any language (e.g. Arabic, English).
focus	Transform ² of the focus point for eye-gaze direction
spine	Transform of one or more Character's skeleton section (head, neck, shoulders, thorax...).
reflexes	means for setting various reflex parameters (blink rate, breathing rate..)
unique_tag	tag identifying data received

² see section [Transforms and Vectors](#)

See some examples in appendix.

Datablock syntax

To maintain consistency, a datablock shall be formatted with a fixed number of the **element separator**, i.e. a semicolon character (;). Hence a datablock shall have 5 element separators.

Data sent to elements shall consist of **commands**. All commands of a datablock shall be sequenced according to the order of the previous table.

Each datablock shall end with EOL.

Blank values

All commands of a datablock shall be sent in one go. However some elements of a datablock may be unspecified: void or whitespace characters between 2 element separators should be interpreted as "no value". In that case, the previous instruction set for this element shall not be modified by the system.

Commands

Commands allow structured values to be passed to modules. A command is a dictionary based structure allowing multiple values to be specified at once.

Commands should only be necessary for modules accepting more than one value, and thus are mostly useful for the reflexes part of a datablock.

Each pair (the key and its respective arguments) shall be separated by the pipe (|) character. Keys and arguments shall be bound with the colon (:) character. Values shall be bound to arguments with the equal (=) character. Several arguments (and their values if any) can be bound to the same key using the ampersand (&) character.

Note: A module may accept only a key, or a key and an argument or a key, argument and value.

Commands components:

Component	Description
key	lowercase label, specific to the module.
argument	string specific to the key.
value	Transform or any other text specifically interpreted by the module.

Note: Complex commands are mostly useful for the reflexes module which uses the key as an identifier to a specific reflex. Refer to [reflexes](#) for more description.

Intensity and Duration

Values

The command parser shall support an optional intensity factor and duration constraint, however this does not imply that all elements (and their relative module) implement these options³.

³ More elements might interpret intensity and duration in further versions.

The Expression element and all Transforms shall support intensity and duration syntax.

Syntax:

- intensity shall be introduced by the star character (*) and stand as a suffix to a value
- duration shall be introduced by the forward slash character (/) and stand after the intensity. A negative duration shall play the animation backwards.

To be more specific, a facial expression playing for 2 seconds can either be sped up or down specifying a different duration. Similarly a facial expression (e.g. raising eyebrows to 0.4) can be more or less intensified specifying a different intensity factor.

Transforms and Vectors

Transforms are 3 dimensional vectors (a set of 3 floats) with values surrounded by characters which define the transformation. Values use the dot character (.) to separate the integer and real part. Vector components shall be separated by a comma character (,).

Commands for the Focus, Spine and Reflexes elements shall specify orientations, rotations, positions and translations using the Transform syntax:

- Rotations are enclosed by a pair of single parenthesis characters ((and))
- Orientations are enclosed by a pair of double parenthesis characters (((and)))
- Translations are enclosed by a pair of single bracket characters ([and])
- Positions are enclosed by a pair of double bracket characters ([[and]])

Vector Space Orientation

When specifying orientation (e.g. AU 65.5), values are expressed in radians using the Cartesian coordinate system, right handed (aka. standard orientation). Also, for rotations, looking from a positive axis back towards the origin, a counter-clockwise rotation will be considered positive.

To summarize, relatively to the character, we have:

- X positive is pointing right
- Y positive is pointing front
- Z positive is pointing up
- X rotation of pi/2 radians orient Y axis towards up
- Y rotation of pi/2 radians orient Z axis towards right
- Z rotation of pi/2 radians orient X axis towards front.

Queuing

Datablocks from the HMS may be sent in as bursts (i.e. series of consistent datablocks received at the same time), hence CHLAS shall allow datablock buffering in a queue, aka. FIFO. As a consequence, dequeuing shall be done whenever possible (see also the sequence diagram).

Interruption

Current datablock processing can be interrupted to give priority to next incoming datablock. Processing interruption shall be achieved sending the INT datablock:

Datablock	Description
INT	interruption identifier for immediate processing of the next incoming datablock

Upon reception of explicit interruption:

- currently processed datablock tag shall be reported as interrupted (see next section)

- queued datablocks shall be flushed
- no status report shall be sent back to the HMS about queued datablocks

To the HMS

Upon completion of a datablock's processing, Dispatch will send an acknowledgement to the HMS:

Reply's elements	Description
unique_tag	tag identifying data received
status	status of CHLAS process for the datablock identified by this tag

Reply shall be one of the following values:

- ACK, meaning datablock has been processed successfully
- NACK, meaning datablock processing was aborted by an error
- INT, meaning datablock processing was interrupted by a newer datablock
- DSC, meaning CHLAS will not be able to send data to the ARAS

If the CHLAS cannot reply with a unique tag (bad datablock or bad system status), CHLAS shall use the question mark character '?' as a tag.

Consequently, the question mark character shall be rejected if used as a datablock's tag.

To the ARAS

The ARAS is responsible for real-time animation (rendering and actuating motors).

The ARAS shall receive low level information from CHLAS:

Request's element	Description
origin	name of module generating the set of s of data
AU_id	identifier of the Action Unit to activate
target_val	normalized target value (float) for an AU activation
attack	time (in ms) for an AU value to reach its target_val.

This data shall be formatted in the following manner:

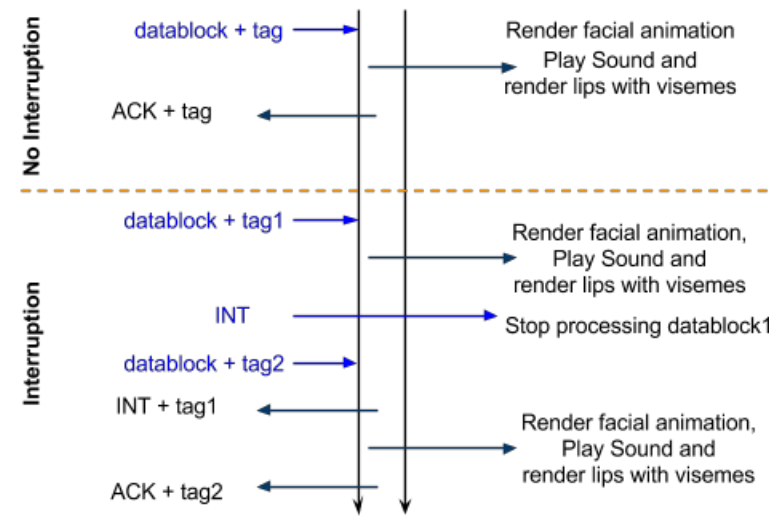
- "origin" + origin + EOL
- a set of (AU_id, target_val, attack + EOL)
- "commit" + EOL

Segmentation is done upon reception of *commit* : buffered data for the last received origin is processed at the same time⁴.

⁴for lightHead (the ARAS), EOL can also be a double ampersand (&&), ensuring process of the both parts of the token at the same time, although this is mainly obsoleted by the transactional nature of the protocol.

Sequence diagram with the HMS

With the concept of target values and attack, CHLAS introduces an "on demand" approach to animation. A key point to reactivity is also to allow for interruption of processing. Even though CHLAS hosts simple processing, queuing (for instance Text-To-Speech) shall be interruptible.



Also, CHLAS processes its queue as soon as possible. As a consequence, a datablock containing only a specific element (e.g. speech) can be processed along another datablock containing only another specific element (e.g. gaze).

Animation

The system is "best-efforts realtime": data is process as soon as possible (i.e.: with no realtime OS support).

Animations are defined using the concept of attack, while sustain and decay are made implicit:

- **attack** sets the duration of a transition from any AU value to a specific AU target value
- *sustain* is the undefined duration between an AU value set at its target, and the time of starting to reach its new target, i.e. the duration when an AU value stays constant.
- *decay* is conceptually inappropriate. Although one can consider this by setting an AU target value of 0 and particular attack time.

Attack time has to be considered with the amplitude of the transition (i.e. difference of target values).

Considering an AU, its target value V and attack time T transiting from states S to S' :

- the larger the absolute value of $(V_{S'} - V_S)$, the faster the transition $S \rightarrow S'$ will appear
- also, the smaller the value of T , the faster the transition $S \rightarrow S'$ will appear.

Negative attack time allows for state recovery:

state S -> play animation A with duration D -> expression transited to S'
 state S' -> play animation A with duration $-D$ -> expression transited back to S

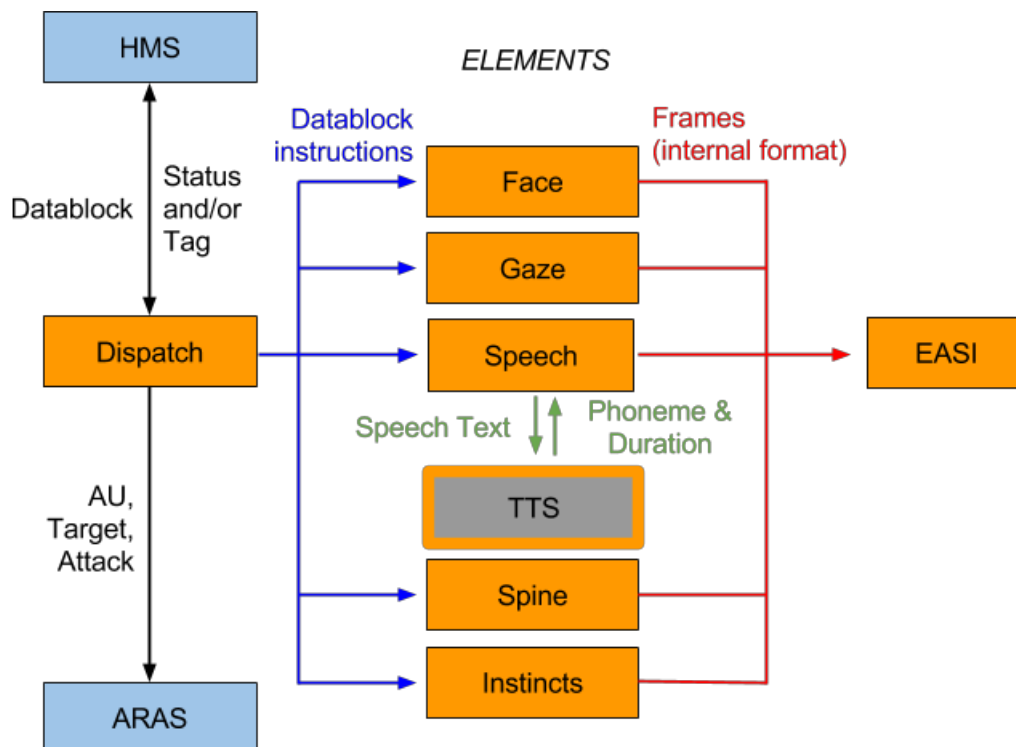
Modules description

CHLAS processing is broken into the several modules, which may themselves be split into submodules. Element modules shall share the same API.

CHLAS shall be implemented with at least the following modules:

- **Dispatch**, dispatching datablock instructions received from the HMS to other modules
- **Face**, translating expression commands into AU animation
- **Speech**, abstracting TTS for generation of speech samples and viseme AU information, playing utterances on time.
- **Gaze**, generating AU information for eye orientation from gaze vectors
- **Spine**, generating AU information for the head, neck and thorax from end-effectors orientation
- **Reflexes**, autonomously generating AU information for unconscious behaviours
- **EASI**, translating internal frames into network packets following the ARAS' protocol.

To summarize, modules of CHLAS process information this way:



Dispatch

Dispatching datablock instructions received from the HMS and sending processed data to the ARAS.

Data incoming to CHLAS contains heterogenous information for animating the character with various modalities. All datablocks received are tagged; this allows maintaining consistency of information during their processing within Dispatch and notifying the HMS of related events.

Also communication errors shall be dealt at this stage or previously:

- failure to establish communication
- failure to transmit or receive

Dispatch shall:

1. check that datablocks received respects protocol specification
2. separate datablocks in elements for each module
3. transmit elements to modules
4. maintain the coherence of elements belonging to the same tag and processed by modules in a timely manner
5. use the EASI module for assembling and sending data to the ARAS
6. send to the HMS the tag relative to the data received and processed along with the processing status of the original datablock.

EASI

Formating animation packets for the Facial Animation System protocol.

EASI stands for Expressive Animation System Interface. It performs final timing analysis and allows abstracting the backend used for facial animation.

Given a set of internal frames it shall:

1. compose the final message that will be sent to the ARAS,
2. determine (and re-compute) the "end time" of their processing.

Face

Generating AU information and attack for facial expressions

Data incoming from Dispatch may contain the following expressive instructions:

- the facial expression identifier
- the intensity of the facial expression

Face shall:

1. Validate instruction
2. Load facial expression translation tables
3. Lookup the frames corresponding to the given facial expression identifier.
4. Weight the AU target values with intensity
5. Update attack times from duration factor
6. Create corresponding frames following the internal format.

Face should:

1. Store facial expression tables in a separated file.

Each AU of the retrieved set has an associated target value. The set of these target values shall be

weighted with the intensity value.

A note on the ARAS: if no new target is received before attack time is elapsed, the ARAS will keep rendering the last state. However, humans usually display a facial expression for a particular amount of time and then shift back towards a neutral face. A similar behavior may be achieved through the HMS: it can send the same element instruction with duration factor -1. Refer to Appendix for examples.

Speech

Managing lip-synching and utterances, and abstracting Text-To-Speech system in use

The speech module transforms HMS' text into raw sound data and its corresponding phonemes/duration. It abstracts the TTS backend in use.

It is also responsible for initiating the playing of the speech samples on time. This might require an internal sound player if this behaviour is not supported by the TTS.

Finally, Speech also makes sure the shape of the mouth (also known as *viseme*) corresponds to the uttered phoneme, a process known as lip-synchronization.

Data incoming from Dispatch contains:

- text to be uttered
- voice/language to use

Speech shall:

1. Accept unicode (UTF-8) text input
2. Convert text input into raw sound data (via a TTS) while blocking.
3. Convert text input into phonemes/duration (via a TTS) while blocking.
4. Load visemes translation tables
5. Manage the playing of raw-sound data, in sync with the lips in **real-time**.
6. Have the ability to interrupt text/speech that is being processed in a "timely manner" every uttered phoneme
7. Have access to sound samples, phonemes and their duration
8. Create the corresponding frames following the internal format.

Speech should:

1. Convert text into raw sound data while *not* blocking.
2. Convert text into phonemes/duration while *not* blocking
3. Allow switching languages and/or voices (e.g. Arabic and English) via the TTS
4. Store phoneme to AU mapping information in a separated file.

Speech may:

1. Support a caching system to allow the playing of scripts without the need for a running TTS
2. Support a speech interruption policy.

The logic being very similar to Face, please read the Face specification of this part of the document.

Note: FACS merges all tongue displays in AU19, which is actually intentional from the FACS' authors. As a consequence an extension is needed. Unfortunately (to the best of my knowledge) no such work is available, hence new Action Units are defined in Appendix (Modifications from FACS).

Gaze

Generating AU information for the eyes

Data incoming from Dispatch contains:

- the focal point's Transform in meters, relative to center of the eyes.

Gaze shall:

1. Validate data
2. Translate eye orientation (AU 61.5 and AU 63.5⁵)
3. Keep eyes orientation values at all times
4. Create eye orientation value for each eye taking into account human capabilities
5. Create the corresponding frames following the internal format.

AU 61.5 is defined for better consistency (an eye is turned only in one direction), hence computation of vergence is required.

Note1: For eye roll, use multiple an expressive animation.

Note2: Saccades should be sent as a batch of datablocks

Note3: A reflex should compute eyelid stretching values from eyes vertical orientation.

Spine

Generating AU information for the head, neck, shoulders and thorax

Data incoming from Dispatch contains:

- head orientation in a triplet of angles (x,y,z in rads):
 - relative to current orientation if the data is enclosed within parenthesis, i.e. '(' and ')'
 - absolute if the data is enclosed within double parenthesis, i.e. '(((' and '))')
- head position in a triplet of normalized values (x,y,z axis):
 - relative to current position if the data is enclosed within brackets, i.e. '[' and ']'

Spine shall:

1. validate data
2. translate head orientation (see footnote) from a Transform.
3. Create the corresponding frames following the internal format.

For Transform translation, the same policy used for Gaze shall be applied.

Reflexes

Generating AU information autonomously

The reflex module allows unconscious behaviour to happen autonomously (i.e. without datablocks coming from the HMS), as well as tuning these behaviours through commands. Breathing, blinking, ect. should be implemented through reflexes that provide tuning parameters.

⁵FACS is somewhat inconsistent defining AU61 for Eyes Turn Left and AU62 for Eyes Turn Right. For the sake of unification these are merged into a single dimension named AU61.5 (the .5 suffix might avoid confusion with FACS).

This method was applied for similar problematic AU definitions such as head orientation.

Data incoming from Dispatch contains commands which keywords are:

- "enable" to enable a reflex
- "disable" to disable a reflex
- the unique identifier of a reflex.

The reflex module itself is only a manager of all available reflexes. It allows the runtime toggle of reflexes and update of their parameters in a single datablock.

For terminology clarification:

- *The* reflex module is the reflex manager
- A reflex module is part of the implemented reflexes.

The Reflex module shall:

1. Reject the datablock if any reflex encountered an error from processing a command's argument(s) (and values if any)
2. Enable or disable a reflex identified by the argument of "enable" and "disable" keywords
3. Distribute each command's argument(s) (and values if any) to reflexes identified by keywords

Because there is no specification on the number of reflexes modules nor implemented behaviour, each reflex module can interpret specific arguments (and values) that may not be supported between different implementations of the same reflex.

A reflex module shall:

1. report to the Reflex Manager any argument (and value) received they do not support
2. make available and maintain the next time of their own activation when enabled
3. generate their own frames in accordance with their maintained timings
4. be able to monitor frames created by other modules

Conflict Resolver

Managing conflicting AU information for a target state

One may need to understand relevant parts of the ARAS specification for a better knowledge of the animation system and its potential side effects. The ARAS protocol uses a transactional approach:

- declaration of the body *section* (i.e. 'gaze', 'face', 'lips', 'head') followed by Target Frames
- (additional sections and their Target frames)
- a final *commit* indicating the application of buffered sections

However different CHLAS modules can create frames involving the same AUs. Typical cases are:

- visemes conflicting with facial expressions (e.g. speaking while smiling)
- eyelids follow gaze; this can conflict with facial expressions (e.g. natural gaze up and frowning)

Hence, overwrites on ARAS' side might occur. Moreover the dynamic nature of these frames requires the state of the animation system is maintained solely by ARAS itself. Thus CHLAS transactions must resolve overwrites conflicts.

This can be done by managing the sequence of *sections* of the frames it communicates to the ARAS.

The following algorithm should resolve conflicts to create the desired final state:

1. Set target state from Gaze
2. Set targets from Face, resolving conflicts from previous targets
3. Set targets from Lips and resolve any conflicts with previous targets
4. Compute state transition.

The Conflict resolver shall:

1. manage the sequence of triplets overwrites

Detailed Design

Implementation of Reflexes

The Reflex module has a regular module interface, however management of actual reflexes brings the following constraints:

- reflexes don't create frames at the same stage as other modules. Rather they create their frames at the last stage, before frames are transferred to Easi
- reflexes shall be able to monitor all frames created by modules so they can react to it
- reflexes shall use functions from elements and/or reflexes but shall not use their internal data
- reflexes shall be called in a sequence built by the Reflex Manager. This shall be achieved from reflexes' declaration of dependency towards other reflexes.
- reflexes shall be able to be triggered on a specific time

As the Reflex Manager cannot know the behaviour (event or time based) of a reflex, the Reflex Manager calls the `get_next_time()` and `pop_frames()` function of each reflex for every frame created by the modules.

This means each reflex shall check in its `get_from_frames()` function if it is appropriate to return its frame, usually by checking time or availability of data.

Internal Format

An dictlet represents the smallest primitive of the CHLAS system part of the protocol with an ARAS.

A dictlet is a mapping of AU to tuple: { action unit : (target, attack), ...}. It represents an instruction to move a certain muscle or group of muscles (i.e. action unit), to a certain value (i.e. target) and within a certain period of time (i.e. attack). dictlets are the base instruction of the RAS.

Also, symmetric Action Units can specify a side suffix (i.e. either 'R' for right or 'L' for left) for asymmetric animation.

For example,

Raising the left eyebrow: ('01L', 0.5, 2)

A subtle twitch of the outer-lips: ('15', 0.4, 1)

ARAS' Target Frames

A target frame is an unordered set of triplets. It represents a collection of muscle movements starting precisely at the same time. Therefore the length of the target frame is simply the length of its longest triplet (i.e. the triplet with the highest attack time).

Appendix

List of Implemented FACS Action Units

This table omits modifications listed in the next table. A comprehensive list of Action Units in use can be obtained by joining the 2 tables.

Rows in gray are not to be implemented, green entries are AUs without Right/Left component.

AU	description	related AUs and comments
01	Inner Brow Raiser	04 (opposed)
02	Outer Brow Raiser	
04	Brow Lowerer	01
06	Cheek Raiser and Lid Compressor	07 (connected)
07	Eye Lid Tightener	05
08	Lips Closer	discarded (use 24 or 28)
09	Nose Wrinkler	10 (implied usually)
10	Upper Lip Raiser	09
11	Nasolabial Furrow Deepener	
12	Lip Corner Puller	14 (), 18 (opposed)
13	Sharp Lip Puller	
14	Dimpler	12
15	Lip Corner Depressor	
16	Lower Lip Depressor	17 (opposed), 25 (see modifications)
17	Chin Raiser	(also acts as Lower Lip Raiser), 16
18	Lip Pucker	14, 20 (opposed)
19	Tongue Show	discarded
20	Lip Stretcher	18
21	Neck Tightener	pressing appears at the center of the lips. other muscles are involved for l
22	Lip Funneler	
23	Lip Tightener	see 16
27	Mouth Stretch	achievable with 26, 16, 25, 10
28	Lips Suck	viseme 'b'

29	Jaw Thrust	(could be used for 'm' viseme, but not much visual from front view).
30	Jaw Sideways	
31	Jaw Clencher	26
32	Bite	
33	Blow	
34	Puff	
35	Suck	
36	Bulge	
37	Lip Wipe	pressing appears at the center of the lips.
38	Nostril Dilator	39
39	Nostril Compressor	38
43	Eye Closure	
46	Wink	

Modifications from FACS

As mentioned previously in this document, the most significant modification from the original FACS is normalization of all AU values. As a consequence, a neutral face is defined with all AUs set at a value of 0.

For other body parts a value of 0 radian corresponds to the rest pose of the model used: standing on joint feet, straight legs and spine, arms opened at right angle with spine and face straight. Hence, absolute angle values can be negative.

Some minor but significant modifications from FACS are also necessary to make the system work in a more consistent way. Modified areas are in blue, those added are in orange:

area	AU	original	modification
Tongue	19	tongue show, defined as tongue moves (see FACS manual note on this)	discarded, use AU 10, 16, 25 to operate lips and AU 26 to open jaw, as well as Tongue specific AUs.
Mouth	24	pressing of lips (status)	0: lips at rest; 1: lips pressed
	25	parting of lips (status)	0: lips at rest; 1: lips parted parting appears at center of lips only. side parting uses AU 10 and 16. Accounts more for detail lip shape and may be removed eventually.
	26	jaw drop (status)	0: upper and lower teeth are touching

			1: jaw opened wide (max)
Eye	05	only specifies raising the eyelid from neutral position (different from rest position)	0: upper eyelid closed; 1: eyelid fully opened
	ePS	undefined	Pupil Stretcher 0: pupil fully contracted, 1: pupil fully dilated
Eye	61.5	undefined but related to 61 & 62	eye orientation on Z axis (pan) Value in radians (0: iris facing straight, positive turns left)
	63.5	undefined but related to 63 & 64	eye orientation on X axis (tilt) Value in radians (positive tilts upwards)
Spine (head)	51.5	undefined but related to 51 & 52 (head turn) and M60	orientation of head Z axis (pan) Value in radians (0: head facing straight, positive pans left)
	53.5	undefined but related to 53 & 54 (head down/up) and M59	orientation of head X axis (tilt) Value in radians (0: head facing straight, positive tilts upwards)
	55.5	undefined but related to 55 & 56 (head tilt)	orientation of head Y axis (roll) badly named tilt sometimes). Value in radians (0: head facing straight, positive rolls right)
	57.5	undefined but related to 57 & 58 (head forward/backward)	position of head on Y axis This is character dependent. -1: most backward, 1: most forward.
	58.5	undefined, NOT related to M59 or M60	position of head on X axis This is character dependent. -1: leftmost, 1: rightmost.
	59.5	undefined, NOT related to M59 nor M60	position of head on Z axis This is character dependent (if applicable). -1: lowest, 0: centered, 1: highest.
(thorax)	TX	undefined	orientation of Thorax X axis (tilt) (consider top of thorax), number of sections is character dependent. value in radians
	TY	undefined	orientation of Thorax Y axis (roll) value in radians
	TZ	undefined	orientation of Thorax Z axis (pan) value in radians
	thB	undefined	Thorax breathing 0: full exhalation, 1: full inhalation

			Belly breathing is another AU
Tongue	93X	undefined	position of tip of tongue on X axis -1: leftmost, 1: rightmost
	93Y	undefined	position of tip of tongue on Y axis. -1: most backward, 1: most forward
	93Z	undefined	position of tip of tongue on Z axis. -1: lowest, 1: highest
	93mZ	undefined	Z position of middle of tongue This is character dependent. -1: lowest, 0: neutral, 1: uppermost
	93bT	undefined	Tongue gutturer This is character dependent. -1: lowest in throat, 1: most front
	94	undefined	Tongue ZX stretcher -1: most horizontal flat, 1: most vertical stretch
	95	undefined	Tongue roller (on Y axis). This is character dependent. 0: flat, 1: most rolled (pipe-like)
	96-99	undefined	
Shoulders	SY	undefined	orientation of Shoulders Y axis (tilt) value in radians (0: sternum-shoulder and spine form a right angle)
	SZ	undefined	orientation of Shoulders Z axis (tilt) value in radians (0: shoulders are in line with spine)
Skin Effects	skB	undefined	Triggers blushing: 0 no blushing, 1: max blushing
	skS	undefined	Triggers sweating: 0 no sweating, 1: max sweating

Also, most 'M' values (e.g. M59, M83..) and numbers for 'gross behavior' (40,50,80-82,84,85,91,92) are not used since they represent movement. It is tempting to use these numbers to extend FACS, but that could lead to confusion. As a consequence it was decided extensions to the system would use an alphabetical labeling convention.

Tongue: The tongue is divided in 3 sections, each having a Z position. These 3 sections are : the tip, the lingual tonsil (most backward area) and the area in the middle. Also, one may note the relative Z position of these sections is enough to create most visible general foldings, while specific foldings have their own AU (e.g. Tongue roller).

Values for positioning (head and tongue) are relative so any design can produce convincing results. However, for absolute positioning (as with IK), an extra component could be used with character-specific parameters to provide the appropriate relative value.

Animation guidelines

There are 2 ways of animating a character: state-based or event-based. State animation ensures all commands of a datablock are processed at the same time, while event animation allows available modules (not processing any command) to process a datablock as long as no command of this datablock requires a non-available module.

State-based animation

State-based animation rely on CHLAS' buffering behaviour. In this manner, sentences should be sent in bursts of commands.

Most sentences are emotionally influenced. For instance, when a robot's chatting about the weather, parts of the sentence may change its attitude:

"The weather is so hot outside but it's so cold inside!"

It is very likely that the facial expression would change during this sentence. One can imagine such transitions:

```
neutral;          "The weather is"; ((0.0, 0.0, 5.0)); ((0.0, 0.0, 0.0)); ; tag_1
surprised*0.3;    "so hot";          ; ((5.3, 2.2, 1.3)); ; tag_2
surprised*-0.3;   "outside";         ((1.3, 0.0, 5.2)); ((0.0, 0.0, 0.0)); ; tag_3
fear* 0.2;        "but it's so cold"; [1.0, .0, .0]; ((-2.0, -1.3,.0)); ; tag_4
neutral;          "inside.";         ; ((.0, .0, .0)); ; tag_5
```

CHLAS would bufferize and acknowledge processing of each datablock in a timely manner.

Event-based animation

Event-based animation relies on CHLAS' scheduling of modules. In this manner, datablocks are sent on time, leaving empty unused elements of a command.

The same example can be processed this way:

```
neutral;          ;          ;          ; ; tag_1
;                ; ((0.0, 0.0, 5.0)); ((0.0, 0.0, 0.0)); ; tag_2
; "The weather is so hot outside but it's so cold inside!"; ; ; tag_3
- delay estimated for the TTS to reach utterance of "so hot" -
surprised*0.3;    ;          ;          ; ; tag_4
;                ;          ; ((5.3, 2.2, 1.3)); ; tag_5
```

and so on ..

As shown with tag_2, state and event animation can be mixed together. In fact their usage is usually mixed since they serve different compatible purposes.

Attack discrepancies in Target Frames

On ARAS side, the following target frame looks like The Hulk getting angry; it plays for 2 seconds:

```
(('07', .9, 2), ('09', .9, .5), ('01', .9, 1.0), ('04', .9, 1.0), ('05', .9, 1.0), ('10', .9, 1.1))
```

In this example, while the frame is being rendered, movements of the quicker triplets end early and may not be updated until completion of the movement of the slowest triplet. This makes sense.. for example, if you make a big grin on your face, your eyes area will "squeeze" inwards immediately and stay suspended but your lips/mouth will continue to stretch.

Appendix C

Active Learning Experiment

“I would like to learn this one”
“could you teach me this one?”
“this one looks interesting”
“now, what about this one?”
“this is interesting”
“em, what about this one?”
“what about this one?”
“I would like to know what this is”
“ok, what do we have here?”
“yes, this looks interesting”
“what about this one?”
“em, I would like to know what this is”

Table C.1: LightHead’s utterances in active learning condition.

number	AL	GG success	AL response	number	AL	GG success	AL response
1	no	0.48	0.3	3	yes	0.72	0.48
2	no	0.64	0.28	4	yes	0.66	0.76
5	no	0.56	0.38	7	yes	0.56	0.38
6	no	0.8	0.36	8	yes	0.62	0.94
11	no	0.6	0.3	9	yes	0.52	0.4
13	no	0.46	0.26	10	yes	0.68	0.46
14	no	0.54	0.44	12	yes	0.64	0.58
15	no	0.68	0.32	16	yes	0.46	0.38
17	no	0.56	0.38	19	yes	0.6	0.46
18	no	0.64	0.22	20	yes	0.62	0.62
22	no	0.54	0.22	21	yes	0.76	0.44
24	no	0.4	0.38	23	yes	0.54	0.46
25	no	0.42	0.38	26	yes	0.62	0.8
29	no	0.62	0.28	27	yes	0.56	0.44
30	no	0.68	0.48	28	yes	0.7	0.44
31	no	0.54	0.3	32	yes	0.7	0.58
33	no	0.48	0.36	34	yes	0.72	0.42
36	no	0.62	0.26	35	yes	0.56	0.38
38	no	0.54	0.32	37	yes	0.62	0.86
40	no	0.7	0.3	39	yes	0.6	0.8
41	no	0.62	0.18				

Table C.2: Detail of the participants' game success and alignment for both active learning and baseline conditions.

Social Robot Teaching Questionnaire

Participant number:	Age:
Gender: F / M	Native speaker: yes / no

Please answer the following questions by placing an 'X' on the spot that best reflects your answer.
Additionally, you can provide comments to elaborate your answers.

1. How do you rate your interaction with the robot?

not satisfactory at all						very satisfactory
comments						

2. How do you rate the robot's behaviour?

not natural at all						very natural
comments						

3. Do you have any experience with robots?

I have no experience with robots						I have a lot of experience with robots
comments						

4. Who was in control of the teaching sessions?

I was in control						the robot was in control
comments						

5. On what basis did you choose the animal examples as topic? Please explain.

--

6. Do you like science fiction (books, film, etc)?

--	--	--	--	--	--	--

I don't like science fiction at all

I very much like science fiction

comments

--

7. How many emotions do you think the robot has?

--	--	--	--	--	--	--

the robot has no emotions

the robot has a lot of emotions

comments

--

8. How smart do you think the robot is?

--	--	--	--	--	--	--

the robot is not smart at all

the robot is very smart

comments

--

9. How many hours per week do you spend using a computer?

hours computer use per week (estimate):

comments

--

10. General comments

--

How I am in general

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which **you agree or disagree with that statement**.

1 Disagree Strongly	2 Disagree a little	3 Neither agree nor disagree	4 Agree a little	5 Agree strongly
---------------------------	---------------------------	------------------------------------	------------------------	------------------------

I am someone who...

- | | |
|--|---|
| 1. _____ Is talkative | 23. _____ Tends to be lazy |
| 2. _____ Tends to find fault with others | 24. _____ Is emotionally stable, not easily upset |
| 3. _____ Does a thorough job | 25. _____ Is inventive |
| 4. _____ Is depressed, blue | 26. _____ Has an assertive personality |
| 5. _____ Is original, comes up with new ideas | 27. _____ Can be cold and aloof |
| 6. _____ Is reserved | 28. _____ Perseveres until the task is finished |
| 7. _____ Is helpful and unselfish with others | 29. _____ Can be moody |
| 8. _____ Can be somewhat careless | 30. _____ Values artistic, aesthetic experiences |
| 9. _____ Is relaxed, handles stress well | 31. _____ Is sometimes shy, inhibited |
| 10. _____ Is curious about many different things | 32. _____ Is considerate and kind to almost everyone |
| 11. _____ Is full of energy | 33. _____ Does things efficiently |
| 12. _____ Starts quarrels with others | 34. _____ Remains calm in tense situations |
| 13. _____ Is a reliable worker | 35. _____ Prefers work that is routine |
| 14. _____ Can be tense | 36. _____ Is outgoing, sociable |
| 15. _____ Is ingenious, a deep thinker | 37. _____ Is sometimes rude to others |
| 16. _____ Generates a lot of enthusiasm | 38. _____ Makes plans and follows through with them |
| 17. _____ Has a forgiving nature | 39. _____ Gets nervous easily |
| 18. _____ Tends to be disorganized | 40. _____ Likes to reflect, play with ideas |
| 19. _____ Worries a lot | 41. _____ Has few artistic interests |
| 20. _____ Has an active imagination | 42. _____ Likes to cooperate with others |
| 21. _____ Tends to be quiet | 43. _____ Is easily distracted |
| 22. _____ Is generally trusting | 44. _____ Is sophisticated in art, music, or literature |

Appendix D

Ethnic Preferences

Experiment

LightHead robotic museum guide

Instructions

Hi! We are conducting a survey on your preferences for a robotic museum guide. Indeed all information collected is entirely anonymous and will only serve to make better robots. What you will see is a prototype and does not reflect an actual product.

There are 4 required and a final optional group of questions. Going through all the questionnaire should take about 25-30 minutes.

We value your opinion!

Please enter your gender and age group details:

Age

51+ years

Gender

Male

Female

Please set to what extent these statements describe you by selecting the number which best correspond to your experience. For example, if the statement is agree strongly, then select 7. If it is only agree slightly, then select 5.

Don't spend too long over any statement, just give the first answer that comes to your mind. There are no right or wrong answers.

I see myself as someone who...

01. ...Is talkative

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

02. ...Tends to find fault with others

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

03. ...Does a thorough job

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

04. ...Is depressed, blue

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

05. ...Is original, comes up with new ideas

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

06. ...Is reserved

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

07. ...Is helpful and unselfish with others

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

08. ...Can be somewhat careless

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

09. ...Is relaxed, handles stress well

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

10. ...Is curious about many different things

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

11. ...Is full of energy

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

12. ...Starts quarrels with others

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

13. ...Is a reliable worker

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

14. ...Can be tense

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

15. ...Is ingenious, a deep thinker

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

16. ...Generates a lot of enthusiasm

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

17. ...Has a forgiving nature

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

18. ...Tends to be disorganized

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

19. ...Worries a lot

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

20. ...Has an active imagination

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

21. ...Tends to be quiet

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

22. ...Is generally trusting

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

23. ...Tends to be lazy

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

24. ...Is emotionally stable, not easily upset

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

25. ...Is inventive

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

26. ...Has an assertive personality

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

27. ...Can be cold and aloof

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

28. ...Perseveres until the task is finished

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

29. ...Can be moody

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
○ ○ ○ ○ ○ ○ ○

30. ...Values artistic, aesthetic experiences

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

31. ...Is sometimes shy, inhibited

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

32. ...Is considerate and kind to almost everyone

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

33. ...Does things efficiently

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

34. ...Remains calm in tense situations

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

35. ...Prefers work that is routine

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

36. ...Is outgoing, sociable

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

37. ...Is sometimes rude to others

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

38. ...Makes plans and follows through with them

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

39. ...Gets nervous easily

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

40. ...Likes to reflect, play with ideas

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

41. ...Has few artistic interests

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

42. ...Likes to cooperate with others

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

42. ...Can reply honestly to a questionnaire

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

43. ...Is easily distracted

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

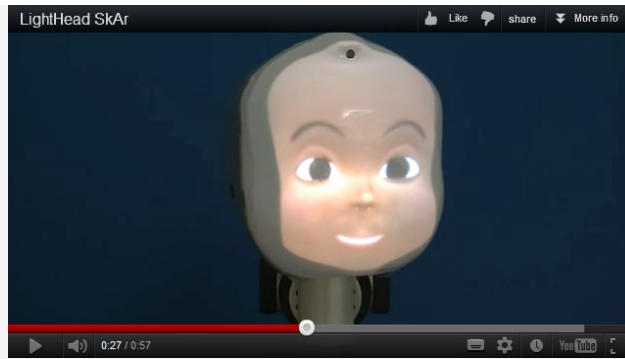
44. ...Is sophisticated in art, music, or literature

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

45. ...Is politically liberal

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Say your local museum decides to deploy a new robot guide, which one would fit you best? Please make sure you fully watched all 5 videos, then rank each robot from most favourite (1) to least favourite (5). Please avoid duplicates.



ranking robot guide #1

1 2 3 4 5

Most Favourite Least Favourite



ranking robot guide #2

1 2 3 4 5

Most Favourite Least Favourite



ranking robot guide #3

1 2 3 4 5

Most Favourite Least Favourite



ranking robot guide #4

1 2 3 4 5

Most Favourite Least Favourite



ranking robot guide #5

1 2 3 4 5

Most Favourite Least Favourite

Now think about your favourite guide of all five robots, and answer the next questions. Do not spend too long over any word-pair. Just give the first answer that comes to your mind. There are no right or wrong answers.

- Non-Humanlike Humanlike
- Stupid Intelligent
- Low Quality High Quality
- Masculine Feminine
- Unengaging Engaging
- Responsible Irresponsible
- Cold Warm
- Weak Strong
- Diligent Lazy
- Impersonal Personal
- Decisive Indecisive

1 2 3 4 5 6 7
 Abnormal Normal

1 2 3 4 5 6 7
 Traditional Contemporary

1 2 3 4 5 6 7
 Serious Fun

1 2 3 4 5 6 7
 Standard Unique

1 2 3 4 5 6 7
 Child Adult

1 2 3 4 5 6 7
 Affordable Expensive

1 2 3 4 5 6 7
 Friendly Unfriendly

1 2 3 4 5 6 7
 Slow Fast

1 2 3 4 5 6 7
 Honest Dishonest

1 2 3 4 5 6 7
 Impolite Polite

1 2 3 4 5 6 7
 Visitor Guide

1 2 3 4 5 6 7
 Active Passive

1 2 3 4 5 6 7
 Unbalanced Balanced

1 2 3 4 5 6 7
 Good Bad

1 2 3 4 5 6 7
 Dishonest Honest

1 2 3 4 5 6 7
 Exciting Boring

1 2 3 4 5 6 7
 Indifferent Interested

1 2 3 4 5 6 7
 Engaged Distracted

1 2 3 4 5 6 7
 Lively Deadpan

1 2 3 4 5 6 7
 I Liked I Disliked

1 2 3 4 5 6 7
 Not as a friend As a friend

1 2 3 4 5 6 7
 Unkind Kind

1 2 3 4 5 6 7
 Trustworthy Untrustworthy

1 2 3 4 5 6 7
 Insensitive Sensitive

What could you say influenced your ranking?

- Head movements
- Age
- Facial appearance
- Geometric design
- Voice
- Blinks
- Eye gaze
- Timing
- Expressivity
- Realism

What sport practices the robot team?

- Chess
- Karate
- Ping Pong
- Football
- Sumo
- Not mentioned

What racial group describes you best?

- American Indian Australian Aborigine or Melanesian
- Caribbean
- Central African or black
- Indian or Bangladeshi
- North African, Arabic, Persian...
- North East Asian (Japanese, Korean, North Chinese, ...)
- South East Asian (Chinese, Vietnamese, ...)
- White Caucasian

You can select up to 2 racial groups. Please note racial affiliation is not related to your nationality, area of living or culture. Moreover, no internationally accepted criteria is possible. If you need help or are interested in the classification used here, check <http://www.racialcompact.com/racesofhumanity.html>

To what other racial group do you belong? (optional)

- American Indian Australian Aborigine or Melanesian
 - Caribbean
 - Central African or black
 - Indian or Bangladeshi
- North African, Arabic, Persian...
- North East Asian (Japanese, Korean, North Chinese, ...)
 - South East Asian (Chinese, Vietnamese, ...)
 - White Caucasian

How familiar are you with computer technology?

- 1 2 3 4 5 6 7
Not at all Very much

How familiar are you with robot technology?

- 1 2 3 4 5 6 7
Not at all Very much

How familiar are you with subspace quantum robot technology?

- 1 2 3 4 5 6 7
Not at all Very much

Thank you for taking part in our 'LightHead' Survey. These are free optional questions; let your voice be heard!

Overall, how do you feel towards being given a tour by a robot?

What changes / additions might you make to the robot to improve its communication / interaction capabilities?

Anything else you'd like to tell us?

References

- Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In A. Esposito, A. Vinciarelli, R. Hoffmann, & V. C. Müller (Eds.), *Cognitive behavioural systems*. Springer. (in press)
- Argall, B. D., & Billard, A. G. (2010). A survey of Tactile Human–Robot Interactions. *Robotics and Autonomous Systems*, *58*(10), 1159–1176. doi: 10.1016/j.robot.2010.07.002
- Arras, K. O., & Cerqui, D. (2005). Do we want to share our lives and bodies with robots? a 2000 people survey. *Autonomous Systems Lab (ASL), Swiss Federal Institute of Technology Lausanne (EPFL), Tech. Rep.*, 0605–001.
- Atienza, R., & Zelinsky, A. (2002). Active Gaze Tracking for Human-Robot Interaction. *Multimodal Interfaces, IEEE International Conference on*, *0*, 261. doi: <http://doi.ieeeecomputersociety.org/10.1109/ICMI.2002.1167004>
- Baillie, J.-C. (2005, aug.). Urbi: towards a universal robotic low-level programming language. In *Intelligent robots and systems, 2005. (iros 2005). 2005 ieee/rsj international conference on* (p. 820 - 825). doi: 10.1109/IROS.2005.1545467
- Baloh, R. W., Sills, A. W., Kumley, W. E., & Honrubia, V. (1975). Quantitative measurement of saccade amplitude, duration, and velocity.

Neurology, 25(11), 1065–1070.

- Bar-Cohen, Y. (2006). Biomimetics using electroactive polymers (EAP) as artificial muscles - A review. *Journal of Advanced Materials*, 38(4), 3–9.
- Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (J. H. Barkow, L. Cosmides, & J. Tooby, Eds.). Oxford University Press.
- Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is The Uncanny Valley An Uncanny Cliff? *ROMAN 2007 The 16th IEEE International Symposium on Robot and Human Interactive Communication, Jeju, Kore*, 368–373. doi: 10.1109/ROMAN.2007.4415111
- Bartneck, C., Reichenbach, J., & Breemen, A. V. (2004). In your face, robot! the influence of a character's embodiment on how users perceive its emotional expressions. In *Technology* (pp. 32–51). Citeseer.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91, 276–292.
- Beira, R., Lopes, M., Praça, M., Santos-Victor, J., Bernardino, A., Metta, G., . . . Saltarén, R. (2006, May). Design of the robot-cub (icub) head. In *Ieee international conference on robotics automation* (pp. 94–100). Orlando.
- Belpaeme, T., & Bleys, J. (2005). Explaining universal colour categories through a constrained acquisition process. *Adaptive Behavior*, 13(4), 293-310.
- Bennewitz, M., Faber, F., Joho, D., Schreiber, M., & Behnke, S. (2005). *Towards a humanoid museum guide robot that interacts with multiple persons* (No. December). IEEE. doi: 10.1109/ICHR.2005.1573603
- Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 401–432). Cambridge

- University Press.
- Beskow, J., & Al Moubayed, S. (2010). Perception of gaze direction in 2D and 3D facial projections. *October*, 10044. doi: 10.1145/1924035.1924051
- Bickel, B., Kaufmann, P., Skouras, M., Thomaszewski, B., Bradley, D., Beeler, T., ... Gross, M. (2012). Physical face cloning. *ACM Transactions on Graphics (TOG)*, 31(4), 118.
- Bickmore, T., Schulman, D., & Yin, L. (2010). Maintaining Engagement in Long-term Interventions with Relational Agents. *Applied artificial intelligence AAI*, 24(6), 648–666. doi: 10.1080/08839514.2010.492259
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2), 293–327. doi: 10.1145/1067860.1067867
- Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C. L., & Lee, D. (2006). The art of designing robot faces: dimensions for human-robot interaction. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction* (pp. 331–332). New York, NY, USA: ACM. doi: 10.1145/1121241.1121301
- Bohus, D., & Horvitz, E. (2011). Multiparty Turn Taking in Situated Dialog : Study , Lessons , and Directions. *Proceedings of the SIGDIAL 2011 the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue(1974)*, 98–109.
- Brå ten, S. (1998). *Intersubjective Communication and Emotion in Early Ontogeny*. Cambridge, UK: Cambridge University Press.
- Bratman, M. E. (1992). Shared Cooperative Activity. *Philosophical Review*, 101(2), 327–341. doi: 10.2307/2185537
- Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA, USA: MIT Press.

- Breazeal, C., Siegel, M., Berlin, M., Gray, J., Grupen, R., Deegan, P., ... McBean, J. (2008). Mobile, dexterous, social robots for mobile manipulation and human-robot interaction. In *Acm siggraph 2008 new tech demos* (pp. 27:1–27:1). New York, NY, USA: ACM. doi: 10.1145/1401615.1401642
- Brooks, A. G., Gray, J., Hoffman, G., Lockerd, A., Lee, H., & Breazeal, C. (2004). Robot’s play: interactive games with social machines. *Computers in Entertainment CIE*, 2(3), 10. doi: 10.1145/1027154.1027171
- Brooks, R. A. (1986). A robust layered control system for a mobile robot. *Robotics and Automation, IEEE Journal of*, 2(1), 14–23.
- Bugmann, G. (2011). What can a personal robot do for you? *Towards Autonomous Robotic Systems*, 360–371.
- Cakmak, M., Chao, C., & Thomaz, A. L. (2010). *Designing Interactions for Robot Active Learners* (Vol. 2) (No. 2). doi: 10.1109/TAMD.2010.2051030
- Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., ... Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In M. N. Huhns & M. P. Singh (Eds.), *Proc of acm siggraph* (pp. 413–420). ACM Press.
- Chen, J. J., Menezes, N. J., Bradley, A. D., & North, T. A. (2011). Opportunities for Crowdsourcing Research on Amazon Mechanical Turk. *Human Factors*, 5, 3.
- Clodic, A., Fleury, S., Alami, R., Chatila, R., Bailly, G., Brethes, L., ... Montreuil, V. (2006). *Rackham: An Interactive Robot-Guide*. IEEE. doi: 10.1109/ROMAN.2006.314378
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals* (Vol. 232; P. Ekman, Ed.) (No. 4). John Murray. doi:

10.1097/00000441-195610000-00024

- Dautenhahn, K. (2004). Robots we like to live with?! - a developmental perspective on a personalized, life-long robot companion. *ROMAN 2004 13th IEEE International Workshop on Robot and Human Interactive Communication IEEE Catalog No04TH8759*, 20, 17–22.
- David S. Bolme, S. O. (2008). *Pyvision - computer vision toolkit*. Website - <http://pyvision.sourceforge.net>. Retrieved from <http://pyvision.sourceforge.net>
- Deboer, M., & Boxer, A. M. (1979). Signal functions of infant facial expression and gaze direction during mother-infant face-to-face play. *Child Development*, vol, 50no4pp1215–1218.
- de Greeff, J., & Belpaeme, T. (2011). The development of shared meaning within different embodiments. In J. Triesch (Ed.), *Proceedings of the joint international conference on developmental learning (icdl) & epigenetic robotics 2011*. Frankfurt, Germany: IEEE.
- de Greeff, J., Delaunay, F., & Belpaeme, T. (2009, June). Human-Robot Interaction in Concept Acquisition: a computational model. In *Ieee 8th international conference on development and learning* (pp. 1–6). Shanghai, China: IEEE. doi: 10.1109/DEVLRN.2009.5175532
- de Greeff, J., Delaunay, F., & Belpaeme, T. (2012). Active robot learning with human tutelage. In *Proceedings of the joint international conference on developmental learning (icdl) & epigenetic robotics*. San Diego, USA: IEEE.
- Delaunay, F., de Greeff, J., & Belpaeme, T. (2009). Towards Retro-projected Robot Faces: an Alternative to Mechatronic and Android Faces. In *Ieee roman 2009 conference*. Toyama, Japan.
- Delaunay, F., de Greeff, J., & Belpaeme, T. (2010). A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *Hri 2010*. Osaka, Japan.

- DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on designing interactive systems: processes, practices, methods, and techniques* (pp. 321–326). New York, NY, USA: ACM. doi: 10.1145/778712.778756
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283–292. doi: 10.1037/h0033031
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry: Interpersonal and Biological Processes*, *32*(1), 88–106.
- Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252. doi: 10.1007/BF00987191
- Ekman, P., & O’Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: the utility of voluntary and involuntary behavior in detecting deception. *Behavioral sciences the law*, *24*(5), 673–686.
- Emery, N. J. (2000, August). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and biobehavioral reviews*, *24*(6), 581–604.
- Engineered Arts Ltd. (2006). *RoboThespian, the original robot actor - humanoid interactive robots to educate and entertain*.
- Fanaswala, I., Browning, B., & Sakr, M. (2011, march). Interactional disparities in english and arabic native speakers with a bi-lingual robot receptionist. In *Human-robot interaction (hri), 2011 6th acm/ieee international conference on* (p. 133 -134).
- Fischer, K., Lohan, K., & Foth, K. (2012). Levels of embodiment: Linguistic analyses of factors influencing hri. In *Proceedings of the ieee/acm international conference on human-robot interaction*. Boston, MA.
- Floridi, L. (2004). Open Problems in the Philosophy of Informa-

- tion. *Metaphilosophy*, 35(4), 554–582. doi: 10.1111/j.1467-9973.2004.00336.x
- Fong, T., Kunz, C., Hiatt, L. M., & Bugajska, M. (2006). The human-robot interaction operating system. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction* (pp. 41–48). New York, NY, USA: ACM. doi: 10.1145/1121241.1121251
- Ford, C., Bugmann, G., & Culverhouse, P. (2013). Modeling the human blink: A computational model for use within human–robot interaction. *International Journal of Humanoid Robotics*, 10(01).
- Ford, C. C., Bugmann, G., & Culverhouse, P. (2010). Eye movement and facial expression in human robot communication. In *Keer2010* (pp. 717–729). INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010.
- Forlizzi, J., & DiSalvo, C. (2006). Service robots in the domestic environment: a study of the roomba vacuum in the home. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction* (pp. 258–265). New York, NY, USA: ACM. doi: 10.1145/1121241.1121286
- Frank, A., & Asuncion, A. (2010). *UCI machine learning repository*.
- Fujita, M. (2001, October). AIBO: Toward the Era of Digital Creatures. *The International Journal of Robotics Research*, 20(10), 781–794. doi: 10.1177/02783640122068092
- Gockley, R., Simmons, R., Wang, J., Busquets, D., & DiSalvo, C. (2004). Grace george: Social robots at aaai. In *Proceedings of aaai'04. mobile robot competition workshop (technical report ws-04-11)* (pp. 15–20).
- Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.
- Gong, L. (2008). The boundary of racial prejudice: Comparing preferences for computer-synthesized White, Black, and robot characters. *Computers in Human Behavior*, 24(5), 2074–2093. doi: 10.1016/j.chb.2007.09.008

- Griffey, J. (2012). Chapter 4: Absolutely fab-ulous. *Library Technology Reports*, 48(3), 21–24.
- Groom, V., Bailenson, J. N., & Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence*, 4(3), 231–248. doi: 10.1080/15534510802643750
- Guizzo, E., & Ackerman, E. (2012). The rise of the robot worker. *Spectrum, IEEE*, 49(10), 34–41.
- Hadjikhani, N., Kveraga, K., Naik, P., & Ahlfors, S. (2009). Early (m170) activation of face-specific cortex by face-like objects. *Neuroreport*, 20(4), 403407. doi: doi:10.1097/WNR.0b013e328325a8e1
- Hall, E. T. (1966). *The hidden dimension*. Bantam Doubleday Dell Publishing Group.
- Hall, J. (1978). Gender effects in decoding nonverbal cues. *Psychological bulletin*, 85(4), 845.
- Hampton, D. M., & Chung, C. (2003). *Interactive toy*. US office of patents.
- Hansen, D. W., & Ji, Q. (2010). In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(1). doi: <http://doi.IEEEcomputersociety.org/10.1109/TPAMI.2009.30>
- Hanson, D. (2005). Expanding the Aesthetics Possibilities for Humanlike Robots. In *Proceedings of iee humanoid robotics conference, special session on the uncanny valley*. Tskuba, Japan.
- Hanson, D., Baurmann, S., Riccio, T., Margolin, R., Dockins, T., Tavares, M., & Carpenter, K. (2009). Zeno: a cognitive character. In *Ai magazine, and special proc. of aaai national conference, chicago*.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3), 335–346. doi: 10.1016/0167-2789(90)90087-6
- Hashimoto, M., & Kondo, H. (2008). Effect of emotional expression to gaze guidance using a face robot. In *Proceedings of the 17th iee interna-*

- tional symposium on robot human interactive communication (roman 2008)* (pp. 91–95). Tamatsu, Y.
- Hashimoto, M., & Morooka, D. (2005). Facial Expression of a Robot using a Curved Surface Display. In *Proceedings of the ieee/rsj international conference on intelligent robots and systems* (pp. 2532–2537).
- Hashimoto, T., Hiramatsu, S., Tsuji, T., & Kobayashi, H. (2007a). Realization and Evaluation of Realistic Nod with Receptionist Robot SAYA. In *Roman 2007 the 16th ieee international symposium on robot and human interactive communication* (pp. 326–331). IEEE. doi: 10.1109/ROMAN.2007.4415103
- Hashimoto, T., Hiramatsu, S., Tsuji, T., & Kobayashi, H. (2007b, aug.). Realization and evaluation of realistic nod with receptionist robot saya. In *Robot and human interactive communication, 2007. ro-man 2007. the 16th ieee international symposium on* (p. 326 -331). doi: 10.1109/ROMAN.2007.4415103
- Hegel, F., Eyssel, F., & Wrede, B. (2010, sept.). The social robot ‘flobi’: Key concepts of industrial design. In *Ro-man, 2010 ieee* (p. 107 -112). doi: 10.1109/ROMAN.2010.5598691
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual review of psychology*, 53(1), 575–604.
- Honda Corporation. (2000). *Honda Corporation Asimo*.
- Hoque, M., Onuki, T., Kobayashi, Y., & Kuno, Y. (2011, may). Controlling human attention through robot’s gaze behaviors. In *Human system interactions (hsi), 2011 4th international conference on* (p. 195 -202). doi: 10.1109/HSI.2011.5937366
- Joachim de Greeff. (2012). *Interactive Concept Acquisition for Embodied Artificial Agents*. Unpublished doctoral dissertation, Plymouth University.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five

Inventory—Versions 4a and 54. *Institute of Personality and Social Research*.

- Jordan, P., & Hernandez-Reif, M. (2009). Reexamination of Young Children's Racial Attitudes and Skin Tone Preferences. *Journal of Black Psychology, 35*(3), 388–403. doi: 10.1177/0095798409333621
- Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., & Hagita, N. (2009). An affective guide robot in a shopping mall. In *Proceedings of the 4th acm/ieee international conference on human robot interaction* (pp. 173–180). New York, NY, USA: ACM. doi: 10.1145/1514095.1514127
- Karahalios, K. G., & Dobson, K. (2005). Chit chat club: bridging virtual and physical space for social interaction. In *Chi '05 extended abstracts on human factors in computing systems* (pp. 1957–1960). New York, NY, USA: ACM. doi: 10.1145/1056808.1057066
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*(0), 22 - 63. doi: 10.1016/0001-6918(67)90005-4
- Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta psychologica, 32*, 101–125.
- Kidd, C. D. (2008). *Designing for Long-Term Human-Robot Interaction Application to Weight Loss*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Kittur, A., Chi, E., & Suh, B. (2008). Crowdsourcing user studies with mechanical turk. In *Proceedings of the twenty-sixth annual sigchi conference on human factors in computing systems* (pp. 453–456).
- Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature, vol, 387*no6635pp767–768.
- Krumhuber, E., Tamarit, L., Roesch, E. B., & Scherer, K. R. (2012). FACSGen 2.0 animation software: Generating three-dimensional FACS-valid facial expressions for emotion research. *Emotion Washington Dc, 12*(2), 0–13. doi: 10.1037/a0026632

- Kühnlenz, K., Sosnowski, S., & Buss, M. (2010). Impact of animal-like features on emotion expression of robot head eddie. *Advanced Robotics*, *24*(8-9), 1239–1255.
- Kuratate, T., Matsusaka, Y., Pierce, B., & Cheng, G. (2011, oct.). Maskbot: A life-size robot head using talking head animation for human-robot communication. In *Humanoid robots (humanoids), 2011 11th IEEE-RAS international conference on* (p. 99 -104). doi: 10.1109/Humanoids.2011.6100842
- Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59. doi: DOI: 10.1016/S1364-6613(99)01436-9
- Lee, M. K., Kiesler, S., & Forlizzi, J. (2010). Receptionist or information kiosk: how do people talk with a robot? In *Proceedings of the 2010 ACM conference on computer supported cooperative work* (pp. 31–40). New York, NY, USA: ACM. doi: 10.1145/1718918.1718927
- Liang, J., Huang, L., Li, N., Huang, Y., Wu, Y., Fang, S., ... Chen, Y. (2012). Electromechanical actuator with controllable motion, fast response rate, and high-frequency resonance based on graphene and polydiacetylene. *ACS nano*, *6*(Xx), 4508–19. doi: 10.1021/nm3006812
- Lincoln, P., Welch, G., Nashel, A., Ilie, A., State, A., & Fuchs, H. (2009). *Animatronic Shader Lamps Avatars* (Vol. 15) (No. 2-3). IEEE. doi: 10.1109/ISMAR.2009.5336503
- Lockerd, A., & Breazeal, C. (2004). *Tutelage and socially guided robot learning* (Vol. 4). IEEE. doi: 10.1109/IROS.2004.1389954
- Ly, O., Lapeyre, M., & Oudeyer, P.-Y. (2011). *Bio-inspired vertebral column, compliance and semi-passive dynamics in a lightweight humanoid robot* (No. 1). IEEE. doi: 10.1109/IROS.2011.6095019
- MacDorman, K. F. (2005). Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it? In *Proceedings of the*

cogsci 2005 workshop: Toward social mechanisms of android science
(pp. 106–118).

- Mahan, J. (1976). Black and White children's racial identification and preference. *Journal of Black Psychology*, 3(1), 47.
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory cognition*, 22(6), 657–672. doi: 10.1080/135062800407202
- Mason, W., & Suri, S. (2011). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1–23. doi: 10.3758/s13428-011-0124-6
- Matsui, D., Minato, T., MacDorman, K. F., & Ishiguro, H. (2005). Generating natural motion in an android by mapping human motion. *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, 3301–3308. doi: 10.1109/IROS.2005.1545125
- Matsusaka, Y. (2008). *Open HRI, Opensource software components for Human Robot Interaction*.
- Maurer, D. (1985). Infant's perception of facedness. In T. Field & M. Fox (Eds.), *Social perception in infants*. Norwood, NJ: Ablex.
- McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2006). Cultural display rules drive eye gaze during thinking. *Journal of Cross-Cultural Psychology*, vol, 37no6pp717–722.
- MDS project at the Personal Robots Group, MIT Media Lab.* (2008).
- Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 491–500.
- Michael Naimark, Nicholas Negroponte, & Chris Schmandt. (1980). *Talking Head Projection*.
- Mine, M., van Baar, J., Grundhofer, A., Rose, D., & Yang, B. (2012, july).

- Projection-based augmented reality in disney theme parks. *Computer*, 45(7), 32 -40. doi: 10.1109/MC.2012.154
- Misawa, K., Ishiguro, Y., & Rekimoto, J. (2012a). Livemask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. In *Proceedings of the international working conference on advanced visual interfaces* (pp. 394–397).
- Misawa, K., Ishiguro, Y., & Rekimoto, J. (2012b). Ma petite chérie: what are you looking at?: a small telepresence system to support remote collaborative work for intimate communication. In *Proceedings of the 3rd augmented human international conference* (pp. 17:1–17:5). New York, NY, USA: ACM. doi: 10.1145/2160125.2160142
- Miyauchi, D., Nakamura, A., & Kuno, Y. (2005). Bidirectional Eye Contact for Human-Robot Communication. *IEICE - Trans. Inf. Syst.*, E88-D(11), 2509–2516. doi: <http://dx.doi.org/10.1093/ietisy/e88-d.11.2509>
- Miyauchi, D., Sakurai, A., Nakamura, A., & Kuno, Y. (2004). Active eye contact for human-robot communication. In *Chi '04: Chi '04 extended abstracts on human factors in computing systems* (pp. 1099–1102). New York, NY, USA: ACM. doi: <http://doi.acm.org/10.1145/985921.985998>
- Monceaux, J., Becker, J., Boudier, C., & Mazel, A. (2009). Demonstration: first steps in emotional expression of the humanoid robot Nao. *International Conference on Multimodal Interfaces*.
- Moore, R. K. (2012, November). A Bayesian explanation of the ‘Uncanny Valley’ effect and related psychological phenomena. *Scientific Reports*, 2(2). doi: 10.1038/srep00864
- Moore, R. K., & Series, R. W. (2002). *Human-machine interface apparatus*. (Patent US 2002/0015037 A1)
- Mori, M. (1970). Bukimi no tani [the uncanny valley] (K. F. MacDorman

- & T. Minato, Trans.). *Energy*, 7(4), 33–35.
- Mumm, J., & Mutlu, B. (2011). Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on human-robot interaction* (pp. 331–338). New York, NY, USA: ACM. doi: 10.1145/1957656.1957786
- Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Human factors* (Vol. 2, pp. 61–68). ACM. doi: 10.1145/1514095.1514109
- Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *Mbti manual: A guide to the development and use of the myers-briggs type indicator* (Vol. 3). Consulting Psychologists Press Palo Alto, CA.
- Nagai, Y., Asada, M., & Hosoda, K. (2006). Learning for joint attention helped by functional development. *Advanced Robotics*, 20(10), 1165–1181. doi: doi:10.1163/156855306778522497
- Neuberg, S. L., Kenrick, D. T., & Schaller, M. (1998). Evolutionary social psychology. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 982–1026). McGraw-Hill. doi: 10.1006/nimg.2002.1213
- Nicolao, M., & Moore, R. (2012). Establishing some principles of human speech production through two-dimensional computational models.
- Noh, J.-y., & Neumann, U. (1998). *A Survey of Facial Modeling and Animation Techniques* (Tech. Rep.). USC Technical Report, 99–705.
- Ochsner, K. N. (2004). Current directions in social cognitive neuroscience. *Current Opinion in Neurobiology*, 14(2), 254 - 258. doi: 10.1016/j.conb.2004.03.011
- Oda, M., & Isono, K. (2008). *Effects of time function and expression speed on the intensity and realism of facial expressions*. doi: 10.1109/IC-SMC.2008.4811429

- Oh, J.-H. O. J.-H., Hanson, D., Kim, W.-S. K. W.-S., Han, Y. H. Y., Kim, J.-Y. K. J.-Y., & Park, I.-W. P. I.-W. (2006). *Design of Android type Humanoid Robot Albert HUBO*. IEEE. doi: 10.1109/IROS.2006.281935
- Onishi, M., Luo, Z., Odashima, T., Hirano, S., Tahara, K., & Mukai, T. (2007, April). Generation of Human Care Behaviors by Human-Interactive Robot RI-MAN. *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 3128–3129. doi: 10.1109/ROBOT.2007.363950
- Oudeyer, P., & Delaunay, F. (2008). Developmental exploration in the cultural evolution of lexical conventions. In *8th international conference on epigenetic robotics: Modeling cognitive development in robotic systems*.
- Oudeyer, P.-Y., & Kaplan, F. (2007). Language evolution as a Darwinian process: computational studies. *Cognitive Processing*, 8(1), 21–35.
- Pantic, M., & Patras, I. (2006, april). Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 36(2), 433–449. doi: 10.1109/TSMCB.2005.859075
- Pantic, M., & Rothkrantz, L. J. M. (2000). Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 18(11), 881–905. doi: DOI: 10.1016/S0262-8856(00)00034-2
- Parent, R. (2012). *Computer animation: Algorithms and techniques*. Morgan Kaufmann.
- Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. In H. Zhang, T.-S. Chua, R. Steinmetz, M. S. Kankanhalli, & L. Wilcox (Eds.), *Multimedia 05 proceedings of the 13th annual acm international conference on multimedia* (pp. 683–689). ACM Press. doi: 10.1145/1101149.1101301

- Peters, C., & Qureshi, A. (2010). A head movement propensity model for animating gaze shifts and blinks of virtual characters. *Computers & Graphics, 34*, 677–687.
- Picard, R. W. (1997). *Affective computing*. Cambridge, MA, USA: MIT Press.
- Picot, A., Bailly, G., Elisei, F., & Raidt, S. (2007). Scrutinizing Natural Scenes: Controlling the Gaze of an Embodied Conversational Agent. In *Iva* (pp. 272–282).
- Pierce, B., Kuratate, T., Maejima, A., Morishima, S., Matsusaka, Y., Durkovic, M., ... Cheng, G. (2012, may). Development of an integrated multi-modal communication robotic face. In *Advanced robotics and its social impacts (arso), 2012 ieee workshop on* (p. 101 -102). doi: 10.1109/ARSO.2012.6213408
- Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., & Salesin, D. H. (1998). Synthesizing realistic facial expressions from photographs. *Proceedings of the 25th annual conference on Computer graphics and interactive techniques SIGGRAPH 98, 2(3)*, 75–84. doi: 10.1145/280814.280825
- Potkonjak, V., Svetozarevic, B., Jovanovic, K., & Holland, O. (2011). Anthropomorphic robot with passive compliance-contact dynamics and control. In *Control & automation (med), 2011 19th mediterranean conference on* (pp. 1059–1064).
- Prendinger, H., & Ishizuka, M. (2005). Human Physiology as a Basis for Designing and Evaluating Affective Communication with Life-Like Characters. *IEICE Trans Inf Syst, E88-D(11)*, 2453–2460. doi: 10.1093/ietisy/e88-d.11.2453
- Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., ... Ng, A. (2009). Ros: an open-source robot operating system. In *Icra workshop on open source software* (pp. 1–8). IEEE.
- Ray, C., Mondada, F., & Siegwart, R. (2008, sept.). What do peo-

- ple expect from robots? In *Intelligent robots and systems, 2008. iros 2008. iee/rsj international conference on* (p. 3816 -3821). doi: 10.1109/IROS.2008.4650714
- Rich, C., Ponsler, B., Holroyd, A., & Sidner, C. L. (2010). Recognizing engagement in human-robot interaction. *2010 5th ACMIEEE International Conference on HumanRobot Interaction HRI* (April), 375–382. doi: 10.1109/HRI.2010.5453163
- Robins, B., Dautenhahn, K., & Dickerson, P. (2009). *From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot*. IEEE. doi: 10.1109/ACHI.2009.32
- Rogers, S., Lunsford, M., Strother, L., & Kubovy, M. (2003, May). The Mona Lisa effect: Perception of gaze direction in real and pictured faces. In *Studies in perception and action* (pp. 19–24). Psychology Press.
- Ruiz-Del-Solar, J., & Loncomilla, P. (2009). Robot Head Pose Detection and Gaze Direction Determination Using Local Invariant Features. *Advanced Robotics*, 23(3), 305–328. doi: 10.1163/156855308X397497
- Saerbeck, M., & Breemen, A. J. N. V. (2007). *Design guidelines and tools for creating believable motion for personal robots*. IEEE. doi: 10.1109/RO-MAN.2007.4415114
- Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., & Hagita, N. (2007). Android as a telecommunication medium with a human-like presence. *Proceeding of the ACMIEEE international conference on Humanrobot interaction HRI 07*, 193–200. doi: 10.1145/1228716.1228743
- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience*, 7(4), 413–22. doi: 10.1093/scan/nsr025

- Scassellati, B. (1998). Building behaviors developmentally: a new formalism. In *Proc. of the 1998 aai spring symposium on integrating robotics research*.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2), 227–256. doi: 10.1016/S0167-6393(02)00084-5
- Schraft, R., Schaeffer, C., & May, T. (1998, aug-4 sep). Care-o-bottm: the concept of a system for assisting elderly or disabled persons in home environments. In *Industrial electronics society, 1998. iecon '98. proceedings of the 24th annual conference of the ieee* (Vol. 4, p. 2476–2481 vol.4). doi: 10.1109/IECON.1998.724115
- Schroeder, B. (2008). *Facial animation: overview and some recent papers*.
- Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *CoRR*, abs/cs/0507056.
- Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expressions. In *The psychology of facial expression* (pp. 295–320). Cambridge University Press.
- Snider, J., & Osgood, C. (1969). *Semantic differential technique: A source-book*. Aldine Publishing Company Chicago.
- Sony. (2003). *SDR-4X QRIO*.
- Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, 1(1), 1–34.
- Steels, L., & Kaplan, F. (2000). Aibo's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1), 3–32.
- Syrdal, D. S., Dautenhahn, K., Woods, S. N., Walters, M. L., & Koay, K. L. (2007). Looking good? Appearance preferences and robot personality inferences at zero acquaintance. In *Proceedings of* (pp. 86–92). AAAI Press.
- Tanaka, F., Cicourel, A., & Movellan, J. R. (2007, November). Socialization

- between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46), 17954–8. doi: 10.1073/pnas.0707769104
- Terveen, L., & McDonald, D. (2005). Social matching: A framework and research agenda. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(3), 401–434.
- Thomas, F., & Johnston, O. (1995). *The illusion of life: Disney animation*. Disney Editions.
- Thomaz, A. L. (2006). *Socially Guided Machine Learning* (Doctor of Philosophy in Media Arts and Sciences No. 1999).
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Lawrence Erlbaum Associates.
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- van Breemen, A., Yan, X., & Meerbeek, B. (2005). iCat: an animated user-interface robot with personality. In *Aamas '05: Proceedings of the fourth international joint conference on autonomous agents and multiagent systems* (pp. 143–144). New York, NY, USA: ACM. doi: <http://doi.acm.org/10.1145/1082473.1082823>
- Wada, K., & Shibata, T. (2007). *Robot Therapy in a Care House - Change of Relationship among the Residents and Seal Robot during a 2-month Long Study*. IEEE. doi: 10.1109/ROMAN.2007.4415062
- Walters, M. L., Dautenhahn, K., Boekhorst, R. T., Koay, K. L. K. K. L., Kaouri, C., Woods, S., ... Werry, I. (2005). *The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment* (Vol. 56). IEEE. doi: 10.1109/ROMAN.2005.1513803
- Waters, K. (1987). A muscle model for animation three-dimensional facial

- expression. In *Siggraph '87: Proceedings of the 14th annual conference on computer graphics and interactive techniques* (pp. 17–24). New York, NY, USA: ACM. doi: <http://doi.acm.org/10.1145/37401.37405>
- Wojdel, A., & Rothkrantz, L. J. M. (2005). Parametric Generation of Facial Expressions Based on FACS. *Computer Graphics Forum*, 4(24), 1–15.
- Wollaston, W. H. (1824). On the apparent direction of eyes in a portrait. *Philosophical Transactions of the Royal Society of London*, 114, pp. 247-256.
- Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M., & Kuzuoka, H. (2008). Precision Timing in Human-Robot Interaction : Coordination of Head Movement and Utterance. *Ratio*, 08(1), 131–139. doi: 10.1145/1357054.1357077
- Yoo, D. H., & Chung, M. J. (2005). A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding*, 98(1), 25–51.
- Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., & Miyamoto, T. (2006). Responsive robot gaze to interaction partner. In *In proceedings of robotics: Science and systems*.
- Yu, C., Aoki, P. M., & Woodruff, A. (2004). Detecting user engagement in everyday conversations. *CoRR*, cs.SD/0410027.
- Zecca, M., Endo, N., Momoki, S., Itoh, K., & Takanishi, A. (2008). *Design of the humanoid robot KOBIAN - preliminary analysis of facial and whole body emotion expression capabilities*. IEEE. doi: 10.1109/ICHR.2008.4755969