LICENTIATE THESIS

Enabling Physical Action in Computer Mediated Communication:

An Embodied Interaction Approach

MUHAMMAD SIKANDAR LAL KHAN

Immersive Interaction Lab (i2Lab)
Department of Applied Physics and Electronics (TFE)
UMEÅ UNIVERSITY, SWEDEN

Umeå, 2015

Enabling Physical Action in Computer Mediated Communication:
An Embodied Interaction Approach

Muhammad Sikandar Lal Khan
Immersive Interaction Lab (i2Lab)
Department of Applied Physics and Electronics (TFE)
Umeå University, Sweden
SE-90748 UMEÅ
Author e-mail: `muhammad.sikandar.lal.khan@umu.se`

ISSN 1652-6295:20
ISBN 978-91-7601-321-2
E-version available at http://umu.diva-portal.org
Printed by Print & Media, Umeå University, Umeå, 2015

ABSTRACT

Despite the current advancement in computer mediated communication (CMC), especially video teleconferencing, CMC is still far from face-to-face (FtF) communication. In FtF communication, humans use multimodal information; this information includes verbal information, visual information, body language, facial expressions and other non-verbal gestures. Whereas, in CMC, humans rely on either mono-modal information such as text-only, voice-only, video-only or bi-modal information by using audiovisual modalities. The comparison between FtF and CMC shows that FtF is better than CMC for conveying emotions and intentions during conversation. So, there is a need to bridge the gap between FtF and CMC and bring CMC as close as possible to FtF communication.

This thesis addresses the factors which make video teleconferencing different from face to face interaction. This thesis also reviews the previously developed hardware and software solutions for improving the video teleconferencing experience. In this work, we have revisited the problem of improving the video teleconferencing experience by considering the limitations of previously developed solutions. We have proposed the usage of embodied interaction concept in video teleconferencing. Embodied interaction concept suggests that the future of interaction can be effectively managed through interfaces which are more visible and available for wider range of engagements. Based on this concept, we have introduced the user's 'embodiment' in video teleconferencing system by including the gesture modality (head gesture in our case) in traditional audio-video communication.

Our embodied interaction based video teleconferencing system is a head/neck robot with a mounted tablet-PC. This thesis presents the design of a biologically inspired electromechanical platform (neck/head robot), which has a similar static and dynamic properties as of a real human neck. In addition, computer vision based motion control algorithm is developed for this electromechanical platform. This algorithm takes human-in-the-loop of feedback motion control system by using a low-cost and low-resolution camera.

Our novel video teleconferencing system is put into real experimental setup to solve the problems in standard video teleconferencing system. The first experiment measures the 'feeling of presence' of remote user among local collaborators and the 'quality of experience' of remote user for comparative scenarios; 1) when remote user is communicating using standard video teleconferencing system (e.g. Skype) and 2) when remote user is communicating through our embodied interaction based video teleconferencing system. The second experiment considers the issue of gaze perception during video teleconferencing. The traditional video teleconferencing system introduces Mona-Lisa gaze effect during conversation. This Mona-Lisa gaze effect can be mitigated by using our interactive video conferencing setup. The user studies of all the experiments validate the effectiveness of our proposed approach.

Keywords: Biologically inspired system, Parallel robot, Neck robot, Head pose estimation, Embodied Interaction, Telepresence system, Quality of interaction, Embodied Telepresence system, Mona-Lisa gaze effect, eye-contact.

Preface

Parts of the contributions presented in this thesis have previously been accepted to peer-reviewed conferences and submitted to journals. A list of these publications is given below.

Main Papers

- **M.S.L. Khan**, Haibo Li, Shafiq Ur Rehman, "Bio-Inspired Computer Aided Model Based Design of TEBoT: a Telepresence Mechatronic Robot," Submitted to Journal of Intelligent and Robotic Systems.
- **M.S.L. Khan**, Lu Zhihan, Shafiq Ur Rehman, Haibo Li., "Head Orientation Modeling: Geometric Head Pose Estimation using Monocular Camera," in IEEE/IIAE International Conference on Intelligent Systems and Image Processing 2013 (ICISIP 2013), Kitakyushu, Japan, 2013, pp. 149-153.
- **M.S.L. Khan**, Haibo Li, Shafiq Ur Rehman, "Embodied tele-presence system (ETS): designing tele-presence for video teleconferencing," in HCI International, Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments, (HCII 2014), Crete, Greece, 2014, pp. 574-585.
- **M.S.L. Khan**, Shafiq Ur Rehman, Pedro La Hera, Feng Liu, Haibo Li, , "A pilot user's perspective in mobile robotic telepresence system," in IEEE, Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA 2014), Siem Reap, Cambodia, 2014, pp. 1-4.
- **M.S.L. Khan**, Haibo Li, Shafiq Ur Rehman, "Gaze Perception and Awareness in Smart Devices," Submitted to International Journal of Human-Computer Studies.

Other Papers

- **M.S.L. Khan**, Shafiq Ur Rehman, Lu Zhihan, Haibo Li., “Tele-embodied agent (TEA) for video teleconferencing,” in Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM ’13), ACM, Luleå, Sweden , 2013.
- Zhihan Lv, Alaa Halawani, **M.S.L. Khan**, Shafiq Ur Rehman, Haibo Li., “Finger in Air: Touch-less Interaction on Smartphone,” in 1st ACM International Conference on Multimedia (ACM MM ’13), Barcelona, Spain, 2013.
- Zhihan Lv, **M.S.L. Khan**, Shafiq Ur Rehman, “Hand and Foot Gesture Interaction for Handheld Devices,” in 21st ACM International Conference on Mobile and Ubiquitous Multimedia (ACM MUM ’13), ACM, Luleå, Sweden , 2013, pp. 621-624.
- Zhihan Lv, Shengzhong Feng,**M.S.L. Khan**, Shafiq Ur Rehman and Haibo Li, “Foot motion sensing: Augmented game interface based on foot interaction for smartphone,” in CHI EA ’14: CHI’ 14 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA: ACM, 2014, pp. 293-296.
- Shafiq Ur Rehman, **M.S.L. Khan**, Liu Li and Haibo Li, “Vibrotactile TV for immersive experience,” in Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA 2014), Siem Reap, Cambodia, 2014.

Acknowledgments

No thesis is ever the work of one person alone, whatever the cover says. I owe a great deal to the various people who helped me, guided me and supported me during my doctoral studies.

Firstly, I would like to express my sincere gratitude to my advisor Dr. Shafiq Ur Rehman and co-advisor Prof. Haibo Li for giving me an opportunity to do research studies under their supervision. Their guidance, motivation and immense knowledge helped me during my research and writing of this thesis. I am truly indebted and thankful for the inspirations and encouragements during these years. Without their help and support I had not been able to pursue my research.

Beside my advisors, I would like to thank my colleagues, Farid, Bo, Alaa, Keni, Per, Karin, Lu, Shahrouz, and Ulrik. Special gratitude goes to Daniel and Pedro for their valuable guidance. I am also grateful to Linda, Lief, Marie and Mona Lisa for their help and technical support during my research. Special thanks to all the Test-subjects, who participated in our experiments. Thanks to all the staff of the department of Applied Physics and Electronics (TFE) for creating an enjoyable and pleasant working environment.

Many thanks to my family: my parents, my brother and sisters for supporting me and for being with me during my hard times. Last but not the least, I would like to thank my friends who made my stay in Umeå wonderful; Bilal, Salik, Mahmood, Evelina, Iyung, Christian, Ali, Zaid and jonathan.

Thank you everyone.

Muhammad Sikandar Lal Khan
Umeå, Sweden, 2015

Contents

Abstract **i**

Preface **iii**

Acknowledgments **v**

1 Introduction **1**
- 1.1 Aim of the thesis . . . 2
- 1.2 Research problem . . . 2
- 1.3 Thesis outline . . . 3

2 Trends in Development of Video Teleconferencing System: An Overview **5**
- 2.1 Software Based Solutions . . . 6
- 2.2 Hardware Based Solutions . . . 8
  - 2.2.1 Non-Anthropomorphic Designs . . . 8
  - 2.2.2 Anthropomorphic Designs . . . 10
- 2.3 Perusal . . . 12

3 An Embodied Interaction Approach **15**
- 3.1 Principles . . . 16
- 3.2 Visibility . . . 17

4 Enabling Physical Action in Computer Mediated Communication: Design **19**
- 4.1 Design . . . 19
  - 4.1.1 Smart Devices . . . 19
  - 4.1.2 Communication Modalities . . . 20
  - 4.1.3 TEBoT . . . 20

4.1.4 Control 20
4.2 TEBoT (Paper I) 22
4.2.1 Scientific Problem 22
4.2.2 Contribution 23
4.3 Head Orientation Modeling (Paper II) 24
4.3.1 Scientific Problem 24
4.3.2 Contribution 24

5 Enabling Physical Action in Computer Mediated Communication: Experiments 27
5.1 Embodied Interaction based Video Teleconferencing System 27
5.1.1 Experimental Setup 28
5.1.2 ETS for Local Collaborators 28
5.1.3 ETS for Remote Collaborator 29
5.2 Embodied Telepresence System (Paper III) 31
5.2.1 Scientific Problem 31
5.2.2 Contribution 31
5.3 Pilot User's Perspective (Paper IV) 32
5.3.1 Scientific Problem 32
5.3.2 Contribution 32
5.4 Gaze Perception and Awareness in Smart Devices 33
5.5 Gaze Perception and Awareness in Smart Devices (Paper V) 35
5.5.1 Scientific Problem 35
5.5.2 Contribution 35

6 Concluding Remarks and Future Work 37

Research Papers 40

I Bio-Inspired Computer Aided Model Based Design of TEBoT, a Telepresence Mechatronic Robot 43
I.1 Introduction 44
I.2 Motion Capturing 46
I.2.1 Sensor Based Approach 46
I.2.2 Vision Based Approach 47
I.3 Head Motion Analysis 47

I.3.1 Frequency Analysis . . . 49
I.3.2 Velocity Analysis and Threshold Calculation . . . 49
I.4 Biologically Inspired Design . . . 51
I.4.1 Biological System: Human Neck . . . 51
I.4.2 Biological Inspired Design . . . 51
I.4.3 Motor Selection . . . 53
I.4.4 Controller Selection . . . 54
I.5 Model Based Design . . . 54
I.5.1 From CAD to TEBoT simulation . . . 54
I.5.2 Robot motion control . . . 55
I.6 Head Pose Estimation Algorithm . . . 57
I.6.1 Face Detection . . . 57
I.6.2 Facial Features Detection . . . 58
I.6.3 Estimating the neck's reference frame . . . 58
I.6.4 Estimating the pose angles . . . 60
I.7 Kalman Filter . . . 61
I.8 System Implementation . . . 62
I.9 Experimental Results . . . 63
I.10 Conclusion . . . 67

II Head Orientation Modeling: Geometric Head Pose Estimation using Monocular Camera 69
II.1 Introduction . . . 70
II.2 Head Orientation Modeling . . . 71
II.2.1 ROI Detection . . . 72
II.2.2 Estimation of Pivot Point . . . 73
II.2.3 Pose Recovery . . . 74
II.3 Experimental Results . . . 75
II.4 Concluding Remarks . . . 76

III Embodied Tele-Presence System (ETS): Designing Tele-Presence for Video Teleconferencing 79
III.1 Introduction . . . 80
III.2 Related Work . . . 81
III.3 EMBODIED TELEPRESENCE SYSTEM (ETS) - A PROOF OS CONCEPT . . . 82

III.3.1 Mechanical Design 83
III.3.2 ETS Software Module 85
III.3.3 ETS System Level Evaluation 85
III.4 PRESENCE: LOCAL COLLABORATORS PERSPECTIVE 86
III.4.1 Feeling of Presence: Measures 87
III.5 USER STUDY 88
III.5.1 Participants and Procedure 88
III.6 RESULTS AND DISCUSSION 89
III.7 CONCLUDING REMARKS 91

IV A Pilot User's Perspective in Mobile Robotic Telepresence System 93

IV.1 Introduction 94
IV.2 DESIGN OF EMBODIED TELEPRESENCE SYSTEM (ETS) FOR PILOT USER 95
IV.2.1 Head Pose Estimation of Pilot User 96
IV.2.2 Design of ETS 96
IV.2.3 Velocity Profile Control for Pilot User 98
IV.3 SYSTEM DEPLOYMENT 98
IV.4 EXPERIMENT 100
IV.4.1 Participants and Procedure 100
IV.5 Experimental Results 101
IV.6 DISCUSSION 102
IV.7 CONCLUSION 103

V Gaze Perception and Awareness in Smart Devices 105

V.1 Introduction 106
V.2 Background and Related Work 111
V.3 Embodied Telepresence System (ETS) 114
V.3.1 Control Mechanism of ETS 115
V.4 Experimental Studies 116
V.4.1 Participants 117
V.4.2 Experimental Setup 117
V.4.3 Procedure 119
V.5 Results 121
V.5.1 Raw Data 121
V.5.2 Regression Analysis 121

V.5.3 Variance Analysis 122
V.5.4 Viewing Position Analysis 122
V.5.5 Gaze Modeling 125
V.5.6 Gaze Awareness 125
V.6 Discussion 127
V.7 Conclusion and Future Work 129

Bibliography **131**

1 Introduction

Distance communication is becoming an important part of human lives because of the current advancement in computer mediated communication (CMC). The CMC is defined as, the communication between distant people with the help of computer technology [87]. The history of CMC dates back to 1960s, when a first prototype email was exchanged [66]. Since then CMC has gone through many transitions. Initially, CMCs were based on text; after some years it has changed to voice and a decade ago audio-video conferencing has gained popularity. These CMCs rely on either mono-modal information such as text-only, voice-only, video-only or bi-modal information based on audio-video conferencing [28]. Compared to CMC, face-to-face (FtF) communication use multimodal information including verbal information, visual information, body language, facial expressions and non-verbal gestures [29]. The research studies show that FtF communication is better than CMC and different researchers and designers consider FtF communication as a benchmark for assessing computer mediated communication [26, 125].

The latest form of CMC is internet-based video teleconferencing in which small groups of geographically distant located people can hold discussions in real-time, during which they hear and see each other and share various other types of data. The studies show that the video teleconferencing is more interactive than voice and text based communication [147]. The video teleconferencing enhances the verbal description based on facial expressions. But, compared to FtF interaction, the video teleconferencing is lacking with respect to the quality of interaction experience. Over the years, researchers have compared video teleconferencing technologies with face-to-face interaction and mainly pointed out the following problems in video teleconferencing;

- Distant located people cannot take turns or do side conversation during conversation [54].
- Distant located people cannot express emotions due to lack of nonverbal modalities

[68].

- Distant located people cannot estimate the gaze of his/her fellow collaborator [56, 106].
- Distant located people cannot make an eye contact [23].

Despite the above specific problems present in standard video teleconferencing, there have been many psychological problems such as level of satisfaction, trust, engagement, involvement, feeling of presence and quality of interaction, etc. [86].

To solve the above-mentioned issues, researchers have proposed different software and hardware solutions as detailed in chapter 2. To this end, the software solutions for improving video conferencing experience use computer graphics, computer vision, augmented reality and virtual reality techniques. Whereas, the hardware solutions use the robotics, embedded system, signal processing and computer vision techniques.

These hardware and software solutions have tried to solve certain problems in video teleconferencing to achieve collaboration as close as possible to face-to-face interaction. Despite these considerable hardware and software improvements, video teleconferencing is still far from face-to-face interaction. The major difference between face-to-face interaction and CMC is fewer numbers of communication modalities (verbal and non-verbal) involve in CMC. Hence, CMC can be improved by increasing these communication modalities. In this work, we have revisited the problem of improving the video teleconferencing experience by considering the limitations of previously developed hardware and software solutions.

1.1 Aim of the thesis

The aim of the thesis is to understand, study and analyze the trends and challenges in computer mediated communication, especially video teleconferencing. In addition, the aim is to propose an intuitive solution for improving video teleconferencing experience by incorporating physical gestures in audio-video conferencing. The long-term vision is to make video teleconferencing as close as possible to face-to-face interaction.

1.2 Research problem

In order to achieve the aim, this research is concerned with developing theories for improving video teleconferencing experience. These theories are based on the Embodied

Interaction Concept given by Paul Dourish, in his book ‘Where the action is’ [47]. To incorporate embodied interaction concept, our research is focused on developing technologies for adding nonverbal modalities in standard audio-video teleconferencing. To accomplish this, we address the following research problems:

- Design of a biologically inspired electromechanical neck/head robot, which can present head gestures along with audio-video conferencing.
- Design of a computer vision based algorithm for estimating the pose angles of the human head using a low-cost web-camera in real-time.
- Incorporating embodied interaction concepts in video teleconferencing system.
- Conducting a user study to collect an empirical data to provide an aftereffects of an embodied interaction based video teleconferencing system as compared to the standard video teleconferencing system.
- Mitigating Mona-Lisa gaze effect introduced by 2D-screen based standard video conferencing setup.

1.3 Thesis outline

The thesis is carefully organized to answer the following questions,

- What are the factors which make video teleconferencing different from face-to-face interaction?
- What researchers have done so far to improve video teleconferencing?
- How to further refine the video teleconferencing experience?
- Which nonverbal modality is important in distance communication?
- How to enhance the interaction experience of the remote and local collaborators during video teleconferencing?
- How to mitigate the Mona-Lisa gaze effect during video teleconferencing?

To answer these questions the thesis is organized as follows. Chapter 2 outlines the previously-developed video teleconferencing techniques, their advantages and drawbacks. Furthermore, chapter 2 presents the highlights of these systems in the form of perusal.

Chapter 3 talks about an embodied interaction concept used for designing our video teleconferencing system. Chapter 4 presents the design of our video teleconferencing system. Chapter 5 presents experimental studies to compare our embodied interaction based video teleconferencing system with standard videoconferencing setup and its effects on remote and local collaborators. Chapter 5 also shows a solution to resolve a Mona-Lisa gaze effect in standard video teleconferencing system.

The concluding remarks and future works are highlighted in chapter 6. Finally, the research papers that provide the basis for this thesis are presented.

2 Trends in Development of Video Teleconferencing System: An Overview

The technical development of video teleconferencing dates back to 1960 when AT & T tried to make the first video telephony [1]. Since then, researchers tried to develop an effective video teleconferencing system, but they were initially restricted by the technological limitations. In 1980s, due to the rapid developments in computing power and digital networks, an improved audio-video transmission became possible, but it was too expensive for general adoption. In early 2000s, video teleconferencing has been adopted by many sectors because of the cheap computers, low cost internet, increased bandwidth and better IP (Internet Protocol) technology. Video teleconferencing tries to establish a real-time connection between the participants located anywhere by synchronous audio and video feed through an internet connection. This connection supports the face, head and voice interaction by sharing 2D screen. The traditional teleconferencing tools available in market are bimodal audio-video teleconferencing software such as, Skype, face-time, tango, etc. These tools are very common among the distant-located families, in work environments, in teleconferencing and meeting scenarios, etc. But, these standard conferencing systems have some inherited problems (as mentioned in previous chapter), which have been a concern of many researchers. To improve the standard conferencing experience, the researchers and designers have proposed several software and hardware solutions.

2.1 Software Based Solutions

Several software methods have been proposed to improve the video teleconferencing experience. The conferencing-experience has already been enhanced by the software technique which adds video to the audio-conferencing. However, there is still a need for further progress. To further elevate the conferencing experience, researchers have proposed large screens for video teleconferencing [2,121]. Moreover, the spatially immersive displays (SID) are also becoming popular for video teleconferencing. These SIDs enable people to interact and communicate while being surrounded by the panorama screens. The most common example of SID is CAVE [41]. The extended version of CAVE has been used by exploiting multiple cameras and projectors in conference and meeting rooms. The blue-c [65] has also utilized the CAVE like projection with 3D projection of a remote person to give an immersion experience during video teleconferencing. Augmented Reality (AR) has also played a part in making effective video conferencing experience [18]. AR technology added virtual non-verbal gestures in distance communication which are missing in traditional audio-video conferencing [57]. AR methods have also employed three-dimensional depth information for creating a feeling of presence of all collaborating participants during teleconferencing [44, 123]. The wearable smart glasses are also being proposed for immersive Group-to-Group teleconferencing [20]. The expressive avatars based video teleconferencing are being used to hide the user's appearance [146].

The mutual gaze awareness or simply an eye-contact is an important factor in human communication [96]. The researchers have pointed out that the video-mediated communication systems are severely hindered by the mutual gaze awareness or an eye-contact [106,143]. The issue of an eye contact makes communication awkward and unnatural. Many software algorithms have been developed to facilitate mutual gaze in distance communication. Kuster et. al. proposed a graphical solution for gaze correction for home video conferencing [89]. This approach uses a Kinect sensor in which RGB and depth images are used to adjust the viewing direction of the face for both communicating sites. A two camera setup is used for the gaze correction of a remote person by attaching them to the left and right (or top and bottom) of the computer screen [39, 166]. They used computer vision algorithm to blend two images from two different cameras to create a cyclopean view. There are also other methods which are trying to solve an eye-contact problem in video teleconferencing by using a single camera [30, 62, 167]. These methods apply image processing techniques on the face of the remote person to re-orient it in 3D space by using face registration and image warping techniques. Some of the software techniques proposed for video teleconferencing are illustrated in figure 2.1.

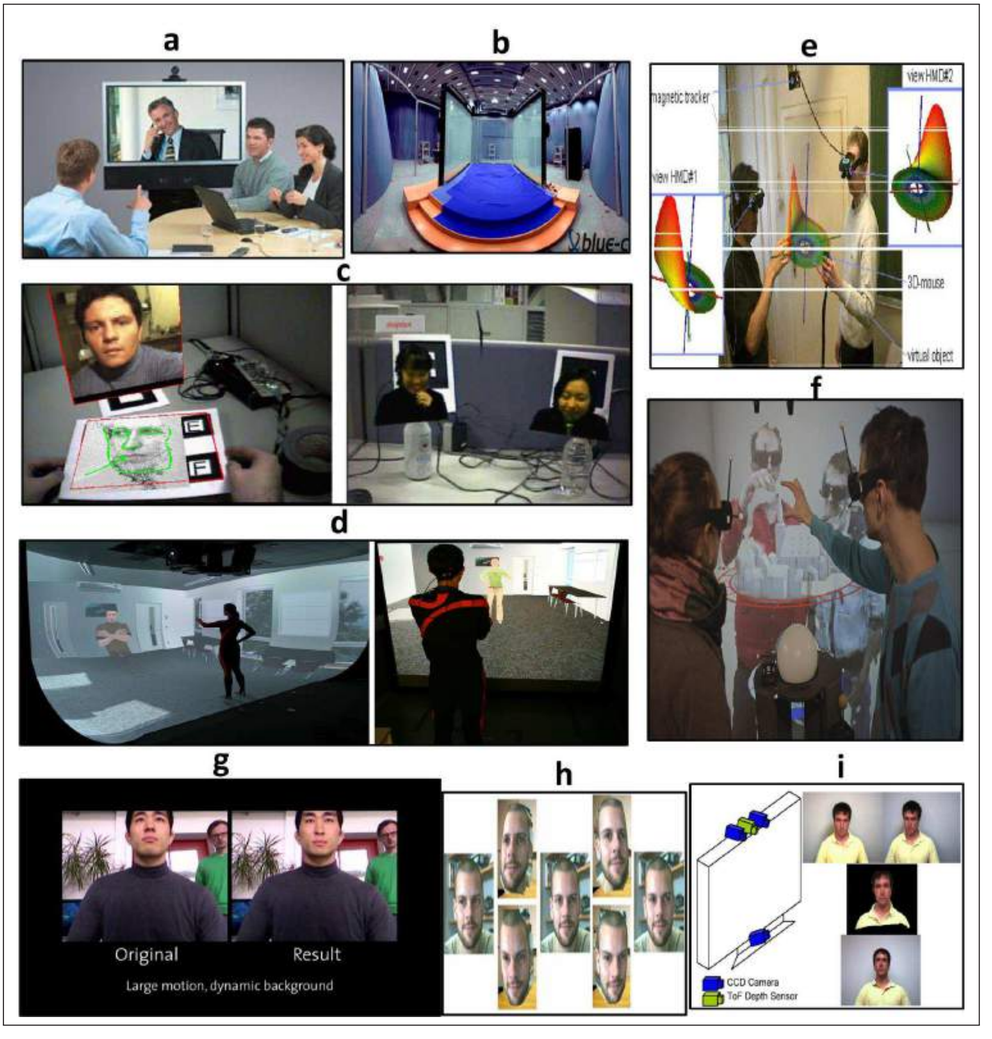


Fig. 2.1: *Software based methods proposed in recent years. (a) a big screen used for video teleconferencing. (b) blue-c with cave like projection with multiple cameras and projectors [65]. (c) Augmented reality based video teleconferencing [18]. (d) The expressive avatars for video teleconferencing [146]. (e) Augmented Reality (AR) technique to share same environment [57]. (f) Augmented Reality (AR) and virtual reality (VR) based wearable glasses for immersive group-to-group conversation [20]. (g) Graphical technique for gaze correction for home video conferencing [89]. (h) gaze correction by using single camera by exploiting image processing techniques [62]. (i) Multiple camera setup for gaze correction in video conferencing [166].*

2.2 Hardware Based Solutions

Recently, there are many hardware solutions proposed to enhance the experience of video teleconferencing and they can be classified into two classes:

1. non-anthropomorphic designs

2. anthropomorphic designs

2.2.1 Non-Anthropomorphic Designs

The non-anthropomorphic video teleconferencing systems (NAVTS) are the audio-video teleconferencing systems with some added capabilities. The most common example of NAVTS is mobile robotics telepresence (MRP) system [86]. The MRPs are the integration of traditional computer mediated communication devices with the mobile robotic platforms. The basic construction of almost all MRPs consist of a mobile robotic base, an LCD screen, a camera, a microphone and some non-verbal gestures like hand gestures, etc. The MRP system is placed in a distant location among the local collaborators and represents the remote collaborator. The remote collaborator is the one who is situated at the remote site and collaborating with the local collaborators through MRP system. The local collaborators are those who are situated at the same physical location as the MRP system. These local collaborators communicate indirectly with the remote collaborator by the direct interaction with the MRP system. The control of a mobile robot base varies from the manual to semi-automatic control, where, the mobile robot base is controlled by the remote user. Two ways have been adopted to control these mobile base; implicit control and explicit control. The remote user controls the robot base explicitly through web-based applications by using keyboard, mouse, joystick and/or other hand held devices [119]. On the other hand, the remote user controls the robot base implicitly by using a real human body movements, for example by the head control [137]. The implicit control possess less cognitive load as compared to explicit control. The audio-video communication in MRP is done by the typical computer mediated communication software, e.g. Skype. Furthermore, some MRP systems have the capability to show some nonverbal gestures. These nonverbal gestures are very important in face-to-face communication and psychologists estimate that 60 to 80 percent of communications between human beings are nonverbal [99]. To express these nonverbal gestures, some MRP systems have added some nonverbal gestures in their systems such as hand gesture, finger pointing

Fig. 2.2: *Non-anthropomorphic video conferencing systems: [(a) Giraff (b) VGO (c) TiLR (d) QB are mobile robotic telepresence system (MRP) system with LCD, webcam, monitor and added capability of mobility [86].] (e) Embodied Social Proxy [ESP] by Microsoft for video teleconferencing [137]. (f) 2DOF of neck robot for video conferencing. (g) PEBBLES for connecting sick child to school [52]. (h) MeBot by MIT with head and hand gesture [10]. (i) MulDiRoH with QDA screen for video teleconferencing [114]. (j) Gaze-2 for parallax free transmission during video teleconferencing [157].*

gesture, head nod, head shake, etc. [10]. Compare to traditional computer based audio-video conferencing tools, these MRPs have achieved greater fame in increasing the video conferencing experience by enhancing the feeling of presence of remote person and by improving the quality of interaction between the local collaborators and the remote person. The physical limitation of presence, i.e., one person cannot be present in two places at the same time has been overturned by these MRP systems. These MRP systems have been used for many applications, but five applications are particularly dominant in the literature: office environments [137], hospitals [158], healthcare [25], education [52] and elderly people assistance [21]. The details of these mobile robotic telepresence (MRP) systems can be found in [86].

There are other non-anthropomorphic video conferencing systems which are different from MRP system. Gaze-2 [157] is a video teleconferencing system which focuses on parallax free transmission of video to every participant in group conversation and the system tries to establish an eye-contact during group conversation. In this system, there are series of cameras looking directly at each user and there is a separate camera that is used for eye tracking. The tracker will tell which camera video should be broadcast to have parallax free transmission of video to other participant. Mebot [10] robot developed by MIT Media Lab is a semi-autonomous robot that gives collaborators a richer way to interact with hand gestures and a head movement capability. Mebot cannot move in distant location, but it can present hand and head gestures of a remote person to the local collaborators. Embodied Social Proxy (ESP) [137] is another telepresence system proposed by Microsoft to represent a remote person. ESP consists of two cameras, one is a fish eye camera for the full view of the meeting room and the other is a PTZ camera for focusing on a single person or other things such as, the white board. Heon-Hui et. al. presented an immersive teleconference system that is used to enhance the immersive interaction capability of remote user [83]. They have proposed a human-robot-avatar interaction model for the remote user. Shiro Ozawa et. al. have also developed a Multi-directional representation of humans (MulDiRoH) teleconferencing system [114]. They used quantized diffusion angle (QDA) screen, one of the newest multi-view display techniques. The advantage of a QDA screen is to show views of remote person from the direction in which the participants face is pointing. This helps other participant to see directly the face of the remote person which helps them to make an eye-contact. Some of the non-anthropomorphic video teleconferencing systems (NAVTS) are shown in figure 2.2.

2.2.2 Anthropomorphic Designs

The characteristics and appearance of anthropomorphic video teleconferencing systems (AVTS) are similar to real-human characteristics and appearance. AVTS has a human-like physical shape, facial expressions and other social cues, as well as natural human-like interaction and communication behaviors (e.g. speech, gaze, gestures). It is reported that humans respond more positively to an artifact that displays human-like behavioral characteristics (emotions, facial expression) [51]. The purpose of AVTS is to devise a mechanism through which distant-human interaction can be facilitated by 'human social' characteristics. The design of AVTS normally focuses on three main areas: AVTS shape, its behavior and its interaction with the humans [46].

These anthropomorphic robots are also termed as humanoid robots and/or androids. Over the years, several humanoid robots have been developed. The physical shape, body structure, facial features, limbs, behavior and way of interaction of these humanoid robots are similar to real humans. HRP-4C [75], Actroid DER3 [8], EveR-3 [35], Geminoid [130] and Aiko [11], etc. are the examples of humanoid robots, which look very similar to real humans and it is hard to differentiate whether it is a robot or a human. On the other hand, there exists set of anthropomorphic robots called androids, which have similar human characteristics but different appearances; for example, ASIMO [129], Telenoid [112], REEM-A [53], HOAP [49], HUBO [117], and Mahru [97] etc. These anthropomorphic robots are helpful for many applications, but, here we are concern about video teleconferencing application. For example, Geminoid HI-1 [130] is a human copy which is developed and used specifically for video teleconferencing. Geminoid HI-1 is a 50 DOF humanoid robot, which can mimic conscious and unconscious human behaviors. The experimental studies have shown that the 'feeling of presence' of a tele-operator through Geminoid is stronger than a man who appears on a video monitor. There are many other anthropomorphic systems, for example telenoids, which are not similar to real-human, but still contains all human features [112]. The concept behind the Telenoid design is a universal representation of all humans (male or female, old or young). For example, a Telenoid in [112] is a 9-DOF system with 3-DOF for neck movement, 1-DOF for mouth movement, 2-DOF for eyes movement and 3-DOF hand/arm movement. The experimental studies show that the robots with minimal human characteristics are also acceptable for distance communication.

Recently, researchers have projected 3D face of a remote person on 3D interfaces such as 3D screens and 3D surfaces. Hiroaki et. al. have presented a telepresence system that projects a 3D face of a remote person on a helium blimp, which is placed among the local collaborators [150]. Furhat [13] is a 3D back projected human head, which is used to remove the Mona-Lisa gaze effect during video teleconferencing. Mona-Lisa gaze effect is common in traditional video teleconferencing system due to the use of 2D screens. The Mona Lisa gaze effect or the Mona Lisa stare is commonly described as an effect that makes it appear as if the Mona Lisa's gaze rests steadily on the viewer as the viewer moves through the room. Similarly, Andrew et al. have presented a 3D display screen, which is able to transmit the 3D face of a remote participant to local collaborators, hence, maintaining an accurate gaze, attention and eye contact among remote and local collaborators [74]. The studies show that these 3D projections are better than 2D projections in providing human like feelings to the local collaborators. Some of the anthropomorphic video teleconferencing systems (NAVTS) are shown in figure 2.3.

2.3 Perusal

The purpose of software and hardware techniques is to improve the experience of video teleconferencing. Every system has some advantages with certain limitations. Due to the lack of comparative studies, one cannot decide which technique is better or the state-of-the-art for video teleconferencing. The decision of which technique is to be used depends on the application and situation at hand. The advantage of a software solution is that it does not require a complex hardware setup for video teleconferencing. Most of the software solutions use virtual reality, augmented reality, computer graphics and computer vision techniques to improve the experience of video teleconferencing. New trends in software solutions are to use 3D technologies. 3D technology is becoming popular in the market, but still current video conferencing setups are restricted by the 2D screens. The use of big screens for video teleconferencing is not a new idea. These screens are expensive, but still they do not require complex setups for video teleconferencing. The combination of 3D technologies with the panoramic 3D screens can be a good setup for future video teleconferencing. Despite the advantages, there are some limitations in software techniques. The most important limitation is that the remote participant is always a virtual agent and is present inside the 2D computer screen. Because of this, the local collaborators do not feel the real presence of the remote person. Furthermore, the software based video conferencing setup only includes the voice and facial expression of a remote person. All the other communication modalities are missing. Hence, the distant located person cannot make an eye contact, cannot do side conversation and cannot perform other common human interactions. Many researchers have compared software based video teleconferencing with a robotic agent based video conferencing. They found that the robotic agents are better for video teleconferencing as some physical entity is representing the remote person [82, 135, 152].

The non-anthropomorphic video teleconferencing systems (NAVTS) have been considered very effective for creating a feeling of presence. The difference between the NAVTS and traditional video conferencing system is the added capability of mobility

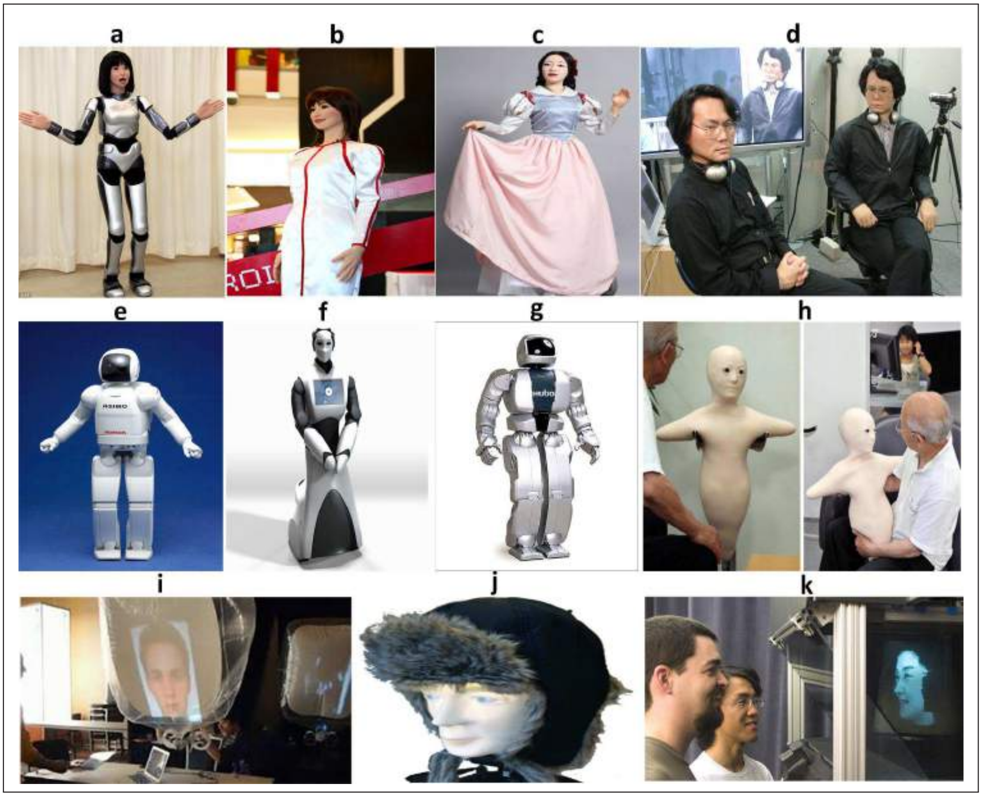


Fig. 2.3: *Anthropomorphic video conferencing systems: (a) HRP-4C [75] (b) Actroid DER3 [8] (c) EveR-3 [35] (d) Geminoid [130] is a duplicate copy of an author whose appearance and characteristic resembles human. (e) ASIMO [129] (f) REEM-A [53] (g) HUBO [117] (h) Telenoid [112] is an android whose appearance is not similar to human being. (i) Floating avatar [150] is a 3D video conferencing system. (j) Furhat [13] 3D face used for distance communication to mitigate mona-lisa effect. (k) 3D display screen for eye contact [74].*

and some gestures at a distant location. These capabilities have been considered very advantageous. These mobile capabilities are beneficial for both the remote and local collaborators. Though, these NAVTS have moved video teleconferencing from the digital world (computers) to the physical world, but NAVTS are still abortive in presenting a real human communication style. During face-to-face communication, humans use a lot of nonverbal communication modalities like head gestures, etc. But, these NAVTS mimic very few nonverbal modalities in distance communication and they have not considered an important question i.e., 'which modalities are important in distance human communication?'

Considering the drawbacks in NAVTS, researchers have proposed AVTS. The human like appearance of AVTS creates a feeling of human-presence among the collaborators. The studies show that these AVTS can be a better representation of a remote person in a distant location. But AVTS are very complex and very expensive for the normal use. Furthermore, the appearance of AVTS represents one person so a new AVTS is required for every other person. The control of both NAVTS and AVTS have limitations. Almost all theses systems are controlled explicitly by the mouse, keyboard and other hand-held devices. Hence, it is not practical to actuate the distant robot and communicate with distant people in parallel.

Based on above discussion, there is a need to revisit the issue of improving the video teleconferencing experience. A system is needed which has the capabilities and appearance of AVTS and simplicity of NAVTS. The system should present nonverbal cues during distance communication. Furthermore, a better control mechanism is needed to actuate the tele-robot at a distant location.

3 An Embodied Interaction Approach

This chapter presents the theoretical framework and foundation, which are used for improving video teleconferencing experience. Theoretically, the solution is based on Embodied interaction concept. Embodied interaction is new form of interaction with computer systems that occupy our world, a world of physical and social reality, and that exploit this fact in how they interact with us. Embodied Interaction is concerned with i) *Interaction* and ii) *Computation*. Interaction deals with the ways in which the new interactive systems are manifested and incorporated in our world. Similarly, computation is an active representational power, embodied in these interactive systems.

The word embodied interaction is proposed by Paul Dourish in 2001 in his book 'Where the Action is' [47]. Embodied interaction is,

"when people interact mentally and physically with technology".

Furthermore, embodied interaction is,

"when humans create and communicate meaning through interaction with the system".

The massive increase in computational power and the expanding context in which we put that power to use - both suggest that we need new ways of interaction with computers. These intuitive ways should be better tuned to our needs and abilities and should be better than tangible and social interaction approaches of last century. One of the well known approaches is 'Embodied Interaction' approach. Paul Dourish pointed out that computing is moving into the social and physical spaces, where new embodied interfaces will result in more effective, expressive and engaging interactions. To contribute to this vision, we present a theoretical argument of embodied interaction based video teleconferencing setup based on the following principles.

3.1 Principles

To incorporate embodied interaction in video teleconferencing setup, we consider the following six principles of embodied interaction design [47]:

1. Computation is a medium.
2. Meaning arises on multiple levels.
3. User, not designers, manage and communicate meaning.
4. User, not designer, manage coupling.
5. Embodied technologies participate in the world they represent.
6. Embodied interaction turns action into meaning.

The standard video teleconferencing setup use *computer* as a medium for audio-video communication. Based on embodied interaction concept, we propose to use *computation* as a medium, where communication is done not simply through audio-video conferencing, but, through the way that *computation* enlivens these audio-video conferencing by communicative non-verbal 'actions'. These communicative non-verbal 'actions' are expressed, conveyed, shared, explored and developed by the *computation*. Furthermore, we propose that embodied interaction based video conferencing framework conveys *multiple meanings* for users according to the different ways that framework is put into the practice. For example, the setup can be used i) to present the head gestures of remote person along with audio-video conferencing, ii) to explore the distant environment and iii) to make an eye-contact with distant-located person, etc. All depends on the way the framework is put into practice by the collaborating parties.

In standard video conferencing system, designers has the responsibility to describe the functionality and set-of-rules to use the system. Considering 3rd and 4th principle, we argue that the functionality of the system is still in the hands of designers but the *set-of-rules* to use the video teleconferencing system should be given in the hands of the users, who can change these rules to create and communicate different meanings.

The standard video teleconferencing systems are confined to 2D computer screen, which separates us from the physical world. Considering principle 5, the goal of embodied interaction in video conferencing should be to take us away from these 2D computer displays and search for better embodied technologies where 'digital' and 'physical' world are seamlessly merged in each other. We propose to apply this principle in video teleconferencing so that a system can actively participates in the physical world during video teleconferencing, . Furthermore, we propose that the physical *actions* of remote users presented by embodied interaction based video conferencing setup should convey meaning to the local collaborators. Thus, converting *nonverbal actions* of remote user into *meaning* for local collaborators.

3.2 Visibility

Embodied Interaction concept further suggests that the future of interaction should be based on the interfaces which are more visible and available for a wider range of engagements and interactions. This concept is in contrast with the idea of invisible computers and invisible user interfaces given by Norman [110] and Fishkin et al. [55]. Embodied Interaction approach advocates that it is not possible to actively engage and interact with a system when it is not there. The role of HCI designers should be to beautify the experience of interaction with computers, like artists-designers who want their designs to be engaging and meaningful. What is notable about their designs is their physicality; their embodiment serves not to render them invisible, but rather to make them more visible and expressive to encourage a deeper engagement. The embodiment of something gives a concrete form to an abstract idea. When we talk about embodiment, we are talking about giving a form to ideas that are usually not physical. Conversation, for example, is not a physical thing and at the same time it is not just speech patterns carried through the air. Conversation is embodied in more ways than just speech patterns. It is embodied in the way that it happens in the world, through the engaged participation of two equally embodied people. So, embodied interaction is an embodied phenomenon that happens in the world and world lends form, substance and meaning to the interaction.

4 Enabling Physical Action in Computer Mediated Communication: Design

This chapter presents the design and control of our electromechanical platform. The contribution of the research papers and references to the papers are also given for detail descriptions.

4.1 Design

4.1.1 Smart Devices

Smart devices such as tablet PC (iPad, galaxy-tab, etc.) and smartphones are driving the technology of daily life. Beyond the ordinary use in daily life, they are being used by people for various advance purposes in scientific area, entertainment, education, medical purpose, communication, gaming, etc. Today, there are more smart-devices than human beings on the planet and these devices are multiplying five times faster than the human population [145]. It is also expected that the smart devices could increase to 10 billion by the end of 2016 [145]. This is the fastest growing trend by any technology from 0 to 10 billion devices in 3 decades. No other technology has impacted the human life like this before. Despite the thousands of useful applications, these smart devices are the biggest source of communication ranging from text, voice and video communication. By the constant increase in mobile network speed and computational power of multimedia capable cellular phones, iPads and PDAs, and portable netbook computers, video communication is becoming more and more popular.

4.1.2 Communication Modalities

The smart devices based video communication involves two communication modalities i.e., verbal and visual modalities. Whereas, face-to-face communication involves multiple communication modalities ranging from verbal modality, visual modality, head movement, hand movement and body language. The previous research analysis show that adding more modalities in distance communication enhance video teleconferencing experience. However, the designers of video teleconferencing systems did not specifically mention which nonverbal modality is important in distance communication. The research in psychology shows that among all non-verbal modalities during human conversation, facial expressions and head movements are the most important for the flow of information [3, 24].

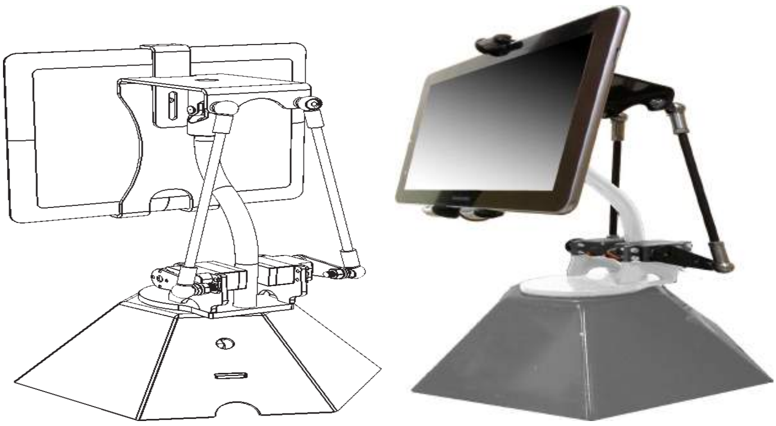
4.1.3 TEBoT

The previous research efforts for improving video teleconferencing experience are focused on developing complex hardware solutions (as detailed in chapter 2). However, there is a need to adopt/incorporate already developed computer devices like tablet-PC and smart phones for video teleconferencing. There are two-fold advantages of using currently developed smart device; one they are commonly used devices for video teleconferencing these days and second they have a high computational power. Furthermore, we plan to add head gesture in smart devices based video communication.

To present the head gestures along with audio-video communication; we have proposed a hybrid model that is a mixture of non-anthropomorphic and anthropomorphic design. A telepresence mechatronic platform - named - TEBoT has been designed which has the characteristics of AVTS and the simplicity of NAVTS. The CAD model of the platform is shown in figure 4.1(a) and the original platform is shown in figure 4.1(b). The TEBoT is bio-inspired neck robot with a tablet PC mounted on it. The tablet-PC is used for audio-video communication and the electromechanical platform is used for presenting the head gestures of a remote participant. These head gestures include head nod, head shake, head yaw, head roll and head pitch movements.

4.1.4 Control

Most of the previously developed telepresence robots are controlled by keyboard, mouse or joystick. These control methods are not a practical solution as communicating and controlling is not possible at the same time. To control the TEBoT, we have swerved



(a) CAD model (b) Original Platform

Fig. 4.1: *Computer Aided Model Based Design of TEBoT*

from a traditional method of controlling by keyboard, mouse or joystick and proposed two implicit control methods where the TEBoT is implicitly controlled by the remote user head,

- Wearable control method
- Non-wearable control method

In wearable control, a head mounted Inertial Measurement Unit (IMU) is used to control the TEBoT (see figure 4.2a). Whereas in non-wearable control method, a computer vision based geometric head pose technique is proposed where the head gestures of a real human are estimated in real-time by the webcam of the computer (see figure 4.2b). These head gestures are then map to TEBoT's mechatronic platform. In the context of head motion capture, computer vision through webcam offers several advantages over wearable devices, as wearable setup is not preferred by the users.

The details of the design and control are presented in paper I and paper II of this thesis.

Paper I

M.S.L. Khan, Haibo Li, Shafiq Ur Rehman, "Bio-Inspired Computer Aided Model Based Design of TEBoT: a Telepresence Mechatronic Robot," Submitted to Journal of Intelligent and robotic systems.

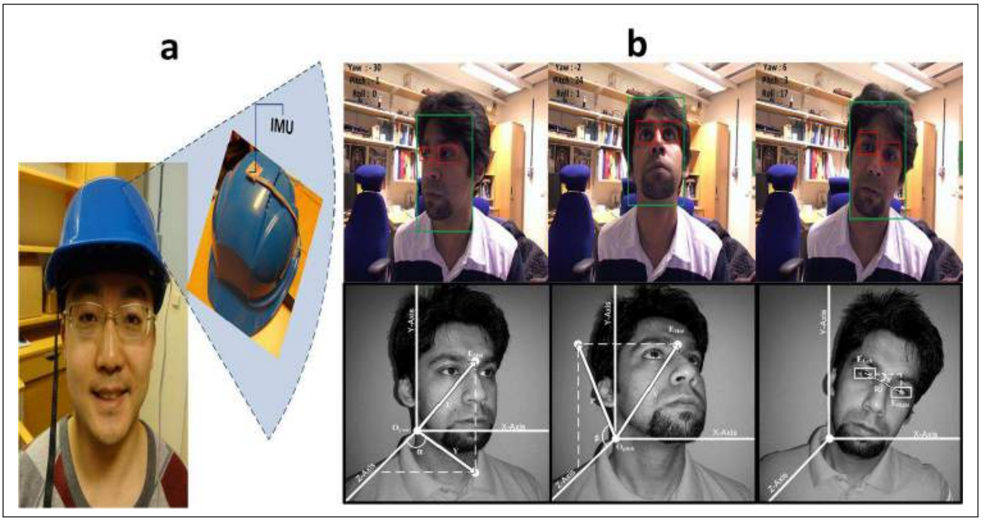


Fig. 4.2: *(a) IMU based head pose estimation. (b) Geometric head pose estimation algorithm by using face and eyes' regions to calculate yaw, pitch and roll angles (details are given in paper I and paper II).*

Paper II

M.S.L. Khan, Lu Zhihan, Shafiq Ur Rehman, Haibo Li., "Head Orientation Modeling: Geometric Head Pose Estimation using Monocular Camera," in 1st IEEE/IIAE International Conference on Intelligent Systems and Image Processing 2013 (ICISIP 2013), Kitakyushu, Japan, 2013, pp. 149-153.

To facilitate the reading of the above papers, here we provide a summary of the problem formulation and contribution of paper I and paper II.

4.2 Bio-Inspired Computer Aided Model Based Design of TEBoT, a Telepresence Mechatronic Robot

4.2.1 Scientific Problem

In this paper we have aimed a complete study of a biologically Inspired Computer Aided Model Based Design, where the following tasks are considered;

- Design of a biologically inspired electro-mechanical neck/head robot which has a similar static and dynamic properties as of a real human neck.

- Devise a computer vision based motion control strategy for electro-mechanical neck/head robot by bringing human-in-the-loop by using a low cost and low resolution camera.

4.2.2 Contribution

To design a neck/head robot, which has a similar characteristics as of real human neck/head, we have started with the *head motion analysis* of seven participants. The experiment is performed to measure the *frequencies*, *velocities* and the *angular ranges* of the real head movements. The *frequency analysis* results are used for the *controller design* of our electro-mechanical neck/head robot. The *velocity analysis* results are used for a *motor selection* of the robot and finally, the *angular ranges* are used in the *mechanical design* of the head/neck robot.

For actual mechanical platform of the neck robot, we have studied the complex anatomical structure of the real human neck. The *human neck study* and *head motion analysis* results are combined to construct a stable, reliable and nice looking biologically-inspired neck platform - named - *TEBoT*. The CAD model of TEBoT is shown in figure 4.1(a) and the actual platform is shown in figure 4.1(b).

For simulation purpose, the CAD model of a TEBoT is imported to simulink/MATLAB by using simmechanics library. To simulate the static and dynamic properties of a TEBoT we have used a model based design approach (MBD). MBD approach is a software driven approach which allows you to investigate your design completely, helps you to find mistake and assists you to explore new concepts without the overhead of the extensive hardware. For controller design, we have solved the inverse kinematic problem of a TEBoT, where the input is the desired yaw, pitch and roll angles and the output is the required torques for three servo motors. In matlab this is achieved with the command *looptune*, that can directly tune the two PID loops to achieve the desired response time with minimal loop interaction and adequate MIMO stability margins.

In contrast with traditional motion planning techniques for controlling biological systems, we have considered a human-in-the-loop of biological motion control, where real-human movements are mapped to biological systems. In literature, the real-human movements are captured by one of the following approaches; (i) sensor based approach (SBA) and (ii) vision based approach (VBA). In this article, we have considered the limitations of previously developed SBA and VBA and proposed a novel vision based system which uses low cost, low bit rate, low resolution and non-wearable webcam of the computer to capture the real human movements. For TEBoT motion planning, we have proposed a *ge-*

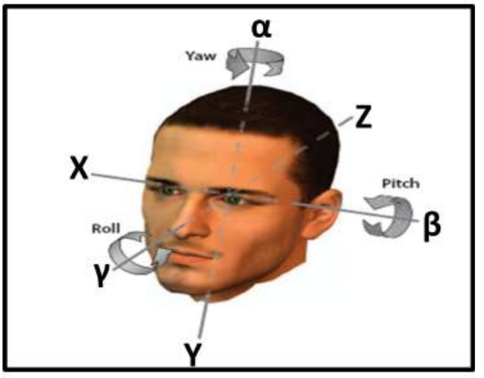


Fig. 4.3: *3 DOF framework for Yaw, Pitch and Roll movement of human head.*

ometric head pose estimation technique, which calculates the pose angles of human head using monocular camera and map these pose angles to TEBoT's mechanical platform.

We provide results of simulations and a real-time experiment to validate our approach.

4.3 Head Orientation Modeling: Geometric Head Pose Estimation using Monocular Camera

4.3.1 Scientific Problem

The scientific problem we have considered for this work is *how to estimate the pose angles of the human head using a low-cost web-camera in real-time*. The problem can be explained graphically by figure 4.3, where we have a human head which can rotate arbitrary around X, Y and Z axis. The goal is to estimate these rotational movements (yaw (α), pitch (β) and roll (γ) angles) using just webcam i.e., 2D video information.

4.3.2 Contribution

The two discrete contributions of this work are;

- Estimation of a Pivot point for the head movement.
- Geometric head pose estimation technique for estimating yaw, pitch and roll angles of human head.

Biologically, there is no such point as the pivot point for the human head because the head movements, such as head nod and head shake, etc., involve all the cervical vertebras and muscles of the neck. For the convenience of the head modeling, a pivot point (center of the origin) of the head movement is assumed to be around C2 cervical vertebrae. This point is calculated by using the eye-separation distance and width and height of the human face.

We have proposed a novel geometric head pose estimation algorithm, which is the main contribution of this research paper. The algorithm has following steps to estimate yaw (α), pitch (β) and roll (γ) angles of a human head.

- *Input: Video from the webcam.*
- Step 1: Face and eyes region detection using Haar-like feature algorithm.
- Step 2: Real time tracking of face and eyes region using Tracking-Learning-Detection (TLD) frame work.
- Step 3: Pivot point estimation using face and eyes region.
- Step 4: Geometric analysis for calculating head rotations (yaw (α), pitch (β) and roll (γ) angles) considering pivot point and eyes region.
- *Output: yaw (α), pitch (β) and roll (γ) angles of human head*.

The proposed algorithm has successfully solved the problem of head pose in real-time using 2D information from normal video camera. Besides training and learning required by TLD and Haar algorithm, our algorithm does not require offline training session and does not require any prior knowledge regarding camera parameters and/or distance of the user from the camera.

5 Enabling Physical Action in Computer Mediated Communication: Experiments

This chapter presents video teleconferencing system based on embodied interaction framework. This chapter also presents the experimental studies to tackle problems in standard video conferencing setup. The contribution of the research papers and references to the papers are also given for detail descriptions.

5.1 Embodied Interaction based Video Teleconferencing System

The embodied interaction concept (as detailed in chapter 3) is incorporated in video teleconferencing by using our telepresence mechatronic robot - TEBoT (as detailed in chapter 4). We named this embodied interaction based video teleconferencing system as Embodied Telepresence System (ETS). ETS integrates standard audio-video conferencing with mechanical embodiment of the head gesture of a remote person to enhance the experience of audio-video conferencing by communicative non-verbal ‘actions’ (head gestures). Furthermore, ETS tries to bring digital distance communication to the physical world and try to make it more visible to have a deeper engagement. The ETS design makes CMC to behave in ways that map more effectively to real world interaction. Drawing on embodied cognition insights, ETS attempts to bring mind and body closer to introduce a new way of video teleconferencing. The theories of tangible, social and embodied interaction together with cognitive science are used in our embodied telepresence system (ETS), which embodies the remote participant's head to give more levelheaded presence in the distinct space and be perceived by his collaborators as being equally present among them.

We have used ETS to solve the following problems, which are inherited in standard video teleconferencing;

- Distant located people cannot take turns or do side conversation during conversation [54].
- Distant located people cannot express emotions due to lack of nonverbal modalities [68].
- Distant located people cannot estimate the gaze of his/her fellow collaborator [56, 106].
- Distant located people cannot make an eye contact [23].

The ETS is put into different settings to solve the above mentioned problems, where different *meaning* arises at different level. Furthermore, the solutions for these problems are presented in the form of experimental studies. The solution for the first 2 problems are presented in this section and the solution for last 2 problems are presented in the next section.

5.1.1 Experimental Setup

An embodied tele-presence system (ETS) is setup in a real scenario (see figure 5.1) to measure the *feeling of presence* of a remote user among local collaborators and to measure the *quality of experience* of remote user for comparative scenarios; 1) when remote user is communicating using *standard* video teleconferencing system (e.g. Skype) 2) when remote user is communicating through our embodied interaction based video teleconferencing system (i.e., *ETS*). The left side of the figure 5.1 is the local collaborators' site and the right side of the figure 5.1 is the remote site. The local collaborators are those who are situated at the same physical location as the ETS and a remote collaborator is a person who remotely connects to his/her local collaborators via the ETS. The experiments were performed for measuring the interaction experience according to i) local collaborators perspective and ii) remote user perspective.

5.1.2 ETS for Local Collaborators

An experiment is conducted to test the hypothesis, i.e. "Embodied Interaction based video teleconferencing system increases the 'sense of presence' and enhances the 'quality of in-

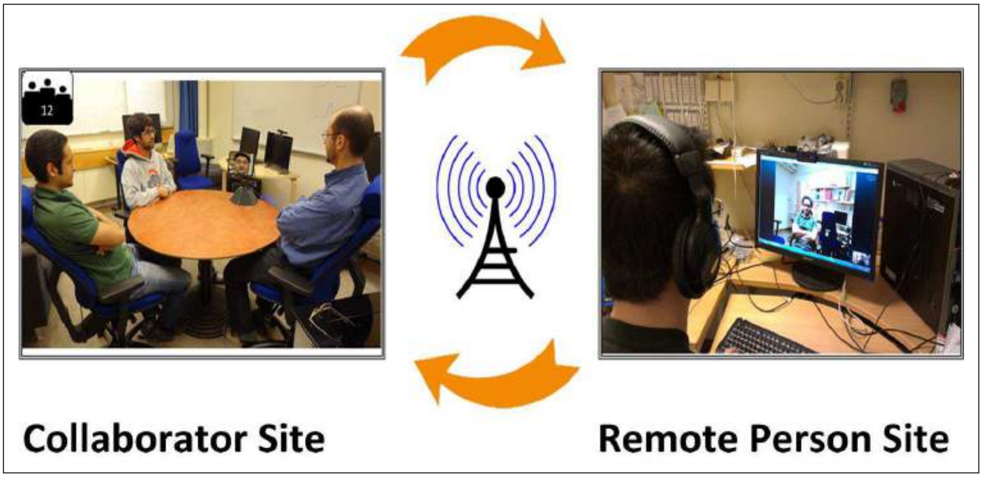


Fig. 5.1: *Application Scenario*

teraction', thus creating a 'feeling of presence' of a remote person among his/her local collaborators". A user study was conducted in which 20 participants from five different countries experienced distant communication using the ETS and Skype. Upon completion of these experiments, the participants were asked to fill in the questionnaires for measuring 'presence' with respect to local collaborator site. The results of perceived 'social' and 'spatial' presence after experiencing standard video conferencing (Skype) and Embodied Tele-presence System (ETS) are shown in figure 5.2a and figure 5.2b. It is clear from the results that the collaborators gave their preference to ETS on all dependent measures. The embodied interaction based video teleconferencing system with precise head gesture has the potential to solve problems, such as 'lack of presence' and 'expression of engagement'.

5.1.3 ETS for Remote Collaborator

Embodied Tele-presence System (ETS) was deployed in novel scenario to enhance the *experience* of remote user during video teleconferencing. The experiment is performed with twenty different participants to measure the remote user collaboration experience by using five 'User satisfaction measures': Interest, Understanding, Comfort, Being there, Quality of Interaction and six 'common human interaction measures': Exploring the distant environment, Side conversation, focus of attention, Eye Contact, Gaze, Turn taking. The results of five user satisfaction measures and six common interaction measures for i)

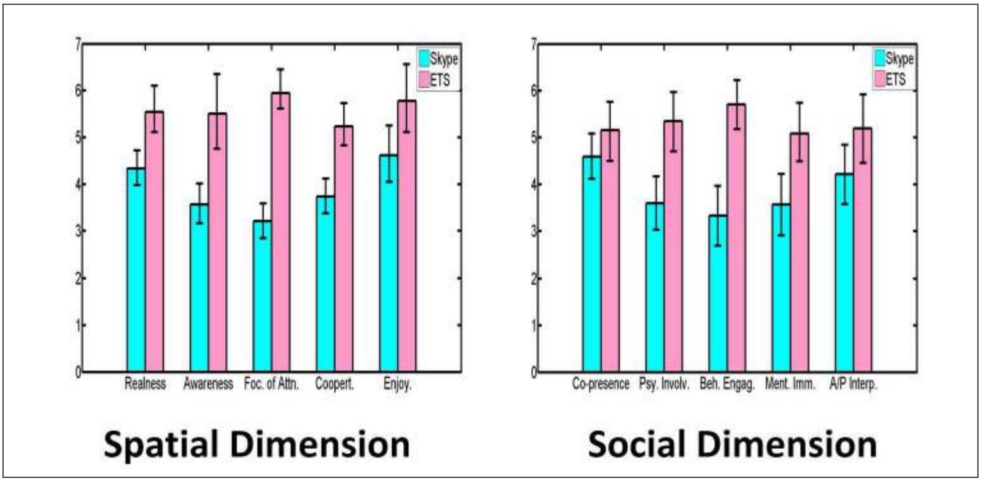


Fig. 5.2: *from left to right– Comparison of Skype Vs ETS on (a) Spatial and (b) social dimensions*

Skype and ii) ETS are shown in figure 5.3a and 5.3b, respectively. It can be concluded from this study that the ETS with the view adjustment capability helped participants to explore the distant environment, allow them to do side conversation and can take turns during video teleconferencing.

The details of the Embodied Telepresence System and the user studies are presented in paper III and paper IV of this thesis.

Paper III

M.S.L. Khan, Haibo Li, Shafiq Ur Rehman, "Embodied tele-presence system (ETS): designing tele-presence for video teleconferencing," in HCI International, Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments, (HCII 2014), Crete, Greece, 2014, pp. 574-585.

Paper IV

M.S.L. Khan, Shafiq Ur Rehman, Pedro La Hera, Feng Liu, Haibo Li, , "A pilot user's perspective in mobile robotic telepresence system," in Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA 2014), Siem Reap, Cambodia, 2014, pp. 1-4.

To facilitate the reading of the above papers here we provide a summary of the problem formulation and contribution of paper III and paper IV.

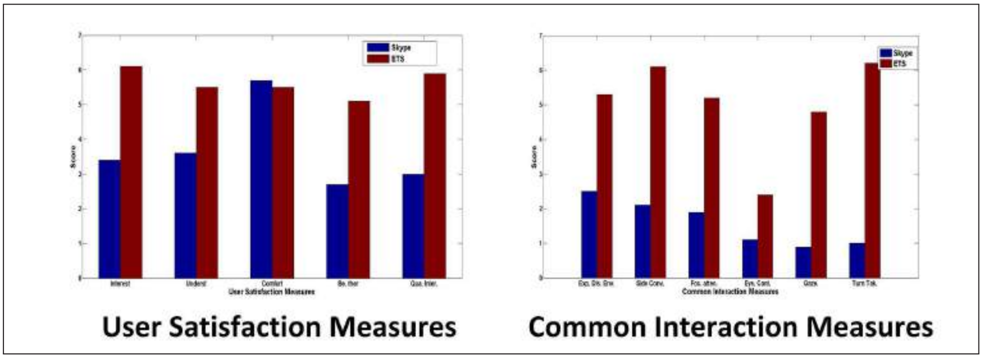


Fig. 5.3: *from left to right– Comparison of Skype Vs ETS on (a) User Satisfaction measures and (b) Common Interaction measures*

5.2 Embodied Telepresence System (ETS): Designing Telepresence for video teleconferencing

5.2.1 Scientific Problem

To bridge the gap between face-to-face communication and distant video communication, two problems are considered in this paper;

- How to introduce embodied interaction concepts in video teleconferencing system.
- How to design a user study to collect empirical data to provide an aftereffects of an embodied interaction based video teleconferencing system as compared to the standard video teleconferencing system.

5.2.2 Contribution

To answer the first scientific question, we have designed and developed a new video teleconferencing system - named - Embodied Telepresence System (ETS). The conceptual design of ETS significantly varies from the standard video teleconferencing systems. Conceptually, ETS incorporates *embodied interaction concept* introduced by Paul Dourish [47]. ETS integrates audio, video and mechanical embodiment of head gesture. ETS not only performs the audio-video communication between the remote and local collaborators but also presents the 'head gesture' of remote user to the local collaborators. Hence, ETS embodies a 'physical representation' of the remote person at local collaborator site. These interactions between the local collaborator(s) and the computer device are influenced by the introduction of ETS and the local collaborators ‘feel’ the physical presence of a remote person at local site.

The other main contribution of the paper is the design of the ‘user study’ to empirically establish the evidence regarding the ‘feeling of presence’ of a remote user among local collaborators for comparative scenarios; 1) when remote user is communicating using standard video teleconferencing system (e.g. Skype) 2) when remote user is communicating through our ETS. We have considered two dimensions of presence for this study 1) social presence and 2) spatial presence. In the user study, the participants were asked to fill in the questionnaires related to social and spatial dimensions of presence. The results of perceived social and spatial presence after experiencing standard video conferencing (Skype) and Embodied Tele-presence System (ETS) are shown in figure 5.2a and figure 5.2b.

The results of this user study show that the Embodied interaction based video teleconferencing system outperforms the standard video conferencing system.

5.3 A Pilot User’s Perspective in Mobile Robotic Telepresence System

5.3.1 Scientific Problem

In this work we have considered the effect and influence of Embodied Telepresence System (ETS) on the quality of experience and interactions of a pilot user during distance communication.

5.3.2 Contribution

In our previous work related to Embodied telepresence system (ETS) studies we have focused on the quality of experience of video teleconferencing for the local collaborators. In this paper, we have used ETS design for an intuitive scenario to improve the experience of pilot user during video teleconferencing. Pilot user is a person who remotely connects to his/her collaborators whereas the local collaborators are those who are situated at the same physical location as the ETS.

We have designed and developed an intuitive velocity-profile control for our ETS. In velocity-profile control the movement of ETS is controlled by the movement of the pilot user’s head. This addition helps a pilot user to explore distant environment, take turns, do side conversations, etc. more naturally.

In this work we have reported the result regarding how accurately the pilot user can take turns, can do side conversation and how accurately it can focus on some specific thing in the meeting room. The participant performed an experiment with 1) Skype and 2) ETS for comparative results. Upon completion of an experiment they answered questions related to User Satisfaction measures and common human interaction measures. The results of five user satisfaction measures and six common interaction measures for i) Skype and ii) ETS are shown in figure 5.3a and 5.3b respectively.

The results suggest that the turn taking capability of Embodied Telepresence System helps pilot user to explore distant environment; and hence improves the quality of experience and interactions.

5.4 Gaze Perception and Awareness in Smart Devices

The remaining two problems, i.e., i) Distant located people cannot estimate the gaze of his/her fellow collaborator and ii) Distant located people cannot make an eye contact, in standard video conferencing are considered in this section.

In face-to-face communication, eye contact and gaze awareness play a significant role for conveying emotions and intentions during conversation. During face-to-face communication, people perceive each other's gaze quite naturally and accurately. However, the gaze awareness/perception are ambiguous in distance communication performed by smart devices (such as iPad, galaxy-tab, smartphones, etc.). It has been reported that the major causes of this ambiguity are the camera position relative to the screen and 2D rendition of 3D human face, i.e., 2D screen is unable to deliver an accurate gaze perception to it users in video conversation (see figure 5.4).

Previously, different hardware setups with complex software algorithms were proposed to solve the problem of gaze awareness/perception in distance communication. The most recent solution for accurate gaze perception employs 3D interfaces such as 3D screens and 3D masks. However, today, more and more users are using smart devices for personal distance communications so there is a need to improve gaze awareness/perception in these smart devices rather than developing new hardwares or 3D interfaces. In this work, we have considered the problem of gaze awareness/perception in distance communication performed by smart devices. Our hypothesis is that, "an accurate gaze perception can be achieved in distance communication by '3D embodiment' of remote user head". We have used our previously developed embodied telepresence system - ETS for 3D embodiment of remote user head. The ETS is 3 DOF neck robot with a mounted tablet-PC (see figure 4.1(a)(b)).

A comparative study is conducted between i) our proposed approach, i.e., ETS and ii) standard 2D stationary screen (2D-SS) based video conferencing setup. The study measures the accuracy of both systems in terms of gaze and eye-contact. Two important gaze related issues are considered in this experiment; i) 'Mona-Lisa gaze effect' in which the gaze is always directed at the person independent of his position in the room (see figure 5.5) and ii) 'Gaze Awareness/Faithfulness' which is the ability to perceive an accurate spatial relationship between the observing person and the object, which is being perceived by the actor. The results of our study were evaluated statistically which confirm that the ETS performs better than the standard video conferencing setup. The ETS not only mitigates the Mona Lisa gaze effect in 2D contents but also supports the 3 levels of gaze faithfulness, i.e., mutual, partial and full gaze faithfulness.

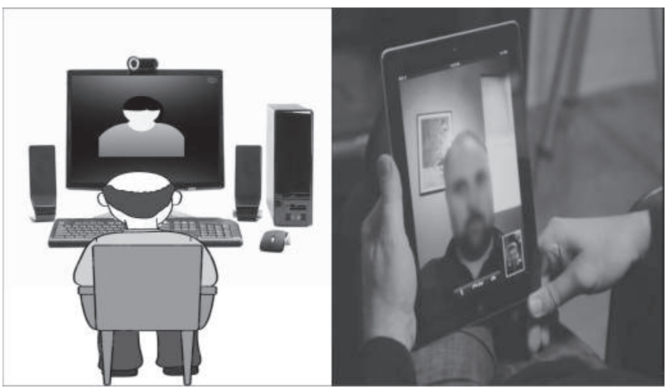


Fig. 5.4: *Standard video conferencing Setups, from left to right (a) consists of a desktop computer with a webcam mounted on top of the LCD/LED screen and (b) in case of smart device a person holds the iPad for video conversation.*

The details of improving gaze perception in distance communication is presented in paper V.

Paper V

M.S.L. Khan, Haibo Li, Shafiq Ur Rehman, "Gaze Perception and Awareness in Smart Devices," in Submitted to International Journal of Human-Computer Studies.

To facilitate the reading of the above paper here we provide a summary of the problem formulation and contribution of paper V.

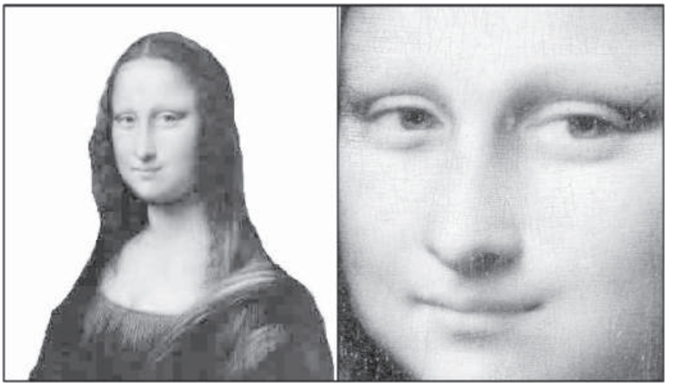


Fig. 5.5: *Mona Lisa Gaze Effect: from left-to-right (a) the original image of mona lisa (b) zoomed image of mona lisa. From these images it is clear that mona lisa is looking directly at you irrespective of your position. (this picture is in the public domain)*

5.5 Gaze Perception and Awareness in Smart Devices

5.5.1 Scientific Problem

The scientific questions we have considered in this work are;

- How to improve a gaze perception during distance communication performed by two dimensional smart screens such as tablet PC, iPad, etc?
- How to curtail mona-lisa gaze effect during video conferencing in these 2D screen based devices?
- How to achieve a gaze faithfulness in smart devices, i.e., mutual, partial and full gaze faithfulness in video conferencing setup?

5.5.2 Contribution

To solve the above scientific problems, researchers are employing complex hardware and software solutions. The most recent solution for accurate gaze perception employs 3D interfaces such as 3D screens and 3D masks. We argue that due to popularity of smart devices (tablet-PC, smart phones, ipad, etc.) for distance communication, there is a need to solve the problem of gaze perception and awareness in these devices rather than developing new hardwares. Our proposed solution to tackle above scientific problems is based on '3D embodiment' of remote user head. 3D embodiment is achieved by our ETS (see figure 4.1(a)(b)), which can rotate around to give an accurate gaze perception of remote person to the local collaborators. Our hypothesis for this work is that, “an accurate gaze perception can be achieved in distance communication by ‘3D embodiment’ of remote user head”.

An experiment is performed with ETS and the gaze perception results are compared with standard 2D-screen (2D-SS) based teleconferencing system. Our experiment considered two important gaze-related issues: i) ‘Mona-Lisa gaze effect’ in which the gaze is always directed at the person independent of his position in the room and ii) ‘Gaze Awareness/Faithfulness’ which is the ability to perceive an accurate spatial relationship between the observing person and the object, which is being perceived by the actor. The following statistical analysis is done on an acquired data of i) ETS and ii) 2D-SS;

- *Regression analysis*: to examine which system (either 2D-SS or ETS) is closer to our expected response, i.e., which system succeed in accurately presenting the gaze of a remote user to his local collaborators.

- *View position analysis*: What effect does the local collaborator’s viewing position play in perceiving the gaze direction of a remote person? To further study the effect of viewing position on decision making of local participants, a pair-wise Pearson correlation is performed.

- *Gaze Modeling*: Gaze modeling is done to see which system (either ETS or 2D-SS) is successful in mitigation the mona-lisa gaze effect in video conferencing setup.

- *Gaze awareness*: It is important to characterize different types of gaze information to analyze video conferencing system. There are three types of gaze information which are used to compare ETS with 2D-SS. 1) mutual gaze or eye-contact is knowing whether someone is looking at you. 2) partial gaze awareness is knowing the general direction someone is looking and 3) full gaze awareness is knowing what someone is looking at.

The results of an experiment confirm that the 3D movement of the 2D screen not only mitigates the Mona Lisa gaze effect in 2D contents but also supports the 3 levels of gaze faithfulness, i.e., mutual, partial and full gaze faithfulness, hence accurately projects the human gaze in distant space.

6 Concluding Remarks and Future Work

The thesis focused on “Enabling Physical Action in Computer Mediated Communication (CMC)”. The action information of standard CMC is confined to 2D display screens. To present the action information of the digital content into a physical world, this thesis has presented a novel video conferencing setup. Our proposed Embodied Telepresence System (ETS) presents the ‘action information’ both digitally and physically during real video conferencing. The ETS mimics the head gestures (physical actions) of a remote person in parallel with audio-video conversation (digital content). Conceptually, ETS is based on ‘embodied interaction concept’ which points out that computer devices should move into social and physical spaces. Overall, the research goal of this thesis is in-line with the goals of embodied interaction that focus on taking us away from these 2D rectangular displays and search for better embodied technologies where ‘digital’ and ‘physical’ world are seamlessly merged in each other.

Contributions of this thesis can be divided into two parts: technical contributions toward the design of telepresence mechatronic robot - TEBoT and the experimental studies in which the ETS is used to improve the flaws in standard video conferencing setup.

In terms of design, we have developed a reliable and nice looking biologically-inspired neck platform (TEBoT) based on real head motion analysis. The TEBoT is compact and self-contained and it fulfils the static and dynamic performances of human neck. To present the head gestures by TEBoT, we have included human-head in the loop where real-human head provides an input to the TEBoT. To capture these real-human head movements we have considered the limitations of previously developed sensor based approaches (SBAs) and vision based approaches (VBAs) and proposed a novel vision based technique which captures the pose angles of human head in real-time without using any wearable sensors and/or markers. Our geometric head pose estimation algorithm calculates the pose angles based on facial feature points and the geometric manipulation of these feature points. This new input control is based on low cost, low resolution, low bit rate and non-wearable webcam of the computer.

For experimental studies, an embodied telepresence system (ETS) is used, which is the combination of embodied interaction concept and our previously designed electromechanical platform (TEBoT). The ETS presents the head gestures of remote user in parallel with audio video conversation in real conferencing setup. We have conducted three experiments where,

1. ETS is used to increase the 'feeling of presence' of remote user among local collaborators.
2. ETS is used by the remote user to 'explore' the local environment.
3. ETS is used to improve the 'gaze perception and awareness' in smart devices based distance communication.

In first experimental study, ETS is used to present the head gestures of remote user among his local collaborators. The result of this experiment shows that Embodied interaction based video conferencing system performs better than the standard video conferencing system in representing nonverbal behaviors (in our case head gesture), thus creating a 'feeling of presence' of a remote person among his/her local collaborators.

In second experimental study, we have given a control of ETS to remote user to explore the local environment. The results show that the remote person is successful in specifying the target of attention, can do side conversation, can take turns, etc.

In third experiment, ETS is used to improve the gaze perception and awareness during distance communication performed by smart devices. The experimental study measures how accurate is ETS in delivering the gaze of a remote person as compared to standard 2D stationary screen (2D-SS) based video conferencing setup. The results of an experimental study show that the local collaborators easily perceived the gaze of remote person projected by the ETS. Furthermore, ETS not only mitigates the Mona Lisa gaze effect in 2D contents but also supports the three levels of gaze faithfulness i.e., mutual, partial and full gaze faithfulness.

ETS was specifically designed for the conferencing purpose, but this system can be used in many different applications. Tele-robots are becoming popular in medical field and our ETS can be used as a co-presenter of a doctor who can remotely assist his staff in daily works and surgeries. ETS can also be used by healthcare attendants and/or family members to check on the person with special needs. ETS can be used by a sick child to attend lectures remotely. Furthermore, ETS can be used for inspection, online training and can be deployed in areas where human access is limited.

In future version of TEBoT, we plan to add arms and hands to our current platform. These mechanical limbs will be controlled by the remote person hands and arms by non-wearable computer vision algorithm. Furthermore, experimental studies will be conducted to evaluate the combination of audio, video, head gesture and hand gesture in video teleconferencing. Currently, we have just considered the head information for gaze perception/awareness. In our future work, we plan to use eye-tracker in parallel with our geometric head pose estimation algorithm to calculate the exact gaze direction of a remote person. Furthermore, we will devise a way to map this gaze information to ETS platform. Finally, we will focus on deployment of smart devices in novel embodied interaction scenarios.

Bibliography

Bibliography

[1] "A collection of several articles on the at & t picturephone." In Bell Laboratories RECORD, Bell Laboratories, 1969, volume 47, pages 134-153 & 160-187.

[2] "Video conference with large optical screens," 2014, http://www.dnp-screens.com/DNP08/Segmenter/Conference-room/Video-conference-screen.aspx.

[3] "Psychologists find that head movement is more important than gender in non-verbal communication," University of Virginia, ScienceDaily., 6 May 2009, www.sciencedaily.com/releases/2009/05/090525105459.htm.

[4] "Serial robot," accessed Aug 2014, http://www.uspto.gov/web/patents/classification/ //uspc901/defs901.htm.

[5] "Cervical fusion mid cervical." Morphopedics. Where Technology and Orthopedics Collides, accessed jan 2012 (2012), http://morphopedics.wikidot.com/cervical-fusion-mid-cervical.

[6] "Vicon Peak: Reflective marker based system," http://www.vicon.com, Accessed Jan 2015.

[7] "Mpu-6050, 6 dof imu board," accessed July 2014, webpage, https://www.sparkfun.com/products/11028.

[8] "Der3, actroid-der series,," July 2014, webpage, http://www.kokoro-dreams.co.jp/english/rttokutyu/actroid.html.

[9] "Non-verbal communication when speaking english," last accessed 2014 (2014), http://www.teacherjoe.us/convskills09.html.

[10] S. O. Adalgeirsson and C. Breazeal, "Mebot: a robotic platform for socially embodied presence," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 2010, pp. 15–22.

[11] Aiko, "Aiko the female robot: Inventor le trung talks about his perfect woman (video)," accessed July 2014, website http://www.unz.org/Pub/HuffingtonPost-2008dec-02433.

[12] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive Behavioural Systems*. Springer, 2012, pp. 114–130.

[13] S. Al Moubayed, J. Edlund, and J. Beskow, "Taming mona lisa: communicating gaze faithfully in 2d and 3d facial projections," *ACM Transactions on Interactive Intelligent Systems*, vol. 1, no. 2, p. 25, 2012.

[14] M. Argyle and M. Cook, "Gaze and mutual gaze." 1976.

[15] K. J. Aström and T. Hägglund, "Pid controllers: theory, design, and tuning," *Instrument Society of America, Research Triangle Park, NC*, 1995.

[16] V. N. Balasubramanian, S. Krishna, and S. Panchanathan, "Person-independent head pose estimation using biased manifold embedding," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 63, 2008.

[17] T. Baltrusaitis, P. Robinson, and L. Morency, "3d constrained local model for rigid and non-rigid facial tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2610–2617.

[18] I. Barakonyi, T. Fahmy, and D. Schmalstieg, "Remote collaboration using augmented reality videoconferencing," in *Proceedings of Graphics interface 2004*. Canadian Human-Computer Communications Society, 2004, pp. 89–96.

[19] V. Baren and W. I. Jsselsteijn, "Measuring presence: A guide to current measurement approaches." Technical report, IST-2001-39237, Information Society Technologies,, Tech. Rep., 2004.

[20] S. Beck, A. Kunert, A. Kulik, and B. Froehlich, "Immersive group-to-group telepresence," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 4, pp. 616–625, 2013.

[21] J. M. Beer and L. Takayama, "Mobile remote presence systems for older adults: acceptance, benefits, and concerns," in *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 2011, pp. 19–26.

[22] G. A. Bekey, *Autonomous robots: from biological inspiration to implementation and control*. MIT press, 2005.

[23] E. Bekkering and J. Shim, "Trust in videoconferencing," *Communications of the ACM*, vol. 49, no. 7, pp. 103–107, 2006.

[24] S. M. Boker, J. F. Cohn, B.-J. Theobald, I. Matthews, T. R. Brick, and J. R. Spies, "Effects of damping head movement and facial expression in dyadic conversation using real–time facial expression tracking and synthesized avatars," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3485–3495, 2009.

[25] I.-L. Boman and A. Bartfai, "The first step in using a robot in brain injury rehabilitation: patients' and health-care professionals' perspective," *Disability and Rehabilitation: Assistive Technology*, no. 0, pp. 1–6, 2014.

[26] P. Bordia, “Face-to-face versus computer-mediated communication: A synthesis of the experimental literature,” *Journal of Business Communication*, vol. 34, no. 1, pp. 99–118, 1997.

[27] E. Boyarskaya and H. Hecht, “The mona lisa effect: Is it confined to the horizontal plane?” *Perception*, vol. 38, 2009.

[28] J. K. Burgoon, “The relationship of verbal and nonverbal codes,” *Progress in communication sciences*, vol. 6, pp. 263–298, 1985.

[29] J. K. Burgoon, J. A. Bonito, A. Ramirez, N. E. Dunbar, K. Kam, and J. Fischer, “Testing the interactivity principle: Effects of mediation, propinquity, and verbal and nonverbal modalities in interpersonal interaction,” *Journal of communication*, vol. 52, no. 3, pp. 657–676, 2002.

[30] T.-J. Cham, S. Krishnamoorthy, and M. Jones, “Analogous view transfer for gaze correction in video sequences,” in *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on*, vol. 3. IEEE, 2002, pp. 1415–1420.

[31] C.-T. Chen, *Linear system theory and design*. Oxford University Press, Inc., 1995.

[32] D.-S. Chen and Z.-K. Liu, “Generalized haar-like features for fast face detection,” in *Machine Learning and Cybernetics, 2007 International Conference on*, vol. 4. IEEE, 2007, pp. 2131–2135.

[33] L. Chen, L. Zhang, Y. Hu, M. Li, and H. Zhang, “Head pose estimation using fisher manifold learning.” in *AMFG*, 2003, pp. 203–207.

[34] M. Chen, “Leveraging the asymmetric sensitivity of eye contact for videoconference,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2002, pp. 49–56.

[35] D. Choi, D.-W. Lee, D. Y. Lee, H. S. Ahn, and H. Lee, “Design of an android robot head for stage performances,” *Artificial Life and Robotics*, vol. 16, no. 3, pp. 315–317, 2011.

[36] A. Clark, “An embodied cognitive science?” *Trends in cognitive sciences*, vol. 3, no. 9, pp. 345–351, 1999.

[37] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” in *Computer Vision - ECCV 98*. Springer, 1998, pp. 484–498.

[38] J. A. Corrales, F. Candelas, and F. Torres, “Hybrid tracking of human operators using imu/uwb data fusion by a kalman filter,” in *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*. IEEE, 2008, pp. 193–200.

[39] A. Criminisi, J. Shotton, A. Blake, and P. H. Torr, “Gaze manipulation for one-to-one teleconferencing,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 191–198.

[40] D. Cristinacce and T. F. Cootes, “Feature detection and tracking with constrained local models.” in *BMVC*, vol. 2, no. 5, 2006, p. 6.

[41] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the cave," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, 1993, pp. 135–142.

[42] F. Delaunay, J. de Greeff, and T. Belpaeme, "A study of a retro-projected robotic face and its effectiveness for gaze reading by humans," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 2010, pp. 39–44.

[43] F. Delcomyn, *Biologically inspired robots*. Bioinspiration and Robotics Walking and Climbing Robots, Maki K. Habib (Ed.), ISBN: 978-3-902613-15-8, 2007.

[44] A. DeVincenzi, L. Yao, H. Ishii, and R. Raskar, "Kinected conference: augmenting video imaging with calibrated depth and audio," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 2011, pp. 621–624.

[45] J. Devore, *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole Publishing Company, 1999.

[46] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler, "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*. ACM, 2002, pp. 321–326.

[47] P. Dourish, *Where the action is: the foundations of embodied interaction*. MIT press, 2004.

[48] M. Dumont, S. Maesen, S. Rogmans, and P. Bekaert, "A prototype for practical eye-gaze corrected video chat on graphics hardware." *SIGMAP*, vol. 2008, pp. 236–243, 2008.

[49] M. Elmogy, C. Habel, and J. Zhang, "Online motion planning for hoap-2 humanoid robot navigation," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 3531–3536.

[50] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 52–73, 2007.

[51] F. Eyssel, F. Hegel, G. Horstmann, and C. Wagner, "Anthropomorphic inferences from emotional nonverbal cues: A case study." in *RO-MAN*, 2010, pp. 646–651.

[52] D. I. Fels, J. K. Waalen, S. Zhai, and P. Weiss, "Telepresence under exceptional circumstances: Enriching the connection to school for sick children," in *Proc. Interact*, 2001, pp. 617–624.

[53] F. Ferro and L. Marchionni, "Reem: A humanoid service robot," in *ROBOT2013: First Iberian Robotics Conference*. Springer, 2014, pp. 521–525.

[54] K. Fischer and T. Tenbrink, "Video conferencing in a transregional research cooperation: Turn-taking in a new medium," *Connecting Perspectives, Videokonferenz: Beiträge zu ihrer Erforschung und Anwendung, Aachen: Shaker Verlag*, 2003.

[55] K. P. Fishkin, T. P. Moran, and B. L. Harrison, "Embodied user interfaces: Towards invisible user interfaces," in *Engineering for Human-Computer Interaction*. Springer, 1999, pp. 1–18.

[56] E. Fox, "The role of visual processes in modulating social interactions," *Visual Cognition*, vol. 12, no. 1, pp. 1–11, 2005.

[57] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. Kramer, "Gestures over video streams to support remote collaboration on physical tasks," *Human-Computer Interaction*, vol. 19, no. 3, pp. 273–309, 2004.

[58] V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun, "Real time motion capture using a single time-of-flight camera," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 755–762.

[59] W. W. Gaver, "The affordances of media spaces for collaboration," in *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. ACM, 1992, pp. 17–24.

[60] A. Gee and R. Cipolla, "Determining the gaze of faces in images," *Image and Vision Computing*, vol. 12, no. 10, pp. 639–647, 1994.

[61] J. Gemmell, K. Toyama, and C. L. Zitnick, "Manipulation of video eye gaze and head orientation for video teleconferencing," 1999.

[62] J. Gemmell, K. Toyama, C. L. Zitnick, T. Kang, and S. Seitz, "Gaze awareness for videoconferencing: A software approach," *IEEE Multimedia*, vol. 7, no. 4, pp. 26–35, 2000.

[63] D. O. Girardo, M. B. Popovic, A. Blumenau, I. Galiana, F. L. H. III, R. D. Howe, L. Jentoft, S. B. Kesner, E. L. Lin, and S. Mandala, "Physics applied to post-stroke rehabilitation shoulder soft robotics brace december 31, 2011 report."

[64] M. S. Grewal and A. P. Andrews, *Kalman filtering: theory and practice using MATLAB*. John Wiley & Sons, 2011.

[65] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang *et al.*, "blue-c: a spatially immersive display and 3d video portal for telepresence," in *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3. ACM, 2003, pp. 819–827.

[66] K. Hafner and M. Lyon, *Where wizards stay up late: The origins of the Internet*. Simon and Schuster, 1998.

[67] M. Hashimoto and D. Morooka, "Robotic facial expression using a curved surface display," *Journal of Robotics and Mechatronics*, vol. 18, no. 4, p. 504, 2006.

[68] C. Heath and P. Luff, "Disembodied conduct: communication through video in a multimedia office environment," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1991, pp. 99–103.

[69] R. M. Held and N. I. Durlach, "Telepresence," *Presence: Teleoper. Virtual Environ.*, vol. 1, no. 1, pp. 109–112, Jan. 1992.

[70] J. L. Hiatt, L. P. Gartner, J. Rohen, and J. L. Gadd, *Textbook of head and neck anatomy*. Lippincott Williams & Wilkins, 2001.

[71] T. Horprasert, Y. Yacoob, and L. S. Davis, "Computing 3d head orientation from a monocular image sequence," in *25th Annual AIPR Workshop on Emerging Applications of Computer Vision*. International Society for Optics and Photonics, 1997, pp. 244–252.

[72] J. Huang, X. Shao, and H. Wechsler, "Face pose discrimination using support vector machines (svm)," in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, vol. 1. IEEE, 1998, pp. 154–156.

[73] T. Jenkin, J. McGeachie, D. Fono, and R. Vertegaal, "eyeview: Focus+ context views for large group video conferences," in *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 2005, pp. 1497–1500.

[74] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec, "Achieving eye contact in a one-to-many 3d video teleconferencing system," in *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3. ACM, 2009, p. 64.

[75] S. Kajita, K. Kaneko, F. Kaneiro, K. Harada, M. Morisawa, S. Nakaoka, K. Miura, K. Fujiwara, E. S. Neo, I. Hara *et al.*, "Cybernetic human hrp-4c: A humanoid robot with human-like proportions," in *Robotics Research*. Springer, 2011, pp. 301–314.

[76] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 7, pp. 1409–1422, 2012.

[77] A. Kendon, "Some functions of gaze direction in social interaction," *Acta Psychologica*, vol. 26, pp. 22 – 63, 1967.

[78] M. S. L. Khan, H. Li, and S. ur Réhman, "Embodied tele-presence system (ets): Designing tele-presence for video teleconferencing," in *Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments*. Springer, 2014, pp. 574–585.

[79] M. S. L. Khan, S. U. Réhman, L. Zhihan, and H. Li, "Head orientation modeling: Geometric head pose estimation using monocular camera," in *The 1st IEEE/IIAE International Conference on Intelligent Systems and Image Processing 2013 (ICISIP2013)*, 2013.

[80] M. S. L. Khan and S. ur Réhman, *Distance Communication: Trends and Challenges and How to resolve them*. Handbook: Strategies for a Creative Future with Computer Science, Quality Design and Communicability Human-Computer Interaction Collection, Blue Herons Editions, Ch. 1. ISBN: 978-88-96471-13-5, 2014.

[81] M. S. L. Khan, S. ur Réhman, P. La Hera, F. Liu, and H. Li, "A pilot user's prospective in mobile robotic telepresence system," APSIPA 2014, Dec 2014.

[82] C. D. Kidd and C. Breazeal, "Effect of a robot on user perceptions," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 4. IEEE, 2004, pp. 3559–3564.

[83] H.-H. Kim, J.-S. Park, J.-W. Jung, and K.-H. Park, "Immersive teleconference system based on human-robot-avatar interaction using head-tracking devices," *International Journal of Control, Automation and Systems*, vol. 11, no. 5, pp. 1028–1037, 2013.

[84] J. Kjeldskov, J. H. Smedegård, T. S. Nielsen, M. B. Skov, and J. Paay, "Eyegaze: enabling eye contact over video," in *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. ACM, 2014, pp. 105–112.

[85] C. L. Kleinke, "Gaze and eye contact: a research review." *Psychological bulletin*, vol. 100, no. 1, p. 78, 1986.

[86] A. Kristoffersson, S. Coradeschi, and A. Loutfi, "A review of mobile robotic telepresence," *Advances in Human-Computer Interaction*, vol. 2013, p. 3, 2013.

[87] T. Kuchi, "Computer mediated communication: Social interaction and the internet," *Online Information Review*, vol. 30, no. 2, pp. 194–195, 2006.

[88] T. Kuratate, Y. Matsusaka, B. Pierce, and G. Cheng, "Mask-bot: a life-size robot head using talking head animation for human-robot communication," in *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*. IEEE, 2011, pp. 99–104.

[89] C. Kuster, T. Popa, J.-C. Bazin, C. Gotsman, and M. Gross, "Gaze correction for home video conferencing," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, p. 174, 2012.

[90] A. Lablack, C. Djeraba, and J. Martinet, *Head pose estimation using a texture model based on Gabor Wavelets*. INTECH Open Access Publisher, 2010.

[91] A. Lablack, F. Maquet, N. Ihaddadene, and C. Djeraba, "Visual gaze projection in front of a target scene," in *IEEE International Conference on Multimedia and Expo, 2009. ICME 2009.* IEEE, 2009, pp. 1839–1840.

[92] S. R. Langton, R. J. Watt, and V. Bruce, "Do the eyes have it? cues to the direction of social attention," *Trends in cognitive sciences*, vol. 4, no. 2, pp. 50–59, 2000.

[93] H. Li, P. Roivainen, and R. Forchheimer, "3-d motion estimation in model-based facial image coding," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 6, pp. 545–555, 1993.

[94] L. Loke and T. Robertson, "Moving and making strange: An embodied approach to movement-based interaction design," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 20, no. 1, p. 7, 2013.

[95] M. Lombard, T. B. Ditton, D. Crane, B. Davis, G. Gil-Egui, K. Horvath, J. Rossman, and S. Park, "Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument," in *Third international workshop on presence, delft, the netherlands*, vol. 240, 2000.

[96] C. N. Macrae, B. M. Hood, A. B. Milne, A. C. Rowe, and M. F. Mason, "Are you looking at me? eye gaze and person perception," *Psychological Science*, vol. 13, no. 5, pp. 460–464, 2002.

[97] Mahru, "Robot maid combats unclean households," accessed July 2014, website http://www.koreatimes.co.kr/www/news/tech/2010/01/13359176.html.

[98] E. N. Marieb, *Essentials of human anatomy & physiology*. San Francisco :Benjamin Cummings, 2006.

[99] A. Mehrabian, *Nonverbal communication*. Transaction Publishers, 1977.

[100] J.-P. Merlet, *Parallel robots*. Springer, 2006, vol. 128.

[101] K. Meyer, H. L. Applewhite, and F. A. Biocca, "A survey of position trackers," in *Presence: Teleoperators and Virtual Environments (ISSN 1054-7460), vol. 1, no. 2, p. 173-200.*, vol. 1, 1992, pp. 173–200.

[102] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 504–513.

[103] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding*, vol. 104, no. 2, pp. 90–126, 2006.

[104] A. F. Monk and C. Gale, "A look is worth a thousand words: full gaze awareness in video-mediated conversation," *Discourse Processes*, vol. 33, no. 3, pp. 257–278, 2002.

[105] Q. Muhlbauer, K. Kuhnlenz, and M. Buss, "A model-based algorithm to estimate body poses using stereo vision," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*. IEEE, 2008, pp. 285–290.

[106] N. Mukawa, T. Oka, K. Arai, and M. Yuasa, "What is connected by mutual gaze?: user's behavior in video-mediated communication," in *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2005, pp. 1677–1680.

[107] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 607–626, 2009.

[108] J. Ng and S. Gong, "Multi-view face detection and pose estimation using a composite support vector machine across the view sphere," in *Recognition, Analysis, & Tracking of Faces & Gestures in Real-Time Systems, IEEE ICCV Workshop on*. IEEE Computer Society, 1999, pp. 14–14.

[109] D. Nguyen and J. Canny, "Multiview: spatially faithful group video conferencing," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2005, pp. 799–808.

[110] D. A. Norman, *The psychology of everyday things*. Basic books, 1988.

[111] N. P. Oess, J. Wanek, and A. Curt, “Design and evaluation of a low-cost instrumented glove for hand function assessment,” *Journal of neuroengineering and rehabilitation*, vol. 9, no. 2, 2012.

[112] K. Ogawa, S. Nishio, K. Koda, G. Balistreri, T. Watanabe, and H. Ishiguro, “Exploring the natural reaction of young and aged person with telenoid in a real world.” *JACIII*, vol. 15, no. 5, pp. 592–597, 2011.

[113] K.-I. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita, “Multiparty videoconferencing at virtual social distance: Majic design,” in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. ACM, 1994, pp. 385–393.

[114] S. Ozawa, S. Mieda, Y. Yao, H. Takada, T. Kawakami, S. Nasu, T. Ishinabe, M. Kano, M. Sasai, T. Uchida *et al.*, “Muldiroh: A multi-view human representation system using a qda screen with multiple cameras,” *Display Technology, Journal of*, vol. 10, no. 2, pp. 87–93, 2014.

[115] Y. Pan and A. Steed, “Preserving gaze direction in teleconferencing using a camera array and a spherical display,” in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*. IEEE, 2012, pp. 1–4.

[116] Y. Pan, W. Steptoe, and A. Steed, “Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction,” in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2014, pp. 1397–1406.

[117] I.-W. Park, J.-Y. Kim, and J.-H. Oh, “Online biped walking pattern generation for humanoid robot khr-3 (kaist humanoid robot-3: Hubo),” in *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*. IEEE, 2006, pp. 398–403.

[118] F. Paterno, *Model-based design and evaluation of interactive applications*. Springer Science & Business Media, 2000.

[119] E. Paulos and J. Canny, “Social tele-embodiment: Understanding presence,” *Autonomous Robots*, vol. 11, no. 1, pp. 87–95, 2001.

[120] C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, and I. Poggi, “A model of attention and interest using gaze behavior,” in *Intelligent virtual agents*. Springer, 2005, pp. 229–240.

[121] PolyCom, “Video, voice and content solutions for face-to-face collaboration,” 2014, http://www.polycom.com/video-collaboration.html.

[122] J. D. Prothero, D. E. Parker, T. Furness, and M. Wells, “Towards a robust, quantitative measure for presence,” in *Proceedings of the conference on Experimental Analysis and Measurement of Situation Awareness*, 1995, pp. 359–366.

[123] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, “The office of the future: A unified approach to image-based modeling and spatially immersive displays,” in *Proceedings*

of the 25th annual conference on Computer graphics and interactive techniques. ACM, 1998, pp. 179–188.

[124] N. Rezzoug, J. Jacquier-Bret, and P. Gorce, "A method for estimating three-dimensional human arm movement with two electromagnetic sensors," *Computer methods in biomechanics and biomedical engineering*, vol. 13, no. 6, pp. 663–668, 2010.

[125] M. Rhoads, "Face-to-face and computer-mediated communication: What does theory tell us and what have we learned so far?" *Journal of Planning Literature*, vol. 25, no. 2, pp. 111–122, 2010.

[126] D. J. Roberts, J. Rae, T. W. Duckworth, C. M. Moore, and R. Aspin, "Estimating the gaze of a virtuality human," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 4, pp. 681–690, 2013.

[127] O. Rudovic, M. Pantic, and I. Patras, "Coupled gaussian processes for pose-invariant facial expression recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 6, pp. 1357–1369, 2013.

[128] D. R. Rutter and G. M. Stephenson, "The role of visual communication in synchronising conversation," *European Journal of Social Psychology*, vol. 7, no. 1, pp. 29–37, 1977.

[129] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent asimo: System overview and integration," in *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, vol. 3. IEEE, 2002, pp. 2478–2483.

[130] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, "Android as a telecommunication medium with a human-like presence," in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*. IEEE, 2007, pp. 193–200.

[131] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab, "Human skeleton tracking from depth data using geodesic distances and optical flow," *Image and Vision Computing*, vol. 30, no. 3, pp. 217–226, 2012.

[132] A. Sellen, B. Buxton, and J. Arnott, "Using spatial cues to improve videoconferencing," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1992, pp. 651–652.

[133] M. Sheets-Johnstone, "The primacy of movement (advances in consciousness research)," 2009.

[134] T. B. Sheridan, "Musings on telepresence and virtual presence," *Presence: Teleoperators and virtual environments*, vol. 1, no. 1, pp. 120–126, 1992.

[135] K. Shinozawa, F. Naya, J. Yamato, and K. Kogure, "Differences in effect of robot and screen agent recommendations on human decision-making," *International Journal of Human-Computer Studies*, vol. 62, no. 2, pp. 267–279, 2005.

[136] N. Sim, C. Gavriel, W. Abbott, and A. Faisal, “The head mouse - head gaze estimation "in-the-wild" with low-cost inertial sensors for bmi use.” IEEE, 2013, pp. 735–738.

[137] D. Sirkin, G. Venolia, J. Tang, G. Robertson, T. Kim, K. Inkpen, M. Sedlins, B. Lee, and M. Sinclair, “Motion and attention in a kinetic videoconferencing proxy,” in *Human-Computer Interaction–INTERACT 2011*. Springer, 2011, pp. 162–180.

[138] F. Solina and R. Ravnik, “Fixing missing eye-contact in video conferencing systems,” 2011.

[139] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot modeling and control*. Wiley New York, 2006, vol. 3.

[140] S. Srinivasan and K. L. Boyer, “Head pose estimation using view based eigenspaces,” in *Pattern Recognition, International Conference on*, vol. 4. IEEE Computer Society, 2002, pp. 40 302–40 302.

[141] C. Stanton, A. Bogdanovych, and E. Ratanasena, “Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning,” in *Proc. Australasian Conference on Robotics and Automation*, 2012.

[142] R. Stiefelhagen, J. Yang, and A. Waibel, “A model-based gaze tracking system,” *International Journal on Artificial Intelligence Tools*, vol. 6, no. 02, pp. 193–209, 1997.

[143] R. Stokes, “Human factors and appearance design considerations of the mod ii picture-phone&# 174; station set,” *Communication Technology, IEEE Transactions on*, vol. 17, no. 2, pp. 318–323, 1969.

[144] A. Sundaresan and R. Chellappa, “Markerless motion capture using multiple cameras,” in *Computer Vision for Interactive and Intelligent Environment, 2005*. IEEE, 2005, pp. 15–26.

[145] C. Systems., “Cisco visual networking index: .” Global mobile data traffic forecast update,, 2011-2016, http://www.cisco.com.

[146] K. Tanaka, S. Onoue, H. Nakanishi, and H. Ishiguro, “Motion is enough: How real-time avatars improve distant communication,” in *Collaboration Technologies and Systems (CTS), 2013 International Conference on*. IEEE, 2013, pp. 465–472.

[147] J. C. Tang and E. Isaacs, “Why do users like video?” *Computer Supported Cooperative Work (CSCW)*, vol. 1, no. 3, pp. 163–196, 1992.

[148] A. M. Tekalp and J. Ostermann, “Face and 2-d mesh animation in mpeg-4,” *Signal Processing: Image Communication*, vol. 15, no. 4, pp. 387–421, 2000.

[149] S. Thayer and W. Schiff, “Gazing patterns and attribution of sexual involvement,” *The Journal of Social Psychology*, vol. 101, no. 2, pp. 235–246, 1977.

[150] H. Tobita, S. Maruyama, and T. Kuzi, “Floating avatar: telepresence system using blimps for communication and entertainment,” in *CHI’11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011, pp. 541–550.

[151] R. Valenti, N. Sebe, and T. Gevers, "Combining head pose and eye location information for gaze estimation," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 802–815, 2012.

[152] A. Van der Hoest and S. McCallum, "Eye contact in leisure video conferencing," *Norsk informatikkonferanse*, vol. 2012, 2012.

[153] P. van der Smagt, M. Grebenstein, H. Urbanek, N. Fligge, M. Strohmayr, G. Stillfried, J. Parrish, and A. Gustus, "Robotics of human movements," *Journal of Physiology-Paris*, vol. 103, no. 3, pp. 119–132, 2009.

[154] T. Vatahska, M. Bennewitz, and S. Behnke, "Feature-based head pose estimation from images," in *Humanoid Robots, 2007 7th IEEE-RAS International Conference on*. IEEE, 2007, pp. 330–335.

[155] G. Venolia, J. Tang, R. Cervantes, S. Bly, G. Robertson, B. Lee, and K. Inkpen, "Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 1049–1058.

[156] R. Vertegaal, G. van der Veer, and H. Vons, "Effects of gaze on multiparty mediated communication," in *Graphics Interface*, 2000, pp. 95–102.

[157] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung, "Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2003, pp. 521–528.

[158] P. M. Vespa, C. Miller, X. Hu, V. Nenov, F. Buxey, and N. A. Martin, "Intensive care unit robotic telepresence facilitates rapid physician response to unstable patients and decreased cost in neurointensive care," *Surgical neurology*, vol. 67, no. 4, pp. 331–337, 2007.

[159] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[160] N. Wang, X. Gao, D. Tao, and X. Li, "Facial feature point detection: A comprehensive survey," *arXiv preprint arXiv:1410.1037*, 2014.

[161] G. Welch and G. Bishop, "An introduction to the kalman filter. university of north carolina, department of computer science," TR 95-041, Tech. Rep., 1995.

[162] G. Welch and E. Foxlin, "Motion tracking survey," *IEEE Computer graphics and Applications*, pp. 24–38, 2002.

[163] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence: Teleoperators and virtual environments*, vol. 7, no. 3, pp. 225–240, 1998.

[164] G. D. Wood and D. C. Kennedy, "Simulating mechanical systems in simulink with simmechanics," *The Mathworks Report*, 2003.

[165] R. Yang and Z. Zhang, "Model-based head pose tracking with stereovision," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 255–260.

[166] R. Yang and Z. Z, "Eye gaze correction with stereovision for video-teleconferencing," in *Computer Vision-ECCV 2002*. Springer, 2002, pp. 479–494.

[167] B. YIP and J. S. JIN, "Face re-orientation in video conference using ellipsoid model." 2003, in OZCHI, 167Ű173.

[168] B. You, D. Kim, C. Kim, Y. Oh, M. Jeong, and S. Oh, "Network-based humanoid mahru as ubiquitous robotic companion," *17th World Congr., Int. Federation Autom. Control, Seoul, Korea*, 2008.

[169] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. rep., Microsoft Research, Tech. Rep., 2010.

[170] L. Zhang, S. Bi, and D. Liu, "Dynamic leg motion generation of humanoid robot based on human motion capture," in *Intelligent Robotics and Applications*. Springer, 2008, pp. 83–92.

[171] H. Zhou and H. Hu, "Human motion tracking for rehabilitation - a survey," *Biomedical Signal Processing and Control*, vol. 3, no. 1, pp. 1–18, 2008.