Implementation and evaluation of multi-avatar communication system for online communication support

Faisal Mehmood

March 2023

Implementation and evaluation of multi-avatar communication system for online communication support

A dissertation submitted to

THE GRADUATE SCHOOL OF ENGINEERING SCIENCE

OSAKA UNIVERSITY

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ENGINEERING

BY

Faisal Mehmood

March 2023

Table of Contents

List of Tables 8

List of Figures 9

Abstract… 13

Chapter 1… 14

Introduction 14

Chapter 2… 19

Attentional behaviors and temporal delays of children with ASD (visitors) in response to different social cues of robotic agents. 19

Aim 19

Abstract 19

2.1 Introduction 20

2.2 Multi-party communication system 22

2.3 Materials and method 23

2.3.1 Method 23

2.3.2 Participants 23

2.3.3 Stimuli 23

2.3.4 Procedure 25

2.3.5 Measurements 25

2.4 Results 26

2.4.1 Latency in shifting attention 26
2.4.2 Number of accepted social stimulus. 27
2.5 Discussion 28
2.6 Conclusion and future work. 29
Chapter 3… 30
Communication apprehension and eye contact anxiety in video conference involving teleoperated robot avatar: a subjective evaluation study. 30
Aim 30
Abstract 31
3.1 Introduction 31
3.2 Robotic video conferencing system for providing effective communication support for people with CA and AEC. 34
3.3 Experiment I 36
3.3.1 Method 36
3.3.2 Participants 37
3.3.3 Apparatus 37
3.3.4 Stimuli 38
3.3.5 Manipulation check 39
3.3.6 Procedure 39
3.3.7 Measurements 39

3.3.8 Results 41

3.4 Experiment II 43

3.4.1 Method 43

3.4.2 Participants 44

3.4.3 Apparatus 44

3.4.4 Stimuli 45

3.4.5 Manipulation check 45

3.4.6 Procedure 45

3.4.7 Measurements 45

3.4.8 Results 45

3.5 Discussion 48

3.6 Conclusion and future work. 50

Chapter 4… 52

Avatars-mediated video conference system for mediating the stress and anxiety of response time of a person with communication difficulties 52

Aim 52

Abstract 53

4.1 Introduction 54

4.2 Proposed robotic video conference system 58

4.2.1 Hardware 58

4.2.2 Software 59
4.2.3 Uttering agent selection 59
4.2.4 Interactive responses 60
4.3 Subjective evaluation experiment I 60
4.3.1 Experimental conditions 61
4.3.2 Experimental design 61
4.3.3 Method 61
4.3.4 Manipulation check 62
4.3.5 Participants 62
4.3.6 Video stimuli 62
4.3.7 Measurements 64
4.3.8 Results 65
4.4 Subjective evaluation experiment II 67
4.4.1 Experimental conditions 68
4.4.2 Experimental design 68
4.4.3 Method 68
4.4.4 Manipulation check 68
4.4.5 Participants 68
4.4.6 Video stimuli 69
4.4.7 Measurements 70

4.4.8 Results ... 70

4.5 Interactive experiment ... 73

4.5.1 Experimental conditions ... 73

4.5.2 Experimental design ... 73

4.5.3 Method ... 73

4.5.4 Participants ... 74

4.5.5 Stimuli ... 74

4.5.6 Procedure ... 74

4.5.7 Measurements ... 75

4.5.8 Results ... 76

4.6 Discussion ... 79

4.7 Conclusion and future work. ... 81

Chapter 5… ... 83

Effect of having and switching multiple avatars ... 83

Aim ... 83

Abstract ... 84

5.1 Introduction ... 84

5.2 Video teleconferencing system involving physical avatars ... 88

5.3 Materials and method ... 91

5.3.1 Method ... 91

5.3.2 Participants 91

5.3.3 Conversational scripts 92

5.3.4 Stimuli 92

5.3.5 Procedure 93

5.3.6 Measurements 94

5.4 Results 97

5.4.1 Right to talk (RoT) 97

5.4.2 Perceived social support 97

5.4.3 Sense of being attended (SoBA) 98

5.5 Discussion 99

5.6 Conclusion and future work 102

Chapter 6… 104

Conclusion and future work 104

6.1 Conclusion: 104

6.2 Future work of Ph.D. 105

6.2.1 Avatar personality design 106

6.2.2 Dialogue flow design 107

6.2.3 Teleoperation load design 108

Acknowledgment 110

List of Publications 111

Journal Paper(s) [published] 111

Journal Paper(s) [Submitted] 111

Conference Paper(s) [published].. 111

References: 112

Appendixes 130

List of Tables

Table I: Post hoc Wilcoxon signed rank tests for latency in shifting attention

Table II: Post hoc Wilcoxon signed rank tests for number of accepted stimuli

List of Figures

Chapter 01:

Figure 1.1: Research map; communicational factors and single-party communication challenges

Figure 1.2: Expanded research map; communicational factors and multi-party communication

Figure 1.3: proposal, double avatar system for the operator

Chapter 02:

Figure 2.1: Multi-party communication system to study the attentional behaviors

Figure 2.2: The sequence of presenting the stimuli

Figure 2.3: Multi-party communication system; (a) visual stimulus, (b) speech stimulus, and (c) motion stimulus

Figure 2.4: latency in shifting attention to robotic partner

Figure 2.5: The number of accepted stimuli from robotic partners

Chapter 03:

Figure 3.1: study II location of research map

Figure 3.2: Robotic video conferencing system for providing effective communication support to people with CA and AEC (Robotic condition).

Figure 3.3: Traditional video conference system of Experiment I, (Human Condition)

Figure 3.4: Experiment-Ⅰ: Human condition (A) and Robot condition (B).

Figure 3.5: Communication apprehension (CA) score.

Figure 3.6: Anxiety in eye contact (AEC) score

Figure 3.7: Sense of being attended (SoBA) score

Figure 3.8: Traditional video conference system of Experiment-II.

Figure 3.9: Experiment-II: Human condition (A) and Robot condition (B).

Figure 3.10: Communication apprehension (CA) score.

Figure 3.11: Anxiety in eye contact (AEC) score.

Figure 3.12: Sense of being attended (SoBA) score.

Chapter 04:

Figure 4.1: Study III location of research map

Figure 4.2: Proposed video conference system for reducing the stress of managing RT in a person.

Figure 4.3: GUIs: (a) for interviewee, (b) for interviewer

Figure 4.4: Video stimuli, (a) conventional robotic video conferencing system with a negative thought bubble of the interviewer, (b) proposed robotic video conferencing system with an angry mood emoji of interviewer (c) happy mood emoji, (d) neutral mood emoji, (e) negative thought bubble, (f) positive thought bubble.

Figure 4.5: Imagined situational communication apprehension measure (*iSCAM*)

Figure 4.6: Imagined fear of negative evaluation (*iFNE*)

Figure 4.7: Imagined sense of being attended (*iSoBA*)

Figure 4.8: Imagined intention to use (*iITU*)

Figure 4.9: Video stimuli (a) conventional video conference system with negative thought bubble. (b) proposed video conference system with angry mood emoji.

Figure 4.10: Imagined situational communication apprehension measure (*iSCAM*).

Figure 4.11: Imagined fear of negative evaluation (*iFNE*).

Figure 4.12: Imagined sense of being attended (*iSoBA*).

Figure 4.13: Imagined intention to use (*iITU*) score.

Figure 4.14: proposed video conference system with a participant

Figure 4.15: Experienced situational communication apprehension measure (*eSCAM*).

Figure 4.16: Experienced fear of negative evaluation (*eFNE*) score.

Figure 4.17: Experienced sense of being attended (*eSoBA*) score.

Figure 4.18: Experienced intention to use (*eITU*) score.

Chapter 05:

Figure 5.1: Study IV location of research map

Figure 5.2: The proposed robotic video conference system

Figure 5.3: Operator’s GUI.

Figure 5.4: Right to talk (RoT) score of operators.

Figure 5.5: Social support (SS) score of operators.

Figure 5.6: Sense of being attended (SoBA) score of operators.

Chapter 06:

Figure 6.1: Research map and future work ideas

Figure 6.2: Future works

Abstract

People with communication apprehension have difficulties in daily life face-to-face and online interactions. They tend to avoid interacting with humans directly and seek other alternative ways of communication e.g., using technologies. Technologies like text, audio, video, and avatar are examples of single-party conversation where some challenges exist e.g., 1) reducing pressure of communication and giving own commitment of communication to a representative, and 2) giving/directing pressure of communication to other party and getting their commitment to conversation. However, dealing with such challenges in single-party communication is very difficult. An alternative to the single-party conversation is multi-party conversation. I studied the phenomenon of multi-party conversation, involving two avatars and a visitor, which revealed that it is better at capturing the attention of visitors alternatively; an intuition that leads our research. The previous literature also provides some clues regarding the probable usefulness of multi-party conversation in dealing with the aforementioned challenges. Therefore, considering such expected usefulness of multi-party communication systems, I proposed its use for operators, i.e., using two avatars to communicate with visitors: a multi-party conversation scenario. By conducting a series of studies, I observed that using multi-party conversation system(s) is better in 1) avoiding pressure of communication and getting commitment of others, 2) avoiding the pressure of communication and giving own commitment of communication to a representative, and 3) getting others commitment in communication and giving pressure of communication. The conducted studies contribute to the literature by adding knowledge regarding the usefulness of proposed methods in dealing with the challenges of the single-party conversation system; as mentioned above.

Chapter 1

Introduction

People with communication apprehension (CA) face difficulties in daily life communication. Such difficulties are not only evident in daily life face-to-face (FtF) interactions [1–4] but are also in online interactions too [5,6]. There are several factors that are the cause of such difficulties. Commitment and pressure in communication are two of them. The commitment to communication can be given to others by appointing an alternative representative of oneself. On the other hand, commitment in communication can be taken by approaching others via multiple representatives. Similarly, by adopting the same methods i.e., appointing an alternative representative of oneself or approaching to others via multiple representatives, pressure in communication can be avoided or directed towards others. Instead of interacting with others FtF, people with CA prefer to use alternative ways of interactions e.g., different technologies [7–11].

To handle factors of commitment and pressure in communication, computer mediated communications (CMC), and robot avatar technologies are explored. CMC basically refers to as "a process of human communication via computers, involving people, situated in particular contexts, engaging in processes to shape media for a variety of purposes," [12]. It includes text-only, audio-only, and video technologies for communication that can be either synchronous or asynchronous. Such technologies are the example of single-party communication, which have different types of pros and cons e.g., text-only technology helps in reducing the pressure of communication by lowering the CA of the users [13–15] but it is very limited in directing pressure of communication towards others because of providing very low social presence [16,17]. On the other hand, audio-only technology fails in reducing the pressure of communication of users [18,19]; as it has audio privacy concerns [20]. However, compared to text-only technology, it is relatively batter in directing pressure of communication towards others because of providing low social presence [16,21]. Similarly, video technology also fails in reducing the pressure of communication of users [22–25]; as it has audio-visual privacy concerns [26–28]. However, compared to text and audio technologies, it is relatively better in directing pressure of communication towards others because of providing social presence [29,30]. Despite of having different pros and cons, CMC technologies can't provide option of giving commitment; because of having no facility of offering alternative representative(s) to users. However, commitment can be taken from others while using CMC (in case of video only) because of having facility of presenting the user itself to others. On the other hand, users can avoid pressure of communication when using the robot technology; as they can give commitment to robot avatar. However, directing pressure towards others, and getting their commitment is difficult. In short, CMC and robot avatar technologies are examples of single party communication where handling of conversational factors e.g., commitment and pressure for users is difficult; please see figure 1.1 representing the research map with axes as conversational factors.

An alternative to single party communication is multi-party communication, which involves more than two peers in conversation. Since multi-party communication involves more than two peers, the attentional focus of each of the peer is expected to be divided in between other peers of communication. To verify such an intuitional fact, a multi-party communication scenario for people with CA was setup; where it was confirmed that twin robot avatars can alternatively capture the user's attention [31]. Such a confirmation of intuitional fact was a motivation for us to start exploring the possible benefits of using multi-party communication scenario. Some benefits of using multi-party communication are reported in previous literature, e.g., 1) it can be passive social in nature, where a user with CA does not need to respond to each stimulus [32,33]; i.e., avoiding pressure of communication for own self and giving commitment to communication via representations, 2) it provides ease to openly disagree with the opinion(s) of other(s) [34]; i.e., avoiding pressure of communication from others, 3) it provides opportunities of getting praise in front of others [35–37] ; i.e., directing pressure of communication towards others and getting their commitment in communication, and 4) it increases chances of having robust communication with others [38–40]; i.e., getting their commitment in communication. The confirmation of intuitional fact, and the inline support from the previous literature provides sufficient motivation to explore the expected benefits of multi-part communication for the users; commonly referred to as the operators.

Therefore, I propose the usage of multi-party communication system i.e., one having multiple representatives for the operators. Such representatives of the operators are referred to as their avatars. The proposal enables us to expand the axes of conversational factors in research map; see figure 1.2. Assume a situation where an operator is involved in an online discission with a visitor while using two avatars, see figure 1.3. The operator can see the remote environment through the video stream available on the monitor screen on his/her side, where a visitor and second avatar are visible to him/her. The operator can control both avatars; one through his/her own live stream and other through the tablet. Such a system enables operator to utter from either of the avatars and conversate with the visitor. Consequently, the visitor also has two options to direct his/her attention focus i.e., towards either of the avatar. Such a situation is expected to provide support to the operator in controlling the conversational factors, i.e., commitment and pressure of communication. To reduce own commitment, and pressure of communication, an operator can utter from the second avatar and shift the attentional focus of visitor towards it. Similarly, to get the visitor's commitment and direct the conversational pressure towards him/her, an operator can alternatively utter from both avatars. To verify such effects a series of studies are conducted. The detailed explanations of each study are provided in the relevant chapters.

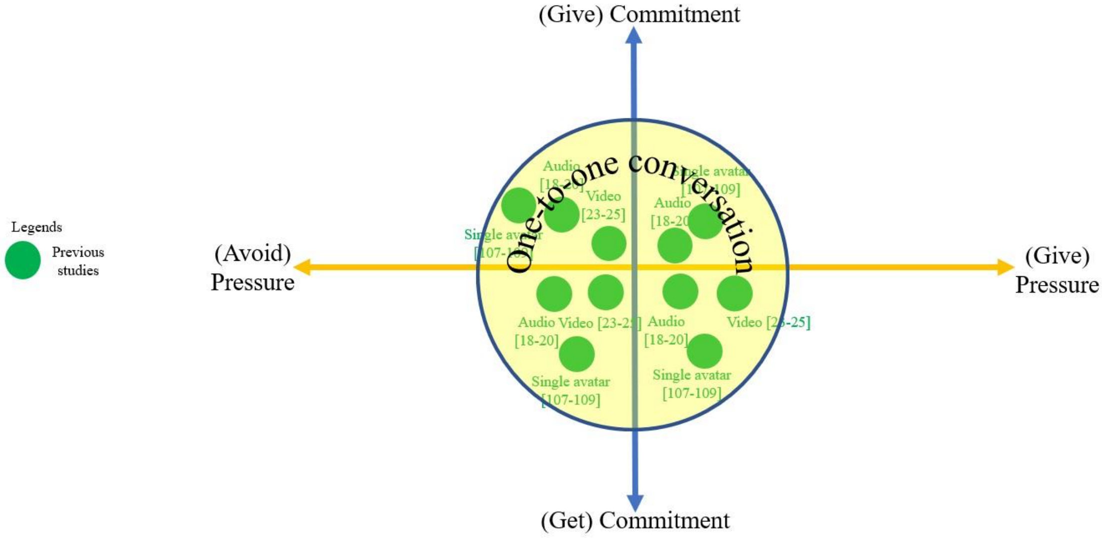

Figure 1.1: Research map; communicational factors and single-party communication challenges

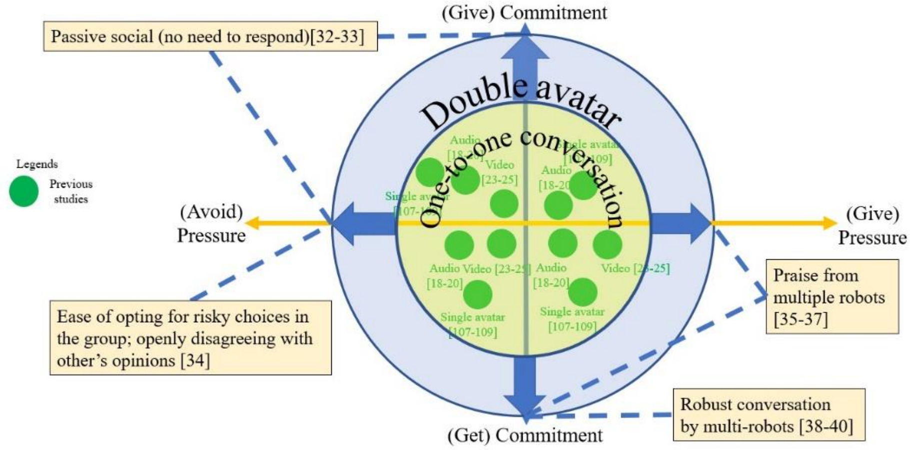

Figure 1.2: Expanded research map; communicational factors and multi-party communication

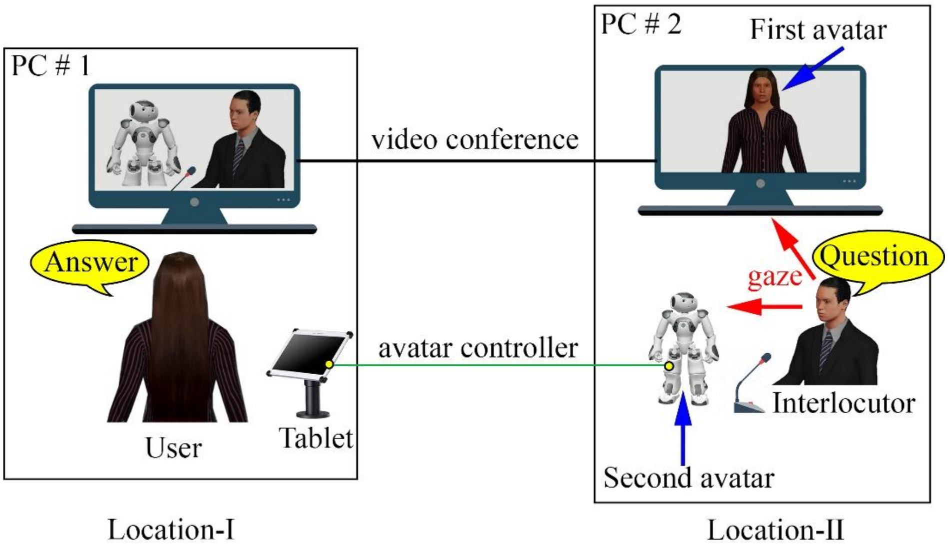

Figure 1.3: proposal, double avatar system for the operator

Chapter 2

Attentional behaviors and temporal delays of children with ASD (visitors) in response to different social cues of robotic agents.

Aim

The study is a(n) motivational/intuitional study which revealed the usefulness of multi-party communication system concerning the division of the attentional focus of visitor and motivates us to explore the advantages of multi-party communication system for users of other end, i.e., the operators which is a core/major focus of our thesis. It is an affective parameter to obtain the commitment from visitor towards either of robot agents in communication.

Abstract

Chapter 02 discusses the attentional behaviors of people with a high level of communication difficulties; while interacting with two robotic agents using three different types of stimuli i.e., visual, speech, and motion. The attention appealing strength of each type of stimuli was assessed by two indexes i.e., 1) latency in shifting attention towards robotic agents producing stimuli, and 2) the number of attentions paid. The speech stimulus showed significantly higher appealing strength for attention compared to visual and motion stimuli in terms of low latency in shifting attention and number of attentions paid. The study contributes to the literature in terms of impact of type of stimuli on the attentional behavior of people with high level of communication difficulties while interacting with two robotic agents.

2.1 Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that influences social skills of children/adolescents with ASD [41]. Verbal and non-verbal communication difficulties are also included among such affected skills. To provide effective curing for autism, its detection at earlier stage is required and several types of strategies, mechanism, and indices are considered so far to achieve such goal [42–44]. Matrices to measure attentional and communicational skills for children/adolescent with ASD can help psychologists in proposing remedy by assessing the quality of such skills. Responses towards positioning, eye contact, waving, smiles, imitation, and calling out their names are examples of such metrics [45]. Further, decreased visual attention to dynamic stimuli is another matric to identify ASD [46]. Apart from matrices, some ASD screening tools are also helpful e.g., Childhood Autism Rating Score (CARS) [47,48] Autism Diagnostic Observation Schedule (ADOS) [49,50], and the Mullen Scales of Early Learning (MSEL) [51,52].

Joint attention is a social skill; an ability of an individual to share the attention on an object with another individual. And ASD affects such skills in children with ASD [53]. In clinical applications, the behavior of the human is influenced by the attention to and/or from the robot [54]. Robots are helpful in increasing attentional skills in children with ASD that eventually modulates their verbal initiations, frequency, and duration of eye contact [55]. On the other hand, type of intervention provided by human caregiver can affect the children with ASD differently [56]. In case of people with dementia, robots are found effective in promoting positive behavior [57] and improving corresponding indicators [58]; while improving the attention of patients with dementia [59]. Using the humans agent, improvement of weak joint attention of children with ASD is examined [60]. If the human agent is trained, then joint attention between the humans agent and ASD child can be increased [61].

Other than human agents, some technologies e.g., robotic agents [62], mobile applications [63], and virtual reality [64] are used to improve the joint attention of the ASD children. The appearance of the robotic agents is an important factor for children with ASD [65]; where the robots with movable eyes have good consequences compare to those who do not [66]. The problem of having weak joint attention is not limited to static scenes but it also exists in dynamic scene [67]. It deteriorates the verbal communication skills of ASD children [68]. The patterns of joint attention of ASD children are unique [69], and such patterns are affected by the type of interacting partner (robot/human) and type of stimuli (e.g., gaze orienting, pointing, vocal instructions, and their combinations) being used [70,71]. A varying level of joint attention is observed in children with ASD when interacting with a single robot agent [72].

Children with ASD have difficulties in orienting the attentional focus from one location to another [73]. Such difficulties causes higher latencies in orienting behaviors of ASD then the TD [74]. In case of interacting with two robotic partners, how such orienting behaviors of ASD children change with respect to type of stimuli? Which modalities are useful in gaining the attention quickly? And how quickly such attentions can be captured while interacting with two robotic partners? Percentage accuracy and duration of eye contact were used to answer the first question but in case of having single interaction partner, i.e., the robot [72]. In case of having multiple robotic interaction partners, such questions are yet required to be answered. So, this research studies the attentional behavior of children with ASD when two interactional partners are involved. The attentional behavior is assessed by using two indexes; latency in orienting the attention, and number of attentions paid.

2.2 Multi-party communication system

Two humanoid robots named NAO were used in the experiment which were teleoperated by the researcher, see figure 2.1. Social stimulus module, attention latency measurement module, and communication modules are run on each of the humanoid robot. Social stimulus module controls the type of stimuli being run on each of the robot while attention latency measurement module records the ASD children’s latency in paying attention; where latency is measured by noticing the time difference i.e., the time at which the stimulus is presented ($t_{Stimulus}$) and the time at which the attention is paid by the ASD children ($t_{Attention}$). Communication module is used to establish real time connections between the robots, and computer.

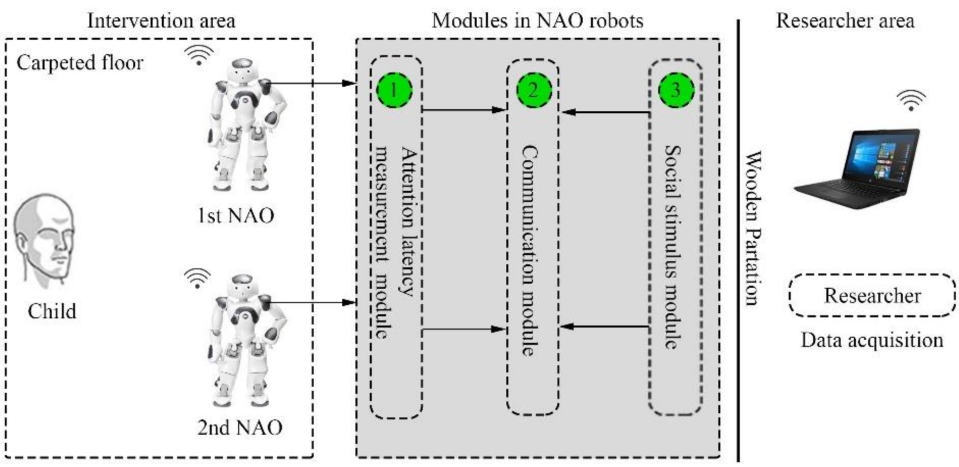

Figure 2.1: Multi-party communication system to study the attentional behaviors

2.3 Materials and method

2.3.1 Method

The participants were required to sit in front of the robotic partners and pay the attention towards the robotic agent delivering the specified stimulus. During the experiment, the log files recorded $t_{Stimulus}$ and $t_{Attention}$. The experiment was conducted for two months and each of the participant was required to attend a total of eight sessions; i.e., one session per week.

2.3.2 Participants

A total of nine ASD children participated in the experiment (age=7.57 years); including two females and seven males. The participants were diagnosed as the individuals having minimal to no ASD symptoms using CARS scale. The study was approved by ethical committee of partner university and autism resource centre.

2.3.3 Stimuli

Three types of social stimuli were used i.e., visual, speech, and motion. In case of visual stimuli, the robot agents change the color of the eyes in cyclic manner. In speech stimulus, robot agents greet by saying "*Hello! nice to meet you*", and in motion stimulus, it waves using the right arm. A total of 120 stimuli were provided in a session in which each stimulus was presented forty times. I assumed that visual stimulus is least effective, motion is most effective, and speech is moderately effective; whereby effectiveness, I mean the attention appealing strength of a stimulus. The sequence of execution of stimuli is presented in figure 2.2.

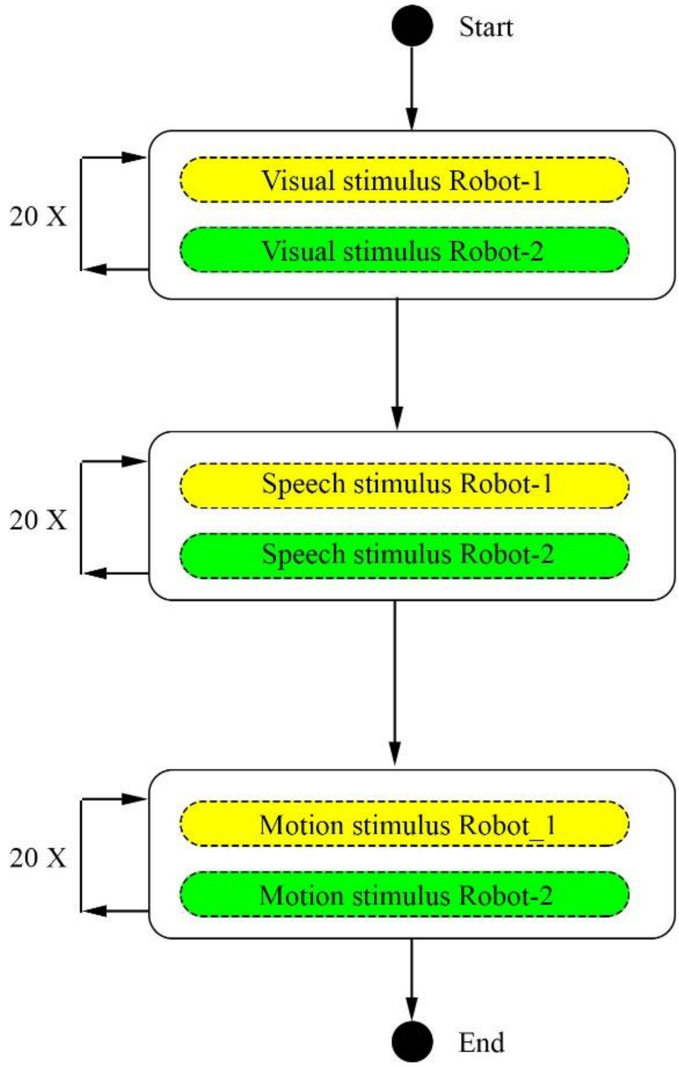

Figure 2.2: The sequence of presenting the stimuli

2.3.4 Procedure

The participants were brought to the experimental room by a therapist and asked to sit in a chair present in front of the robot. The researcher teleoperated the robot agents which were presenting stimuli in an alternative manner, see figure 2.3. The participants were free to leave the experiment at any time; in case they feel uneasy.

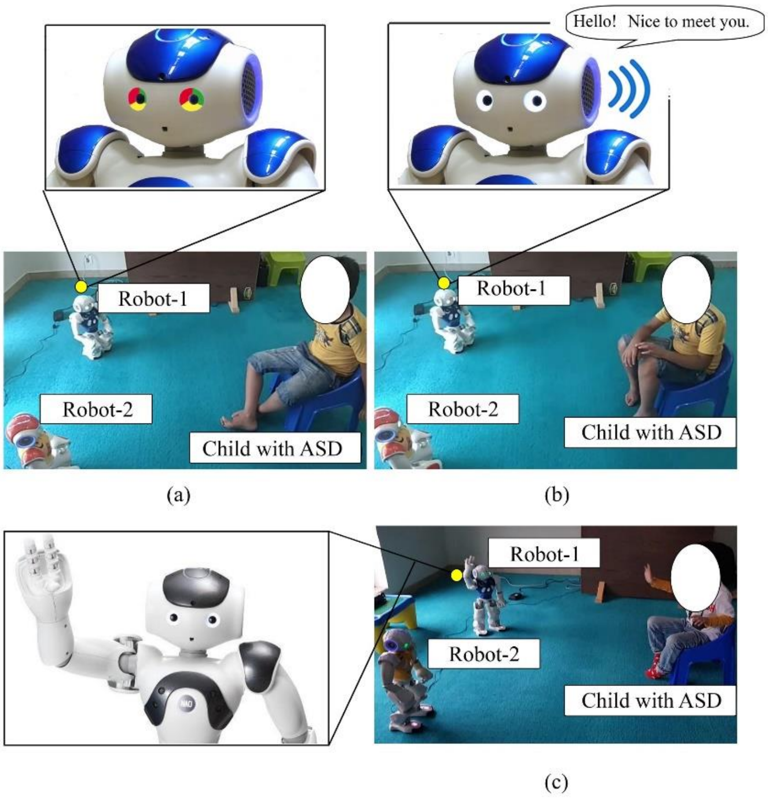

Figure 2.3: Multi-party communication system; (a) visual stimulus, (b) speech stimulus, and (c) motion stimulus

2.3.5 Measurements

I used two indexes to explore the attentional behaviors of the children with ASD i.e., the latency in paying attention towards a robotic partner presenting stimulus, and number of accepted social stimuli.

2.4 Results

2.4.1 Latency in shifting attention

Friedman test was conducted to identify the effect of type of social stimuli on the latency of in orienting the attention towards the robotic partner presenting the stimulus. There was a significant effect of type of stimulus on the latency of orienting the attention; $X^2(2) = 9.55$, p=0.008. The average latency values associated with visual, speech, and motion stimulus are 3.44 seconds, 3.27 seconds, and 3.73 seconds respectively, see figure 2.4. The post hoc analyses with Bonferroni correction are presented in table I.

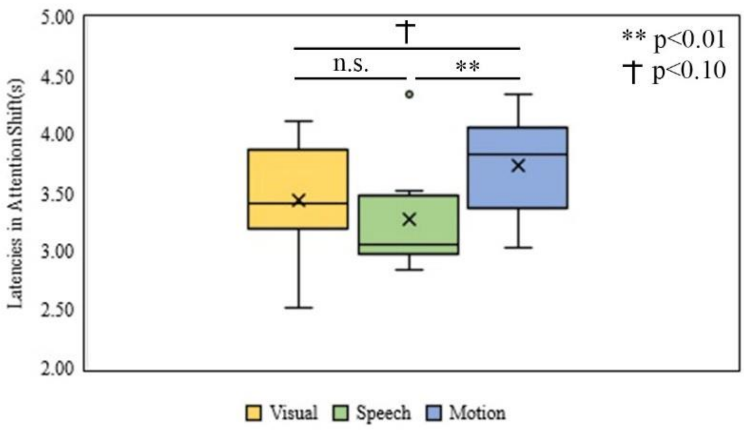

Figure 2.4: Latency in shifting attention to robotic partner

Table I: *Post hoc* Wilcoxon signed rank tests for latency in shifting attention					
Sr.	Stimulus I	Stimulus II	Z-value	p-value (2 tailed)	*r* value
1	visual (*Mdn* = 3.41)	speech (*Mdn* = 3.05)	−1.244	$0.214^{n.s.}$	−0.29
2	speech (*Mdn* = 3.05)	motion (*Mdn* = 3.84)	−2.666	0.008^{**}	−0.62
3	visual (*Mdn* = 3.41)	motion (*Mdn* = 3.84)	−1.897	$0.058^{\dagger}$	−0.45

2.4.2 Number of accepted social stimulus.

Friedman test was conducted to identify the effect of type of social stimuli on the number of accepted social stimuli from the robotic partner. There was a significant effect of type of stimulus on the latency of orienting the attention; $X^2(2) = 9.56$, p=0.008. The average number of accepted stimulus values associated with visual, speech, and motion stimulus are 11.10 times, 17.53 times, and 14.32 times respectively, see figure 2.5. The post hoc analyses with Bonferroni correction are presented in table II.

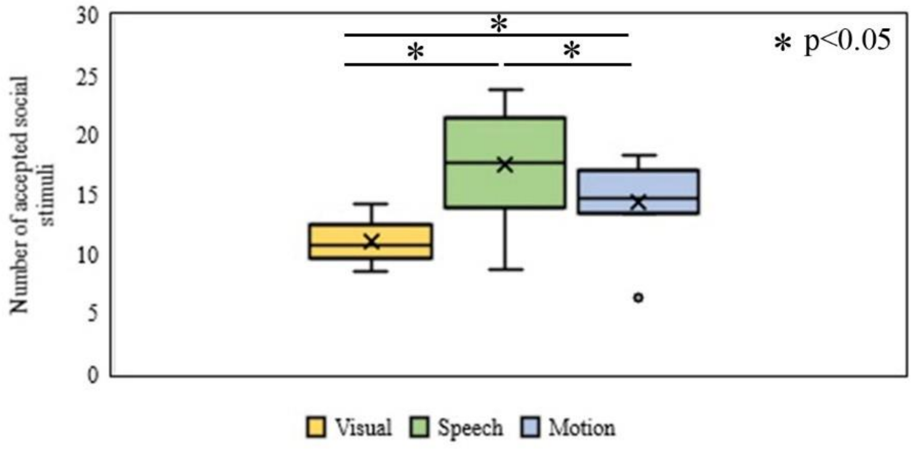

Figure 2.5: The number of accepted stimuli from robotic partners

Table II: *Post hoc* Wilcoxon signed rank tests for number of accepted stimuli					
Sr.	Stimulus I	Stimulus II	Z-value	p-value (2 tailed)	*r* value
1	visual (*Mdn* = 10.75)	speech (*Mdn* = 17.63)	−2.547	0.011*	−0.60
2	speech (*Mdn* = 17.63)	motion (*Mdn* = 14.75)	−2.429	0.015*	−0.57
3	visual (*Mdn* = 10.75)	motion (*Mdn* = 14.75)	−2.192	0.028*	−0.52

2.5 Discussion

In the current study I observed the effect of type of stimuli of robotic partners on the latency in orienting the attention towards robotic partners, and on number of accepted stimuli. In other words, I observed the attention appealing strength of stimuli of robotic partners. These types of robotic stimuli were visual, speech, and motion and they were presented in least-to-most (LTM) effective order. The results confirmed that attentional behaviors of the children with ASD depends upon the type of the stimuli being used. The speech stimulus is found most effect in terms of having lowest latency in capturing the attention and higher number of acceptances compared to visual and motion stimuli.

The main reason of finding the effectiveness of speech stimulus is that ASD children prefer to interact with speech generative devices in comparison to picture exchange (motion), and manual signs (visual) modalities [75]. Such a reason indicates their preference for speech stimulus based communication. The main reason of capturing the attention slowly for motion stimulus could be the abnormal motion perception in ASD children [76]. Similarly, the main cause of having least effectiveness for visual stimulus could be the 1) the presence of sensory issues in children with ASD [77]; which makes them sensitive towards stimuli other than visual, and 2) the unfamiliarity with visual stimuli of the robot; as they are not being human like.

Although differences in the attentional behaviors of the ASD children concerning the type of stimuli were observed. However, there are some limitations. The speech stimulus used 100% volume, and visual stimulus used 100% intensity of LEDs. The motion stimulus was completed in 5 seconds. Moreover, the order of the presentation of stimuli remained same for all of subjects. The change in such parameters may bring the change in attention behaviors of the children with ASD. The concluded results are also limited by small number of samples as its very difficult to get a huge number of sample of children with ASD. It would be worthwhile to examine the effect of combined social stimuli on the latency and orienting the attention for ASD children.

2.6 Conclusion and future work.

In current study, the attentional behaviors of ASD children in a multi-party communication scenario were observed. The effect of type of stimuli on the attentional behavior was confirmed. To assess the attentional behaviors, I observed the latencies in orienting the attention towards a robotic agent and number of accepted stimuli. A total of nine participants, including two females, and seven males, took part in the study. They were diagnosed as ASD children with minimal to no symptoms of autism. The study lasts for two months where each participant had to participate eight times; once per week. Significant effects of type of stimuli were found on the latencies and number of accepted stimuli. Post hoc analyses revealed that speech stimulus is the most effect one in capturing the attention of the ASD children in terms of having lowest latency and highest number of acceptances compared to visual and motion. The study contributes to the literature in terms of showing the possible effect of adopting multi-party communication scenario.

Chapter 3

Communication apprehension and eye contact anxiety in video conference involving teleoperated robot avatar: a subjective evaluation study.

Aim

The study focusses on the avoiding the pressure of communication and getting the commitment for operators; see figure 3.1. The pressure in the communication is avoided by reducing the communication apprehension, and eye contact anxiety. While commitment in communication is obtained by diverting the visitor's attention towards own self while explaining the reasons to him/her for choosing a particular yes/no option as answer to his/her question.

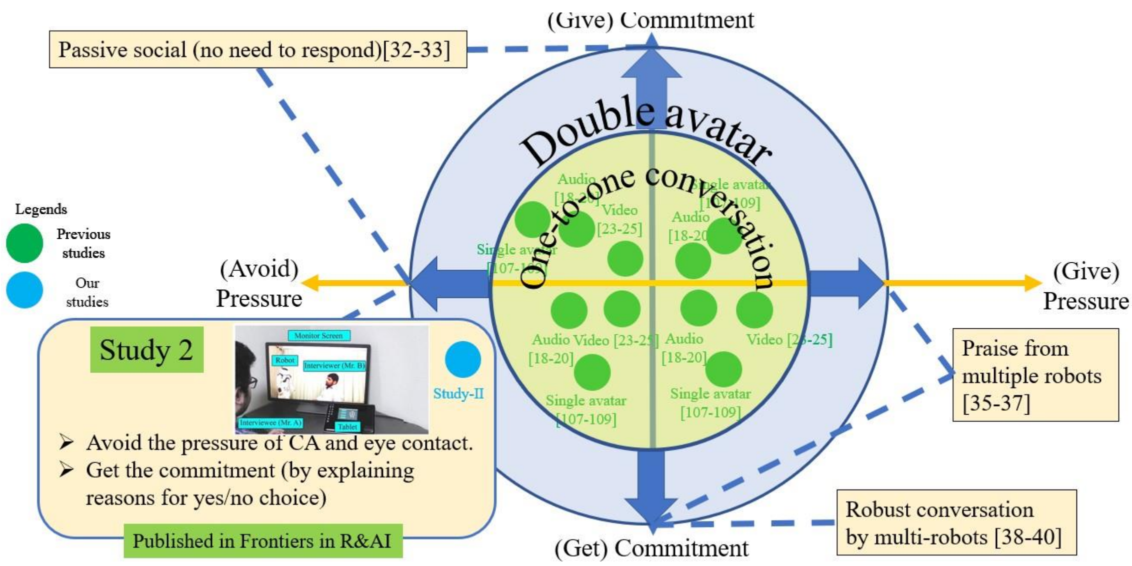

Figure 3.1: study II location of research map

Abstract

Chapter 03 discusses whether using a teleoperated robot avatar in a video conference can provide effective communication support to people with communication apprehension (CA) and anxiety in eye contact (AEC); where the effectiveness of communication support was assessed by CA, AEC, sense of being attending (SoBA) and intention to use (ITU) indexes. Two imagination-based surveys were conducted where recruited participants watched video stimuli with and without the proposed system. Later, they were asked to imagine themselves as an interviewee (a character of video stimuli) and rate their impressions. A significant decrease in expected CA and AEC was observed in both experiments while a significant increase in SoBA was observed in the second experiment. This study contributes to the literature in terms of the impact of using teleoperated robot avatar on the CA and AEC of the individuals.

3.1 Introduction

Communication apprehension (CA) is "an individual's fear or anxiety associated with either real or anticipated communication with another person or persons" [78]. The CA of an individual influences the quality of communication in face-to-face [2,4], and online interactions [5,6]. Therefore, the effectiveness of communication is compromised [11], and he/she is perceived as a less positive communication partner by others [10]. On the other hand, an individual's fear or feeling of discomfort while being stared at by others is referred to as anxiety in eye contact (AEC) [79]. The major cause of the generation of AEC in an individual is his/her social anxiety [80], which not only reduces the frequency and duration of eye contact [81] but also influences face-to-face (FTF) [82,83] and online communications [84].

As an alternate to FTF interactions, few audio, and text-only technologies are present e.g., cell phones, social websites, text messages [13], audio calls, voice, and electronic mail [85], and computer mediate communications [86]. These technologies have the potential to moderate the social anxiety of an individual [14] and preferences for them as also been observed [87]. However, these technologies eliminate the eye contact opportunities completely, hence making communication non-vivid only. Apart from this, such technologies also reduce the social presence of their users [88,89], where the social presence can be defined as the perception of an individual's presence in a communication event [90]. Reduced social presence causes failure in keeping the sense of being attending (SoBA) of an individual; where SoBA can be defined as "*the feeling of an individual when he/she is listened to, given attention, focused, or questioned/answered by others in conversations*" [91].

Another famous alternative to FTF interactions is video conference technology which not only reduces the CA and AEC of users [92–94] but also keeps their social presence [29,30,95]. People prefer to use video technology because it provides verbal and non-verbal information about the interactees e.g., details of attentional focus of remote partner of communication [96], which contributes to mutual understanding among interactees [97]. However, video conference also provides unnecessary eye contact opportunities which generate anxieties [98], fear-relevant features [99], gaze avoidance behaviors [100], and interrupted dialogues among partners of communication [101]. One simple solution to reduce the AEC problem in a video conference is to instruct the interlocutor to avert his/her gaze while interacting but the anxiety of participant cannot be regulated by simply averting the gaze [102], however, averting gaze will reduce the social presence [103], which is not the favorable tradeoff.

Avatars are "*an interactive and social representation of a user*" [104] or representation of a user in a given physical medium for experiencing the physical environment [105]. Robot avatar assists in online communication situations, e.g., education [106,107], virtual tours [108], and family communications [109]. Physical robot avatars can hide the identity of the users [110,111], which is expected to contribute to the reduction of CA and AEC. Further, they also provide an enriched social presence to their users [112,113] which is a key element to make communication successful. Considering such advantages of physical robot avatars, if the robot avatar is placed beside the interlocutor in a video conference, CA and AEC of the user will be reduced while simultaneously increasing his/her SoBA. Reduction in AEC is expected because of the diverted attentional focus of the interlocutor which will be towards the robot avatar instead of the user of a robot. Similarly, a reduction in CA is expected because of the availability of robots as an alternative communication channel. While, on the other hand, an increase in SoBA is expected because of the existence of the robot avatar in the direction of the attentional focus of the interlocutor, which is expected to be felt like an avatar of him/herself by the user.

In this chapter, a robotic system integrated with video conferences is proposed to support the people with CA and AEC in the conversation in telecommunication. Assume a situation where an interlocutor and a robot avatar are present in front of each other. The interlocutor is communicating through video conference with a user (i.e., a person with CA and AEC) who can control the robot avatar present at the interlocutor's side. In such an arrangement of the system, the user has two ways to utter: i.e., utter by him/herself through video conference or utter through robot avatar present in front of the interlocutor. Therefore, the interlocutor will also be having two options for directing his/her attention;

either towards the robot avatar or monitor screen showing the user’s profile. Such situations are expected to reduce the CA and AEC of the user as he/she will have an alternative medium of communication (robot avatar) alongside decreased pressure of being focused by the interlocutor. In parallel, his/her SoBA is also expected to be kept high because he/she is expected to feel the robot as an avatar of him/herself. To verify such effects, two different video evaluation-based experiments were conducted where recruited participants watched the video scenes of telecommunication with and without the proposed system. after watching the video stimuli, they were asked to imagine themselves as a user of robot avatar in the video stimuli and evaluate the expected feelings of CA, AEC, and SoBA. In experiment I, the video of the proposed method was compared with the one including ordinary online conversation, where the attentional focus of the interlocutor is kept directed towards the profile of the user, to show the positive effect of the proposed method in terms of expected feelings of CA and AEC. In experiment-II, it was compared with the one including ordinary online conversation, where the attentional focus of the interlocutor is kept away from the profile of the user, to show the positive effect of the proposed method in terms of expected feelings of SoBA.

3.2 Robotic video conferencing system for providing effective communication support for people with CA and AEC.

The proposed robotic video conferencing system consists of a desktop computer, a tablet, and a humanoid robot, see Fig 3.2. Using a desktop computer, an online discussion session was arranged for the interviewer and interviewee physically present at different locations; Room-1 and Room-2. The robot was present in front of the interviewer in Room-1 which was controllable through a tablet present beside the interviewee in Room-2. The interviewee was able to see the environment of Room-1 through commercial software for an online video conference where the teleoperated robot and the interviewer were present. The humanoid robot used was NAO, which is bipedal, 58 cm tall with 25 degrees of freedom, fully programmable, and capable of interacting with people through visual, speech, and motion stimuli. Throughout the discussion session, the robot remained in a standing position with idling movements; it gently keeps moving its body to and for in a horizontal direction without changing the position of its feet on the table. It also keeps on changing attentional focus between interviewer and interviewee alternatively by turning its head. It looks on the camera on the monitor of Room-1 which is perceived by the interviewee in Room-2 as it is looking at him/her. The GUI on the tablet consists of four buttons: “yes”, “no”, “I do not know”, and “exit”. The server-client architecture of the transmission control protocol (TCP) was used to exchange the commends between the tablet and teleoperated robot. As soon as the button on the table is pressed by the interviewee, the robot stops its idling movements, turns its head to the camera of the monitor, nod twice, turns its head back to the interviewer, and utter the corresponding answer. The possible utterances were “yes, I think I will go for that”, “no, I think I will not go for that”, and “I do not know” corresponding to “yes”, “no”, and “I do not know” buttons respectively. The “exit” button was for terminating the discussion session, but it was not used in this experiment. During the discussion session with the availability of such a system for the interviewee, the interviewer asked a yes/no and in-depth question while focusing on the robot. The interviewee was required to answer the yes/no question through the robot by pushing the button on the tablet, and in-dept questions in his/her own voice. The yes/no question was deliberately asked first as asking a yes/no question is expected to be an easier step for the interviewee to reveal his/her point of view, especially when it is difficult for him to answer concisely what he/she is thinking.

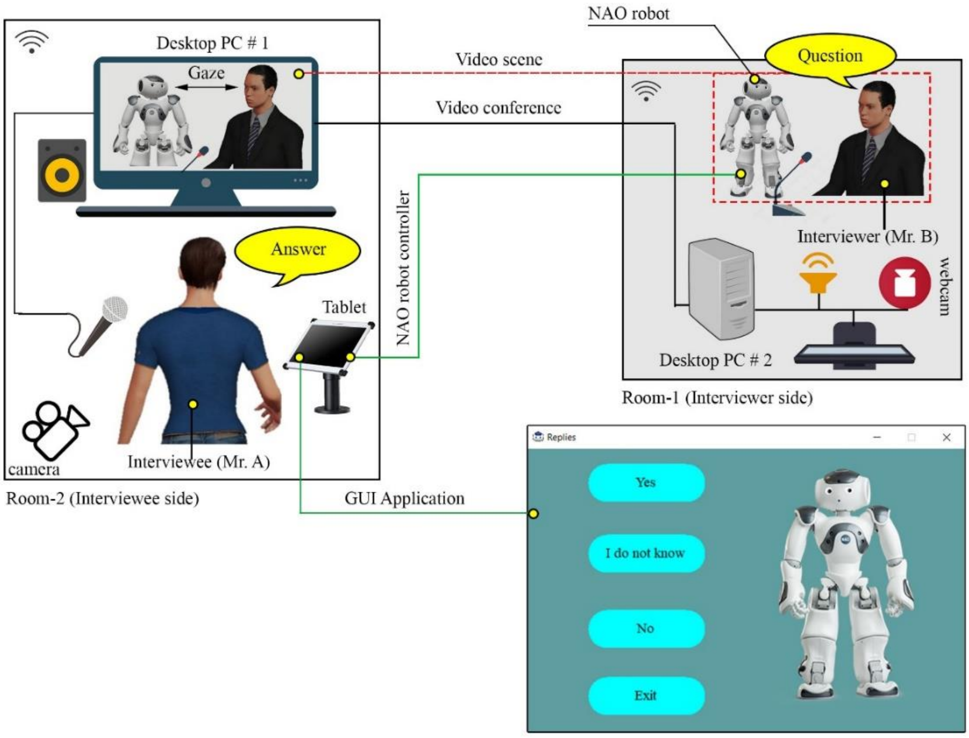

Figure 3.2: Robotic video conferencing system for providing effective communication support to people with CA and AEC (Robotic condition).

3.3 Experiment I

3.3.1 Method

The impression for the conversation using the proposed system (hereafter Robot condition (see Fig. 3.2)) from the interviewee's point of view was compared to one without it (hereafter Human condition (see Fig. 3.3)). In Human condition, the interviewer kept directed towards the interviewee so that the interviewee perceives him/her directed towards him/herself. In this experiment, instead of inviting the participants to experience the system, an imagination-based survey was conducted where recruited participants were asked to watch the video stimuli corresponding to Robot and Human conditions and later rate their impressions by considering themselves as an interviewee (a character of video stimuli). Type of the condition (Human vs Robot was an independent variable while CA, AEC, SoBA, and ITU were the dependent variables.

Figure 3.3: Traditional video conference system of Experiment I, (Human Condition)

3.3.2 Participants

A total of 200 participants (*M*=32.73, *SD*=8.96 years) were recruited from the internet, including 158 males and 42 females, having no serious issues with CA, and AEC. They were divided into two groups G1 and G2, depending on their day of birth (*even*=113, *odd*=87).

3.3.3 Apparatus

A web browser interface was used by the participants for watching the video stimuli of both conditions and for answering the questionnaire too.

3.3.4 Stimuli

In both conditions, conversations between experimenters related to topics of earning unfair money and paying taxes were recorded. In Human condition, the ordinary video conference system namely Zoom (Zoom video communications Inc. 2011) was used, where the gaze of the interviewer was towards the monitor having web camera. In such a situation, the interviewee in Room-2 would perceive the interviewer's gaze was directed to him (see figure 3.4 (A)). While, in Robot condition, the gaze of the interviewer was directed at the robot throughout the conversation except when interviewer looked at interviewee for inviting him to answer the in-depth questions (see figure 3.4 (B)). The durations of the video stimuli were 38 seconds for Human condition and 51 seconds for Robot condition respectively. The duration of video stimulus for Robot condition was larger compared to Human condition because of robot's delay to utter yes/no answers. however, the sequence of the utterances in both video stimuli remained constant. The interviewer asked a yes/no question followed by an in-depth question. The questions of interviewer and corresponding answers of interviewee are given in appendix.

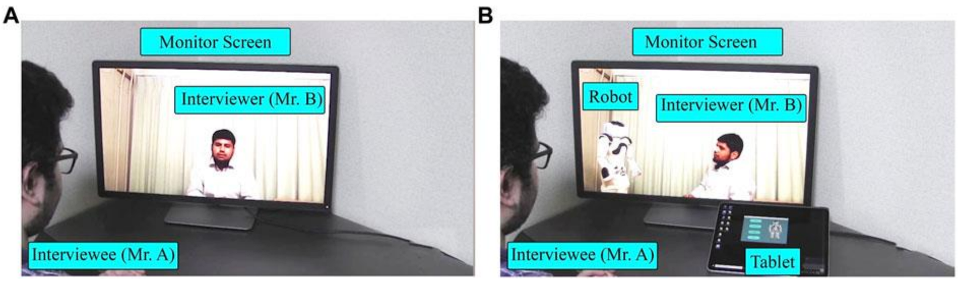

Figure 3.4: Experiment-Ⅰ: Human condition (A) and Robot condition (B).

3.3.5 Manipulation check

In order to verify that whether the participants carefully watched and understood the content of both video stimuli of both condition, two manipulation checks were performed. The data of the participants who verified the criterion were used for further analysis.

3.3.6 Procedure

The participants were required to complete an online survey form comprises six parts. In Part-I, each participant was required to carefully read and agree to the content of a web-based informed consent. Later, each of them was required to provide some personal details e.g., age, gender, and daily life Ca and AEC in parts II and III respectively. The information about daily life CA (*M*=16.85, *SD*=4.57) and AEC (*M*=44.18, *SD*=25.15) was obtained to see the serious issues in participants; if any. The participants of the group G1 watched the Human condition in Part-IV (see figure 3.4 A) and Robot condition in Part-V (see figure 3.4 B). Immediately after watching stimulus for each condition, they were required to imagine and rate their perceived CA, AEC, and SoBA. On the other hand, the order was reversed for the participants of the group G2. Finally, the participants were asked to tell about their preference for using Human and Robot conditions when their interlocutor would be their boss, teacher, doctor, psychologist, or stranger.

3.3.7 Measurements

3.3.7.1 Expected communication apprehension

The responses of the participants to CA questionnaire were recorded three times in web-based survey i.e., Parts III, IV, and V, using interpersonal sub-score of personal report of communication apprehension-24 (PRCA-24) [114]. A 1-5 Likert-type point scale was used (strongly disagree, disagree, neither agree not disagree, agree, and strongly agree).

3.3.7.2 Expected anxiety and making/avoiding eye contact

The responses of the participants to AEC questionnaire were recorded three times in web-based survey i.e., Parts III, IV, and V, using gaze anxiety rating scale (GARS) [80]. A 0-3 Likert type point scale was used (none, mild, moderate, and sever); where the total score was obtained after summing the ratings.

3.3.7.3 Expected sense of being attended (SoBA)

A scale named SoBA that quantifies the feelings of an individual of being listened, focused upon, or questioned/answered by others in conversation; see appendix. The participants were asked to imagine their self as interviewer and rate how much SoBA they were expected to have. It was recorded two times in a web-based survey i.e., Parts IV and V, using a 1-5 Likert-type point scale. The internal consistency of the scale is reported in the results section.

3.3.7.4 Intention to use the system

The participants evaluated their intention to use the video conferencing system in Robot condition while responding to intention to use (ITU) questionnaire [115] at the end of the web-survey i.e., Part VI. ITU questionnaire has 1-5 Likert-type point scale.

3.3.7.5 Preference to use the system

The participants provided their preference to use the video conference system in Robot condition; where they considered the interlocutor their own boss, teacher, doctor, psychologist, or the stranger. Such preference was recorded by simply asking about their degree of agreement in using robotic video conference system in each situation on a 1-5 Likert type point scale.

3.3.8 Results

3.3.8.1 Expected communication apprehension

The Wilcoxon signed rank test was conducted to identify the effect of the type of the video conferencing system (Human vs Robot conditions) on the expected CA of the participant. It showed that mean rank of the expected CA of the participant in Robot condition (*Mdn*=16) was significantly less compared to mean rank of Human condition (*Mdn*=17), (n=200, Z=3.71, p=2.08×10^{-4}, r=0.18), Figure 3.5. The reported p-values are two tailed.

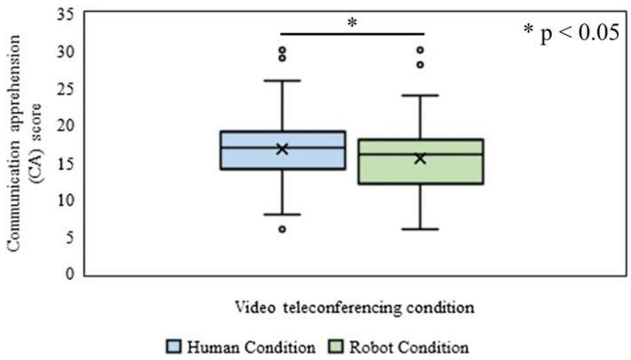

Figure 3.5: Communication apprehension (CA) score.

3.3.8.2 Expected anxiety in eye contact

The Wilcoxon signed rank test was conducted to identify the effect of the type of the video conferencing system (Human vs Robot conditions) on the expected AEC of the participant. It showed that mean rank of the expected AEC of the participant in Robot condition (*Mdn*=44) was significantly less compared to mean rank of Human condition (*Mdn*=49), (n=200, Z=3.37, p=7.27×10^{-4}, r=0.17), Figure 3.6.

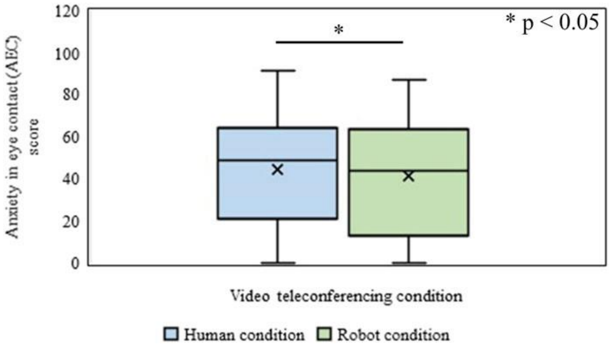

Figure 3.6: Anxiety in eye contact (AEC) score

3.3.8.3 Expected sense of being attended

The Wilcoxon signed rank test was conducted to identify the effect of the type of the video conferencing system (Human vs Robot conditions) on the expected SoBA of the participant. It showed that mean rank of the expected SoBA of the participant in Robot condition (*Mdn*=16.5) was not significantly different compared to mean rank of Human condition (*Mdn*=17), (n=200, Z=0.44, p=0.65, r=0.022), figure 3.7. The internal consistency of the SoBA scale was high (α=0.81).

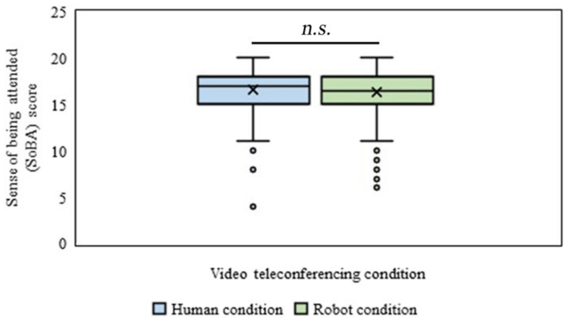

Figure 3.7: Sense of being attended (SoBA) score

3.3.8.4 Intention to use the system

The Wilcoxon signed rank test for single sample with *hypothesized Mdn*=3.0 (the center of the scale) showed a significantly higher tendency of the participants to use the Robot condition (*Mdn*=4.0), (n=200, Z=6.51, $p=7.36\times10^{-11}$, r=0.46).

3.3.8.5 Preference to use the system

The Friedman's test showed no significant effect of type of roles of interviewer on the preference of the interviewee to use the Robot condition; $\chi^2(4, n=200) = 9.44$, $p=0.051$.

3.4 Experiment II

3.4.1 Method

In Experiment I, positive effect of the proposed system on the CA and AEC of participants has been observed. Such positive effect could be caused by the averting gaze pattern of the interviewer in Robot condition; but it was not controlled in Experiment I. If averting gaze pattern of the interviewer is the possible cause of the reduction of CA and AEC of the interviewee, then such a simple strategy would be sufficient for reducing the CA and AEC of interviewee. However, reduction of the SoBA of the interviewee is a drawback associated with the usage of such a simple strategy.

Experiment-II was conducted with controlled direction of the gaze of the interviewer to verify that the observed effect was actually because of the proposed system (i.e., Robot condition) not because of the change in the direction of the gaze of the interviewer. In the new Human (averted) condition, the direction of the gaze of the interviewer was away from the monitor with web camera; to whom the interviewee perceived as the interviewer was looking away from him/her (see Figure 3.8, 3.9 (A)). The relative angle of the gaze of the interviewer was controlled in Human (averted) and Robot conditions (see Figure 3.9 (B)).

Figure 3.8: Traditional video conference system of Experiment-II.

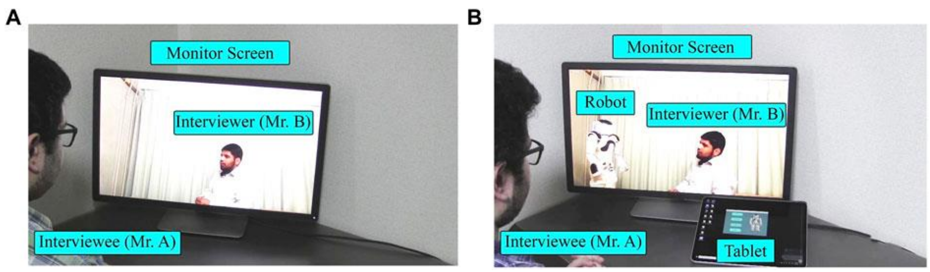

Figure 3.9: Experiment-II: Human condition (A) and Robot condition (B).

3.4.2 Participants

Another set of 200 participants (*M*=32.66, *SD*=9.29 years) was recruited from the internet: including 148 males and 52 females, having no serious CA (*M*=17, *SD*=3.63), and AEC (*M*=49.88, *SD*=24.23) issues. The participants were divided into groups G1 and G2, concerning to their day of birth (*even*=128, *odd*=72).

3.4.3 Apparatus

The participants were required to use the web-browser interface for watching the video stimuli of both conditions, and later they answered the questionnaires.

3.4.4 Stimuli

The content of conversation remained same in-between conditions of Experiment-I and II, however not the gaze pattern. The time duration was 39 seconds and 51 seconds for Human (averted) and Robot conditions, respectively.

3.4.5 Manipulation check

The manipulation checks used in Experiment-II were same as that of Experiment-I. The data of the participants, who passed the manipulation checks, were considered for further analysis.

3.4.6 Procedure

The procedure for Experiment-II was identical to that of Experiment-I. however, the video stimulus used for Human (averted) condition was different.

3.4.7 Measurements

The measurements used in Experiment-II were same as that of Experiment-I.

3.4.8 Results

3.4.8.1 Expected communication apprehension

The Wilcoxon signed rank test was conducted to identify the effect of the type of the video conferencing system (Human (averted) vs Robot conditions) on the expected CA of the participant. It showed that mean rank of the expected CA of the participant in Robot condition (Mdn=17) was significantly less compared to mean rank of Human (averted) condition (Mdn=17.5), (n=200, Z=3.38, p=7.2×10^{-4}, r=0.17), Figure 3.10.

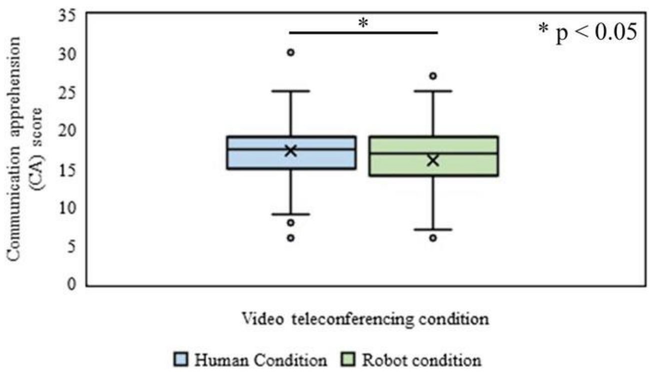

Figure 3.10: Communication apprehension (CA) score.

3.4.8.2 Expected anxiety in eye contact

The Wilcoxon signed rank test was conducted to identify the effect of the type of the video conferencing system (Human (averted) vs Robot conditions) on the expected AEC of the participant. It showed that mean rank of the expected AEC of the participant in Robot condition (*Mdn*=52) was significantly less compared to mean rank of Human (averted) condition (*Mdn*=53), (*n*=200, *Z*=2.04, *p*=0.040, *r*=0.10), Figure 3.11.

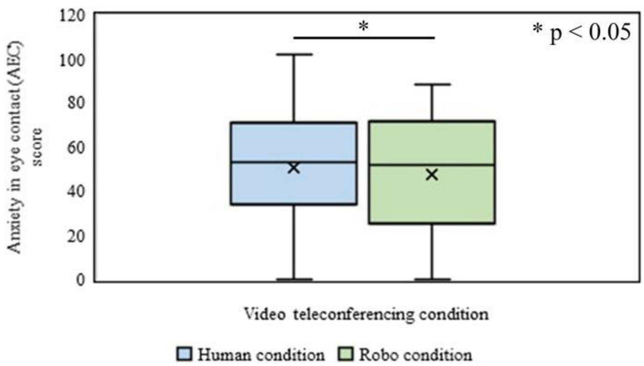

Figure 3.11: Anxiety in eye contact (AEC) score.

3.4.8.3 Expected sense of being attended

The Wilcoxon signed rank test was conducted to identify the effect of the type of the video conferencing system (Human (averted) vs Robot conditions) on the expected SoBA of the participant. It showed that mean rank of the expected SoBA of the participant in Robot condition (*Mdn*=17) was significantly higher compared to mean rank of Human (averted) condition (*Mdn*=16), (*n*=200, *Z*=2.39, *p*=0.016, *r*=0.12), Figure 3.12.

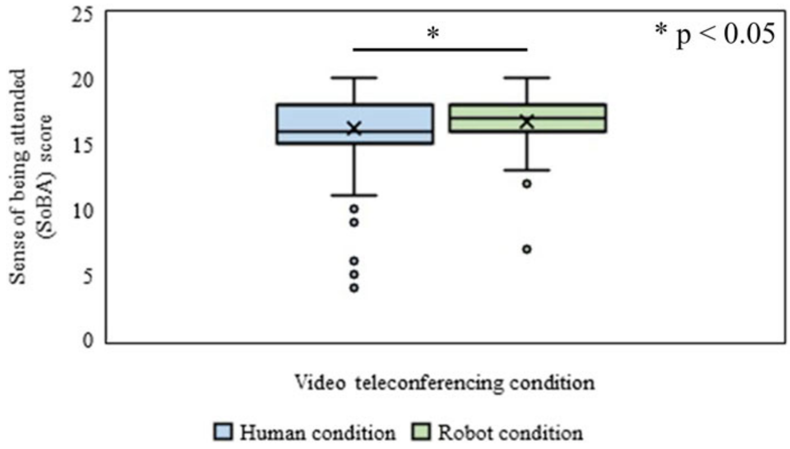

Figure 3.12: Sense of being attended (SoBA) score.

3.4.8.4 Intention to use the system

The Wilcoxon signed rank test for single sample with *hypothesized Mdn*=3 (the center of the scale) showed a significantly higher tendency of the participants to use the Robot condition (*Mdn*=4.0), (*n*=200, *Z*=8.18, $p=5.81\times10^{-18}$, *r*=0.58).

3.4.8.5 Preference to use the system

The Friedman's test showed a significant effect of type of roles of interviewer on the preference of the interviewee to use the Robot condition; $\chi^2(4, n=200) = 16.30$, *p*=0.003. Multiple Wilcoxon signed rank tests with Bonferroni correction revealed participants' significant preference for using the robot condition for communicating with boss (*Mdn*=4.0, *SE*=0.070) over doctor (*Mdn*=4.0, *SE*=0.073) (*n*=200, *Z*=-2.249, *p*=0.025, *r*=-0.11); teacher (*Mdn*=4.0, *SE*=0.081) over doctor (*Mdn*=4.0, *SE*=0.073) (*n*=200, *Z*=-2.708, *p*=0.007, *r*=-0.14); and teacher (*Mdn*=4.0, *SE*=0.081) over stranger (*Mdn*=4.0, *SE*=0.074) (*n*=200, *Z*=-2.220, *p*=0.026, *r*=-0.11).

3.5 Discussion

Significant reduction in participants' expected CA and AEC was observed in Robot condition of Experiment-I. However, it was not clear that whether the utilization of robot avatar or the averted gaze of the interviewer was actual cause of it. Experiment-II was conducted to verify such ambiguity where Robot condition was compared with human (averted gaze) condition. The results of Experiment-II clarified that usage of the robot avatar was the main cause of reduction of CA and AEC of participants, not the mere shift the attentional focus of interviewer. The mere shift of the attentional focus of interviewer interferes with the social presence of interviewee as it reduced the SoBA of the interviewee. Therefore, to provide communication support to interviewee by reducing the anxieties while keeping the SoBA, use of robot avatar is suggested.

In socially anxious people, perception of direct gaze generates fear relevant features [99]. In Robot condition of Experiment-II, the decrease in CA and AEC of participants could be explained by the fact that interviewer's attentional focus was directed to a different agent (i.e., robot) causing reduction in fear-relevant features. Conversely, shared gaze towards a specified area in a scene increases the engagement among the participants in online interaction [116]. In Robot condition of Experiment-II, both interviewer and interviewee had consistent opportunities to share theirs gazes in the scene while focusing on robotic agent. Moreover, the events observed through avatars are perceived as operator's own experiences [117], so when participants watch the eye contact between interviewer and their avatar, it would be perceived as their own direct eye contact with interviewer without apprehension, causing increased engagement in scene. Moreover, perception of averted gaze of interviewer activates interaction avoidance behaviors in observers [118]. However, the shared gaze and enhanced experience of eye contact through avatar would contribute to increased SoBA of the participants that further motivates them to actively communicate with interviewer.

There are some limitations of current study. The significant statistical differences do not necessarily mean significant improvement. The effects were observed in pre-recorded videos, so it's not necessarily guaranteed to be reproduced in real-world. The participants were asked to imagine themselves as interviewee of the video scene they watched, so the results were their imagination-based evaluations. However, the degree to which they could imagine themselves as the character in a scene was not controlled. Moreover, the recruited participants were not having sever CA and AEC issues. Therefore, to overcome the limitations, interactive experiments, with individuals having serious CA and AEC issues, using proposed system are required to observe the actual potential of the system in real world to draw more affirm conclusions. For the proposed system to be effective for real world usage, it is necessary to be accepted by not only the interviewees but also the interviewers as well: as interviewers would be the individuals who might suffer because of the CA and AEC difficulties of interviewee in conversation. For simplicity, I only focused on interviewee's side, however, it is equally worthwhile to have look into the possible expected effects at interviewer's side as well. Moreover, I also supposed that providing limited number of pre-defined yes/no answers is a supportive way for interviewees having CA and AEC to immediately respond to interviewer's questions. However, existence of such feature simultaneously limits the freedom of conversation. To provide conversation freedom to individuals with CA and AEC, it's worth examining to see the effect of the integration of automatic mechanism that helps them by predicting the next probable word for the input sentences like chat bot [119],

[120]. Apart from limitations, integration of the proposed system in the real world could also be challenging. In the beginning, it would not be easy to find the suitable candidates having issues of CA and AEC, and later to provide them proper training about how to use the system: as I need to decide about which interaction modality should be used to communicate with them for training. Further, at subsequent stages, it might be challenging for many individuals with CA and AEC to endure the cost of the deployment, and maintenance of the system along with several unforeseen technical and non-technical issues for which they will be completely relaying on the providers of the service(s).

3.6 Conclusion and future work.

In this study, the usage of teleoperated robot avatar in a video conference is proposed to provide communication support to people with CA and AEC. The recruited participants watched the videos of an interview scene with and without proposed system. and later provided their imagination-based evaluations while considering themselves as an interviewee. in the proposed system the interviewee had two options to provide answers: utterance by a robot avatar co-present with interviewer or utterance by self in won voice. the proposed system i.e., video conference integrated with teleoperated robot avatar was compared with two ordinary video conference systems: first where interviewer's gaze directed towards interviewee and second where interviewer's gaze diverted from interviewee. Positive effects of the proposed method were observed on the expected CA, AEC, and social presence of the interviewee. Present study contributes to literature in terms of examining the expected impact of using teleoperated robot avatar in video conference to provide communication support to people with CA and AEC. In the future, I will examine whether using a teleoperated robot avatar in video conference provides communication support to individuals with severe CA and AEC, with different cultural and linguistic backgrounds.

Chapter 4

Avatars-mediated video conference system for mediating the stress and anxiety of response time of a person with communication difficulties

Aim

The study focuses on avoiding the pressure of communication and giving commitment to communication; see figure 4.1. The pressure in communication is avoided by reducing the stress of managing response time in communication. While commitment in communication is given by diverting responsibility towards the supporter agent instead of the operator for replying late in communication.

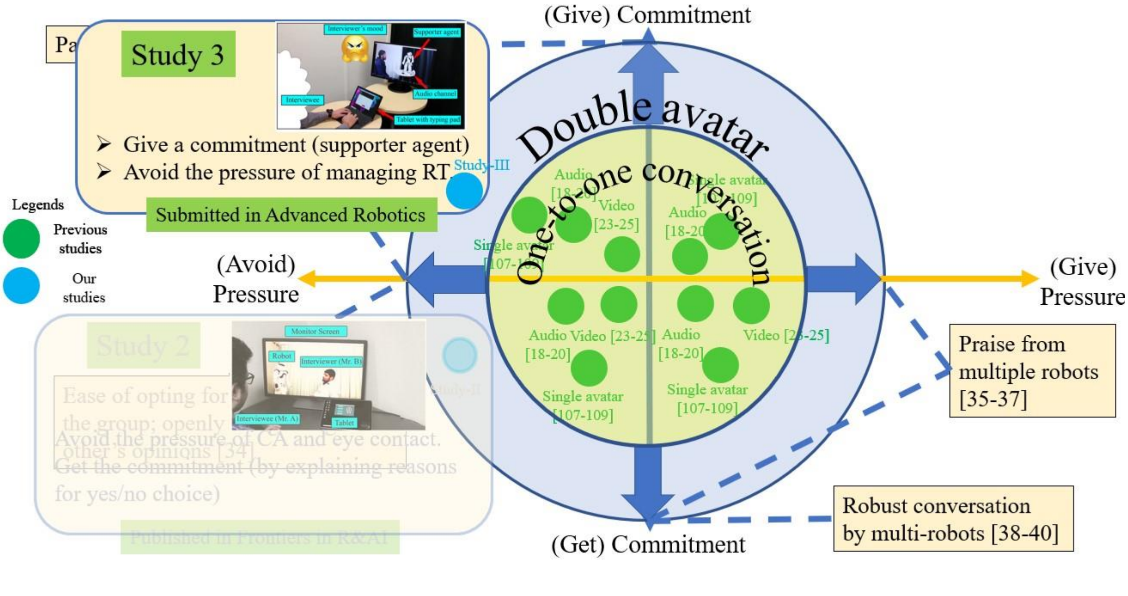

Figure 4.1: Study III location of research map

Abstract

Chapter 04 discusses whether using a supporter and avatar agents having interactive responses adaptive to RT variations in a video conference can reduce the stress of response time (RT) management of a person with a communication difficulty. The reduction of stress of RT management was assessed by situational communication apprehension measure (SCAM), fear of negative evaluation (FNE), SoBA, and ITU indexes. Three experiments were conducted to examine the effect: two subjective evaluation-based experiments and one interactive experiment. In the subjective evaluation-based experiments, recruited participants watched video stimuli with and without the proposed system. Later, they were asked to imagine themselves as an interviewee (a character of video stimuli) and provide their responses accordingly. In comparison to the video stimuli of the first subjective evaluation experiment, the video stimuli used in the second subjective evaluation experiment are more brushed-up to further tighten the experimental controls.

On the other hand, in the interactive experiment, recruited participants experienced both systems of video stimuli. A significant decrease in imagined SCAM, imagined FNE, and a significant increase in imagined SoBA and ITU was observed in the imagination-based surveys. Similarly, a significant decrease in experienced SCAM, experienced FNE, and a significant increase in experienced SoBA was observed in the interactive experiment. This study contributes to the literature in terms of the impact of using a teleoperated avatar and supporter agents having interactive responses adaptive to RT variations of a person with communication difficulties.

4.1 Introduction

People with communication difficulties have problems while communicating with others not only in face-to-face (FTF) interactions [1–3] but also in online interactions too [5,121,122]. Stress and anxiety in communication are among the causes of disrupted, and delayed talks, inability to communicate problems, resulting in reduced effectiveness of communication. The presence of stress and anxiety in person also generates failure in managing response time (RT) in communication [123,124]; where RT is defined as the total time required to a respondent to produce response for a given stimulus [125]. It is the sum of times required for: 1) planning the response to questions or given choice, 2) shifting turn for talk, and 3) executing the planned response [126]. It depends on several other elements in a communication event, e.g., sequence and timings of turn-taking [127], type and characteristic of question(s), and the topic of conversation [128–130]. On the other hand, apprehension in communication is one of the difficulties; commonly known as communication apprehension (CA) specifically defined as "*an individual's fear or anxiety associated with either real or anticipated communication with another person or persons*" [114]. CA at a particular time in a specific situation is referred to as situational communication apprehension. It tells how a person apprehended in a recent communication event; quantified by situational communication apprehension measure (SCAM) [131]. Presence of CA affects the effectiveness of communication [11]. The anxiety, distress, or apprehension of an individual about his/her negative evaluation by others is called fear of negative evaluation (FNE) [3], and it is positively correlated with CA [132].

To reduce the stress, anxiety, or apprehension of a person in communication, different types of technologies have been examined so far; namely text-only, audio-only and video (synchronous or asynchronous) technology. More specifically, such technologies include online social websites, cell phones, text/instant messaging [13], audio telephonic call, voice mail, electronic mail, and computer-mediated communications (CMC) [133]; acting as communication channels. Individuals having social anxiety and CA prefer using text, and audio technologies [13,87] that moderate not only their social anxiety but also their CA as well [14,94]. Asynchronous video technology also moderates the social anxiety and CA of individuals [92,93,134]; where asynchronous video technology is a type of video technology that provides one-way offline interaction opportunities to users. On the other hand, synchronous video technology lacks in moderating social anxiety and CA of individuals [22–24]; where synchronous video technology is a type of video technology that provides two-way online interaction opportunities to users. Although technologies moderate anxieties of users (as presented in [13,14,87,92,93]), however the aspect of social presence of users of such technology requires consideration, which is defined as a user's perception about the presence of his/her peer(s) in communication [90]. It is considered to be an important factor contributing to their feelings of belonging and connectedness [135–137] as well as sense of being attending (SoBA), which can be felt when a subject person is properly listened to, focused, given attention, or questioned/answered while attending conversation [91]. Text and audio-only channels offer reduced social presence to users [88,89,138] in comparison to video conference channel [95,139].

Another type of technology to reduce stress, anxiety, or apprehension of a person in communication is robot avatar technology. Avatar defined as "*an interactive, social* *representation of a user*" [104] or "*representation of the self in a given physical medium*" [105] is another channel of communication to efficiently convey social presence. Use of physical robot avatars in tele-communication is found effective in hospitals [140], virtual tours [108], family communication [109], and education [106,107]. It does not only help in hiding the identity of the users [110,111] which is expected to reduce the stress and anxieties, hence CA, and FNE, but also provide high social presence to users [112,113], so expected to keep their SoBA as well. Also, an avatar can be used for persuading its interlocutors when it is physically co-present with them [141] and receiving positive responses from its interlocutors [142]. However, mere use of robot avatar does not resolve the issues of RT management and its associated stress and anxieties in a person with communication difficulty. Methods to support users of tele-communication with avatars against such stress and anxieties are expected to be established, which will contribute to increasing effectiveness of communication.

To allow a user to communicate with reduced stress and anxiety about RT, I propose a video conferencing system by placing an avatar agent and its supporter agent in a remote place with an interlocutor. For simplicity, I adopted a simplified avatar agent using only a loudspeaker to convey the user's voice and a monitor to display the user's face as well as a humanoid robot as the supporter agent to cooperate with the avatar agent. Namely, the humanoid robot is present in front of interlocutor while the loudspeaker is placed near to web-camera and monitor screen and invisible from the user (see Fig. 5.2). The user can use a tablet device to input his/her utterance. The typed utterance is conveyed to interlocutor by either channel of communication: the loudspeaker or the humanoid robot. The selection of the channels depends on user's RT for interlocutor's utterance. Quick and slow user's responses are produced by the loudspeaker and the humanoid robot, respectively. Consequently, the interlocutor has two options for his/her attentional focus, the supporter agent or the avatar agent which consists of the monitor screen displaying user's face and the loudspeaker and located near to the web-camera. These situations are expected to decrease the SCAM and FNE of user by averting the interlocutor's attentional focus from the user hence reducing the interlocutor's pressure when the user fails in responding to the interlocutor within the acceptable delay. Moreover, the SoBA of the user is also expected to be maintained when he/she feels the first agent as avatar of him/herself and second agent as his/her supporter. Two experiments, a video-based experiment and an interactive experiment, were conducted to verify such effects. In both experiments, communication with the proposed system was compared to one with a conventional system where the same robot as the supporter robot in the proposed system attended to the online conversation in the same position but did not show any response to the interlocutor or the user's avatar agent. In the video-based experiment, participants firstly watched the video clip of online communication with and without proposed system and then imagined themselves as a user in the video and evaluated the expected feelings about SCAM, FNE, and SoBA. While in the interactive experiment, participants experienced online communication with or without proposed method and rated their SCAM, FNE, and SoBA. In short, a video conferencing system, consisting of an avatar and its corresponding supporter agent with interactive responses for interviewer adaptive to RT variation of user, was evaluated if it could reduce stress and anxiety of user related to RT in its communication.

4.2 Proposed robotic video conference system

4.2.1 Hardware

The proposed robotic video conference system consists of two desktop computers, a tablet, a loudspeaker, and a humanoid robot, see figure 4.2. Desktop computers are used to arrange an online discussion session between an interviewee and an interviewer in different rooms, named Room-1 and Room-2, respectively. The teleoperated avatar and supporter agents are available for interviewee to convey his/her utterances to interviewer in Room-2; where interviewee can control those by tablet, present in Room-1. NAO robot was used as the supporter agent. It was 58 cm tall with 25 degrees of freedom, fully programmable in different languages, and capable of interacting through speech, and motion stimuli.

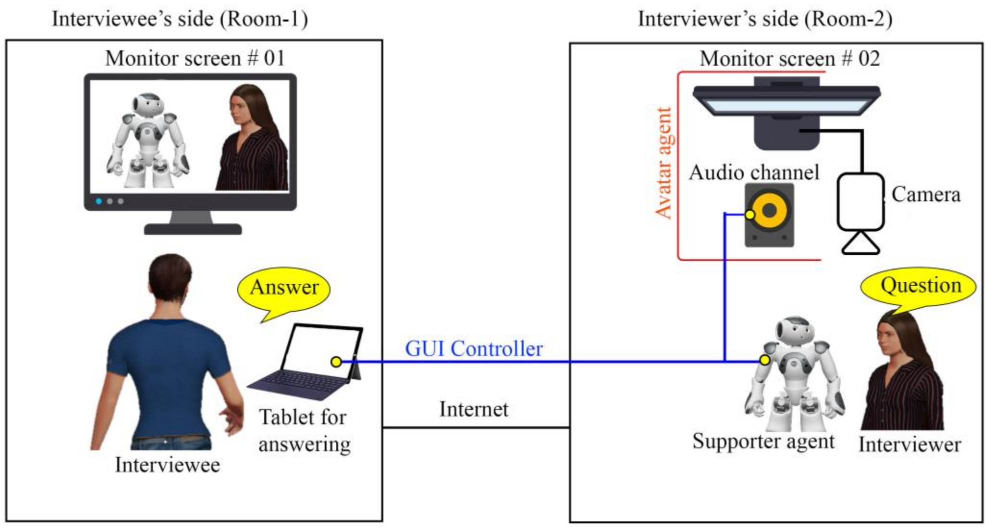

Figure 4.2: Proposed video conference system for reducing the stress of managing RT in a person.

4.2.2 Software

The tablet runs the GUI to control the avatar, and supporter agents. It consists of two text areas: one for typing answers and second for showing the last typed answer; see figure 4.3 (a). The interviewee is required to type the message on the GUI running on the table. Upon pressing the enter button, the typed message is uttered either from avatar or supporter agent depending upon RT. Similarly, another GUI was running on the second desktop computer. It consists of only one button named "*question*"; see figure 4.3 (b). The interviewer is also required to press the "*question*" button as soon as he/she finishes asking the question.

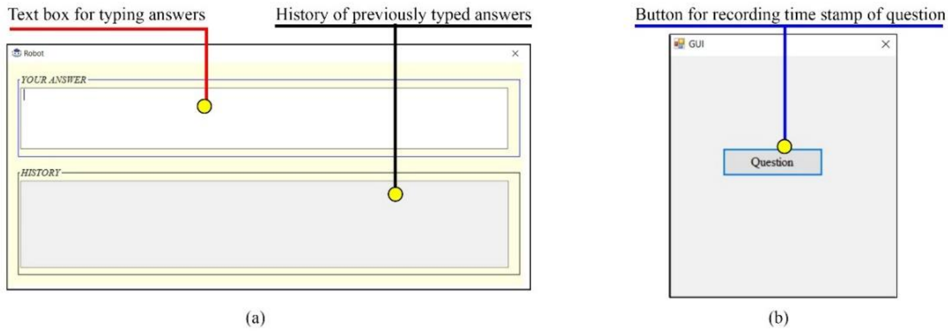

Figure 4.3: GUIs: (a) for interviewee, (b) for interviewer

4.2.3 Uttering agent selection

The selection of the uttering agent was based on the time taken by the interviewee to produce the response for the questions asked by the interviewer. When interviewer press the "*question*" button, the time is noted; named as $t_{question}$. Similarly, when interviewee press the enter button, again the time is noticed; named as $t_{answer.}$ Subtracting both times provides us the total time consumed by the interviewee to produce the response for interviewer i.e., RT. If the RT is less than seven seconds, then avatar agent utters the message. While, on the other hand, if the RT is equal to or greater than seven seconds, then supporter agent utters the message.

4.2.4 Interactive responses

During the discussion, the supporter agent remained in a standing position while performing idling movements i.e., moving body (torso) horizontally in left/right direction without changing its standing position on the table. It also alternately changes its attentional focus between the interviewer and the interviewee by turning its head. It looks at the web camera placed near the avatar agent which interviewee perceives a direct gaze towards him/herself. When supporter agent utters the answer of interviewee, it mentions interviewee in front of interviewer using third person pronoun i.e., “He/She said, [typed answer]” but when utterance comes from avatar agent, it is delivered as it is i.e., “[typed answer]”. In former case, upon reception of answer from interviewee, supporter agent turns its head towards monitor screen, nod once, turns its head back to interviewer and delivers answer. After listening answer, interviewer nods twice by saying “Right! Right!”. On the other hand, in later case upon receiving answer, both supporter agent and the interviewer turn their heads toward monitor screen, listen to answer from avatar agent, nod once by saying “Right!”, and turns their heads back to each other. I deliberately used a supporter agent for RT greater than seven seconds; as increased RT will raise the stress and anxiety of RT in the interviewee and at that instant of time, where the supporter agent plays its vital role as third person to reduce it.

4.3 Subjective evaluation experiment I

Experiment I is a subjective evaluation based experiment to explore the potential of the proposed system by analyzing the imagined feelings of the participants.

4.3.1 Experimental conditions

There are two experimental conditions, named as conventional and proposed video conference systems. In conventional video conference system, only avatar agent was available. While, in proposed video conference system, avatar and supporter agents were available. The avatar agent talks as first person perspective, while the supporter agent talks as third person perspective. For each of the condition, a video stimulus was recorded.

4.3.2 Experimental design

The experimental design was repeated measure. The type of video conference system was independent variable while SCAM, SoBA, FNE, and ITU were dependent variables.

4.3.3 Method

Video stimuli of conventional and proposed systems were included in a web-based survey. It further includes the questionnaires; SCAM, FNE, SoBA, and ITU indexes. In the beginning of survey, a web-based informed consent was obtained from participants. Then participants were required to provide some personal information e.g., gender, age, day of birth, daily life SCAM and FNE. After that they were required to watch video stimuli and answer the questionnaire while imagining themselves as interviewee, a character of video stimuli.

4.3.4 *Manipulation check*

Three manipulation checks were inserted to ensure that subjects watched the video stimuli and understood the content of those. The data of the subjects who passed at least two manipulation checks were considered.

4.3.5 *Participants*

A total of fifty one participants (M = 33.20, SD = 7.58 years) passed the manipulation check criterion in which thirty seven were males and fourteen were females. The participants were divided into groups A and B depending on the date of birth (even = 32, odd = 19) to counterbalance the conditions.

4.3.6 *Video stimuli*

In online discussion session, two video stimulus were recorded corresponding to conventional and proposed video conferencing system; see figure 4.4 (a) and (b). The topic of discussion was money related. The interviewer asked two yes/no and two in-depth questions. Depending on the RT of the participant, both video stimuli comprise different responses. Further, video stimuli also include the impact of the RT of the participant on interviewer and it was shown through thought bubbles and emojis, see figure 4.4 (c) to (f). The presentation timings of though bubbles and emojis were not identical in both video stimuli, as conventional video conference system was taking more time in producing response compared to proposed video conference system. However, the order of presentation was identical. To represent the interviewer's mood, I used positive (happy), negative (angry) and neutral (neither happy nor angry) types of emojis. While, for representing the interviewer's thoughts, I used positive and negative thought bubbles, see figure 4.4 (e) and (f). In video stimuli of the conventional robotic video conference system, only angry and neutral mood emojis with negative thought bubbles were used for the interviewer. While in video stimuli of the proposed robotic video conference system, happy and angry mood emojis with positive and negative thought bubbles were used for the interviewer. The interviewee was completely visible in each stimulus and his behavior of acknowledging the interviewer's questions was not controlled. Similarly, the gazing behavior of the interviewer was also not controlled.

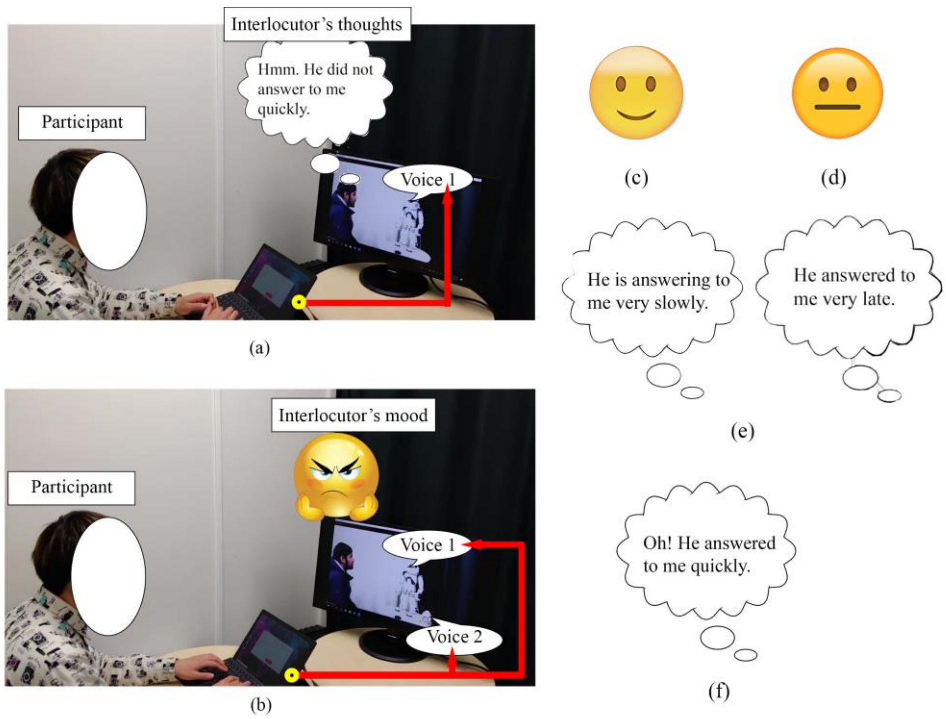

Figure 4.4: Video stimuli, (a) conventional robotic video conferencing system with a negative thought bubble of the interviewer, (b) proposed robotic video conferencing system with an angry mood emoji of interviewer (c) happy mood emoji, (d) neutral mood emoji, (e) negative thought bubble, (f) positive thought bubble.

4.3.7 Measurements

The type of video conference system was independent variable while SCAM, SoBA, FNE, and ITU were dependent variables of the experiment.

4.3.7.1 SCAM

SCAM [131] measures apprehension of an individual in communication at a particular time in a specific situation. It was rated on 1-to-7-point Likert-type scale (extremely accurate, moderately accurate, somewhat accurate, neither accurate nor inaccurate, somewhat inaccurate, moderately inaccurate, and extremely inaccurate).

4.3.7.2 FNE

FNE measures an individual's anxiety, distress, or apprehension about his/her negative evaluation by others; brief version of fear of negative evaluation scale (Brief-FNE) [143]. It was rated on 1 to 5-point Likert-type scale (extremely characteristic of me, very characteristic of me, moderately characteristic of me, slightly characteristic of me, not at all characteristics of me).

4.3.7.3 SoBA

SoBA scale [91] quantifies the feelings of being attended of an individual in a conversation. It was rated on 1 to 5-point Likert-type scale (strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree).

4.3.7.4 ITU

The participant's intention to use the conventional and proposed systems was obtained using ITU (intention to use) questionnaire [115] rated on a 1-5-point Likert-type scale.

4.3.8 Results

The participants imagined themselves as a character of video stimuli so the SCAM, FNE, SoBA, and ITU scores will be imagined ones.

4.3.8.1 Imagined SCAM (iSCAM)

The range of overall *iSCAM* score varies in between -120 to +120; after subtracting the daily life SCAM score from the *iSCAM* scores for conventional and proposed video conference systems. Wilcoxon singed rank test revealed that mean rank of *iSCAM* score for conventional video conference system (*Mdn* = 3) was significantly high compared to proposed video conference system (*Mdn* = 0); $n = 51$, $Z = 2.018$, $p = 0.043$ (2-tailed), $r = 0.20$, see figure 4.5.

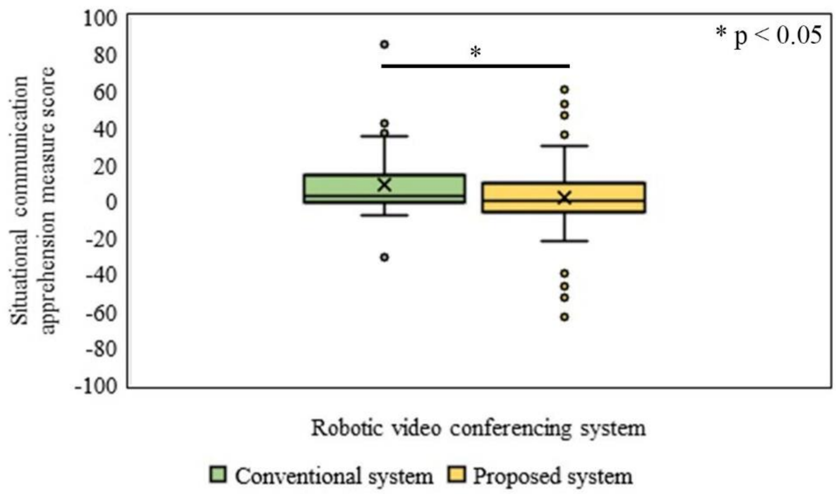

Figure 4.5: Imagined situational communication apprehension measure (*iSCAM*)

4.3.8.2 Imagined FNE (iFNE)

The range of overall *iFNE* score varies in between -48 to +48: after subtracting the daily life FNE score from the *iFNE* scores for conventional and proposed video conference system. Wilcoxon singed rank test revealed that mean rank of *iFNE* score for conventional video conference system (*Mdn* = 2) was significantly high compared to proposed video conference system (*Mdn* = -1); $n = 51$, $Z = 2.018$, $p = 0.043$ (2-tailed), $r = 0.20$, see figure 4.6.

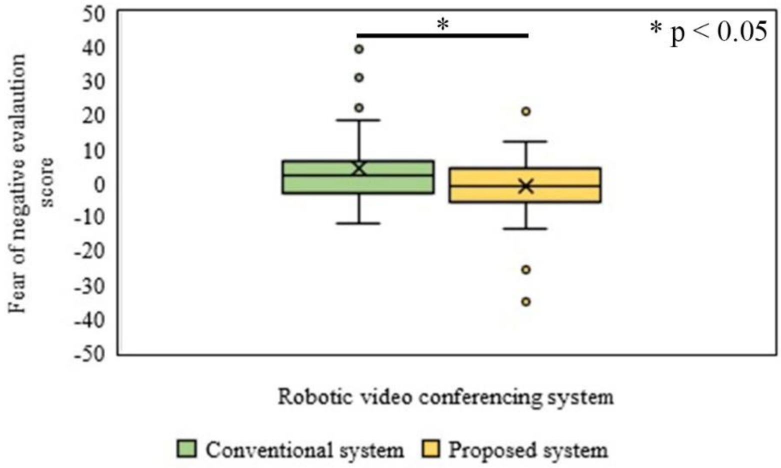

Figure 4.6: Imagined fear of negative evaluation (*iFNE*)

4.3.8.3 Imagined SoBA (iSoBA)

The range of *iSoBA* score varies in between 1 to +5. Wilcoxon singed rank test revealed that mean rank of *iSoBA* score for conventional video conference system (*Mdn* = 3.0) was significantly lower compared to proposed video conference system (*Mdn* = 4.0); $n = 51$, $Z = 2.00$, $p = 0.044$ (2-tailed), $r = 0.20$, see figure 4.7. SoBA is our own developed index consisting of four items that quantifies an individual's feelings of being attended in a conversational event. The internal consistency of the SoBA scale was high i.e., $\alpha = 0.91$.

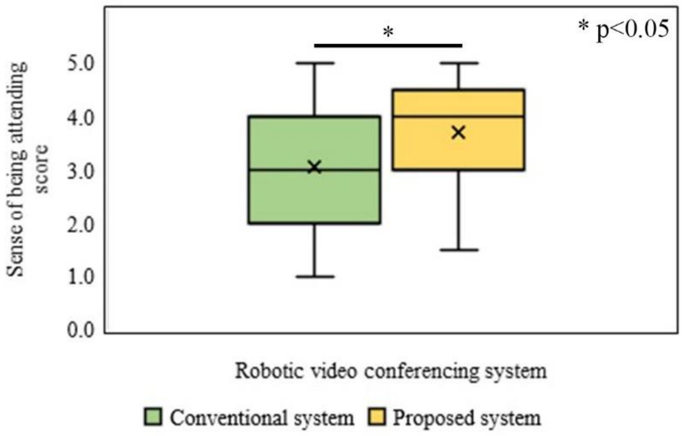

Figure 4.7: Imagined sense of being attended (*iSoBA*)

4.3.8.4 Imagined ITU

The range of *iITU* score varies in between 1 to +5. Wilcoxon singed rank test revealed that mean rank of *iITU* score for conventional video conference system (*Mdn* = 3.67) was significantly lower compared to proposed video conference system (*Mdn* = 4.0); $n = 51$, $Z = 2.08$, $p = 0.03$ (2-tailed), $r = 0.21$, see figure 4.8.

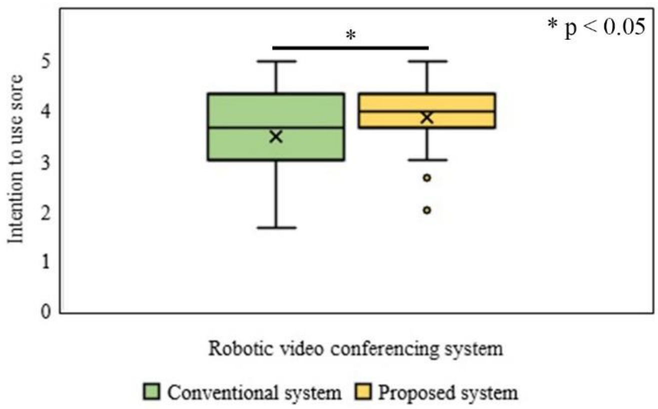

Figure 4.8: Imagined intention to use (*iITU*)

4.4 Subjective evaluation experiment II

Experiment II is also a subjective evaluation based experiment to explore the potential of the proposed system by analyzing the imagined feelings of the participants.

Although the results of experiment I were promising but it was hard to determine the actual cause of the obtained results due to lack of experimental control. Therefore, experiment II was conducted where the video stimuli were brushed-up to further tighten the experimental controls.

4.4.1 Experimental conditions

Same as that of subjective evaluation experiment I.

4.4.2 Experimental design

Same as that of subjective evaluation experiment I.

4.4.3 Method

Same as that of subjective evaluation experiment I.

4.4.4 Manipulation check

Two manipulation checks were inserted to verify that participants understood the content and watched the video clip for each system; namely, topics related to of the given conversation in it were questioned. The data of the participants, who passed both manipulation checks, were considered for further analysis.

4.4.5 Participants

A total of twenty seven participants were recruited online (M=32.26, SD=10.36 years); including eighteen males and nine females, were considered for further analysis after processing manipulation check. Participants were divided into two groups *G1* and *G2* based on their day of birth and given stimuli in different orders; where *G1* includes participants having even day of birth ($n=10$) while *G2* includes participants having odd one ($n=17$).

4.4.6 Video stimuli

In both video stimuli for proposed and conventional video conferencing systems, two male experimenters appeared as an interviewer and an interviewee and talked about a serious issue, namely, something about "unfair money" or "paying tax". In the conversation, interviewer asked two yes/no and two in-depth questions. The questions asked by interviewer and answers given by interviewee are given in appendix. To clearly express the mood and feelings of the interviewer to participants who watched the video stimuli, a negative thought bubble and an angry face emoji were used in each video stimuli; where thought bubble contains interviewer's thinking about interviewee e.g., "he is replying to me late." see figure. 4.9 (a) and (b)).

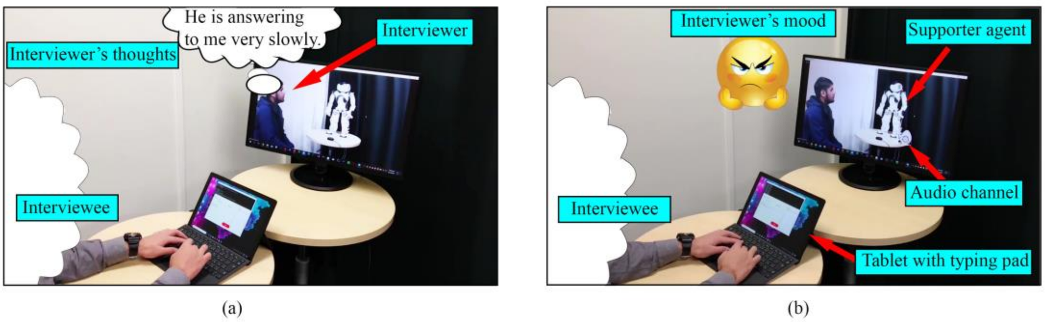

Figure 4.9: Video stimuli (a) conventional video conference system with negative thought bubble. (b) proposed video conference system with angry mood emoji.

Note that the same sequence of utterances and expressions by thought bubble and emoji were used in both video stimuli; hence they are balanced in-between conditions. The interviewee was not visible in either of the stimulus, and also not his behavior of acknowledging the interviewer's questions. Similarly, the gazing behavior of the interviewer was also controlled i.e., identical in between conditions.

4.4.7 Measurements

The measurements were same as that of experiment I.

4.4.8 Results

4.4.8.1 Imagined SCAM (iSCAM)

The baseline SCAM score of each participant was subtracted from the imagined SCAM score to evaluate their apprehension regarding the given situation. To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the imagined SCAM of the participant, Wilcoxon signed-rank test was conducted. The median rank of imagined SCAM of participants for the proposed system (*Mdn*=-1) was significantly less than that of the conventional system (*Mdn*=3) (*Z*=-2.06, *p*=0.039, *r*=0.28), see figure 4.10.

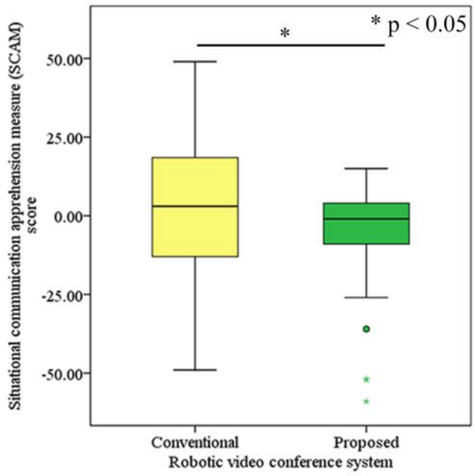

Figure 4.10: Imagined situational communication apprehension measure (*iSCAM*).

4.4.8.2 Imagined FNE (iFNE)

The baseline FNE score for each participant was subtracted from the imagined FNE score. To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the imagined FNE of the participant, Wilcoxon signed-rank test was conducted. It was revealed that the median rank of imagined FNEs of participants for the proposed system (*Mdn*=1) was significantly lower than that of conventional system (*Mdn*=2) (*Z*=-2.36, *p*=0.018, *r*=0.32), see figure 4.11.

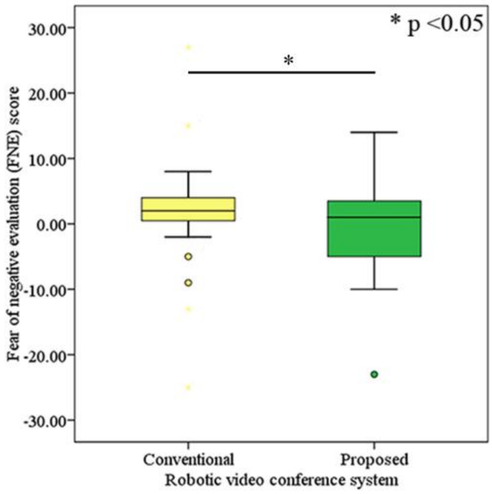

Figure 4.11: Imagined fear of negative evaluation (*iFNE*).

4.4.8.3 Imagined SoBA (iSoBA)

To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the imagined SoBA of the participant, paired sample t-test was conducted. It was revealed that the mean imagined SoBA of participants for the proposed system (*M*=4.4, *SD*=0.43) was significantly higher in comparison of conventional system (*M*=4.1, *SD*=0.59) (*t (*26*)* =-2.10, *p*=0.046, *d*=0.40), see figure 4.12.

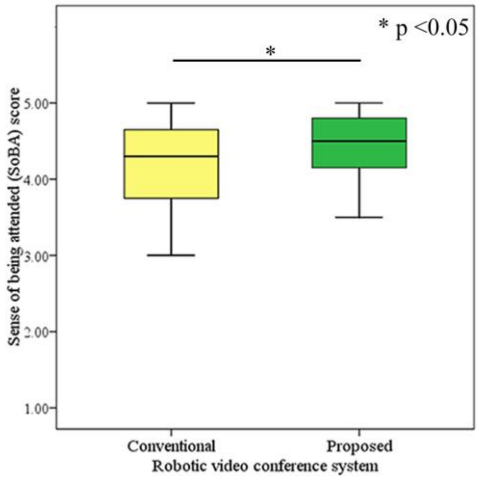

Figure 4.12: Imagined sense of being attended (*iSoBA*).

4.4.8.4 Imagined ITU

To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the ITU of the participant, Wilcoxon signed-rank test was conducted. It was revealed that the median rank of the ITU of participants for the proposed system (*Mdn*=4.0) was significantly higher than conventional system (*Mdn*=3.0) (Z=-1.99, p=0.047, r=0.27), see figure 4.13.

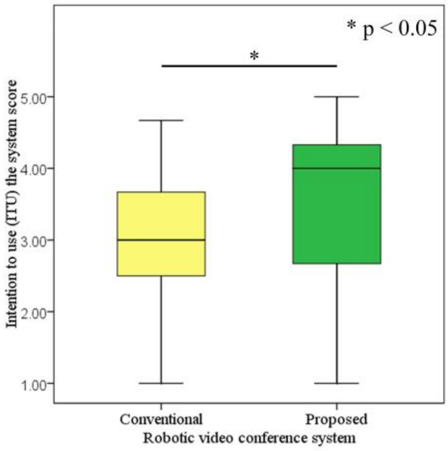

Figure 4.13: Imagined intention to use (*iITU*) score.

4.5 Interactive experiment

Although the results of the subjective evaluation experiments were promising but still, I was unable to see the actual potential of the proposed system in real-life situations. Therefore, in order to see the potential of the proposed video conference system in real life, I conducted field experiments.

4.5.1 Experimental conditions

The experimental conditions were two i.e., proposed and conventional video conferencing systems.

4.5.2 Experimental design

The experimental design was repeated measure.

4.5.3 Method

The proposed video conference system was compared to conventional video conference system by asking participants to attend to conversations using either of the systems, which followed the same scenario used in subjective evaluation experiments I and II where the participants played the role of the interviewee.

4.5.4 Participants

Nineteen native Japanese participants (*M*=23.15, *SD*=2.60 years), including ten males and nine females, were recruited. Participants were randomly divided into two groups G1 (*n*=9) and G2 (*n*=10) where they alternately experienced conversations with either system in different order.

4.5.5 Stimuli

Two experiences of interactive conversation were given to the participants where they were interviewed by a Japanese female experimenter acting as interviewer. To talk with the experimenter in each conversation, the participant teleoperated an avatar and a supporter agent in proposed system while only avatar agent in conventional system. In each conversation, the same conversation flows used in Experiment I were reproduced where the interviewer gave same questions and responses in the same order (see appendix) while thought bubbles and emojis were not used. The average duration of experience for each type of system was round about 5 minutes.

4.5.6 Procedure

Participants were required to attend the session consisting of three steps. In step I, they were requested to read and agree to the content of the written informed consent. Then, some demographic information, e.g., age, gender, and how they had apprehension and fear of evaluation by others in daily life, were obtained. At the end of the step I, the experimenter explained to the participant about how both types of systems work. They were also instructed to answer with short phrases for easy question (yes/no). While for difficult questions, they were asked to provide detailed (long) answers. Participants of group G1 experienced conversation with conventional system in step II and one with proposed one in step-III. In each step, immediately after the conversation with either type of systems, SCAM, FNE, SoBA, and ITU scores of participants were obtained. The order of conversations with conventional and proposed system was reversed for participants of group G2. The participant was free to end the session any time by informing the interviewer if he/she felt uncomfortable. The setup of the proposed system is shown in figure 4.14.

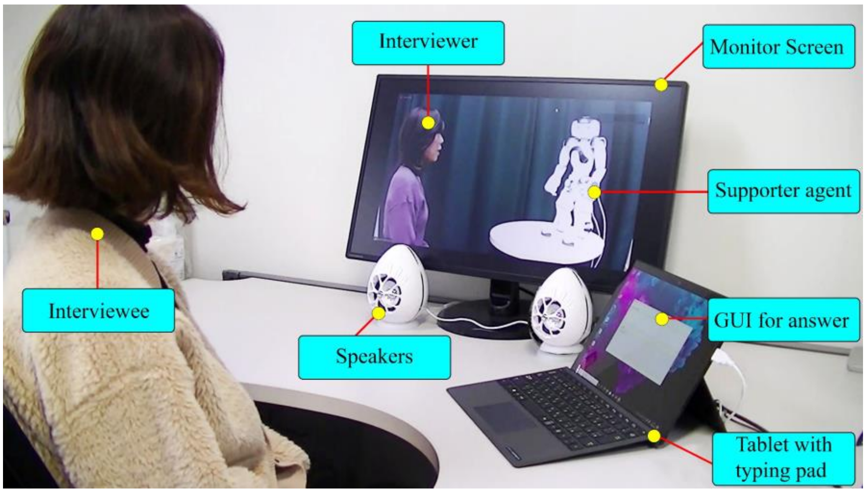

Figure 4.14: proposed video conference system with a participant

4.5.7 Measurements

The measurements used in interactive experiment were identical to those used in subjective evaluations experiments I and II. However, the scales were translated into Japanese. Note that, the participants were asked to score their experienced SCAM, FNE, and SoBA by focusing on the given experiences of conversation that they attended as an interviewee in each experimental condition.

4.5.8 Results

4.5.8.1 Experienced situational communication apprehension measure (eSCAM)

The baseline SCAM score for each participant was subtracted from the SCAM score to calculate the experienced SCAM score. To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the participants' experienced SCAM, a paired-samples t-test was conducted. To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the participants' experienced SCAM, a paired-samples t-test was conducted. Significant decrease in the mean experienced SCAM of participants for the proposed system (M=-17.89, SD=20.31) compared to conventional system (M=-13, SD=22.93) was observed (t (18) =2.13, p=0.047, d=0.48), see figure 4.15.

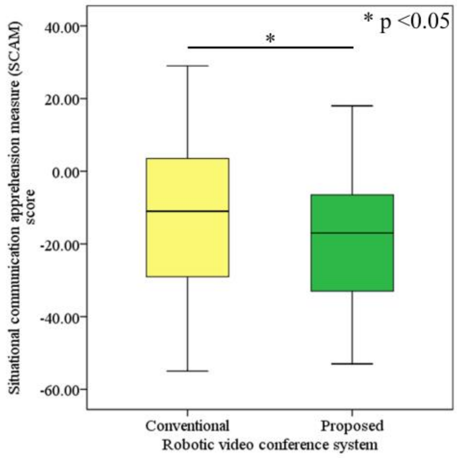

Figure 4.15: Experienced situational communication apprehension measure (*eSCAM*).

4.5.8.2 Experienced fear of negative evaluation (eFNE)

The baseline FNE score of each participant was subtracted from the FNE score to calculate the experienced FNE. To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the FNEs of the participants, a paired-samples t-test was conducted. Significant decrease in the mean FNE score of the participants in the proposed system (*M*=-8.47, *SD*=9.98) compared to the conventional system (*M*=-2.68, *SD*=8.62) was observed ($t\,(18)$ =2.15, p=0.045, d=0.49), see figure 4.16.

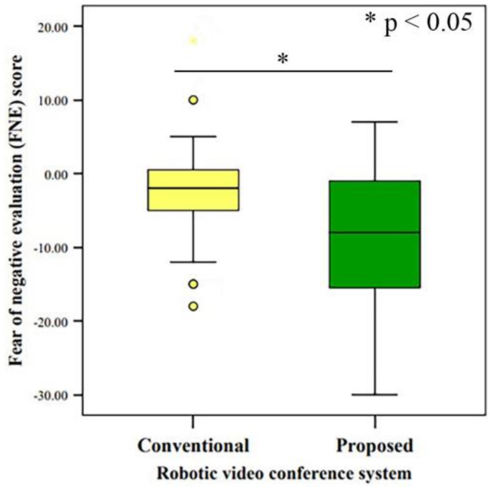

Figure 4.16: Experienced fear of negative evaluation (*eFNE*) score.

4.5.8.3 Experienced sense of being attended (eSoBA)

A paired-samples t-test was conducted to identify the effect of the type of robotic video conference system (conventional vs. proposed) on the SoBA of the participant. The mean SoBA of participants for the conventional system (*M*=3.60, *SD*=0.61) was significantly less than that of the proposed system (*M*=3.92, *SD*=0.68) ($t\,(18)$ =-2.15, p=0.045, d=0.49), see figure 4.17.

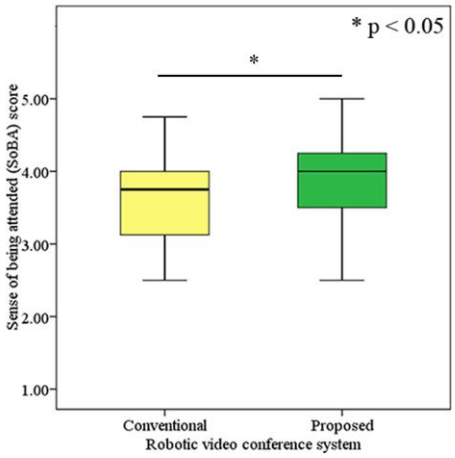

Figure 4.17: Experienced sense of being attended (*eSoBA*) score.

4.5.8.4 Experienced intention to use (eITU) of the system

To identify the effect of the type of robotic video conference system (conventional vs. proposed) on the ITU of the participant, paired-samples t-test was conducted. The mean ITU of the participants for the conventional system (*M*=2.65, *SD*=0.95) was not significantly different from that of the proposed system (*M*=2.60, *SD*=1.05) (*t (*18*)* =0.31, *p*=0.76, *d*=0.071), see figure 4.18.

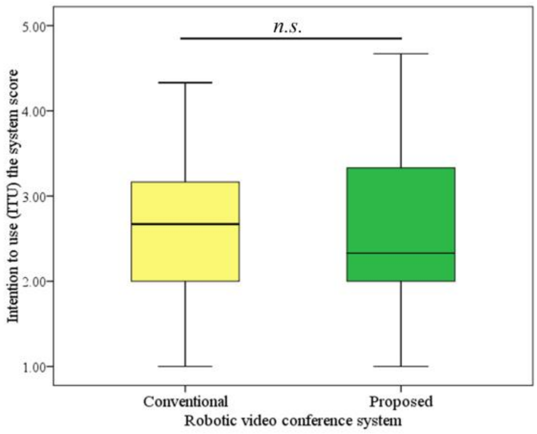

Figure 4.18: Experienced intention to use (*eITU*) score.

4.6 Discussion

A video evaluation based, and a real world-based experiments were conducted to see the potential of proposed system. In Experiment I, significant reduction in imagined SCAM, FNE and significant increase in imagined SoBA were observed in proposed system in comparison to conventional system. However, it was not clear whether the effect found was limited to the video experiment, including the thought bubbles and emojis attached. Experiment II clarified that it was not limited to the video experiment, but the effects could be revealed in the interactive real-word environment. Therefore, the proposed RT management policy was useful for both imagined and experienced feelings of SCAM, FNE, and SoBA for participants.

The main reason behind the success of the proposed RT management policy was the intelligent switching between available communication mediums (avatar vs supporter). In the proposed system, the fast answers were uttered from avatar agent while the slow answers were uttered from supporter agent. Such situation was advantageous for participant as he/she might feel that he/she was not responsible for slow answers but only for fast answers. Similarly, participant might also feel that he/she is not responsible for the interviewer's waiting time and the consequent probable anxiety, in which the interviewer might drought whether the participant was going to reply in a slow way, or the interviewer should stop waiting and say something to continue the conversation. The presence of such feelings in participant is considered to not only reduce his/her SCAM but also the FNE during the interaction. Further, higher ratio of quick replies through avatar agent throughout interaction also causes the increased ration to receive positive responses from interviewer for fast replies, hence further contributing to reducing SCAM and FNE scores. The lack of perceived social acceptance in a person predicts his or her social anxiety [144] but RT regulations were handled in such a way that it was intentionally excluding participant (interviewee) from the responsibility of slow responses in conversation and increasing his/her perception about how much the interviewee was socially accepted by interviewer, contributing to decrease in SCAM, and FNE.

The SoBA of participants also become better and there could be two main reasons for it. 1) All the fast replies were given to interviewer from avatar agent, while the slow replies were given from supporter agent. It might lead participants to feel more presence near interviewer and even higher ITU. 2) When supporter agent and interviewer look together at camera for listing user's typed utterance from avatar, it generates feelings of being focused in interviewee by two agents simultaneously: hence increasing SoBA. Having shared gaze to a specific area in a scene increases engagement among participants in online interaction [116], therefore the shared gaze of the interviewer and supporter robot agent to the monitor (to look at avatar agent answering the fast replies) might increase the feeling of being engaged for the participant and hence SoBA.

Although results are in the favor of our proposed system, however, there are some limitations to our findings. The threshold used for RT management policy was constant in both studies, however, the level of sensitivity of the participants about feeling anxious for their RT may vary. Therefore, gathering such data and studying about human's feelings of anxiety associated with RT is required to develop an adaptive RT threshold management mechanism for the proposed method. Also, the scenario of the interaction was limited in the experiments where the topic of the conversation along the verbal and non-verbal behaviors of the interviewer were restricted. To further explore the potential of the system in daily life usage, such restrictions should be removed by considering more casual, non-anxious topic of conversation within free chatting environment. The other shortage was that the participants with severe SCAM and FNE issues were not recruited, who were eager to be supported in real-world online interactions. Therefore, the effects observed in experiments I (video evaluation) and II (interactive) are not necessarily guaranteed to be observed for such users in real life. Further, in case of recruiting the participants with sever SCAM and FNE issues, the system should be tuned up in accordance with the guidelines of a specialist dealing in anxiety related therapies; as it could also become a challenging task to overcome unforeseen technical and non-technical issues. The demographic profiles' information of participants e.g., cultural, and linguistic backgrounds were neither considered nor controlled in the experiments which could make it difficult to observe the found effects in a specific cultural and linguistic background(s).

4.7 Conclusion and future work.

In this research, the potential of the proposed system with RT management policy for providing communication support to users in online video conferences was explored. It consists of teleoperated avatar and supporter agents having interactive responses. The user can utter from any of these agents concerning his/her RT variations in conversation. While the user's slow answers were uttered from the supporter agent, the fast ones were done by the avatar agent. Such a real time intelligent switching in between communication channels is expected to make the user feels responsible only for the avatar agent's utterances but not for the supporter agent's ones. This situation causes reduction of user's stress of RT in online video conference-based communications. To verify the expected effects, two experiments were conducted: imagination-based video experiment and real-word interaction experiment. In the imagination-based experiment, participants watched the video stimuli with and without proposed system and imagined themselves as interviewee for providing their impressions. While in the interactive experiment, they experienced both systems in a real-word interactive environment. The experimental results showed the positive effect of proposed method in reducing the user's SCAM, FNE, and increasing SoBA. This is considered as the successful result of RT management policy in reducing the feeling of responsibility about the slow answers which is expected to provide communicational support for the users in the video conference. In future, I will examine whether using proposed system helps in reducing stress of RT for users with severe SCAM and FNE.

Chapter 5

Effect of having and switching multiple avatars

Aim

The study focusses on directing/giving the pressure of communication to other; see figure 5.1. The pressure in the communication is directed/given to others by increasing the right to talk and providing the social support.

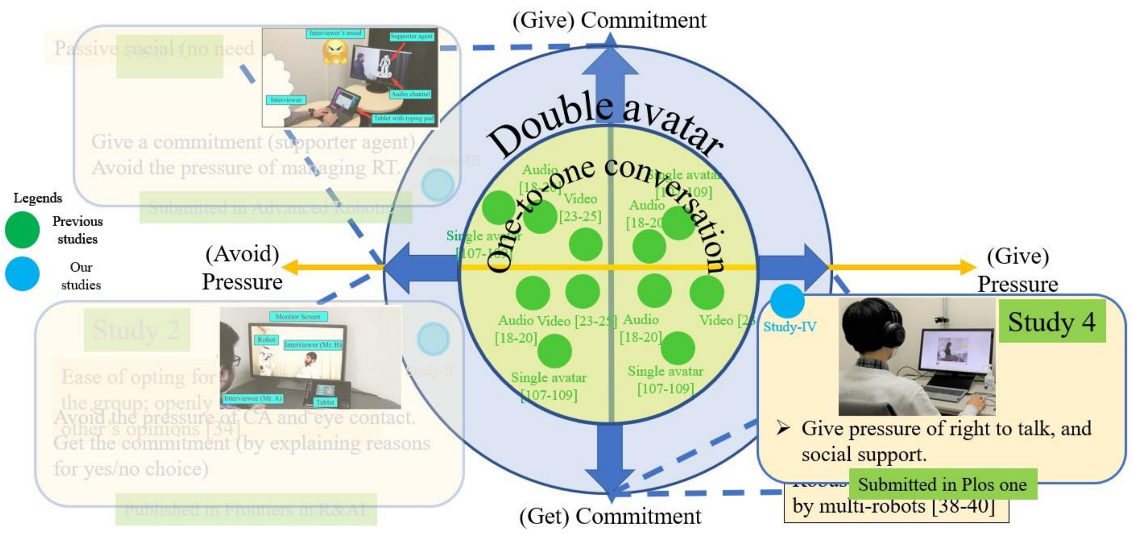

Figure 5.1: Study IV location of research map

Abstract

People with communication difficulties have problems in daily life online interactions e.g., less right to talk (RoT), low social support (SS), and low sense of being attended (SoBA). Computer mediate technologies are limited in resolving such problems because of their limited capacity of transferring verbal and non-verbal cues between interactees. In this study, to resolve the problems of less RoT, low SS, and low SoBA, I proposed a robotic video conference system having two teleoperated robot avatars. The proposed system was compared with another robotic video conference system having only one teleoperated robot avatar. In a field experiment, a total of thirty-seven participants took part in two discussion sessions while using each type of system; where RoT, SS, and SoBA were the measured indices. The proposed system was able to increase the feelings of RoT and SS of users significantly compared to other robotic video conference system. The study contributes to the literature in terms of showing the effect of the type of robotic video conference system on the user's feelings about RoT, SS, and SoBA.

5.1 Introduction

Individuals with communication difficulties have problems in daily life online interactions; e.g., less right to talk (RoT), lack or absence of social support (SS), and low social presence (SP) in communication. The first problem is less RoT; defined as an individual's feelings concerning the provision of equitable speaking and opinion-expressing opportunities with respect to the peers of conversation. Such feelings are influenced by different types of elements in peers of conversation, e.g., the number of conversational turns, duration to listen and talk [145], number of utterances, inter utterances pauses, and back-channel responses [146]. In an ideal case, equality is required among all such types of elements for each peer of conversation as general rules of talk. However, in reality, such rules of talk are very prone to violations that cause vocal interruptions, speech, and social anxiety in peers of communication [147]. The chances of violations of rules of talk can be reduced to a great extent if peers of conversation socially support each other. The second problem is the lack or absence of SS; defined as "information leading the subject to believe that he is cared for and loved, esteemed, and a member of a network of mutual obligations" [148]. In conversation, the presence of social support increases the willingness to communicate with individuals [149], and mediates anxiety and depression [150]. The third problem is low SP; defined as an individual's perception of his/her own presence in a conversation [90]. It contributes to feelings of belonging and connectedness in conversation [135,136]. Further, such belonging and connectedness influence an individual's sense of being attended (SoBA); defined as the experience of feelings of being focused, attentive to, listened to, and queried/answered in conversation [151]

In order to resolve the problems of less RoT, lack or absence of SS, and low SP issues for individuals with communication difficulties, computer-mediated communication (CMC) technologies have been examined. CMC includes text, audio, and video interaction technologies. Although for the people with communication difficulties, each CMC technology provides a huge support to establish a communication and have a communication with others, however these technologies are failed to assure the RoT of users because of the limited capability of transferring verbal and non-verbal information that influences the abiding by rules of talk [152]; hence demanding users to repair RoT by mutual cooperation [153,154]. Further, CMC technologies have a limited potential of offering SS to users [155] e.g., text-only technology is very limited in offering SS [156]. However, audio-only technology can provide limited SS [157]; as alongside verbal cues, it also provides limited non-verbal cues as well e.g., vocal tone variation associated with discrete emotions [158]. On the other hand, video technology is relatively better at providing SS to users [159]; as alongside verbal cues, it also provides limited non-verbal cues e.g., awareness of attentional focus, ease in speaking turns, facial expressions [160], [96]. Moreover, CMC technologies are also limited in providing a social presence to users [88]. Text-only technology provides a very low social presence to users [88,161], while audio-only technology provides a low social presence; better compared to text-only technology [16]. On the other hand, video technology is relatively rich in providing a social presence to users [29,30]. However, due to the presence of video streaming, the communication apprehension of the users increases that eventually decreases the ease of talk [24,25,162]. To prevent such problems while keeping the social presence, other available technologies are required to be explored. Considering the performance of text-only, and audio-only technologies concerning the provision of SP to users, it is expected that they will also be very limited in providing SoBA to users. However, video technology can provide a reasonable amount of SoBA to users: because of providing rich social presence [91]. In conclusion, text, audio, and video technologies are limited in solving the aforementioned problems.

Robot avatar technology has been examined concerning SP but the issues of RoT and SS in communication are required to be studied yet. There are two types of users of robot avatar technology; those interacting through avatars hereafter referred to as operators and those interacting with avatars hereafter referred to as visitors. In social interactions, using a single physical avatar facilitates operators to communicate with visitors, e.g., in education [106,107] and family interactions [109,163]. However, such interactions are dyadic in nature so it's difficult to increase operator's RoT; as it would be defined in accordance with dyadic rules of talk. So, 1) whether the perception of the operator about the RoT is same as that of dyadic interaction? 2) is there any reduction in the perception of RoT? If yes then, 3) how can I improve the perception of RoT in avatar-mediated communication? Such questions are yet required to be explored. The experiences through avatars are considered as operators' own experiences [117], and when communicating with a visitor via an avatar, there would be no chance for the operator to avail SS from his/her own robot avatar. Since, there is dyadic interaction with the visitor through avatar so the SS that operator can avail will be limited. The feelings of receiving SS are very important in interaction, especially for the people with communication difficulties [164] so alternative method(s) are required to be searched yet. A human subject could have high SoBA by watching a video scene of a conversation, where a visitor was interacting with an avatar of a side-participant [151]. It implies that an operator feels being supported in conversation via avatar when another avatar cares for his/her avatar. In other words, it would be worth examining the effect of using a second avatar on the SS of the operator communicating via an avatar. The visitor talking with two avatars of operator would be a triadic interaction situation which is also expected to increase the RoT of operator as per rule of talk for triadic interaction. Therefore, in this study, I am proposing a system consisting of two avatars, controlled by an operator to talk with the visitor for experiencing higher RoT and SS.

In the proposed system (see figure 6.2), the operator's utterances are produced from either of the teleoperated avatars. The choice of the speaking avatar and production of the backchannel responses from the other avatar are processed randomly. Sometimes, the second avatar takes a speaking turn from the current avatar to talk about the same opinion as that of the first one. Such turn-taking and backchanneling behavior of the second avatar are expected to make the operator feels supported in communication. Consequently, the visitor also has to switch the attention towards the speaking robot avatar throughout the conversation. Such treatment of visitor would be evident to operator by video feed on monitor. Furthermore, the proposed system is expected to provide another merit of increased RoT because an operator is expected to attend a multi-party conversation via two agents. Humans usually tend to equalize conversational turns, time to listen to and talk [145], as well as the number of utterances, inter utterance pauses, and back channel responses in conversation [146]. It reveals that humans tend to expect an equal RoT for each participant in the conversation. As experiences through the avatars are considered as the operator's own experiences [117], so the assigned RoT of each avatar is expected to be perceived by the operator as his/her own RoT in multi-party conversation. Therefore, in the proposed system, the operator is expected to perceive double amount of RoT at most. While on the other hand, at least more than one expected to be perceived when communicating through a single avatar.

5.2 Video teleconferencing system involving physical avatars

The schematic diagram of the proposed system is shown in figure 5.2. It consists of a computer, a headset with a microphone, two semi-humanoid robots, and a web camera. Using a computer and web camera, an online interaction session was arranged between the operator and visitor; physically present at different locations namely location-I and location II, respectively. The robots are physically present in front of the visitor at location II, and both were avatars of the operator. I used the CommU robot which is developed via collaboration between Osaka University and Vstone Co., Ltd., Japan. It is a semi-humanoid robot with clear eyes; having 14 degrees of freedom in total; 31 cm of height; programmable using JavaScript language, and capable of interacting through visual, speech, and motion stimuli.

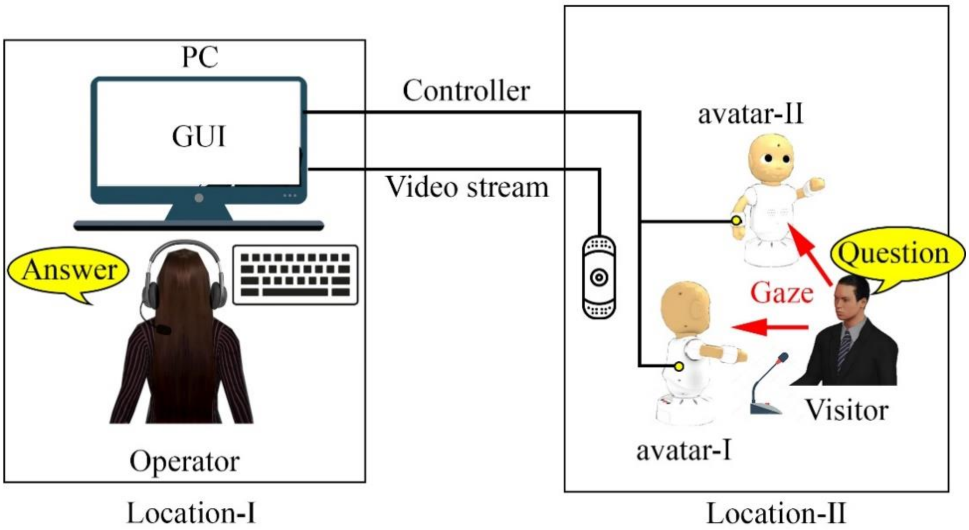

Figure 5.2: The proposed robotic video conference system

To control the robot avatars over a wide area network in real-time, a locally build GUI was used by operators; see figure 5.3. The GUI consists of two sections; named sections I and II. Section I consists of visual feedback while section II consists of utterance-related handling options. The visual feedback section was designed using web-RTC and it provides a real-time view of the visitor's environment to the operator, referred to as section I in figure- 5.3. It also displays the detected spoken answers of the operator as dynamically added buttons at the bottom, in the middle of the section. On the other hand, the utterance section consists of a text field with three buttons, referred to as section II in figure- 5.3. It provides several facilities to the operator e.g., typing new answers, editing previously detected answers, deleting answers, and enabling/disabling speech recognition.

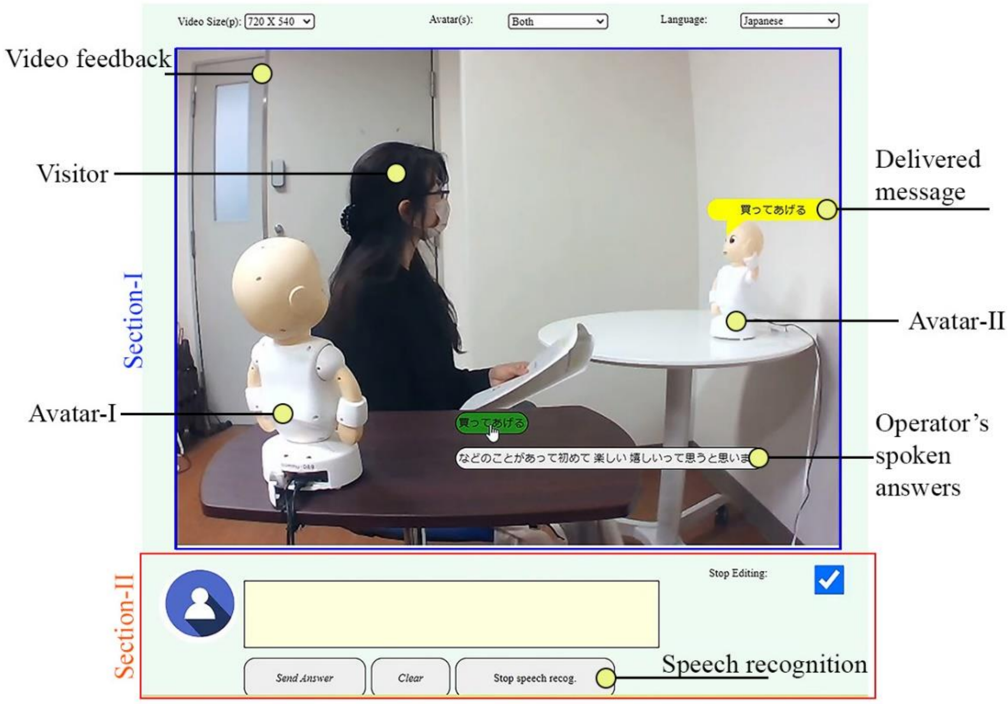

Figure 5.3: Operator’s GUI.

In an online interaction session, the proposed system behaves in two different ways concerning whether an operator provides the answer(s) to the question(s) of the visitor or not. In a case, where the operator does not provide the answer(s), avatar-I remains still with a visual focus towards the visitor, while avatar-II keeps on performing idling motions i.e., keeps switching visual focus between the visitor and the avatar-I by turning its head and torso. On the other hand, in a case where the operator provides the answer(s), one of the avatars utters in synthesized voice in front of the visitor while raising the left arm. Meanwhile, the other avatar shifts its visual focus to uttering avatar and acknowledges by nodding; pretending that the provided answer was accurate and acceptable. To deliver the answer of the operator, the system randomly chooses the avatar.

While conversing through such a system, the visitor asks questions from the avatar who delivered the answer, and the operator is required to answer the questions either by speaking or typing. The proposed system also manages conversational turns between the operator and visitor. When the visitor speaks, the operator gets the information from real-time video feedback. While, when the operator types or speaks the answer, the visitor gets the information from the glowing (red) cheeks of the robot(s).

5.3 Materials and method

5.3.1 Method

The impression of conversation between the operator and visitor was evaluated by using two types of systems namely conventional and proposed. The conventional system is a video conference system integrated with a single teleoperated avatar (hereafter referred to as single avatar condition) while the proposed system is a video conference system integrated with double teleoperated avatars (hereafter referred to as double avatar condition). The recruited participants were asked to visit the experimental site in person and attend four online conversation sessions with a visitor i.e., two practice and two experimental sessions. Both practice and experiment sessions have two conditions: a single avatar condition and a double avatar condition. The number of avatars was an independent variable of the study while the RoT, SS, and SoBA were dependent variables.

5.3.2 Participants

A total of thirty-seven native Japanese-speaking participants (M=21.68 years, SD= 2.13 years), involving 21 males and 16 females, were recruited. They were randomly divided into two groups i.e., G1 and G2. Group G1 experienced the single avatar condition first and then the double avatar condition. While, for group G2, the experiencing sequence was the opposite. Please note that the practice and experimental sessions were counterbalanced.

5.3.3 Conversational scripts

5.3.3.1 Practice sessions

Two short conversational scripts for practice sessions and asked the participants for their recommendations were chosen. In first short conversation script, recommendations were related to type of food while in the second short conversation script, recommendations were related to club activities during the schooling period.

5.3.3.2 Experimental sessions

To choose conversational scripts for experimental sessions, a separate subjective evaluation experiment was conducted in which the recruited participants read, and evaluated four different conversational scripts concerning RoT and SS. Later, I chose two conversational scripts having equal RoT and SS. The topics of chosen conversational scripts were “whether a person should choose love or money to live a better life?” and “whether a person should save the life of a child or the lives of two old persons in a car accident?”.

5.3.4 Stimuli

In both experimental sessions, an operator (i.e., participant) and a visitor talked about two topics namely, “whether a person should choose love or money to live a better life?” and “whether a person should save the life of a child or the lives of two old persons in a car accident?” The content of the topics is presented in appendix. Please note that conversational topics were also counterbalanced between experimental sessions. In a single avatar-based conversation session, an avatar agent (labeled as avatar-I, see figure 6.3) was placed in front of the visitor and teleoperated by the operator to convey the answers to the visitor. During the conversation, the visitor directed his/her attentional focus to the avatar agent meanwhile, both visitor and the teleoperated avatar agent were visible to the operator through the monitor of the video conference system. While in a double avatar-based conversation session, two teleoperated avatar agents were placed; one on the left side, and the other in front of the visitor, see figure 5.3. The operator's answers were produced by either of the robot avatars randomly and conveyed to the visitors in a synthesized voice. During the conversation, the visitor kept changing his/her attentional focus by turning his/her head and torso to the speaking avatar agent. Meanwhile, both the visitor and the teleoperated avatar agents were visible to the operator through the monitor of the video conference system. In both experimental sessions, the visitor not only asked the questions from the operator but was also providing logical reasoning so that operator think about changing his/her opinion. The sequence of asking questions, and provision of logical reasoning remained the same in both conditions. The duration of each practice session was 2 to 3 minutes approximately. While for each of the experimental sessions, it was 10 to 12 minutes approximately. The language of conversation practice and experimental sessions was Japanese.

5.3.5 Procedure

The participants were required to visit the experimental site, where they read and agreed to the content of the written consent form. Meanwhile, they were randomly assigned to a group; either G1 or G2. In the beginning, participants were required to complete the two short practice sessions; where they practiced the usage of both systems, i.e., single, and double avatar conditions. They were briefed on the functionalities of each element of the GUI controller and later filled out questionnaire forms. After completing practice sessions, the participants were then required to complete two experimental sessions. In the experimental session, participants of group G1 were briefed again regarding the single avatar condition. In the briefing, the functionalities of all the elements of GUI, and the topic of conversation were explained to them. They were also instructed not to rush and provide very long answers. Instead, try to use the system peacefully, reply with calm, and try to give as many short answers as they want. After that, the first experimental session was arranged where the participants experienced the conversation using a single robot condition and filled out the questionnaire form. Similarly, a second experimental session was arranged where the participants experienced the conversation using double robot conditions and later filled out another questionnaire form. On the other hand, for the participants of group G2, the sequence of experience of conditions was the opposite.

5.3.6 *Measurements*

5.3.6.1 Right to talk (RoT)

RoT is an individual's feelings concerning the provision of equitable speaking and opinion-expressing opportunities with respect to the peers of conversation. I developed a new scale in the Japanese language to quantify such feelings of the participants in the conversational scenario of our experimental setup. A 1–7 Likert-type point scale was used (strongly disagree, somewhat disagree, disagree, neither agree nor disagree, somewhat agree, agree, strongly agree) where ratings are summed to yield the operator's total scores for both conditions. The corresponding English translation of the questionnaire is given in appendix.

5.3.6.2 Validity and reliability of RoT scale

The validity and reliability of the RoT scale were assessed by conducting a separate subjective evaluation experiment. After obtaining the data from subjective evaluations, exploratory factor analysis was carried out. The Kaiser-Meyer-Olkin (KMO) test revealed the sample adequacy (KMO=0.91) and Bartlett's tests of sphericity revealed the factorability of the covariance matrix ($X^2(15) = 966.47$; $p<0.05$). The principal component analysis is used for the factor extraction method. In accordance with the Kaiser criterion of factor(s) retention, only one factor was retained explaining 87.61% of the total variance. All items of the RoT scale were significantly loaded for the factor retained, see appendix. The reliability of the RoT scale was measured by Cronbach's alpha which turns out to be very high i.e., $\alpha=0.97$.

5.3.6.3 Social support (SS)

Social support is an "*information leading the subject to believe that he is cared for and loved, esteemed, and a member of a network of mutual obligations*" [148]. I developed another new scale in the Japanese language to quantify such feelings of the participants in the conversational scenario of our experimental setup. A 1–5 Likert-type point scale was used (strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree), where ratings are summed to yield the operator's total scores for both conditions. The corresponding English translation of the questionnaire is given in appendix.

5.3.6.4 Validity and reliability of SS scale

The validity and reliability of the SS scale were also assessed by conducting a separate subjective evaluation experiment. After obtaining the data from subjective evaluations, exploratory factor analysis was carried out. The Kaiser-Meyer-Olkin (KMO) test revealed the sample adequacy (KMO=0.85) and Bartlett's tests of sphericity revealed the factorability of the covariance matrix ($X^2(6) = 582.63$; $p<0.05$). The principal component analysis is used for the factor extraction method. In accordance with the Kaiser criterion of factor(s) retention, only one factor was retained explaining 91.77% of the total variance. All items of the SS scale were significantly loaded for the factor retained, see appendix. The reliability of the SS scale was measured by Cronbach's alpha which turns out to be very high i.e., α=0.96.

5.3.6.5 Sense of being attended (SoBA)

SoBA is a scale used to quantify the feelings of a participant concerning being listened to, attended to, focused upon, or questioned/answered by an individual in a conversational scenario; developed by [151]. I updated the SoBA questionnaire according to our experimental setup while keeping the essence of the original scale and translated it into the Japanese language; obtained reliability α=0.84. A 1–5 Likert-type point scale was used (strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree), where ratings are summed to yield the operator's total scores for both conditions. The corresponding English translation of the questionnaire is given in appendix.

5.4 Results

5.4.1 Right to talk (RoT)

The Wilcoxon signed rank was conducted to identify the effect of the type of condition (single avatar vs double avatars) on the RoT feelings of the operator. It was revealed that the median value of the RoT for the operator of the double avatars condition (*Mdn*=35) was significantly higher than the single avatar condition (*Mdn*=34), (Z = -1.99, p = 0.047, r = 0.23), see figure 5.4.

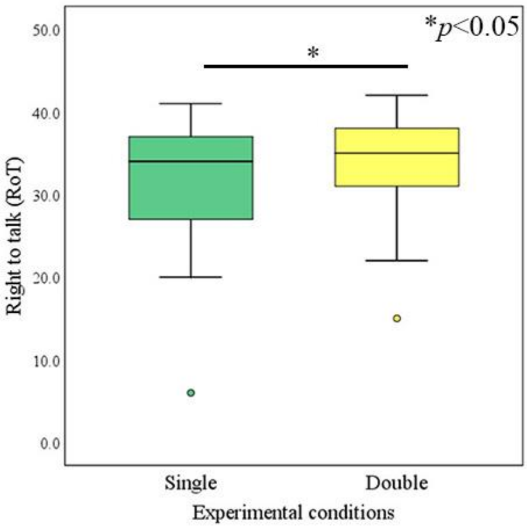

Figure 5.4: Right to talk (RoT) score of operators.

5.4.2 Perceived social support

The Wilcoxon signed rank was conducted to identify the effect of the type of condition (single avatar vs double avatars) on the SS of the operator. It was revealed that the median value of the SS for the operator of the double avatars condition (*Mdn*=15) was significantly higher than the single avatar condition (*Mdn*=15), (Z =-2.11, p = 0.034, r = 0.24), see figure 5.5.

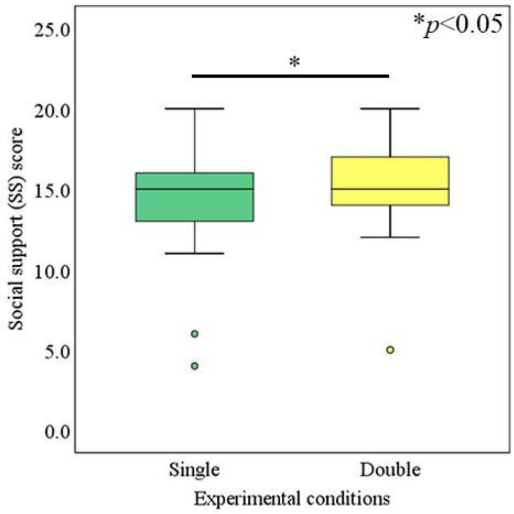

Figure 5.5: Social support (SS) score of operators.

5.4.3 Sense of being attended (SoBA)

Two paired sample t-test was conducted to identify the effect of the type of condition (single avatar vs double avatars) on the SoBA of the operator. The mean value of experienced SoBA for double avatars condition (*M*=18.29, *SD*=4.26) was not significantly higher than single avatar condition (*M*=18.13, *SD*=4.06), ($t(36) = -0.33$, $p = 0.74$, $d = 0.054$), see figure 5.6.

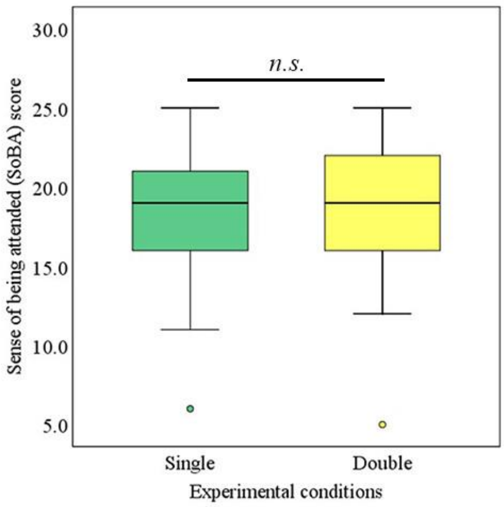

Figure 5.6: Sense of being attended (SoBA) score of operators.

5.5 Discussion

Experimental results revealed that operators talking through double avatars experiencing relatively higher communication support compared to those talking through a single avatar; where communication support is assessed by RoT, perceived SS, and SoBA indices. The operators using double-robot avatars perceived significantly higher SS in communication compared to operators using a single-robot avatar. Similarly, the operators using double-robot avatars felt significantly more RoT in conversation compared to operators using a single-robot avatar. However, there was no significant difference between the SoBA of operators using single or double robot avatars. It is suggested that the use of double avatars provides communication support to operators by manipulating their feelings of RoT and increasing their perception of being supported socially in conversation.

The main reason for a significant increase in RoT was the operator's feeling of ownership of individual RoTs of his/her two remote representations i.e., avatar agents. Humans tend to follow the rules of talk in conversation [145,146]. In our conversation setup, where a visitor interacts with two avatars simultaneously, the talk is perceived as a triadic interaction scenario by operators. In such a triadic interaction scenario, equal RoT is received by each peer of the conversation i.e., visitor and avatars. Now, since the operators teleoperated the avatars, and experiences through avatars are considered his/her own experiences [117] so, the individual RoT of each avatar would eventually become the operators' RoT; the main cause of observed effect, i.e., significantly increased RoT feelings of operators. Such an accumulative RoT would be a maximum of up to two folds. The main reason for a significant increase in SS was the operator's feeling of being supported by avatar agents in front of the visitor throughout the conversation; even at the time when there was a difference of opinions with the visitor. Social support is the information transferred to a subject that leads him/her to believe that he/she is cared for and loved, esteemed, and a member of a network of mutual obligations [148]. How and in which specific time frame such information is being transferred, are very important elements in influencing the beliefs of subjects [165]. Generally speaking, such information can be transferred to the subject by showing verbal, non-verbal, or both i.e., verbal, and non-verbal behaviors in a specific time frame. In our case, the robot avatars were transferring such information by showing non-verbal supportive behaviors to the operators in front of the visitors specifically at the time when they finish expressing their opinions. Such supportive behaviors were shown to the operators throughout the conversation session; the major cause of observed effect, i.e., significantly high perceived SS.

The operators talking through double avatars were expected to have significantly higher SoBA compared to operators talking through a single avatar. However, such an effect was not observed. Instead, a small increase in SoBA was observed. There could be two possible reasons concerning why such an effect was not observed for operators: 1) differences in the treatments of the visitors, and 2) the effect of partial occlusion in the field of view for the operator. In remote interactions, SoBA is more related to the treatment of the visitors towards the operator's remote representations i.e., avatar robot(s). Such treatments were expected to be perceived by operators as treatments to themselves. However, it seems that operators did not receive similar types of treatments due to varying degrees of visitors in terms of reproducing similar behaviors as instructed. So, it's quite natural to accept such a fact because the degree to which the given instructions are strictly followed varies from human to human [166–168]. Moreover, one of the avatars was placed in the line of sight of the operator i.e., in front of the camera in such a way that its rear side of the head and torso was evident; causing a small degree of visual occlusion in the field of view of the operator. For operators, such a visual occlusion hindered the process of having direct visual attention from visitors, and eventually influenced the quality of interaction [169]. So, the presence or absence of direct visual attention from visitors has an influence on the SoBA of operators [151].

Despite the communication support through double physical avatars, however, there are some limitations. Firstly, the degree to which each participant perceives a lack of SS and RoT in communication was not controlled. Further, the degree to which each participant prefers a specific type of SS and a specific timing of receiving SS was also not controlled. Secondly, I did not recruit participants with a severe lack of perceived SS, and RoT issues in communication. Thirdly, all of the material of the experiment was translated into Japanese; a specific linguistic and cultural background. Therefore, the observed effects do not necessarily guarantee reproducibility in the real world. To overcome such limitations of mere significant results with participants without severe lack of perceived SS and RoT issues, interactive experiments with individuals affected with severe lack of perceived SS and RoT issues using the proposed system in a more controlled way are required to observe the actual potential of the system in real-life and to draw more affirm conclusions.

Another major limitation of our study is the usage of non-verbal behaviors of avatar robots to influence the perceived SS of operators. I did not explore the effects of using verbal or a combination of verbal, and non-verbal behaviors of avatar robots on the perceived SS of operators. Moreover, I also did not explore the effect of time of provision of SS to operators. Further, whether the effect of RoT will keep on increasing as the number of avatar robots kept on increasing is also not explored. For simplicity, I only focused on the effects on the operators' side in the current manuscript. However, usage of the proposed system in real life would also require acceptance from the visitors; as their chances of being affected, due to operators' severe lack of SS and RoT in communication, are higher.

Besides limitations, some challenges would hinder the integration of the proposed system into daily life. In the beginning, it might be a challenging task to find the appropriate individuals with a severe lack of perceived SS and RoT issues in communication, and later train them to use such a system in daily life independently. Further, in subsequent stages, it might also be challenging to endure the cost of deployment of the system and later bear the maintenance cost along with multiple unforeseen technical and non-technical issues for which individuals with a severe lack of perceived SS and RoT will be completely dependent on service providers.

5.6 Conclusion and future work

In this study, I illustrated that a robotic video conference system, having two teleoperated robot avatar, increases the operator's feelings of RoT, and SS significantly in online conversations. While having remote experiences through both robot avatars simultaneously, the operator can speak through any of the robot avatars. Since talking through two teleoperated robot avatar will eventually become a triadic interaction scenario on the visitor's side, so he/she will be required to abide by the rules of triadic conversation. Such a situation is advantageous for the operators because, in the end, the individual RoT of each of the robot avatar will become operator's RoT. Similarly, the SS provided by the robots to each other will also become SS to operator; in front of the visitor throughout the conversation. Moreover, the operator's SoBA is also expected to be increased as the visitor has to be more attentive towards the operator through robot avatars. To verify such expected effects a field experiment was conducted with RoT, SS, and SoBA as measured indices. The experimental results showed the positive effect of using two avatars on the operator's RoT and SS while not on SoBA. In the future, I will examine the effect of using the proposed system for people with sever lack of perceived SS, and RoT issues in communication.

Chapter 6

Conclusion and future work

6.1 Conclusion:

In my research work, I explored few advantages of using multi-party communication scenarios; specifically related to controlling the commitment in communication and pressure of communication for operators. The results revealed that the usage of double avatars system helps in controlling such conversational factors for the operator. In an intuitional study i.e., study I, I implemented a multi-party communication system and got the intuition about the controlling of aforementioned factors. It was observed that attentional aspect of visitor can be manipulated, and such a phenomenon is expected to manipulate the conversational factors. Such an intuition motivated us to start exploring the advantages of using multi-party communication system for the operators; hence the following three studies carried out.

In study II, I explored that how to avoid the pressure of communication and get the commitment of visitors for operators; where avoidance of pressure of communication was assessed with the reduction of CA and AEC. While commitment in communication is obtained by diverting the visitor's attention towards own self while explaining the reasons for choosing a particular yes/no option. In study III, I explored that how to avoid the pressure of communication and give your own commitment of communication; where avoidance of pressure of communication was assessed with the reduction of stress of managing RT in communication, while commitment of communication is given by diverting responsibility towards the supported agent instead of operator for replying late in communication. In study IV, I explored that how to direct/give a pressure of communication towards visitors and get their commitment of communication; where directing/ giving the pressure of communication was assessed with the increased RoT and SS in communication. In the future, I will be exploring further possibilities for manipulating the proposed conversational factors for operators.

6.2 Future work of Ph.D.

The future work of the Ph.D. is focused on additional factors that can affect the proposed conversational factors i.e., giving and getting commitment, and avoiding and giving or directing pressure of communication. The proposed axes are further expanded to explore the possibilities, and directions of future work in four quadrants' regions, see figure 6.1. I have indicated a few such factors and provided a concise explanation.

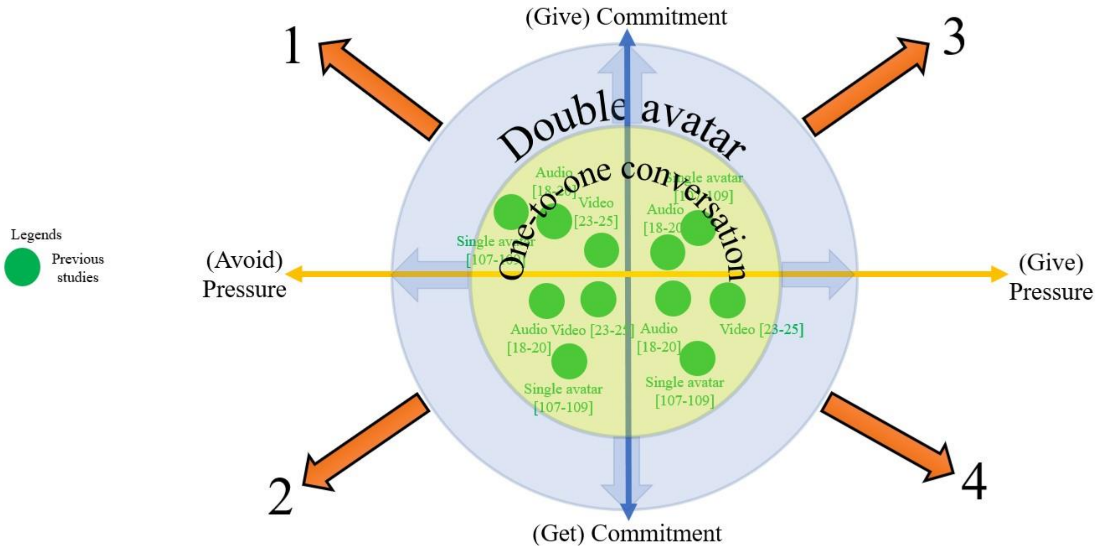

Figure 6.1: Research map and future work ideas

6.2.1 Avatar personality design

Personality plays a vital role in forming interpersonal relations with others [170]. The personalities of avatar can be designed or altered [171]. The degree to which the personality of avatar is designed depends upon an individual's own personality traits and preferences [172]. People prefer to change the personality of the avatars based on the context of conversation [173–175] and avatars personality affects the quality of the interaction [176–178]. Designing of personalities of avatar also incorporates avatar realism; defined as the degree to which an avatar is similar to its operator in terms of appearance, behavior, and personality. Increased degree of realism of avatars increases the naturalness of communication [179,180]. Avatar personality designing is a useful factor that can influence the conversational factors I am interested in e.g., getting/giving commitment, and directing/giving the pressure of communication, see figure 6.2. Considering the previous literature there could be two possible hypotheses that can be tested. They are as follows:

Hypothesis I: The operators who talk through avatars having competitive personalities perceive high communication competence of themselves compared to those who talk through avatars having non-competitive personalities.

Such a hypothesis will contribute towards the controlling the conversational factors i.e., (give) pressure, and (give) commitment. Imagine a situation where an operator has two avatars to talk with the visitor, where one avatar has competitive personality. The avatar with competitive personality will generate a sense of competition with visitor which will help in giving the pressure of conversation hence increasing the operator' feelings of competence in communication. While, on the other hand, in case of having a competition that visitors do not like, then operator can easily give the commitment to the avatar with competitive personality.

Hypothesis II: The operators who talk through avatars having cooperative personalities have a high willingness to communicate compared to those who talk through avatars having non-cooperative personalities.

Such a hypothesis will contribute towards (avoid) pressure, and (get) commitment factors of conversation. Imagine a situation where an operator has two avatars to talk with the visitor, where both avatars have cooperative personality. The presence and availability of avatars with cooperative personality will reduce the likelihood of seeking escape from the communication for apprehensive operators; hence increasing their willingness to communicate that will be avoiding pressure of communication. Furthermore, cooperating personality avatars will indirectly affect the visitors as well to get in line with the avatars and eventually becomes cooperative towards operators; (get) commitment.

6.2.2 Dialogue flow design

Designing the flow of the dialogue is an important factor that can affect the conversational factors I am interested in e.g., getting/giving commitment, and avoiding the pressure of communication, see figure 6.2. The design of flow of the dialogue can be controlled by integrating the empathy factor in it and establishing positive relations with the visitors [38,181–183]. Presence of empathy makes the communication effective [184], [185], while its absence generates frustration [186]. Considering the previous literature there could be one possible hypothesis that can be tested. It is as follows:

Hypothesis: The operators who talk through empathetic avatars can form better interpersonal relations compared to those who talk through non-empathetic avatars.

Such a hypothesis will contribute towards (avoid) pressure, and (get) commitment factors of conversation. Imagine a situation where an operator has two avatars to talk with the visitor, and both are empathetic towards visitors and operators. The presence of such empathetic avatars will help in calming down the situations where difference of opinions exists between two parties i.e., visitors and operators; (avoid) pressure. Sometimes, the empathetic avatars will talk in line with visitors that will encourage them to share similar type of experiences with details in communication [187]. Such a situation will be advantageous for the operators because they would like to take the credit; (get) commitment.

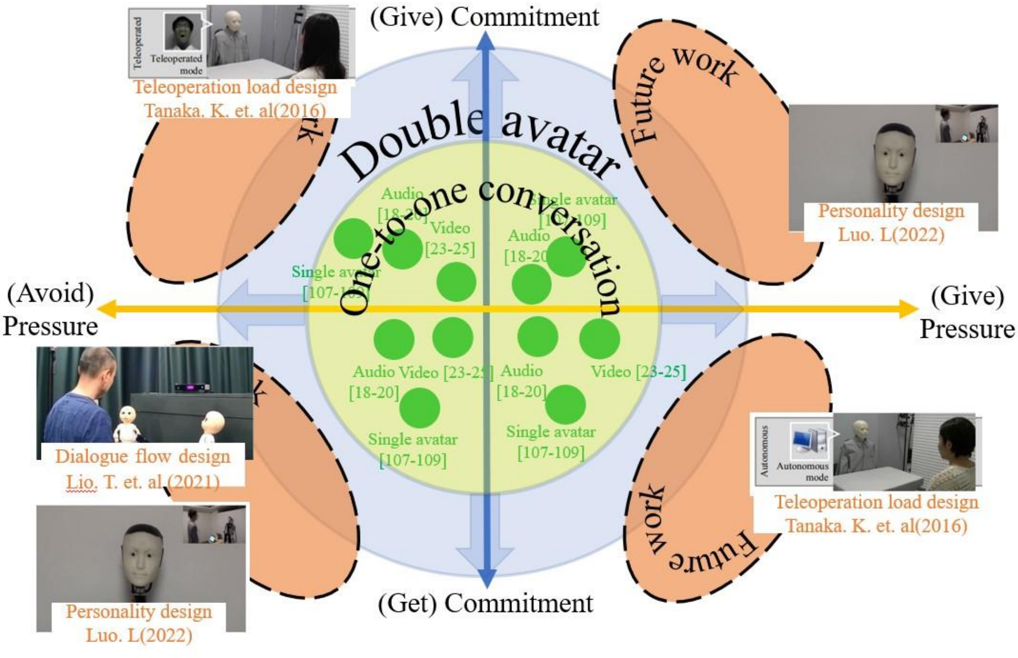

Figure 6.2: Future works

6.2.3 Teleoperation load design

The teleoperation load is an important factor that can affect the conversational factors I am interested in e.g., getting/giving commitment, and avoiding the pressure of communication, see figure 6.2. The teleoperation load can be controlled by manipulating the degree of autonomy of avatars. It influences the ease in communication for operators [188]. However, the change in teleoperation load of operators cannot be judged by the visitors [189]. Considering the previous literature there could be one possible hypothesis that can be tested. It is as follows:

Hypothesis: The operators having a low teleoperation load will abide by the rules of talk more compared to those operators having a high teleoperation load.

Such a hypothesis will contribute towards (avoid) pressure, (get) commitment, and (give) pressure, and (get) commitment factors of conversation.

Imagine a situation where an operator has two avatars to talk with the visitor; where one of the avatars has high teleoperation load while the other avatar has low teleoperation load. In such a situation, operators will get busy and will be unable to abide by the rules of talk so ease of talk will reduce. In that case, operator can give the commitment to the avatar for not properly following the rules of talk (e.g., unequal number of utterances, and unmatched speech convergence etc.) and hence (avoid) pressure.

On the other hand, imagine a situation where an operator has two avatars to talk with the visitor, and both have low teleoperation load. In such a situation, operators will be relieved and hence will be able to abide by the rules of talk so ease of talk will increase. In that case, operator can give the pressure of communication to visitor and consequently get the commitment for following the rules of talk more accurately.

Acknowledgment

I am very thankful to the prof. Ishiguro for providing me an opportunity to attain the highest level of education of my life; the one I started dreaming for years ago! I am also very thankful to my co-supervisors, prof. Yoshikawa, and prof. Hamed for providing me enormous guidance about the studies and help regarding the life in Japan throughout of doctoral course. I would also like to thank to prof. Iiguni, and prof. Sato for their constructive and valuable comments concerning the improvement of my research work. I am also very thankful to all the members of the lab; especially the sectaries, and students who helped me in the time when I faced different types of difficulties in daily life matters. Indeed, I have very nice memories with many! I also acknowledge and thank the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for providing me financial support and giving me an opportunity to experience a wonderful country Japan. Indeed, I felt Japan as second home.

Also, I am very thankful to my previous teachers i.e., prof. Yasar Ayaz, Dr. Sara Ali, and Ms. Haleema Sadia for providing me encouragement, and help in continuing the work even being separated geographically.

At the end, I am also very much thankful to my family who kept on providing me emotional and social support, to achieve another milestone in my life!

List of Publications

Journal Paper(s) [published]

1. **Mehmood, F**., Mahzoon, H., Yoshikawa, Y., Ishiguro, H., Sadia, H., Ali, S., & Ayaz, Y. (2021). Attentional Behavior of Children With ASD in Response to Robotic Agents. *IEEE Access*, 9, 31946-31955.
2. **Mehmood, F**., Mahzoon, H., Yoshikawa, Y., & Ishiguro, H. (2021). Communication Apprehension and Eye Contact Anxiety in Video Conferences Involving Teleoperated Robot Avatars: A Subjective Evaluation Study. *Frontiers in Robotics and AI, 8*.

Journal Paper(s) [Submitted]

1. **Mehmood, F.**, Mahzoon, H., Yoshikawa, Y., & Ishiguro, H. (202X). Avatars-mediated video conference system for mediating the stress and anxiety of response time of a person with communication difficulties. *Advanced Robotics*.
2. **Mehmood, F.**, Mahzoon, H., Yoshikawa, Y., & Ishiguro, H. (202X). Effect of having and switching multiple avatars on the operator's right to talk and social support. *Plos One*.

Conference Paper(s) [published]

1. **Mehmood, F.,** Mahzoon, H., Yoshikawa, Y., & Ishiguro, H. (2021, August). An interactive response strategy involving a robot avatar in a video conference system for reducing the stress of response time management in communication. In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN) (pp. 969-974). IEEE.

References:

1. Thomas GF, Tymon Jr WG, Thomas KW. Communication apprehension, interpretive styles, preparation, and performance in oral briefing. J Bus Commun. 1994;31(4):311–26.
2. Blume BD, Baldwin TT, Ryan KC. Communication apprehension: A barrier to students' leadership, adaptability, and multicultural appreciation. Acad Manag Learn Educ. 2013;12(2):158–72.
3. Drinkwater M, Vreken N. Communication apprehension as a factor influencing the quality of life of people. Inst Distance Educ. 1998;
4. Elwood TD, Schrader DC. Family communication patterns and communication apprehension. J Soc Behav Pers. 1998;13(3):493.
5. Punyanunt-Carter NM, Cruz JJD La, Wrench JS. Analyzing college students' social media communication apprehension. Cyberpsychology, Behav Soc Netw. 2018;21(8):511–5.
6. Ho SS, McLeod DM. Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. Communic Res. 2008;35(2):190–207.
7. McCROSKEY JC, RICHMOND VP, DALY JA, COX BG. THE EFFECTS OF COMMUNICATION APPREHENSION ON INTERPERSONAL ATTRACTION. Hum Commun Res. 1975;2(1).
8. McCroskey JC. The effects of communication apprehension on nonverbal behavior. Commun Q. 1976;24(1):39–44.
9. Remland MS, Jones TS. The Effects of Nonverbal Involvement and Communication Apprehension on State Anxiety, Interpersonal Attraction, and Speech Duration. Commun Q. 1989;37(3).
10. McCroskey JC, Richmond VP. The effects of communication apprehension on the perception of peers. West J Commun (includes Commun Reports). 1976;40(1):14–21.

11. Freimuth VS. The effects of communication apprehension on communication effectiveness. Hum Commun Res. 1976;2(3):289–98.

12. Thurlow C, Lengel L, Tomic A. Computer mediated communication. Julia Hall, Jamilah Ahmed, Fabienne Pedroletti, editors. Oliver’s Yard, 55 city road, London, UK: Sage; 2004. 1–231 p.

13. Pierce T. Social anxiety and technology: Face-to-face communication versus technological communication among teens. Comput Human Behav. 2009;25(6):1367–72.

14. High AC, Caplan SE. Social anxiety and computer-mediated communication during initial interactions: Implications for the hyperpersonal perspective. Comput Human Behav. 2009;25(2):475–82.

15. Keaten JA, Kelly L. “Re: We really need to talk”: Affect for communication channels, competence, and fear of negative evaluation. Commun Q. 2008;56(4):407–26.

16. Wu L. Investigating social presence in audio and text online interaction for language learning. Chinese J Appl Linguist. 2020;43(2).

17. Kehrwald B. Understanding social presence in text-based online learning environments. Distance Educ. 2008;29(1).

18. Rosen S, Tesser A. Fear of negative evaluation and the reluctance to transmit bad news. J Commun. 1972;22(2):124–41.

19. Tesser A, Rosen S, Waranch E. Communicator mood and the reluctance to transmit undesirable messages (the mum effect). J Commun. 1973;23(3):266–83.

20. Nautsch A, Jiménez A, Treiber A, Kolberg J, Jasserand C, Kindt E, et al. Preserving privacy in speaker and speech characterisation. Comput Speech Lang. 2019;58.

21. Le T Van, Cunningham U, Watson K. The relationship between willingness to communicate and social presence in an online English language course. JALT CALL J. 2018;14(1).

22. Monson SJ, Wolcott LL, Seiter JS. Communication Apprehension in Synchronous Distance Education. 1999;

23. Dupagne M, Stacks DW, Giroux VM. Effects of video streaming technology on public speaking students' communication apprehension and competence. J Educ Technol Syst. 2007;35(4):479–90.

24. Campbell J. Media richness, communication apprehension and participation in group videoconferencing. J Information, Inf Technol Organ. 2006;1.

25. der Zwaard R Van, Bannink A. Video call or chat? Negotiation of meaning and issues of face in telecollaboration. System. 2014;44(1).

26. Xu H, Cai Z, Takabi D, Li W. Audio-Visual Autoencoding for Privacy-Preserving Video Streaming. IEEE Internet Things J. 2022;9(3).

27. Venkatesh MV, Zhao J, Profitt L, Cheung SCS. Audio-visual privacy protection for video conference. In: Proceedings - 2009 IEEE International Conference on Multimedia and Expo, ICME 2009. 2009.

28. Neustaedter C, Greenberg S, Boyle M. Blur filtration fails to preserve privacy for home-based video conferencing. ACM Trans Comput Interact. 2006;13(1).

29. Borup J, West RE, Thomas RA, Graham CR. Examining the impact of video feedback on instructor social presence in blended courses. Int Rev Res Open Distance Learn. 2014;15(3).

30. Cukor P, Baer L, Willis BS, Leahy L, O'Laughlen J, Murphy M, et al. Use of videophones and low-cost standard telephone lines to provide a social presence in telepsychiatry. Telemed J. 1998;4(4).

31. Mehmood F, Mahzoon H, Yoshikawa Y, Ishiguro H, Sadia H, Ali S, et al. Attentional Behavior of Children With ASD in Response to Robotic Agents. IEEE Access. 2021;9:31946–55.

32. Hayashi K, Sakamoto D, Kanda T, Shiomi M, Koizumi S, Ishiguro H, et al. Humanoid robots as a passive-social medium-a field experiment at a train station. In: 2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2007. p. 137–44.

33. Sakamoto D, Hayashi K, Kanda T, Shiomi M, Koizumi S, Ishiguro H, et al. Humanoid robots as a broadcasting communication medium in open public spaces. Int J Soc Robot. 2009;1(2):157–69.

34. Iio T, Yoshikawa Y, Ishiguro H. Pre-scheduled turn-taking between robots to make conversation coherent. In: HAI 2016 - Proceedings of the 4th International Conference on Human Agent Interaction. 2016.

35. Shiomi M, Okumura S, Kimoto M, Iio T, Shimohara K. Two is better than one: Social rewards from two agents enhance offline improvements in motor skills more than single agent. PLoS One. 2020;15(11):e0240622.

36. Shiomi M, Hagita N. Do synchronized multiple robots exert peer pressure? In: Proceedings of the fourth international conference on human agent interaction. 2016. p. 27–33.

37. Tae MI, Ogawa K, Yoshikawa Y, Ishiguro H. Using Multiple Robots to Increase Suggestion Persuasiveness in Public Space. Appl Sci. 2021;11(13):6080.

38. Iio T, Yoshikawa Y, Ishiguro H. Double-meaning agreements by two robots to conceal incoherent agreements to user's opinions. Adv Robot. 2021;1–11.

39. Iio T, Yoshikawa Y, Ishiguro H. Retaining human-robots conversation: comparing single robot to multiple robots in a real event. J Adv Comput Intell Intell Informatics. 2017;21(4):675–85.

40. Nishio T, Yoshikawa Y, Ogawa K, Ishiguro H. Development of an Effective Information Media Using Two Android Robots. Appl Sci. 2019;9(17):3442.

41. Association DSAP, Association AP, others. Diagnostic and statistical manual of mental disorders: DSM-5. Vol. 5. American psychiatric association Washington, DC; 2013.

42. Barbaro J, Halder S. Early identification of autism spectrum disorder: Current challenges and future global directions. Curr Dev Disord Reports. 2016;3(1):67–74.

43. Dawson G, Bernier R, Ring RH. Social attention: a possible early indicator of efficacy in autism clinical trials. J Neurodev Disord. 2012;4(1):1–12.

44. Zwaigenbaum L, Bryson S, Garon N. Early identification of autism spectrum disorders. Behav Brain Res. 2013;251:133–46.

45. Barbaro J, Dissanayake C. Early markers of autism spectrum disorders in infants and toddlers prospectively identified in the Social Attention and Communication Study. Autism. 2013;17(1):64–86.

46. Kou J, Le J, Fu M, Lan C, Chen Z, Li Q, et al. Comparison of three different eye-tracking tasks for distinguishing autistic from typically developing children and autistic symptom severity. Autism Res. 2019;12(10):1529–40.

47. Schopler E, Reichler RJ, Renner BR. The childhood autism rating scale (CARS). WPS Los Angeles; 2010.

48. Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). J Autism Dev Disord. 1980;

49. Bastiaansen JA, Meffert H, Hein S, Huizinga P, Ketelaars C, Pijnenborg M, et al. Diagnosing autism spectrum disorders in adults: the use of Autism Diagnostic Observation Schedule (ADOS) module 4. J Autism Dev Disord. 2011;41(9):1256–66.

50. Molloy CA, Murray DS, Akers R, Mitchell T, Manning-Courtney P. Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. Autism. 2011;15(2):143–62.

51. Ozonoff S, Goodlin-Jones BL, Solomon M. Evidence-based assessment of autism spectrum disorders in children and adolescents. J Clin Child Adolesc Psychol. 2005;34(3):523–40.

52. Landa R, Garrett-Mayer E. Development in infants with autism spectrum disorders: a prospective study. J child Psychol psychiatry. 2006;47(6):629–38.

53. Falck-Ytter T, Fernell E, Hedvall ÅL, Hofsten C Von, Gillberg C. Gaze performance in children with autism spectrum disorder when observing communicative actions. J Autism Dev Disord. 2012;42(10):2236–45.

54. Bharatharaj J, Huang L, Krägeloh C, Elara MR, Al-Jumaily A. Social engagement

of children with autism spectrum disorder in interaction with a parrot-inspired therapeutic robot. Procedia Comput Sci. 2018;133:368–76.

55. Chung EY. Robotic intervention program for enhancement of social engagement among children with autism spectrum disorder. J Dev Phys Disabil. 2019;31(4):419–34.

56. Jones EA, Carr EG, Feeley KM. Multiple effects of joint attention intervention for children with autism. Behav Modif. 2006;30(6):782–834.

57. Liang A, Piroth I, Robinson H, MacDonald B, Fisher M, Nater UM, et al. A pilot randomized trial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc. 2017;18(10):871–8.

58. Lane GW, Noronha D, Rivera A, Craig K, Yee C, Mills B, et al. Effectiveness of a social robot,“Paro,” in a VA long-term care setting. Psychol Serv. 2016;13(3):292.

59. Moyle W, Jones CJ, Murfield JE, Thalib L, Beattie ERA, Shum DKH, et al. Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. J Am Med Dir Assoc. 2017;18(9):766–73.

60. Hansen SG, Raulston TJ, Machalicek W, Frantz R. Caregiver-mediated joint attention intervention. Behav Interv. 2018;33(2):205–11.

61. Kourassanis-Velasquez J, Jones EA. Increasing joint attention in children with autism and their peers. Behav Anal Pract. 2019;12(1):78–94.

62. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng. 2012;14:275–94.

63. Stephenson J, Limbrick L. A review of the use of touch-screen mobile devices by people with developmental disabilities. J Autism Dev Disord. 2015;45(12):3777–91.

64. Bellani M, Fornasari L, Chittaro L, Brambilla P. Virtual reality in autism: state of the art. Epidemiol Psychiatr Sci. 2011;20(3):235–8.

65. Robins B, Dautenhahn K, Dubowski J. Does appearance matter in the interaction

of children with autism with a humanoid robot? Interact Stud. 2006;7(3):479–512.

66. Kumazaki H, Yoshikawa Y, Yoshimura Y, Ikeda T, Hasegawa C, Saito DN, et al. The impact of robotic intervention on joint attention in children with autism spectrum disorders. Mol Autism. 2018;9(1):1–10.

67. O'Hearn K, Lakusta L, Schroer E, Minshew N, Luna B. Deficits in adults with autism spectrum disorders when processing multiple objects in dynamic scenes. Autism Res. 2011;4(2):132–42.

68. Adamson LB, Bakeman R, Suma K, Robins DL. An expanded view of joint attention: Skill, engagement, and language in typical development and autism. Child Dev. 2019;90(1):e1--e18.

69. Zhao S, Uono S, Yoshimura S, Kubota Y, Toichi M. Atypical gaze cueing pattern in a complex environment in individuals with ASD. J Autism Dev Disord. 2017;47(7):1978–86.

70. David DO, Costescu CA, Matu S, Szentagotai A, Dobrean A. Developing joint attention for children with autism in robot-enhanced therapy. Int J Soc Robot. 2018;10(5):595–605.

71. Admoni H, Scassellati B. Social eye gaze in human-robot interaction: a review. J Human-Robot Interact. 2017;6(1):25–63.

72. Ali S, Mehmood F, Ayaz Y, Khan MJ, Sadia H, Nawaz R. Comparing the Effectiveness of Different Reinforcement Stimuli in a Robotic Therapy for Children with ASD. IEEE Access. 2020;8.

73. Rinehart NJ, Bradshaw JL, Moss SA, Brereton A V, Tonge BJ. A deficit in shifting attention present in high-functioning autism but not Asperger's disorder. Autism. 2001;5(1):67–80.

74. Townsend J, Harris NS, Courchesne E. Visual attention abnormalities in autism: Delayed orienting to location. J Int Neuropsychol Soc. 1996;2(6):541–50.

75. McLay L, Schäfer MCM, van der Meer L, Couper L, McKenzie E, O'Reilly MF, et al. Acquisition, preference and follow-up comparison across three AAC modalities taught to two children with autism spectrum disorder. Int J Disabil Dev

Educ. 2017;64(2):117–30.

76. Bertone A, Mottron L, Jelenic P, Faubert J. Motion perception in autism: a “complex” issue. J Cogn Neurosci. 2003;15(2):218–25.

77. Cheung PPP, Siu AMH. A comparison of patterns of sensory processing in children with and without developmental disabilities. Res Dev Disabil. 2009;30(6):1468–80.

78. McCroskey JC. Oral communication apprehension: A reconceptualization. Ann Int Commun Assoc. 1982;6(1):136–70.

79. Schulze L, Renneberg B, Lobmaier JS. Gaze perception in social anxiety and social anxiety disorder. Front Hum Neurosci. 2013;7:872.

80. Schneier FR, Rodebaugh TL, Blanco C, Lewin H, Liebowitz MR. Fear and avoidance of eye contact in social anxiety disorder. Compr Psychiatry. 2011;52(1):81–7.

81. Moukheiber A, Rautureau G, Perez-Diaz F, Soussignan R, Dubal S, Jouvent R, et al. Gaze avoidance in social phobia: objective measure and correlates. Behav Res Ther. 2010;48(2):147–51.

82. Hodge RL. Interpersonal classroom communication through eye contract. Theory Pract. 1971;10(4):264–7.

83. Argyle M, Dean J. Eye-contact, distance and affiliation. Sociometry. 1965;289–304.

84. Howell AN, Zibulsky DA, Srivastav A, Weeks JW. Relations among social anxiety, eye contact avoidance, state anxiety, and perception of interaction performance during a live conversation. Cogn Behav Ther. 2016;45(2):111–22.

85. Rice RE. Media appropriateness: Using social presence theory to compare traditional and new organizational media. Hum Commun Res. 1993;19(4):451–84.

86. Georgakopoulou A. Computer-mediated communication. Pragmat Pract. 2011;9:93.

87. Jr NLR, Lewis P V. Communication apprehension as a determinant of channel

preferences. J Bus Commun. 1984;21(3):53–61.

88. Oh CS, Bailenson JN, Welch GF. A systematic review of social presence: Definition, antecedents, and implications. Front Robot AI. 2018;5:114.

89. Borup J, West RE, Graham CR. Improving online social presence through asynchronous video. Internet High Educ. 2012;15(3):195–203.

90. Calefato F, Lanubile F. Communication media selection for remote interaction of ad hoc groups. In: Advances in Computers. Elsevier; 2010. p. 271–313.

91. Mehmood F, Mahzoon H, Yoshikawa Y, Ishiguro H. An interactive response strategy involving a robot avatar in a video conference system for reducing the stress of response time management in communication. In: 2021 30th IEEE International Conference on Robot \& Human Interactive Communication (RO-MAN). 2021. p. 969–74.

92. Leeds EM, Maurer RA. Using digital video technology to reduce communication apprehension in business education. INFORMS Trans Educ. 2009;9(2):84–92.

93. Sautter EP, Zúñiga MA. The video cover letter: embedded assessment of oral communication skills. Qual Assur Educ. 2018;

94. Scott CR, Timmerman CE. Relating computer, communication, and computer-mediated communication apprehensions to new communication technology use in the workplace. Communic Res. 2005;32(6):683–725.

95. Ko C-J. The effect of task types on foreign language learners' social presence in synchronous computer mediated communication (SCMC). JALT CALL J. 2016;12(2):103–22.

96. Daly-Jones O, Monk A, Watts L. Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. Int J Hum Comput Stud. 1998;49(1):21–58.

97. Isaacs EA, Tang JC. What video can and cannot do for collaboration: a case study. Multimed Syst. 1994;2(2):63–73.

98. Bohannon LS, Herbert AM, Pelz JB, Rantanen EM. Eye contact and video-

mediated communication: A review. Displays. 2013;34(2):177–85.

99. Wieser MJ, Pauli P, Alpers GW, Mühlberger A. Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. J Anxiety Disord. 2009;23(1):93–103.

100. Weeks JW, Howell AN, Goldin PR. Gaze avoidance in social anxiety disorder. Depress Anxiety. 2013;30(8):749–56.

101. O'Malley C, Langton S, Anderson A, Doherty-Sneddon G, Bruce V. Comparison of face-to-face and video-mediated interaction. Interact Comput. 1996;8(2):177–92.

102. Langer JK, Rodebaugh TL. Social anxiety and gaze avoidance: Averting gaze but not anxiety. Cognit Ther Res. 2013;37(6):1110–20.

103. Bondareva Y, Meesters L, Bouwhuis D. Eye contact as a determinant of social presence in video communication. In: Proceedings of the 20th International Symposium on Human Factors in Telecommunication. 2006.

104. Meadows MS. I, avatar: The culture and consequences of having a second life. New Riders; 2007.

105. Castronova E. Theory of the Avatar. Available SSRN 385103. 2003;

106. Shimaya J, Yoshikawa Y, Kumazaki H, Matsumoto Y, Miyao M, Ishiguro H. Communication support via a tele-operated robot for easier talking: case/laboratory study of individuals with/without autism spectrum disorder. Int J Soc Robot. 2019;11(1):171–84.

107. Børsting J, Culén AL. A Robot Avatar: Easier Access to Education and Reduction in Isolation? 2016;

108. Cheung CW, Tsang IT, Wong KH. Robot Avatar: A Virtual Tourism Robot for People With Disabilities. Int J Comput Theory Eng. 2017;9(3):229–34.

109. Lee JK, Stiehl WD, Toscano RL, Breazeal C. Semi-autonomous robot avatar as a medium for family communication and education. Adv Robot. 2009;23(14):1925–49.

110. Straub I, Nishio S, Ishiguro H. Incorporated identity in interaction with a teleoperated android robot: A case study. In: 19th International Symposium in Robot and Human Interactive Communication. 2010. p. 119–24.

111. Choi JJ, Kwak SS. Who is this?: Identity and presence in robot-mediated communication. Cogn Syst Res. 2017;43:174–89.

112. Tanaka K, Nakanishi H, Ishiguro H. Physical embodiment can produce robot operator's pseudo presence. Front ICT. 2015;2:8.

113. Gleason B, Greenhow C. Hybrid education: The potential of teaching and learning with robot-mediated communication. Online Learn J. 2017;21(4).

114. McCroskey JC. An introduction to rhetorical communication. Routledge; 2015.

115. Heerink M, Kröse B, Evers V, Wielinga B. Assessing acceptance of assistive social agent technology by older adults: the almere model. Int J Soc Robot. 2010;2(4):361–75.

116. Maurer B, Lankes M, Tscheligi M. Where the eyes meet: Lessons learned from shared gaze-based interactions in cooperative and competitive online games. Entertain Comput. 2018;27:47–59.

117. Morita T, Mase K, Hirano Y, Kajita S. Reciprocal attentive communication in remote meeting with a humanoid robot. In: Proceedings of the 9th international conference on Multimodal interfaces. 2007. p. 228–35.

118. Hietanen JK, Leppänen JM, Peltola MJ, Linna-aho K, Ruuhiala HJ. Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. Neuropsychologia. 2008;46(9):2423–30.

119. Chakrabarti C, Luger GF. Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics. Expert Syst Appl. 2015;42(20):6878–97.

120. Ashktorab Z, Jain M, Liao QV, Weisz JD. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019. p. 1–12.

121. Brown SA, Fuller RM, Vician C. Who's afraid of the virtual world? Anxiety and computer-mediated communication. J Assoc Inf Syst. 2004;5(2):2.

122. Rockwell S, Singleton L. The effects of computer anxiety and communication apprehension on the adoption and utilization of the Internet. Electron J Commun. 2002;12(1):2.

123. Yoshida A, Yamaguchi T, Yamazaki K. Quantitative Study of Human Behavior in Virtual Interview Sessions for the Development of the Hyper Hospital–A Network Oriented Virtual Reality Based Novel Medical Care System–. IEICE Trans Inf Syst. 1994;77(12):1365–71.

124. Yoshida A, Hagita Y, Yamazaki K, Yamaguchi T. Which do you feel comfortable, interview by a real doctor or by a virtual doctor? A comparative study of responses to inquiries with various psychological intensities, for the development of the Hyper Hospital. In: Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication. 1993. p. 370–4.

125. Mulligan K, Grant JT, Mockabee ST, Monson JQ. Response latency methodology for survey research: Measurement and modeling strategies. Polit Anal. 2003;11(3):289–301.

126. Levinson SC. Action formation and ascription. Handb Conversat Anal. 2013;1:103–30.

127. Torreira FJ, Bögels S, Levinson SC. Breathing for answering. The time course of response planning in conversation. 2016;

128. Strömbergsson S, Hjalmarsson A, Edlund J, House D. Timing responses to questions in dialogue. In: INTERSPEECH. 2013. p. 2584–8.

129. Yan T, Tourangeau R. Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. Appl Cogn Psychol Off J Soc Appl Res Mem Cogn. 2008;22(1):51–68.

130. Olson K, Smyth JD. The effect of CATI questions, respondents, and interviewers on response time. J Surv Stat Methodol. 2015;3(3):361–96.

131. Richmond VP. The relationship between trait and state communication

apprehension and interpersonal perceptions during acquaintance stages. Hum Commun Res. 1978;4(4):338–49.

132. Deffenbacher JL, Payne DM. Relationship of apprehension about communication to fear of negative evaluation and assertiveness. Psychol Rep. 1978;

133. Herring SC. Computer-mediated communication: Linguistic, social, and cross-cultural perspectives. Vol. 39. John Benjamins Publishing; 1996.

134. Ayres J, Ayres FE, Baker AL, Colby N, Blasi C De, Dimke D, et al. Two empirical tests of a videotape designed to reduce public speaking anxiety. 1993;

135. Hara N, Bonk CJ, Angeli C. Content analysis of online discussion in an applied educational psychology course. Instr Sci. 2000;28(2):115–52.

136. Harasim LM. Networlds: Networks as social space. Glob networks Comput Int Commun. 1993;15–34.

137. Kitchen D, McDougall D. Collaborative learning on the Internet. J Educ Technol Syst. 1999;27(3):245–58.

138. Short J, Williams E, Christie B. The Social Psychology of Telecommunications [Internet]. Wiley; 1976. Available from: https://books.google.co.jp/books?id=GMdQPwAACAAJ

139. Keil M, Johnson RD. Feedback channels: Using social presence theory to compare voice mail to e-mail. J Inf Syst Educ. 2002;13(4):295.

140. Logan DE, Breazeal C, Goodwin MS, Jeong S, O'Connell B, Smith-Freedman D, et al. Social robots for hospitalized children. Pediatrics. 2019;144(1).

141. Li J. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int J Hum Comput Stud. 2015;77:23–37.

142. Bainbridge WA, Hart JW, Kim ES, Scassellati B. The benefits of interactions with physically present robots over video-displayed agents. Int J Soc Robot. 2011;3(1):41–52.

143. Leary MR. A brief version of the Fear of Negative Evaluation Scale. Personal Soc

Psychol Bull. 1983;9(3):371–5.

144. Teachman BA, Allen JP. Development of social anxiety: Social interaction predictors of implicit and explicit fear of negative evaluation. J Abnorm Child Psychol. 2007;35(1):63–78.

145. Kanfer FH, Bass BM, Guyett I. Dyadic speech patterns, orientation, and social reinforcement. J Consult Psychol. 1963;27(3).

146. Bilous FR, Krauss RM. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. Lang Commun. 1988;8(3–4).

147. Natale M, Entin E, Jaffe J. Vocal interruptions in dyadic communication as a function of speech and social anxiety. J Pers Soc Psychol. 1979;37(6).

148. Cobb S. Social support as a moderator of life stress. Psychosom Med. 1976;38(5).

149. MacIntyre PD, Baker SC, Clément R, Conrod S. WILLINGNESS TO COMMUNICATE, SOCIAL SUPPORT, AND LANGUAGE-LEARNING ORIENTATIONS OF IMMERSION STUDENTS. Stud Second Lang Acquis. 2001;23(3).

150. Dour HJ, Wiley JF, Roy-Byrne P, Stein MB, Sullivan G, Sherbourne CD, et al. Perceived social support mediates anxiety and depressive symptom changes following primary care intervention. Depress Anxiety. 2014;31(5).

151. Mehmood F, Mahzoon H, Yoshikawa Y, Ishiguro H. Communication Apprehension and Eye Contact Anxiety in Video Conferences Involving Teleoperated Robot Avatars: A Subjective Evaluation Study. Front Robot AI. 2021;8.

152. Walther JB, Loh T, Granka L. Let me count the ways the interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. Vol. 24, Journal of Language and Social Psychology. 2005.

153. Jepson K. Conversations—and negotiated interaction—in text and voice chat rooms. Lang Learn Technol. 2005;9(3):79–98.

154. González-Lloret M. Conversation analysis in computer-assisted language learning.

Vol. 32, CALICO Journal. 2015.

155. Chen Y, Xu Y. Exploring the effect of social support and empathy on user engagement in online mental health communities. Int J Environ Res Public Health. 2021;18(13).

156. Apgar D. The use of group text messaging to enhance social support of social work students. Soc Work Educ. 2020;39(7).

157. Rudy RR, Rosenfeld LB, Galassi JP, Parker J, Schanberg R. Participants' Perceptions of a Peer-Helper, Telephone-Based Social Support Intervention for Melanoma Patients. Health Commun. 2001;13(3).

158. Johnstone T, Scherer KR. Vocal communication of emotion. Handb Emot. 2000;2:220–35.

159. Thorsteinsson EB, James JE, Elizabeth Gregg M. Effects of video-relayed social support on hemodynamic reactivity and salivary cortisol during laboratory-based behavioral challenge. Heal Psychol. 1998;17(5).

160. O'Conaill B, Whittaker S, Wilbur S. Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. Human–Computer Interact. 1993;8(4).

161. Tu CH. The impacts of text-based CMC on online social presence. J Interact Online Learn. 2002;1(2).

162. Campbell JA, Campus L. Communication apprehension and participation in videoconferenced meetings. In: Proceedings of the Tenth Australasian Conference on Information Systems. 1999. p. 160–70.

163. Lee JK, Toscano RL, Stiehl WD, Breazeal C. The design of a semi-autonomous robot avatar for family communication and education. In: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN. 2008.

164. Oommen D. The Relationships among Perceptions of Social Support, Intercultural Communication Apprehension (ICA), and Conflict Management Preferences in the Context of Cultural Adaptation. J Intercult Commun Res. 2014;43(3).

165. Jacobson DE. Types and timing of social support. J Health Soc Behav. 1986;250–64.

166. Lindwall O, Ekström A. Instruction-in-interaction: The teaching and learning of a manual skill. Hum Stud. 2012;35(1):27–49.

167. Dunham S, Lee E, Persky AM. The psychology of following instructions and its implications. Am J Pharm Educ. 2020;84(8).

168. Amerine R, Bilmes J. Following instructions. Hum Stud. 1988;327–39.

169. Radmard S, Moon Aj, Croft EA. Impacts of Visual Occlusion and Its Resolution in Robot-Mediated Social Collaborations. Int J Soc Robot. 2019;11(1):105–21.

170. Borkenau P, Mauer N, Riemann R, Spinath FM, Angleitner A. Thin Slices of Behavior as Cues of Personality and Intelligence. J Pers Soc Psychol. 2004;86(4).

171. Luo L, Ogawa K, Ishiguro H. Identifying Personality Dimensions for Engineering Robot Personalities in Significant Quantities with Small User Groups. Robotics. 2022;11(1).

172. Dunn RA, Guadagno RE. My avatar and me--Gender and personality predictors of avatar-self discrepancy. Comput Human Behav. 2012;28(1):97–106.

173. Triberti S, Durosini I, Aschieri F, Villani D, Riva G. Changing avatars, changing selves? The influence of social and contextual expectations on digital rendition of identity. Cyberpsychology, Behav Soc Netw. 2017;20(8):501–7.

174. Martey RM, Stromer-Galley J, Banks J, Wu J, Consalvo M. The strategic female: gender-switching and player behavior in online games. Information, Commun \& Soc. 2014;17(3):286–300.

175. Velez JA, Loof T, Smith CA, Jordan JM, Villarreal JA, Ewoldsen DR. Switching schemas: Do effects of mindless interactions with agents carry over to humans and vice versa? J Comput Commun. 2019;24(6):335–52.

176. Peña J, Yoo S-C. Under pressure: Avatar appearance and cognitive load effects on attitudes, trustworthiness, bidding, and interpersonal distance in a virtual store. Presence. 2014;23(1):18–32.

177. Fong K, Mar RA. What does my avatar say about me? Inferring personality from avatars. Personal Soc Psychol Bull. 2015;41(2):237–49.

178. Elliott R, Bohart AC, Watson JC, Greenberg LS. Empathy. Psychotherapy. 2011;48(1):43.

179. Herrera F, Oh SY, Bailenson JN. Effect of behavioral realism on social interactions inside collaborative virtual environments. Presence Teleoperators Virtual Environ. 2020;27(2).

180. Bailenson JN, Yee N, Merget D, Schroeder R. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. Presence Teleoperators Virtual Environ. 2006;15(4):359–72.

181. Paiva A, Leite I, Boukricha H, Wachsmuth I. Empathy in virtual agents and robots: A survey. ACM Trans Interact Intell Syst. 2017;7(3).

182. Leite I, Pereira A, Mascarenhas S, Martinho C, Prada R, Paiva A. The influence of empathy in human-robot relations. Int J Hum Comput Stud. 2013;71(3).

183. Niculescu A, van Dijk B, Nijholt A, Li H, See SL. Making social robots more attractive: the effects of voice pitch, humor and empathy. Int J Soc Robot. 2013;5(2):171–91.

184. Valente F. Empathy and Communication: A Model of Empathy Development. J New Media Mass Commun. 2016;3(1).

185. Gompertz K. The Relation of Empathy to Effective Communication. Journal Q. 1960;37(4).

186. Derksen F, Hartman TCO, van Dijk A, Plouvier A, Bensing J, Lagro-Janssen A. Consequences of the presence and absence of empathy during consultations in primary care: A focus group study with patients. Patient Educ Couns. 2017;100(5).

187. Ickes W, Stinson L, Bissonnette V, Garcia S. Naturalistic Social Cognition: Empathic Accuracy in Mixed-Sex Dyads. J Pers Soc Psychol. 1990;59(4).

188. Osawa M, Okuoka K, Takimoto Y, Imai M. Is automation appropriate? semi-

autonomous telepresence architecture focusing on voluntary and involuntary movements. Int J Soc Robot. 2020;12(5):1119–34.

189. Tanaka K, Yamashita N, Nakanishi H, Ishiguro H. Teleoperated or autonomous?: How to produce a robot operator's pseudo presence in HRI. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2016. p. 133–40.

Appendixes

Chapter 03: The SoBA questionnaire and conversational scripts of experiment.

Questionnaire to assess the sense of being attended (SoBA) to of the interviewee.

Item	Questions
	The interviewee felt that the interviewer:
1	carefully listened to his answer?
2	was interested in his answer?
3	was attentive to his answer?
4	was trying to understand his answer.

* Internal consistency α = 0.81.

Human condition (Human in Experiment-I and Human (averted) in Experiment-II):

In human conditions (see Fig. 3.3 and 3.8), the participants were discussing earning money through unfair means and paying taxes. "*Q*" represents the interviewer's question and "*A*" represents the interviewee's answer in his/her voice.

Interviewer**:** *Q1*: If you get a chance to earn a money that is not 100 % fair near you, will you go for that?

Interviewee**:** *A1*: No, I will not go for that.

Interviewer**:** *Q2*: Why? Why will you not go for that?

Interviewee**:** *A2*: Well, I think it is ethically not correct and it is important for me. So, this is the reason I will not go for that.

Robot condition:

In the Robot condition of Experiment-I (Fig. 3.4 (B)) and II (Fig. 3.9 (B)), the participants were discussing the same issue as in the Human conditions. "Q" represents the interviewer's question and "A" represents the interviewee's answer, where the first answer is given by the teleoperated robot avatar, whereas the second answer is the interviewer's own voice.

Interviewer: *Q1*: If you get a chance to earn a money that is not 100 % fair near you, will you go for that?

Interviewee: *A1*: No, I think I do not. (*Robot avatar utterance*).

Interviewer: *Q2*: Why? Why will you not go to this? You (pointing gesture toward the online conference monitor with web camera so that interviewee perceived that the pointing gesture was toward him/herself) can think with him (pointing gesture to robot avatar) and propose some answer.

Interviewee *A2*: Well, I think it is ethically incorrect and it is important for me. This is the reason I will not go for that.

Chapter 04: Conversational scripts of both subjective evaluation experiments.

Subjective evaluation experiment I:

Questions asked by the interlocutor in video stimuli of the conventional and proposed video conference system. Here, "*Q*" is a question and "*A*" is the answer. The participant (i.e., interviewee) answered all questions by typing on the tablet. The language of the conversation was English.

Interlocutor: *Q1*: If you get a chance to earn money that is not 100 % fair near you, will you go for that?

Participant: *A1*: [participant's answer].

Interlocutor: *Q2*: Why? Why will you not go for that? Please think about it.

Participant: *A2*: [participant's answer].

Interlocutor: *Q3*: Do you think that paying for tax actually reduces the overall household income of a person?

Participant: *A3*: [participant's answer].

Interlocutor: *Q4*: Why? Why do you not think so? Please think about it.

Participant: *A4*: [participant's answer].

Subjective evaluation experiment II:

Conventional System:

In the conventional system (see Fig. 4.9 (a)), participants discussed earning money unethically and paying taxes. "*Q*" represents a question asked by the interviewer, and "*A*" represents an answer, given by the interviewee through an avatar agent (robot).

Interviewer*:Q1*: If you get a chance to earn money that is not 100 % fair near you, will you go for that?

Interviewee*:A1*: No, I will not go for that.

Interviewer*:Q2*: Why? Why will you not go for that? Please think about it.

Interviewee*:A2*: because it is not fair in my opinion. So, I will not go for that.

Interviewer*:Q3*: Do you think that paying for tax reduces the overall household income of a person?

Interviewee*:A3*: No, I do not think so.

Interviewer*:Q4*: Why? Why do you not think so? Please think about it.

Interviewee*:A4*: Because it is the money that all the people should pay.

Proposed system:

In the proposed system (see Fig. 4.2, and Fig 4.14), participants discussed the same money-related issue. "*Q*" represents a question asked by the interviewer, and "*A*" represents an answer given by the interviewee either through the teleoperated avatar agent or supporting agent.

Interviewer*:Q1*: If you get a chance to earn money that is not 100 % fair near you, will you go for that?

Interviewee*:A1*: No, I will not go for that [from avatar agent].

Interviewer*:Q2*: Why? Why will you not go for that? Please think about it.

Interviewee*:A2*: He said, "because it is not fair in my opinion. So, I will not go for that." [from supporter agent]

Interviewer*:Q3*: Do you think that paying for tax reduces the overall household income of a person?

Interviewee*:A3*: No, I do not think so. [from avatar agent]

Interviewer*:Q4*: Why? Why do you not think so? Please think about it.

Interviewee*:A4*: He said, “because it is the money that all the people should pay.” [from supporter agent]

Chapter 05: The conversational scripts of experiment and RoT and SS scales.

Conversation Script # 1: Love VS Money

＊灰色の部分は相手ロボットを手のひらで指しながら、話してください。

相手: こんにちは。

自分：…….

相手: 今日は「人は、より良い人生を送るために、愛を選ぶべきか、お金を選ぶべきかの選択を迫られたらどちらを選ぶのか」というトピックで、私と議論をしていただきたいと思います。よろしいでしょうか。

自分: …….

相手: このトピックについてあなたはどういうご意見をおもちですか?

自分の回答を以下から選んでください。

●自分：私は、愛を選びます。[分岐 1]

●自分：私は、お金を選びます。[分岐 2]

[分岐１]

自分: 私は、愛を選びます。

相手: なるほど！その意見ももっともですね。しかし、私はもう一つの意見も重要だと思っています。私はより良い人生を送るためには、愛ではなくお金を選ぶべきだと思います。お金でしか手にできないものはとても多いと思うのですが、そうは思いませんか？

自分：.......

相手: それでも将来の人生を考えますと、お金を選ぶ方が有益だと思うのですが、そうではないでしょうか。

自分：.......

相手: そうでしょうか。実際は、ほぼ全てのものがお金で買えると思います。人生においても、例えば家を買ったり、車を買ったり、その他の贅沢品もお金で買えると思います。賛成されませんか？

自分：.......

相手: 愛は後からどうにでもなると思います。というのも、お金があれば、経済的な面でも安定しますし、生活においてお金が足りないかもしれないという不安からも解放されます。そうすれば、愛についても真剣に考えたり時間を割いたりできると思います。そうではないでしょうか。

自分：.......

相手: 今の段階で、あなたは先ほどのご自分の意見（お金よりも愛を選ぶ）を変えたいとは思いませんか？

自分の回答を以下から選んでください。

- はい 、意見を変えます。[分岐 1-1]
- いいえ、意見は変えません。[分岐 1-2]

[分岐 1-1]

自分: はい 、意見を変えます。

相手: なぜあなたは意見を変えたのですか？

自分：…….

相手: つまり、愛よりもお金を選ぶべきである、という事でしょうか？

自分：…….

相手: そうですね！ちなみに、自分の実際の生活で、愛を選ぶべきか、お金を選ぶべきかの選択を迫られたら、今ならあなたはどうしますか？

自分：…….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分：…….

相手: どうしてそう思うのですか?

自分：…….

相手: 分かりました。お時間をいただき、ありがとうございました。

[分岐 1-2]

自分: いいえ、意見は変えません。

相手: どうしてそう思うのですか?

自分：…….

相手: つまり、お金を選ぶよりも愛を選ぶ方が重要だという事でしょうか。

自分：…….

相手: そうですね！ちなみに、自分の実際の生活で、愛を選ぶべきか、お金を選ぶべきかの選択を迫られたら、今ならあなたはどうしますか？

自分：…….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分：…….

相手: どうしてそう思うのですか?

自分：…….

相手: 分かりました。お時間をいただき、ありがとうございました。

[分岐 2]

自分：私はお金を選びます。

相手: なるほど！その意見ももっともですね。しかし、私はもう一つの意見も重要だと思っています。私はより良い人生を送るためには、お金よりも愛を選ぶのが重要だと思います。愛でしか手に入らないものはとても多いと思うのですが、そうは思いませんか？

自分：…….

相手: それはそうですが、普通に暮らしていればある程度のお金は入ってくるはずだと思います。なので、愛を優先しても、精神的にも健康な人生を送れるはずだし、それは結局収入も安定してくることに繋がると思うのですが、 そうは思いませんか？

自分：…….

相手: それでも、そのようなお金を用意するために求められるだろう困難な過程でもまた、サポートしてくれるパートナーがいたほうが、より幸せな自分を送れるのではないかと思うのですが、これについては同意されませんでしょうか？

自分：…….

相手: しかし早いうちからパートナーがいると、一人では経験できない人生の色々な困難や楽しみ、達成感や社会的な側面などを早いうちから経験でき、その結果将来的に親になるとしても自分たちは影響力があるしっかりとした保護者になれるかと思うのですが、そうではないですか？

自分：…….

相手: 今の段階で、あなたは先ほどのご自分の意見（愛よりもお金を選ぶ）を変えたいとは思いませんか？

自分の回答を以下から選んでください。

- はい、意見を変えます 。[分岐 2-1]
- いいえ、意見を変えません。[分岐 2-2]

[分岐 2-1]

自分: はい、意見を変えます 。

相手: なぜあなたは意見を変えたのですか？

自分：…….

相手: つまり、愛の方がお金よりも重要という事でしょうか？

自分：…….

相手: そうですね！ちなみに、自分の実際の生活で、愛を選ぶべきか、お金を選ぶべきかの選択を迫られたら、今ならあなたはどうしますか？

自分：…….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分：…….

相手: どうしてそう思うのですか?

自分：…….

相手: 分かりました。お時間をいただき、ありがとうございました。

[分岐 2-2]

自分: いいえ、意見を変えません。

相手: どうしてそう思うのですか?

自分：…….

相手: つまり、お金は愛よりも重要であるという事ですか？

自分：…….

相手: そうですね！ちなみに、自分の実際の生活で、愛を選ぶべきか、お金を選ぶべきかの選択を迫られたら、今ならあなたはどうしますか？

自分：…….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分：…….

相手: どうしてそう思うのですか?

自分：…….

相手: 分かりました。お時間をいただき、ありがとうございました。

Conversation Script # 2: Car Accident

＊灰色の部分は相手ロボットを手のひらで指しながら、話してください。

相手:こんにちは。

自分: ………….

相手: 今日は「交通事故で、『子供一人』と『老人二人』のどちらかを救えるとしたら、どちらを救うべきか」というトピックについて、私と議論をしていただきたいと思います。よろしいでしょうか。

自分: ………….

相手: このトピックについてあなたはどういうご意見はどうですか?

ご自分の回答を以下から選んでください。

- 自分: 子供一人を救うべきだと思います。[分岐 1]
- 自分: 老人二人を救うべきだと思います。[分岐 2]

[分岐 1]

自分: 子供一人を救うべきだと思います。

相手: なるほど！その意見ももっともですね。 しかし、私はもう一つの意見も重要だと思っています。私は、子供一人の代わりに、老人二人を助ける方が良いと思います。二つの命と一つの命を天秤にかけたらやはり二人の方が大事なのではないでしょうか。

自分: ………….

相手: しかし、いくらなんでも二つの命に勝るものでしょうか。とてもじゃないですが、そうは思えないのですが。

自分: ………….

相手: 子供の方が年齢が若く、寿命までまだ遠いのはわかります。しかし、後で裁判になったとしても同じことが言えるでしょうか。より多くの命をなぜ救わなかったのかと批判を浴びる可能性があると思うのですが、これには賛成しませんか？

自分: ………….

相手: そうでしょうか。この考えは他の多くの人から支持されると思いますし、そう考えるとあなたも精神的にだいぶ安らぐのではないでしょうか。

自分: ………….

相手: 今の段階で、あなたは先ほどのご自分の意見（子供一人を救うべき）を変えたいとは思いませんか？

ご自分の回答を以下から選んでください。

- ●はい 、意見を変えます。[分岐 1-1]
- ●いいえ、意見は変えません。[分岐 1-2]

[分岐 1-1]

自分: はい 、意見を変えます。

相手: なぜあなたは意見を変えたのですか？

自分: ………….

相手: つまり、二つの命を救う方が、一つの命を救う事よりも、例えそれが子供だったとしても、重要であるという事ですか。

自分: ………….

相手: そうですね！ちなみに、自分の実際の生活で、交通事故があって、『子供一人』と『老人二人』のどちらかしか救えない選択を迫られたら、今ならあなたはどうしますか？

自分: ………….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分: ………….

相手: どうしてそう思うのですか?

自分: ………….

相手: 分かりました。お時間をいただき、ありがとうございました。

[分岐 1-2]

自分: いいえ、意見は変えません。

相手: どうしてそう思うのですか?

自分: ………….

相手: つまり、子供であれば、老人二人と比べると優先されるべきである、という事ですか。

自分: ………….

相手: そうですね！ちなみに、自分の実際の生活で、交通事故があって、『子供一人』と『老人二人』のどちらかしか救えない選択を迫られたら、今ならあなたはどうしますか？

自分: ………….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分: ………….

相手: どうしてそう思うのですか?

自分: ………….

相手: 分かりました。お時間をいただき、ありがとうございました。

[分岐 2]

自分: 老人二人を救うべきだと思います。

相手: なるほど！ その意見ももっともですね。 しかし、私はもう一つの意見も重要だと思っています。私は、子供を救う方が老人二人を救う事よりも優先されるべきであると思います。子供の命の重みは老人と比べると違うと思いますので。

自分: ………….

相手: 一般的に、ひとりひとりの命の価値が同等なのはそうだと思いますが、子供は自分の家族にとってたった一人の子供かもしれません。その場合、その子はその家族の将来の重要な支えとなるはずで、社会的な意味もだいぶ変わってきます。

自分: ………….

相手: たしかに老人でも子供でも、家族にとってはかけがえのない命です。しかし、国や国民にとっては、将来的な展望としても子供の方が重要で偉大な人物に、将来的になっていたかもしれません。そのような観点からみてもやはり子供を救う意義は大きいのではないでしょうか。

自分: ………….

相手: 少し言いにくいですが、子供の家族は、その家族の未来である子供を助けてくれたとして、生涯あなたをサポートしてくれるかもしれませんよ。老人の場合、そういうことはあまりないのではないでしょうか。

自分: ………….

相手: 今の段階で、あなたは先ほどのご自分の意見（老人二人を救うべき）を変えたいとは思いませんか？

自分の回答を以下から選んでください。

- はい、意見を変えます 。[分岐 2-1]
- いいえ、意見を変えません。[分岐 2-2]

[分岐 2-1]

自分: はい、意見を変えます 。

相手: なぜあなたは意見を変えたのですか？

自分: ………….

相手: つまり、子供一人を救う方が、二人の老人を救う事よりも大事である、という事ですか？

自分: ………….

相手: そうですね！ちなみに、自分の実際の生活で、交通事故があって、『子供一人』と『老人二人』のどちらかしか救えない選択を迫られたら、今ならあなたはどうしますか？

自分: ………….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分: ………….

相手: どうしてそう思うのですか?

自分: ………….

相手: 分かりました。お時間をいただき、ありがとうございました。

[分岐 2-2]

自分: いいえ、意見を変えません。

相手: どうしてそう思うのですか?

自分: ………….

相手: つまり、例え老いた命であったとしても、二つの命は一つの若い命と比べて優先されるべきであるという事ですか？

自分: ………….

相手:　そうですね！ちなみに、自分の実際の生活で、交通事故があって、『子供一人』と『老人二人』のどちらかしか救えない選択を迫られたら、今ならあなたはどうしますか？

自分: ………….

相手: なるほど。その選択は、現実的にも考えて、できそうだと思いますか？

自分: ………….

相手: どうしてそう思うのですか?

自分: ………….

相手: 分かりました。お時間をいただき、ありがとうございました。

The Right to talk scale and its validity analysis:

Right to talk (RoT) scale English version:

How did you feel during the conversation? Please evaluate it using the following:

About your right to talk:

1. I could talk normally.
2. I had a fair order to talk.
3. I could talk sufficiently during the conversation.
4. I could express my opinion.
5. The order of the conversation was assigned to me properly.
6. I could express all my sentences without any interruption.

Right to talk (RoT) scale Japanese version:

雑談中にどのように感じていたと思うか、以下で評価を行なってください。 自分の話す権利について [options: 全くそう思わない, そう思わない, どちらかといえばそう思わない, どちらでもない, どちらかといえばそう思う, そう思う, とてもそう思う]

1. 私は普通に話せていた。
2. 私には公平に話す順番が与えられていた。
3. 私は雑談で十分話せていた。
4. 私は自分の意見を言えていた。
5. 話す順番がちゃんと私にも回って来ていた。
6. 私の話はさえぎられず、ちゃんと最後まで話せた。

Right to talk (RoT) scale factor analysis:

Exploratory Factor analysis table of right to talk (ROT) scale			
Sr.	Items	Factor-I Loadings	Commonalities
1	私は普通に話せていた。	0.967	0.935
2	私には公平に話す順番が与えられていた。	0.922	0.850
3	私は雑談で十分話せていた。	0.973	0.947
4	私は自分の意見を言えていた。	0.938	0.881
5	話す順番がちゃんと私にも回って来ていた。	0.969	0.938
6	私の話はさえぎられず、ちゃんと最後まで話せた。	0.840	0.706
	Eigen Value	5.25	
	Percentage of the total variance	87.61%	
	Total variance	87.61%	

Method: PCA; Rotation: Varimax

The social support scale and its validity analysis:

Social support (SS) scale English version:

How did you feel during the conversation? Please evaluate it using the followings:

About social support:

1. They cared for me properly.
2. The member(s) of the conversation expressed affection for me.
3. The member(s) of the conversation respected me.
4. I was treated properly as a member of the conversation.

Social support (SS) scale Japanese version:

雑談中にどのように感じていたと思うか、以下で評価を行なってください。 社会的なサポートについて [options: 全くそう思わない, そう思わない, どちらでもない, そう思う, とてもそう思う]

1. 私はちゃんと気遣ってもらえていた。
2. 会話のメンバーは私に愛着の念を抱いていた。
3. 会話のメンバーは私に尊敬の念を抱いていた。
4. 私は会話のメンバーの一員としての適切な扱いをうけていた。

Social support (SS) scale Factor Analysis:

Exploratory Factor analysis table of social support (SS) scale			
Sr.	Items	Factor-I Loadings	Commonalities
1	私はちゃんと気遣ってもらえていた。	0.969	0.938
2	会話のメンバーは私に愛着の念を抱いていた。	0.964	0.929
3	会話のメンバーは私に尊敬の念を抱いていた。	0.930	0.865
4	私は会話のメンバーの一員としての適切な扱いをうけていた。	0.969	0.939
	Eigen Value	3.67	
	Percentage of the total variance	91.77%	
	Total variance	91.77%	

Method: PCA; Rotation: Varimax

Sense of being attended to (SoBA) scale- English version:

To what extent were you attended to during the conversation? Please evaluate this using the following:

1. I could convey my opinion to the interlocutor.
2. The interlocutor was interested in my opinions.
3. The interlocutor understood my answers.
4. The interlocutor paid attention to me through my avatar (the robot).
5. I could communicate with the interlocutor naturally.

Sense of being attended to (SoBA) scale- Japanese version:

会話にどれほど参加できたかを以下で教えてください（以下で「相手」とはロボットと対話していた人のことです）。[options: 全くそう思わない, そう思わない, どちらでもない, そう思う, とてもそう思う]

1. 私は相手に自分の意見を伝えることができた。
2. 相手は私の意見に興味を示していた。
3. 相手は私の答えを理解していた。
4. 相手は私のアバター（ロボット）を通して私に注目していた。
5. 相手と自然に会話ができた。