HIGH SOCIAL ACCEPTANCE OF HEAD GAZE LOOSELY SYNCHRONIZED WITH SPEECH FOR SOCIAL ROBOTS

A Dissertation

by

VASANT SRINIVASAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Robin R. Murphy
Committee Members,	Dezhen Song
	Dylan A. Shell
	Thomas K. Ferris
	Cindy L. Bethel
Head of Department,	Nancy M. Amato

August 2014

Major Subject: Computer Engineering

ABSTRACT

This research demonstrates that robots can achieve socially acceptable interactions, using loosely synchronized head gaze-speech, without understanding the semantics of the dialog. Prior approaches used tightly synchronized head gaze-speech, which requires significant human effort and time to manually annotate synchronization events in advance, restricting interactive dialog, and requiring the operator to act as a puppeteer. This approach has two novel aspects. First, it uses affordances in the sentence structure, time delays, and typing to achieve autonomous synchronization of head gaze-speech. Second, it is implemented within a behavioral robotics framework derived from 32 previous implementations. The efficacy of the loosely synchronized approach was validated through a 93-participant 1 x 3 (loosely synchronized head gaze-speech, tightly synchronized head gaze-speech, no-head gaze-speech) between-subjects experiment using the "Survivor Buddy" rescue robot in a victim management scenario. The results indicated that the social acceptance of loosely synchronized head gaze-speech is similar to tightly synchronized head gaze-speech (manual annotation), and preferred to the no head gaze-speech case. These findings contribute to the study of social robotics in three ways. First, the research overall contributes to a fundamental understanding of the role of social head gaze in social acceptance, and the production of social head gaze. Second, it shows that autonomously generated head gaze-speech coordination is both possible and acceptable. Third, the behavioral robotics framework simplifies creation, analysis, and comparison of implementations.

DEDICATION

To my Mom and Dad, this is for you.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Robin Murphy for her patient guidance, enthusiastic encouragement, and useful critiques of this research work. I would like to thank my committee members Cindy Bethel, Dylan Shell, Dezhen Song, and Thomas Ferris for providing me with feedback, advice, and assistance throughout my doctoral education.

A big shout out to the Survivor Buddy family: Zack, Jessica, Aaron, and Jesus for their hard work and help in making this possible. I was fortunate to have the best team anyone could hope for, and this research work just wouldn't be the same without them.

I would like to thank my lab members Brittany Duncan and Joshua Peschel for their encouragements and feedback on this research work. I owe special thanks to Victoria Groom, Cliff Nass, and Leila Takayama for providing valuable help and guidance throughout this journey.

Thanks also go to my government and corporate sponsors. This work was supported by the National Science Foundation Grant IIS - 0905485 "The Social Medium is the Message," and Microsoft External Research.

My parents Srini and Charu, and my sister Rekha deserve special thanks for their unwavering love and support.

Finally, my wife and best friend Kripa, thank you for believing in me and rocking my world.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION *

A key component of social interaction between a robot and human(s) is social head gaze, which serves five important functions - *look interested in human(s)* [1–18], *engage in a verbal conversation with human(s)* [1, 2, 6, 9–11, 13, 16, 18–23], *convey general liveliness and awareness* [4, 8, 23], *show various mental states* [4, 21, 24–26], and *referential gaze* to objects in the environment [1, 2, 4, 8, 16, 27–32]. Two of these functions, *engage in a verbal conversation with human(s)* and *referential gaze* to objects in the environment, require tight synchronization, or precise timing between the speech utterance and the activation of the corresponding head gaze act. While *engaging in a verbal conversation* with a human, the robot generates *fixate* and *avert* head gaze acts that are tightly synchronized with speech, to facilitate turn-taking. If the topic of the discussion is an object in the environment, the robot uses *referential gaze* to *fixate* toward the object 800 msec to 1 sec, before it utters the object's name. The tight synchronization between head gaze and speech (TSHG-S) has been well modeled in human-human literature [33–36], and ensures high quality communication between humans. However, social robots using models for turn-taking in conversations [33, 34] and referential gaze for looking at objects in the environment [35, 36] suffer from three limitations. First, TSHG-S requires manual annotation and semantic content understanding. This requires significant human effort and time. Second, if the robot uses a preset library to select appropriate head gaze behaviors, the head gaze cannot be generated in open-ended, interactive scenarios. This is problematic when the social robot's verbal responses cannot be

*Reprinted with permission from "Evaluation of Head Gaze Loosely Synchronized with Synthetic Speech for Social Robots" by Srinivasan, V., Bethel, C.L., Murphy, R.R, 2014. IEEE Transactions on Human-Machine Systems, Accepted for Publication, Copyright [2014] by Vasant Srinivasan.

anticipated *a priori*. Third, tight synchronization of head gaze and speech that mimics human gaze may not be feasible due to limitations of the robot. These might include the absence of a high degree of motor control, flexibility of joint movements, and/or velocity limits.

This work proposes a more practical implementation alternative to TSHG-S, that of a loosely synchronized head gaze, with real-time synthetic speech (LSHG-S). In LSHG-S, the timing between the speech utterance and the activation of the corresponding head gaze act is flexible, that is, the activation of the head gaze act can lead, lag, or occur at the onset of the speech utterance. A large-scale 93-participant human-user study was conducted to evaluate the social acceptance of LSHG-S in human-robot interaction. This section begins with the primary and secondary research questions, which are discussed in Section 1.1. Section 1.2 defines social head gaze and discusses the importance of using social head gaze in a social interaction. Section 1.3 details LSHG-S, a more practical implementation alternative to TSHG-S. Contributions of this research are presented in Section 1.4. Finally, an outline of the organization of this dissertation work is presented in Section 1.5.

1.1 Research Questions

The primary research question that this work addresses is: *What is a computational theory of social head gaze for social agents?*

The goal of this research is to capture a computational theory of social head gaze as a programmable framework, so that head gaze-speech acts can be autonomously generated. Existing approaches in human-robot interaction requires significant human effort and time to manually annotate synchronization events in advance, restricts interactive dialog, or requires that the operator acts as a puppeteer. The lack of autonomous or consistent implementation of social head gaze is a barrier to conducting

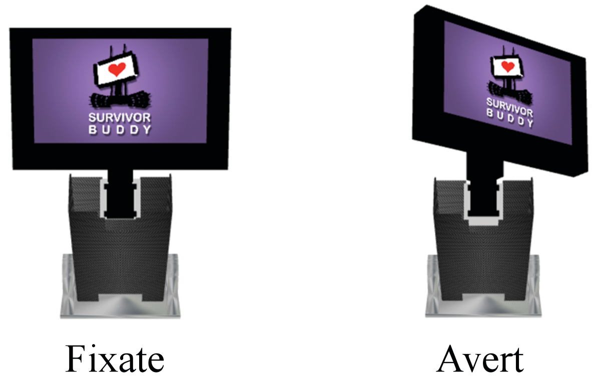


Figure 1.1: Survivor Buddy *Engages in a Verbal Conversation with a Human* Using Fixate and Avert Head Gaze Acts.

reproducible research in the area of affective computing. The fundamental research question poses four related questions:

1. *What is the appropriate set of social head gaze behaviors required for a naturalistic human-robot interaction?*

   Identifying an appropriate set of percepts, corresponding affordances from the sentence structure, time delays, and typing, head gaze acts, behaviors, and coordination function simplifies implementation, and ensures that the robot will not generate unintended social consequences due to head gaze. As one step towards creating a programmable framework, this research surveys 32 distinct human-robot interaction studies of social head gaze in Section 2, and identifies a set of head gaze acts and percepts that have been successful in making a robot comforting, more socially consistent, and predictable to the user(s). Sections 3 and 4 detail the substitution of affordances for linguistic and internal percepts of head gaze, so that the state of an interaction can be inferred autonomously. Affordances are conditions or objects that are directly perceivable without any memory, inference, or interpretation [37].

2. *How can social head gaze be expressed as behaviors or schemas, which are common representations in both psychology and robotics?*

   Expressing social head gaze in behavioral robotics terms translates the qualitative understanding of head gaze into a well-known, tangible implementation framework. Drawing on the literature and observations discussed in Section 2, social head gaze is mapped unto a behavioral robotics framework in Section 3. The behavioral robotics framework expresses the social head gaze phenomena as behaviors using well-established conventions in artificial intelligence by Arkin [38] and Murphy [37], and enables a robotic implementation in Section 4. The behavioral robotics framework captures the commonalities, essence, and experience of a collection of systems through mining and generalization of their implementations [39].

3. *Is it possible to evaluate through sound experimental methods the effectiveness and appropriateness of the head gaze acts generated using the behavioral robotics framework?*

   This question is addressed in Sections 5 and 6 through a large-scale 93-participant experiment that was conducted with a semi-anthropomorphic robot, "Survivor Buddy" (Fig. 1.1) for a victim management application. Five hypotheses are evaluated to assess the social acceptance of LSHG-S and TSHG-S, and to determine if LSHG-S is adequate for human-robot interaction. The five hypotheses are:

   (a) *Hypothesis 1 (H1)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

(b) *Hypothesis 2 (H2)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

(c) *Hypothesis 3 (H3)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

(d) *Hypothesis 4 (H4)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

(e) *Hypothesis 5 (H5)*: The LSHG-S condition improvements over the NHG-S condition will be comparable to those of the TSHG-S condition.

4. *Does the level of synchronization between gaze acts and speech impact the naturalistic perception of the social interaction?*

   Section 7 presents an analysis of the results which suggests that LSHG-S elicits high levels of social acceptance when compared to NHG-S, and is adequate for human-robot interaction. The section discusses four factors that may impact user perception of head gaze-speech synchronization – (1) *Gesture comprehension is temporally more flexible than gesture production*, (2) *Expectation of gaze in a semi-humanoid robot*, (3) *Importance of synchronization at the start and end of turns is greater than at the middle of turns*, and (4) *Absence of lips.*

1.2 Social Head Gaze

Social head gaze is defined as the pattern of head and body orientation that expresses engagement in the current context. It is important for many human-robot interaction applications where a naturalistic interaction is desired such as: healthcare [24, 26], victim management [14, 40], robot guides [8, 10, 11, 13, 41], entertainment [1,2,4,5,9,16], telepresence [42,43], and fundamental research [3,6,7,15,17,20,27–29]. The six known benefits of social head gaze for human-robot interaction are:

1. Increased task performance [2, 4–9, 13, 15, 25, 28, 41].
2. Increased engagement [5–8, 11, 15, 25, 41].
3. Improved perception of a robot's physical, social and intellectual characteristics [1–3, 6, 8, 9, 13, 15, 20, 25, 41, 44].
4. Increased attributions of mind and intentionality to the robot [1, 4, 6, 8, 25].
5. Increased positive affective state [6, 8, 12, 45].
6. Improved attentiveness to the robot and task [8, 13, 27, 28].

1.3 Loosely Synchronized Head Gaze-Speech: A Practical Alternative to Tightly Synchronized Head Gaze-Speech

This research examines the use of affordances to generate LSHG-S for human-robot interaction. The occurrence of turn events and semantics in dialog that activate head gaze acts can be substituted with affordances from the sentence structure of dialog, time delays, and typing. These affordances are computationally trivial, support autonomous generation, are independent of semantics, and are useful for interactive, open-ended conversations. If a robot is an autonomous agent and can generate dialog, the sentence structure and time delays will be transparently available to the robot, and the proposed method can be used. In the case of a tele-operated robot [43, 46] or Wizard-of-Oz experiment [47], the proposed method can be utilized if the robot operator can provide the dialog, from which the sentence structure, time delays, and typing can be determined.

However, there are two potential problems with using LSHG-S:

1. The loosely synchronized gaze acts may not precisely match the dialog presented because there is no semantic understanding in the proposed approach. The lag between the robot's speech and gaze acts may annoy or confuse the human.

2. A social robot that interacts with a human in a realistic scenario will possibly use all five functions of head gaze. The change in synchronization for two functions such as *engage in a verbal conversation with human(s)* and *referential gaze* may impact the human's perception of the robot's overall head gaze during the interaction.

Therefore, a human-robot interaction experiment was conducted to evaluate the social acceptance of the proposed LSHG-S method, and determine if it was adequate for human-robot interaction.

1.4 Contributions

This research provides seven contributions to the *social robotics community*. These contributions are categorized in the areas of fundamental science, social benefits, and economic benefits, arranged in the order of abstraction (abstract to implementation-specific).

1. *Fundamental Science:* This work makes four contributions to science:

   (a) The findings contribute to a fundamental understanding of the role of social head gaze in social acceptance, particularly with regard to how social head gaze can be produced, the question of when less competence is tolerable, and the importance of the speech and gaze synchronization for the listener.

(b) It shows that autonomously generated head gaze-speech coordination is both possible and acceptable. Researchers and practitioners do not have to manually annotate every situation using the Wizard-of-Oz approach [47].

(c) It provides social robotics researchers and practitioners with a formal vocabulary for social head gaze, enabling future implementations that are autonomous, consistent, repeatable, and natural.

(d) This work contributes five new measures for victim management - Person at Ease, Robot Empathy, Robot Integrity, Robot Loyalty, and Robot Caring.

2. *Social Benefits:* The robot can generate socially acceptable head gaze behaviors in real-time for very open-ended, interactive scenarios. These advantages are very important in situations where robot responses cannot be anticipated *a priori* (e.g. personal robots for eldercare).

3. *Economic Benefits:* The economic impact relates primarily to the amount of labor involved and costs required for the modification of existing robots. This research contributes a novel mechanism for inferring percepts from sentence structure, time delays, and typing that is independent of the semantics of dialog. The method reduces the workload of researchers, since they are no longer required to tediously hand code every scenario, Wizard-of-Oz style [47]. The behavioral robotics framework is applicable to a wide variety of robots (anthropomorphic, non-anthropomorphic).

4. *Implementation Benefits:* The behavioral robotics framework simplifies creation, analysis, and comparison of social head gaze implementations.

1.5 Organization of Thesis

The dissertation is organized as follows: Section 2 provides a brief summary of 32 studies that use some aspect of social head gaze in human-robot interaction and analyzes each study in terms of: *Head Gaze-Speech Synchronization*, *Percepts*, *Head Gaze Acts*, *Robots Used*, *Group Configuration*, *Implementation Styles*, *Tasks*, and *Measures*. Section 3 introduces affordances, describes the behavioral robotics framework, and sets the foundation for the computational theory. Section 4 presents the implementation of LSHG-S on a rescue robot for victim management. Section 5 describes the details of an experiment conducted with 93 participants, designed to evaluate the social acceptance of LSHG-S and TSHG-S. Section 6 presents the data analysis and results that demonstrate the social acceptance of LSHG-S and TSHG-S. Section 7 interprets the results of the experiment, discusses the factors that might influence the naturalness of head gaze-speech synchronization, and the limitations of the experiment. Section 8 reaffirms that both the LSHG-S and TSHG-S implementations elicited high levels of social acceptance, and that LSHG-S is adequate for human-robot interaction. It also provides a summary of the contributions of the research, design implications, and directions for future work.

2. RELATED WORK

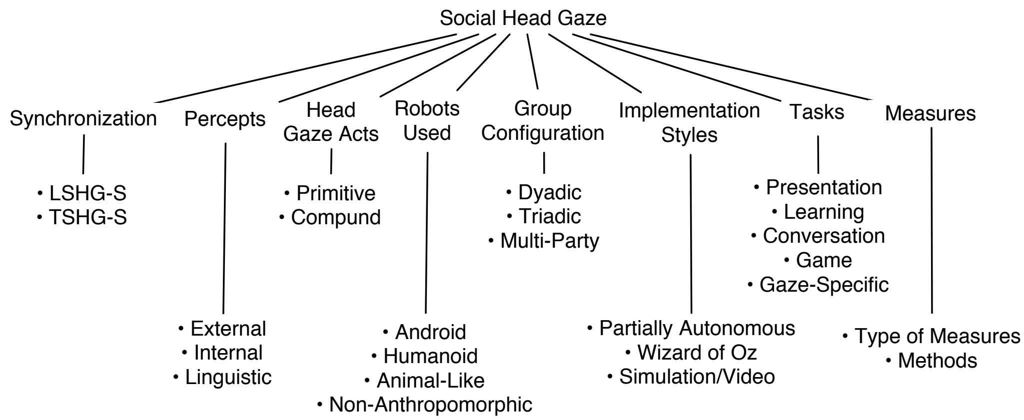


Figure 2.1: Taxonomy of Social Head Gaze.

Social head gaze involves a coordination of head and body orientation that is sensitive to the social context [48]. A review of the literature on human-robot interaction reveals that 32 distinct studies [1–11, 13–20, 23–32, 40–43, 49] have addressed some aspect of the social head gaze generation of robotics.

The human-robot literature on generation of head gaze for conversation [1, 2, 7–11, 13, 16, 18–20, 25, 28, 29, 31] is concerned with TSHG-S. Without exception, each of the 16 major studies identified [1, 2, 7–11, 13, 16, 18–20, 25, 28, 29, 31] implement models of human-human interaction for conversational turn-taking [33, 34, 50] and referential head gaze [35, 36], or rely on a human-human interaction experiment, conducted specifically for determining the synchronization events [13, 20, 25]. Since the synchronization events for TSHG-S require interpretation or inference, manual annotation is the most popular method adopted for the generation of TSHG-S.

The purpose of this review is to: (1) examine the current state-of-the-art for head gaze-speech synchronization in social robotics, (2) identify an appropriate set of percepts and head gaze acts that have been successful in making a robot comforting, more socially consistent, and predictable to the user(s), and (3) examine prior implementations and evaluations to guide experimental design and validation of this work. This section provides a summary of social head gaze in human-robot interaction. Each of the 32 studies are then analyzed in terms of: *synchronization*, *percepts*, *head gaze acts*, *robots used*, *group configuration*, *implementation styles*, *tasks*, and *measures* (Fig. 2.1). This section concludes with a summary of the gaps and opportunities. The scope of this work is currently limited to addressing head gaze in robotic systems, and not eye gaze. This exclusion is due to the higher complexity introduced by eye gaze. Head gaze, alone, has significant value without the need for eye gaze, and is additionally widely applicable (very few robots have movable eyes).

2.1 Summary of Social Head Gaze in Human-Robot Interaction

A brief summary of each of the 32 distinct studies is presented with reference to *study design*, *purpose of the study*, *robot used*, and *implementation style*, and arranged chronologically.

Imai, Ono, and Ishiguro (2001) [27] used a 12 participant between-subjects experiment (gaze and hand pointing, only hand pointing) to evaluate human perception of the robot's joint attention with an object. Five subsystems: Sensor System, Perceptual System, Dialogue Mechanism, Joint Attention Mechanism, and Action Executive were used to incorporate head gaze on a humanoid robot "Robovie." The Joint Attention Mechanism subsystem used simple rules to autonomously direct the gaze of the human towards the poster on the wall, when providing an explanation about it. This study had a simple but effective experimental design and validation.

Imai, Kanda, Ono, Ishiguro, and Mase (2002) [3] used a 36 participant within-subjects experiment to investigate the orientation of the robot's head needed for a person to perceive the robot's head gaze. The humanoid robot, "Robovie," generated autonomous head gaze using a rule-based method. Although IR cameras were used to capture head movements, this data was not analyzed. Moreover, the authors infer from measures that are marginally significant ($p < .1$) or insignificant ($p > .1$).

Matsusaka, Fujie, and Kobayashi (2001) [19] used a humanoid robot "Robita," to implement conversational strategies for a multi-party scenario. Five subsystems: Sensor System, Perceptual System, Reasoning System, Cognitive System, Action Executive were used to generate autonomous rule-based head gaze. While the authors did not conduct a human-user study to validate the system, nor present statistically significant results, this effort is one the earliest investigations on modeling head gaze for robots that takes into consideration triadic scenarios and real world occurrences such as interruptions.

Breazeal, Kidd, Thomaz, Hoffman and Berlin (2003) [4] used a 21 participant between-subjects experiment (implicit and explicit behaviors, explicit behaviors) to explore the impact of non-verbal cues and behavior on task performance (teaching the robot to push buttons in a given order) by a human-robot team. The robot "Leonardo" was an animal-like robot designed for social interaction. The robot implemented autonomous head gaze using associative learning. The authors do not back-up several claims using inferential statistics (for example, inferential statistics for data collected from the video).

Fincannon, Barnes, Murphy, and Riddle (2004) [40] examined video data of seven Urban Search and Rescue personnel utilizing a tele-operated non-anthropomorphic robot "Inuktun," during a confined space training exercise. Since the video was collected in an opportunistic manner and purely observational, there were no experimental manipulations. The authors found that humans interacted with a robot that appeared to be attending to them, and used the same social rules for maintaining eye contact as they would with another human.

Minato, Shimada, Ishiguro, Itakura (2004) [5] used a 18 participant within-subjects experiment (human girl, android with gaze behavior, still android with no gaze behavior) to investigate the acceptance of appearance and head movements of an android robot "Replee R1," during a simple quiz game. The head gaze was generated manually using the Wizard-of-Oz technique [47]. The limitation of the experiment was that the head motions of the android were randomly generated for the android with the good gaze behavior condition.

Sakamoto, Kanda, Ono, Kamashima, Imai, and Ishiguro (2004) [6] used a 50 participant within-subjects experiment (human, cooperative movements, no movement) to examine the importance of cooperative behaviors such as nodding and face-to-face communication in a route understanding situation, where a human gave directions to a humanoid robot "Robovie." Three subsystems: Sensor System, Communicative Units, and Action Executive, and the Wizard-of-Oz approach [47] were used to generate appropriate cooperative behaviors. The study analysis was thorough and had a good mix of subjective measures, video analysis, and body movement analysis.

Sidner, Kidd, Lee, and Lesh (2005) [8] presented a system architecture to initiate and maintain engagement in an interaction using head gaze. The architecture comprised of four subsystems: Sensor System, Perceptual System, Conversation Model, and Action Executive. The study used a 37 participant between-subjects experiment (mover condition with gaze and gesture enabled, talker condition with no gaze or gestures) to evaluate the architecture. The implementation used an animal-like robot "Mel," and a product demonstration task, to capture the turn-taking phenomena from human-human communication. The head gaze was generated using partially autonomous conversation planning strategies.

Macdorman, Minato, Shimada, Itakura, Cowley, and Ishiguro (2005) used two studies to investigate the relationship between the appearance of an android robot "Repliee," and its head gaze behaviors. In both the studies, the participants interacted with the robot in a between-subjects (think questions, know questions) simple question-answer experiment. In the first experiment, eight participants were informed that the robot operated in the autonomous mode, while in the second experiment, seven participants were told that the robot operated in the teleoperation mode. However, in both the experiments, the head gaze was generated manually using the Wizard-of-Oz approach [47].

Kozima et al. [24] used a longitudinal study to evaluate joint attention and emotions for autism therapy and other disorders such as Down's syndrome. The animal-like robot "Keepon," used four subsystems: Perceptual System, Attention Map, Habituation Mechanism, and Emotion Expression to generate emotions such as "happy" or "sad," and initiate and maintain joint attention with the human. The head gaze was generated manually using the Wizard-of-Oz approach [47].

Bennewitz, Faber, Joho, Schreiber, and Behnke (2005) [41] used a proof-of-concept experiment to evaluate the performance of a humanoid museum guide robot for multi-model interactions (including head gaze behaviors) with multiple people. The Behavior System subsystem generated autonomous direct head gaze for a multi-party scenario and emotions such as "joy," "surprise," "fear," and "anger" using simple rules. While this is one of the few studies that addresses a multi-party situation, the results from these experiments, including the human-like perception of the robot, and accurate recognition of different emotional states by the robot were not tested for statistical significance.

Mutlu, Hodgins, and Forlizzi (2006) [9] used a 20 participant between-subjects experiment (look at the participant 80% of time, look at the participant 20% of time) to design head gaze behaviors for conversation in "Asimo," a humanoid robot. The head gaze implementation captured the turn-taking phenomena observed in human-human systems [22] for a story-telling task and used the Wizard-of-Oz method [47]. This was the first human-robot interaction study to use middle of turn synchronization events for conversation.

Kuno, Sadazuka, Kawashima, Yamazaki, Yamazaki, and Kuzuoka(2007) [10] used a 12 participant within-subjects experiment (random gaze, proposed gaze) to assess the perception of head gestures during explanation of exhibits in a museum. The robot used was the humanoid robot, "Robovie." The implementation used a model of turn-taking [34], popular in human-human communication to generate head gaze using the Wizard-of-Oz method [47]. The was one of the few studies that used a random gaze condition and determined that head turning at turn relevant places is important.

Yamazaki, Yamazaki, Kuno, Burdelski, Kawashima, and Kuzuoka (2008) [11] used a 46 participant between-subjects experiment (unsystematic mode, systematic mode) to investigate the timing of speech and gaze in human-robot interaction, using a museum tour guide humanoid robot, "Robovie." The social science model for the implementation was developed following observation of human-human communication in an experiment, and then implemented on the robot. The head gaze was implemented using the Wizard-of-Oz method [47], with the locations of the head turns predetermined. This was a follow-up to a previous study [10], and established that participants are likely to display non-verbal actions, and do so with precision timing, when the robot turns its head at turn relevant places (significant points in interaction) than any other place in the interaction. This work addressed head gaze-speech synchronization, and established that synchronization is particularly important at turn relevant places.

Staudte and Crocker (2009a,b) [28, 29] conducted two studies [28, 29] using a humanoid robot "Peoplebot" to study whether people exploited head gaze when listening to a robot that made statements about the shared visual environment. The implementation utilized models of referential gaze from human-human interactions [35, 36]. The first study used 48 participants in a within-subjects experiment (ambiguous utterance and gaze, unambiguous utterance and gaze). The second study was a 36 participant mixed factorial experiment (statement validity [true, false] vs gaze congruency [congruent, incongruent, no robot gaze]). Both the experiments used videos and head gaze was hard-coded.

Mutlu, Shiwa, Takayuki, Ishiguro, and Hagita (2009a) [13] used a 72 participant between-subjects experiment (two addressees, an addressee and a bystander, or an addressee and an overhearer) to investigate gaze cues for regulation of multiparty conversational interactions in robots. The humanoid robot used in the study "Robovie," played the role of a travel guide, and provided information to the human. For this implementation, the authors developed rules for conversational footing, using both existing models [33, 51, 52] and observations of human-human communication. Head gaze was implemented using the Wizard-of-Oz method [47].

Mutlu, Yamaoka, Kanda, Ishiguro, and Hagita (2009b) [25] used a 26 participant mixed factorial experiment (robots [Robovie, Geminoid] vs gaze cues [gaze, no gaze]) to investigate whether humans can accurately perceive a robot's projected mental state (intentions) from gaze, and whether the physical design of the robot affects these inferences. The authors used two semi-humanoid robots, Robovie and Geminoid, which had two different pitched voices (higher and lower). Head gaze was generated using the Wizard-of-Oz approach [47].

Ishi, Liu, Ishiguro, and Hagita (2010) [20] used a 10 participant within-subjects experiment (model 1 vs model 2, shifted vs model 1, original vs model 2) to evaluate a model for the generation of head nods in humanoid ("Robovie") and android ("Repliee") robots. The model was based on an analysis of the relationship between head motion and speech dialogue acts in human-human conversation. The experiment used videos instead of actual robots and the subjective measures were all single items. Moreover, no statistical analysis of the results was presented.

Bethel and Murphy (2010) [14] used a 128 participant mixed factorial experiment (robot mode [emotive mode, standard mode] vs robot type [Inuktun, Packbot]) to assess the impact of robot orientation (amongst other social behaviors), for victim management in a simulated disaster scenario. Since the robots were non-anthropomorphic and did not have eyes or a head, they oriented towards the human to indicate attention.

Heerink, Krose, Evers, and Wielinga (2010) [26] used a 40 participant between-subjects experiment (more social, less social) to examine the effect of direct head gaze, head nods, and other prosocial behaviors in the user acceptance of "iCat," a small animal-like robot, when caring for the elderly. The head gaze was generated using the Wizard-of-Oz approach [47].

Shimada, Yoshikawa, Asada, Saiwaki and Ishiguro (2010) [15] used a 30 participant between-subjects experiment (with direct head gaze, without direct head gaze) to investigate whether non-verbal interactions by a human during interviews (for example, direct head gaze) made an android robot ("Repliee") appear more acceptable to the human. The head gaze was generated using the Wizard-of-Oz method [47].

Holroyd, Rich, Sidner, and Ponsler (2011) [1] developed the "Human Robot Collaboration" architecture to support human-robot collaboration and engagement in humanoid robots. Five subsystems: Collaboration Manager, Turn Policy, Reference Policy, Response Policy, and Maintenance Policy were used to incorporate turn-taking [33, 34] and referential gaze [36] models from human-human communication. The experiment used a 29 participant between-subjects experiment (operational condition, degraded condition) and a humanoid robot "Melvin" to evaluate the implementation of the architecture. The head gaze was partially autonomous and generated using a rule based method. The control condition, "degraded condition" was contrived, with the robot looking away and exhibiting no gaze behaviors. This is unnatural for an interaction with the robot, as it is expected to face the person.

Sirkin and Ju [42] used a 200 participant between-subjects online experiment (facial expressions alone, physical motion of the robot, combined expression and motion) to explore how embodied telepresence robots can support better communication in distributed teams. The participants rated videos of pre-scripted head movements, such as head nods, no, thinking carefully, short glances, and surprise on a two degree of freedom embodied robot.

Liu, Ishi, Ishiguro, and Hagita (2012) [43] proposed a model for the generation of head tilting and nodding in tele-operated robots from speech signals. The model was implemented and evaluated on three humanoid robots, "Geminoid," "Robovie," and "Telenoid," using a 38 participant between-subjects experiment (nod only vs nod and tilt, nod and tilt vs original, nod only vs original). The experiment used videos and head gaze, and was hard-coded using the proposed model.

Huang and Mutlu (2012) [2, 16] proposed the "Robot Behavior Toolkit" architecture for the generation of social behaviors in human-like robots. Four subsystems: Cognitive System, Behavior Selection System, Activity Model, and Social Behavior Knowledge Base utilized turn-taking [33,34] and referential gaze [35,36] models from human-human communication. The study used a 32 participant between-subjects experiment (human like, delayed, incongruent, no gaze), and the humanoid robot "Wakamaru" to evaluate the implementation of the architecture. The head gaze was partially autonomous and generated using a rule-based method.

Admoni, Hayes, Seifer, Ullman, and Scassellati (2013) [17] used a 53 participant mixed factorial experiment (group size [four, six, or eight] vs gaze duration [zero, one, three, or six seconds]) to investigate the use of short frequent glances and long, less frequent stares, and to determine which behavior was better at conveying a robot's visual attention. The experiment used "Keepon," an animal-like robot; the head gaze behaviors were pre-scripted by hand.

Pitsch et al. [49] used a 59 participant between-subjects experiment (action-related gaze, random gaze, static gaze) to investigate the use of head gaze for tutoring children with a humanoid robot. The experiment used "Asimo," a humanoid robot and included head gaze behaviors such as direct head gaze at the human and anticipatory head gaze toward objects. The head gaze was partially autonomous and generated using a rule-based method.

Andrist, Tan, Gleicher, and Mutlu (2014) [23] used a 30 participant between-subjects experiment (static gaze, bad timing, good timing) to evaluate a model of conversational gaze aversion for humanoid robots. The details for length, timing, and frequency of aversions were extracted from a human-human experiment. Unlike previous work [33,34], the proposed model does not synchronize conversational aversions with speech. The model was implemented on the "NAO" robot, with the participant stepping through the conversational turns by pressing a button on the robot. The experiment conducted does not compare the head gaze generated by the model with a standard condition that uses human-human models to realize conversational gaze aversion. This comparison is important to understand the social acceptance of the head gaze generated by the model. Additionally, the length, timing, and frequency data for conversational aversions was obtained from a human-human communication scenario that lasted for five minutes. It is unclear if the same timing distribution can be used for situations that are longer than five minutes, as expected in a real-world scenario.

Admoni, Dragan, Srinivasa, and Scassellati (2014) [30] used a 32 participant mixed factorial experiment (timing [delay, no delay] vs head gaze [social, non-social]) to evaluate the effect of deliberate delays and head gaze on perception of handover behaviors. The head gaze was hard-coded and implemented on a humanoid robot, "HERB" using the Wizard-of-Oz approach [47].

Huang and Mutlu (2014) [18] used a 29 participant between-subjects experiment (learning based, no gaze, random, conventional) to evaluate a learning based model for the generation of multimodal behaviors (including head gaze) for human-like robots. The model was implemented on a humanoid robot "Wakamaru," for a conversation task. The implementation was partially autonomous because the speech features used as an input to this model were manually annotated and tagged with pre-scripted gestures.

Moon, Troniak, Gleeson, Pan, Zheng, Blumer, Maclean, and Croft (2014) [32] used a 102 participant between-subjects experiment (no gaze, shared attention, turn-taking) to investigate the effect of gaze cues on timing and perceived quality of handover events. The head gaze behaviors were hard-coded on "PR2," a humanoid robot, and implemented using the Wizard-of-Oz approach [47].

Sauppe and Mutlu (2014) [31] used a 24 participant within-subjects experiment to explore how different deictic gestures affect communication under different environmental conditions. Each participant observed the robot for 46 rounds of references made by the robot (deictic gesture [pointing, presenting, touching, exhibiting, grouping, sweeping, minimally articulated, and fully articulated] x environment [neutral, distance from referrer, clustered objects, noise, no visibility, and ambiguity]). The head gaze behavior was hard-coded and implemented on the "NAO" humanoid robot using the Wizard-of-Oz method [47].

2.2 Head Gaze-Speech Synchronization

Sixteen major studies have been conducted to date that synchronize head gaze with speech [1, 2, 7–11, 13, 16, 18–20, 25, 28, 29, 31] (Table 2.1). These studies address social head gaze for *engaging in a verbal conversation with human(s)* or *referential gaze* at objects in the environment, where communication occurs across two different but highly interdependent channels: head gaze and speech. For example, in everyday face-to-face conversation, humans routinely use head gaze acts like fixate and avert in coordination with their speech. The remaining three functions: *look interested in human(s)*, *convey general liveliness and awareness*, and *show various mental states* do not use the speech channel, and hence studies focused on these functions do not address head-gaze speech synchronization. However, a social robot that interacts with a human in a realistic scenario will possibly use all five functions of head gaze. The change in synchronization for two functions such as *engage in a verbal conversation with human(s)* and *referential gaze* at objects in the environment may impact the human's perception of the robot's overall head gaze during the interaction. Therefore, in order synthesize an appropriate set of percepts and head gaze acts that can be used to represent social head gaze, and gain an understanding of successful experimental design and validation, studies that focus on head gaze functions: *look interested in human(s)*, *convey general liveliness and awareness*, and *show various mental states* have been considered for analysis in Sections 2.3 - 2.9.

Each of the 16 major studies [1, 2, 7–11, 13, 16, 18–20, 25, 28, 29, 31] use TSHG-S. This method is precise, and implements models of human-human interaction for conversational turn-taking [33, 34, 50] and referential head gaze [35, 36], or relies on a human-human interaction experiment, conducted specifically for determining the

Studies	**Year Of Study**	**Synchronization Method**	**Type of Speech**	**Generation Mechanism**	**Model**
Matsusaka *et al.* [19]	2001	Autonomous Identification of End of Turn Event	Synthetic, Real-Time	Production Rules	Turn-Taking [34, 50]
Sidner *et al.* [8]	2005	Manual Annotation of Turn Events	Synthetic, Real-time	Production Rules	Turn-Taking [33]
Macdorman *et al.* [7]	2005	Manual Annotation	Human, Pre-Recorded	Pre-Scripted Motion	Human-Human Experiment
Mutlu *et al.* [9]	2006	Manual Annotation of Linguistic Events	Human, Pre-Recorded	Production Rules	Turn-Taking [22]
Kuno *et al.* [10]	2007	Manual Annotation of Turn Events	Synthetic, Human, Pre-Recorded	Production Rules	Turn-Taking [34]
Yamazaki *et al.* [11]	2008	Manual Annotation of Turn Events	Synthetic, Pre-Recorded	Production Rules	Turn-Taking [34]
Staudte & Crocker [28]	2009	Manual Annotation of Linguistic Events	Synthetic, Pre-Recorded	Pre-Scripted Motion	Referential Gaze [35, 36]
Staudte & Crocker [29]	2009	Manual Annotation of Linguistic Events	Synthetic, Pre-Recorded	Pre-Scripted Motion	Referential Gaze [35, 36]
Mutlu *et al.* [13]	2009	Manual Annotation of Turn Events and Footing Cues	Unspecified, Pre-recorded	Production Rules	Human-Human Experiment, Turn-Taking [51, 52]
Mutlu *et al.* [25]	2009	Manual Annotation of Linguistic Events	Human, Pre-Recorded	Pre-Scripted Motion	Human-Human Experiment
Ishi *et al.* [20]	2010	Manual Annotation of Linguistic Events	Human, Pre-Recorded	Production Rules	Human-Human Experiment
Holroyd *et al.* [1]	2011	Manual Annotation of Turn Events	Synthetic, Real-Time	Production Rules	Turn-Taking [33, 34]
Huang & Mutlu [2]	2012	Manual Annotation of Turn & Linguistic Events	Human, Pre-Recorded	Production Rules	Turn-Taking [33, 34], Referential Gaze [35, 36]
Huang & Mutlu [16]	2013	Manual Annotation of Turn & Linguistic Events	Human, Pre-Recorded	Production Rules	Turn-Taking [33, 34], Referential Gaze [35, 36]
Huang & Mutlu [18]	2014	Manual Annotation of Turn & Linguistic Events	Human, Pre-Recorded	Learning	Human-Human Experiment
Sauppe & Mutlu [31]	2014	Manual Annotation of Linguistic Events	Pre-Recorded	Pre-Scripted Motion	Referential Gaze [35, 36]

Table 2.1: Tight Synchronization of Head Gaze with Speech in Literature

synchronization events [13, 20, 25]. TSHG-S has been shown to have many benefits for humans, such as increased task performance [8, 9, 13, 25], increased engagement [1, 7, 8, 25], improved understandability [1, 8], improved likeability [9, 13, 25], and increased positive feelings [8, 45]. However, to date, the use of LSHG-S has not been investigated.

Manual annotation is the most popular method for TSHG-S. There are six different synchronization events discussed in the literature related to manual annotation. The first three are turn events: *Start of Turn* [1, 2, 9–11, 19, 22], *Middle of Turn* [9, 22], and *End of Turn* [1, 2, 9–11, 22]. The last three are linguistic events: *First Word in Rheme* [9, 22], *First Word in Theme* [9, 22], and *Utterance of Object* [2, 16, 18, 28, 29, 31]. The theme specifies the topic of a sentence, i.e., what the sentence is all about, while the rheme specifies what is new or interesting about the topic [53]. A person identifies and marks the synchronization events in pre-recorded audio files containing dialog [1, 2, 7, 9–11, 13, 16, 20, 25, 28, 29] or text used to generate synthetic speech [1, 8]. Manual annotation requires significant human time and effort. It requires that the robot's head gaze behaviors be pre-scripted [25, 28, 29] or selected from a preset library using production rules [1, 2, 7–11, 13, 16, 18, 20, 31]. The use of pre-recorded audio limits the interactivity of the dialog to the extent of the preset library; therefore the robot is not able to adapt its dialog to the needs of the current interaction. Synthetic speech is advantageous for interactive conversations because speech can be generated in real-time based on a textual input, and the robot can adapt to situations through the generation of dynamic dialog. However, without any methods that support autonomous annotation, the original limitations on interactivity still persist.

The only research that uses autonomous head gaze-speech synchronization for conversation is by Matsusaka *et al.* [19], which models conversational strategy for robot participation in a group conversation. However, of the three possible turn events in conversations (*Start of Turn*, *Middle of Turn*, and *End of Turn*), this work only identifies the start of a turn. The implementation does not use any linguistic events, and has not been validated in an experiment.

Findings from Kirchhof & Ruiter [54] suggest that gesture comprehension is temporally and semantically more flexible than gesture production in humans, and a higher tolerance exists for modeling gestures in robots. The implementations of social head gaze are still at the manual Wizard-of-Oz stage, and no work has considered formal methods that transfer findings to support autonomous implementations. This work examines the effect of LSHG-S, using close approximations of turn and linguistic events from human-human interaction models, to determine if it is adequate for human-robot interaction.

2.3 Percepts

Social contexts are situations that arise as a result of humans interacting with each other. A social context consists of one or more percepts that robots can use to have a social interaction with a human. Three types of percepts were identified after generalization and mining of the literature: *external*, *linguistic*, and *internal*.

1. *External* percepts are visible states of the external world. They typically require inference and/or interpretation of sensor data. For example, such percepts include:

    - *Human Shows Initial Interest* [1–16, 24, 26, 40].
    - *Presence of Human* [1–16, 24, 26, 40].
    - *Listening to Human* [1, 8, 20, 26, 42].
    - *Presence of Object* [1, 2, 4, 8, 16, 27–32].

2. *Linguistic* percepts occur in the robot's dialog (text or audio). They include "rheme" and "theme" semantic units, turn events, and object references. These percepts are:

   - *First Word in Theme* [9,13].
   - *First Word in Rheme* [9,13].
   - *Start of Turn* [1,2,6,9–11,13,16,18,20–23].
   - *Middle of Turn* [9,13].
   - *End of Turn* [1,2,6,9–11,13,16,18,20–23].
   - *Onset of Speech Utterance* [1,2,4,8,16,27–32].

3. *Internal* percepts are self-perception of an internal state of the robot. This is based on the robot's beliefs about the people and objects in the world. This includes expression of emotions (e.g., confusion) or intent. Three internal percepts were used by the behavioral robotics framework.

   - $\textit{Internal State}_{\textit{mental state}}$ such as $\textit{Internal State}_{\textit{confused}}$ [4], $\textit{Internal State}_{\textit{happy}}$ [24], $\textit{Internal State}_{\textit{sad}}$ [24], $\textit{Internal State}_{\textit{surprise}}$ [41], $\textit{Internal State}_{\textit{fear}}$ [41], and $\textit{Internal State}_{\textit{anger}}$ [41].
   - $\textit{Internal State}_{\textit{aliveness}}$ [4].
   - $\textit{Internal State}_{\textit{acknowledge}}$ [1,8,20,26].

2.4 Head Gaze Acts

Head gaze acts are "head" movements used to generate social head gaze. Of the six head gaze acts described below, three (*fixate*, *avert*, and *concurrence*) are considered computational primitives, and the others (*short glance*, *confusion*, and *scan*), are considered as compound head gaze acts.

1. *Fixate*, is a head gaze persisting on a target: person, object, or location in space [1–16, 24, 26, 40]. If the person or object is moving, the fixation tracks and maintains gaze with the target [9, 21, 28, 29].

2. *Avert,* is a head gaze away from a person or a look away from the person toward the environment [1, 2, 6, 9–11, 13, 16, 20–22].

3. *Concurrence,* is a repetitive vertical movement greater than 10° or a horizontal movement greater than 25° of the head, which interrupts fixation [21]. Head nodding has been used only in conjunction with fixation [1, 8, 20, 26, 42].

4. *Short glance,* is a fixation persisting 0.77 sec to 1 sec [13]. Short glances are often used in a multi-party situation, when the robot needs to acknowledge the presence of bystanders [13, 41].

5. *Confusion,* is a series of rapid shifts back and forth, accompanied with a roll of the head as amplification [4]. It should be noted that confusion is only one example of an emotional state indicated by head gaze. Other emotional states like joy, surprise, or sadness may require additional head gaze acts [24, 26, 41].

6. *Scan,* is a short glance to a series of random points in space [8].

2.5 Robots Used

Eighteen robots were used across the 32 studies. The robots used to investigate the social head gaze phenomena can be categorized as *android*, *humanoid*, *animal-like*, or *non-anthropomorphic*. Fig. 2.2 summarizes the robots used in the study.

Two different androids Geminoid and Repliee, were used in six studies [5, 7, 15, 20, 25, 43]. These robots have very high fidelity human-like features such as head, eyes, lips, mouth, and skin.

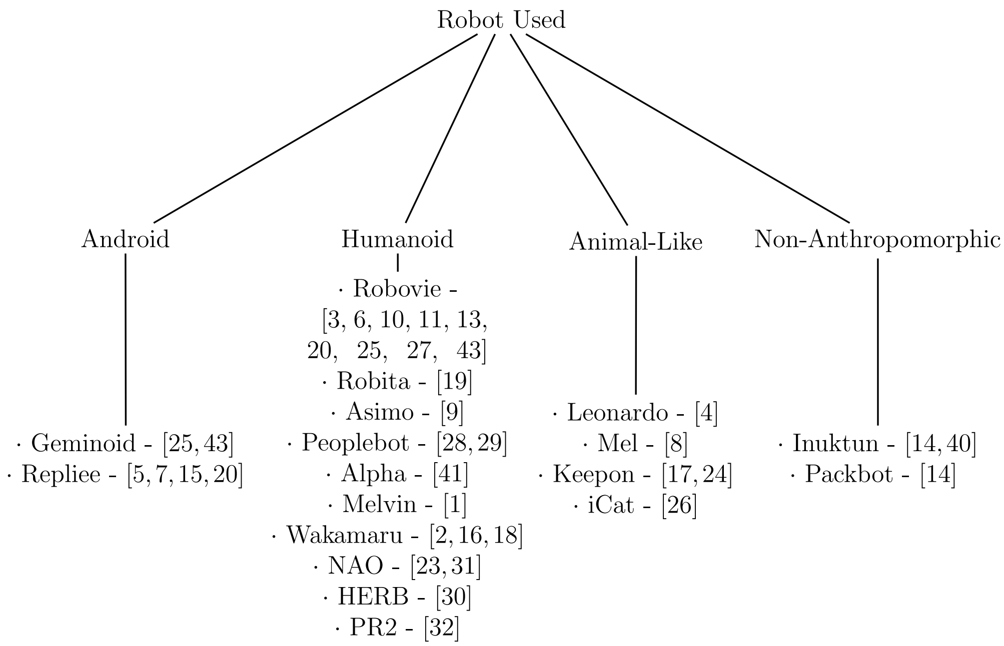


Figure 2.2: Robots Used to Evaluate Social Head Gaze

Humanoid robots are the most common robots used in head gaze research, with 10 robots (Robovie, Robita, Asimo, Peoplebot, Alpha, Melvin, Wakamaru, NAO, HERB, and PR2) being used in 22 studies [1–3, 6, 9–11, 13, 16, 18–20, 23, 25, 27–32, 41, 43]. Humanoid robots have a human-like appearance such as static eyes and mouth, but these features have very low fidelity when compared to the android robots.

Another type of robot used in gaze research is an animal-like robot. Four animal-like robots: Mel, Leonardo, Keepon, and iCat were used in five studies [4, 8, 17, 24]. Animal-like robots have a creature-like appearance, for example Mel looked like a penguin, and Leonardo was designed by professional artists to look like a fanciful creature.

Non-anthropomorphic robots are designed for function and do not have human-like or animal-like features, such as eyes or a head. Two non-anthropomorphic robots, "Inuktun" and "Packbot" have been used in two studies [14, 40].

2.6 Group Configuration

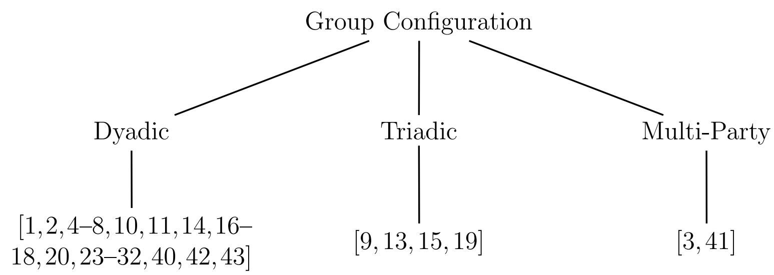


Figure 2.3: Group Configuration

A group can consist of a human and robot (*dyad*) or two humans and robot (*triad*) or one robot interacting with a large number of people (*multi-party*). The different group configurations used in the literature are summarized in Fig. 2.3.

The dyadic two party interactions were used in 27 studies [1, 2, 4–8, 10, 11, 14, 16–18, 20, 23–32, 40, 42, 43] making it the most popular group configuration used for studying social head gaze. Dyadic interactions are typically easier to model and follow simple rules of turn-taking.

Four studies [9, 13, 15, 19] modeled head gaze in a triadic group configuration. Triadic group configurations are more difficult to model because of more complex turn exchange policies and interruptions.

Only two studies [3, 41] used a multi-party interaction. Out of those two studies, one study [3] modeled only one way speech communication, that is, the participants communicated with the robot, however the robot responded only by change in head gaze direction.

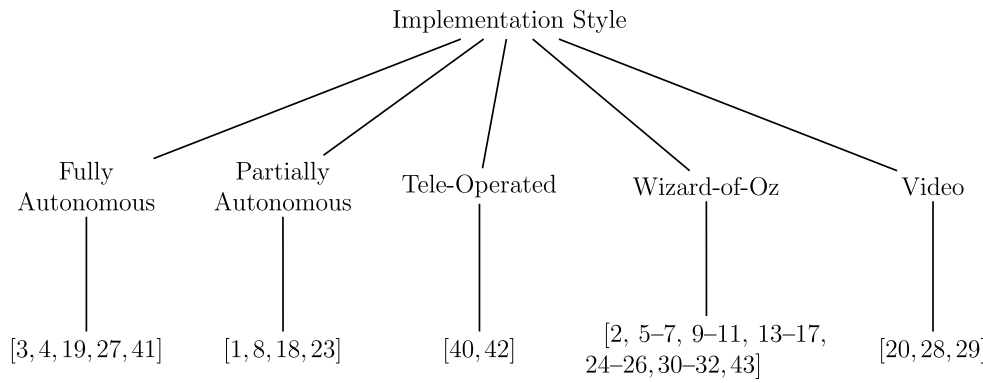


Figure 2.4: Implementation Style

2.7 Implementation Style

Five implementation styles were used to study social head gaze in the literature: *fully autonomous*, *partially autonomous*, *tele-operated*, *Wizard-of-Oz*, and *video*. Fig. 2.4 summarizes the implementation styles used in the study.

Five fully autonomous implementations used object recognition, gesture recognition, speech synthesis, and face recognition algorithms to socially interact with humans [3, 4, 19, 27, 41]. In addition, if the speech channel was present, these implementations autonomously synchronized head-gaze with speech [19].

The four partially autonomous implementations [1, 8, 18, 23] are very similar to the fully autonomous implementations in terms of object or speech recognition capabilities. However, they were categorized as partially autonomous because the head gaze-speech synchronization was manual [1, 8, 18] or the implementation required the presence of a human to step through conversational turns [23].

Two studies [40, 42] implement head gaze using tele-operation, where the robot operator puppeteers the robot. Additionally, the person interacting with the robot is aware that a robot operator is controlling the robot.

Wizard of Oz studies [47] use hard-coded head gaze acts synchronized with speech, and do not use object or speech recognition systems. The person interacting with the robot is not aware that a "Wizard"/robot operator is controlling the robot, and instead attributes agency to the robot. These studies have been popular for studying the head gaze phenomena in robots, and have been used in 19 studies [2, 5–7, 9–11, 13–17, 24–26, 30–32, 43].

Video studies use a video of a robot that is physically embodied, or in simulation to study head gaze. Only three studies [20, 28, 29] used this approach.

2.8 Tasks

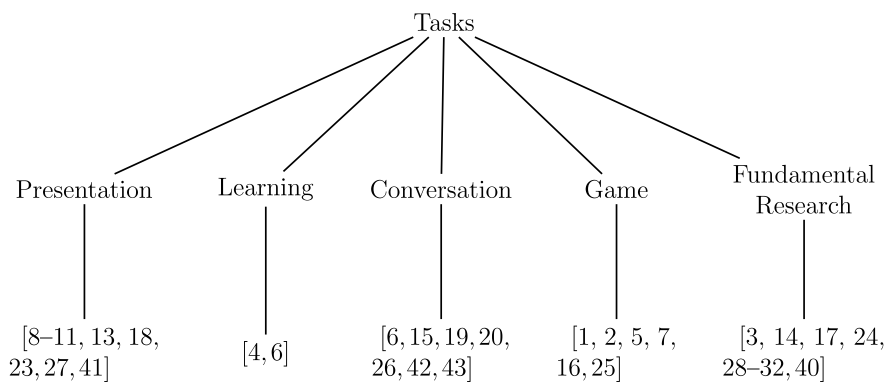


Figure 2.5: Tasks Used to Evaluate Social Head Gaze

There are five types of tasks used by researchers to study social head gaze in human-robot interaction. They are: *presentation*, *learning*, *conversation*, *game*, and *fundamental research*. The different tasks used in the literature are summarized in Fig. 2.5.

The tasks which involve presentation of content such as posters, product demonstration, or story-telling by the robot to the human are categorized as presentation tasks. These tasks are useful for validating elements of social head gaze in a triadic or multi-party scenario. Eight studies used presentation tasks to evaluate social head gaze [8–11, 13, 18, 23, 27].

In a learning task, the robot learns appropriate responses from the human in real-time during the task. Learning tasks are uniquely suited to tailor the behavior of the robot to a specific individual. Two studies use a learning task to achieve a common shared goal [4, 6].

During conversational tasks, the robot uses head gaze behaviors to facilitate conversational turn-taking. The conversation could be purposeful and goal oriented, or small talk. Seven studies used this type of task to gauge the benefit of social head gaze in conversations [6, 15, 19, 20, 26, 42, 43].

The robot and the human are engaged in a game task if they play a game together. Six studies used game tasks to measure performance [1, 2, 5, 7, 16, 25].

Fundamental research tasks are specifically designed to evaluate the performance and direction of gaze or some other aspect of a social interaction that includes gaze. 10 studies used fundamental research tasks [3, 14, 17, 24, 28–32, 40].

2.9 Measures

Three types of measures (*subjective*, *objective*, and *behavioral*) were used to quantify and determine the impact of using social head gaze in robots (Figs. 2.6, 2.7, 2.8). Four different methods (*questionnaires*, *video coding*, *interview*, and *eye-tracker*) were used to evaluate the three types of measures listed previously.

Subjective measures directly assess social gaze by asking the participants to rate their own or observed interpretations of the interaction on an anchored scale (for example, the Self Assessment Manikin (SAM) [55] questionnaire). They are straightforward and easy to administer and are hence widely adopted. Subjective measures are used in 23 studies, as shown in Fig. 2.6, and can be categorized as measures

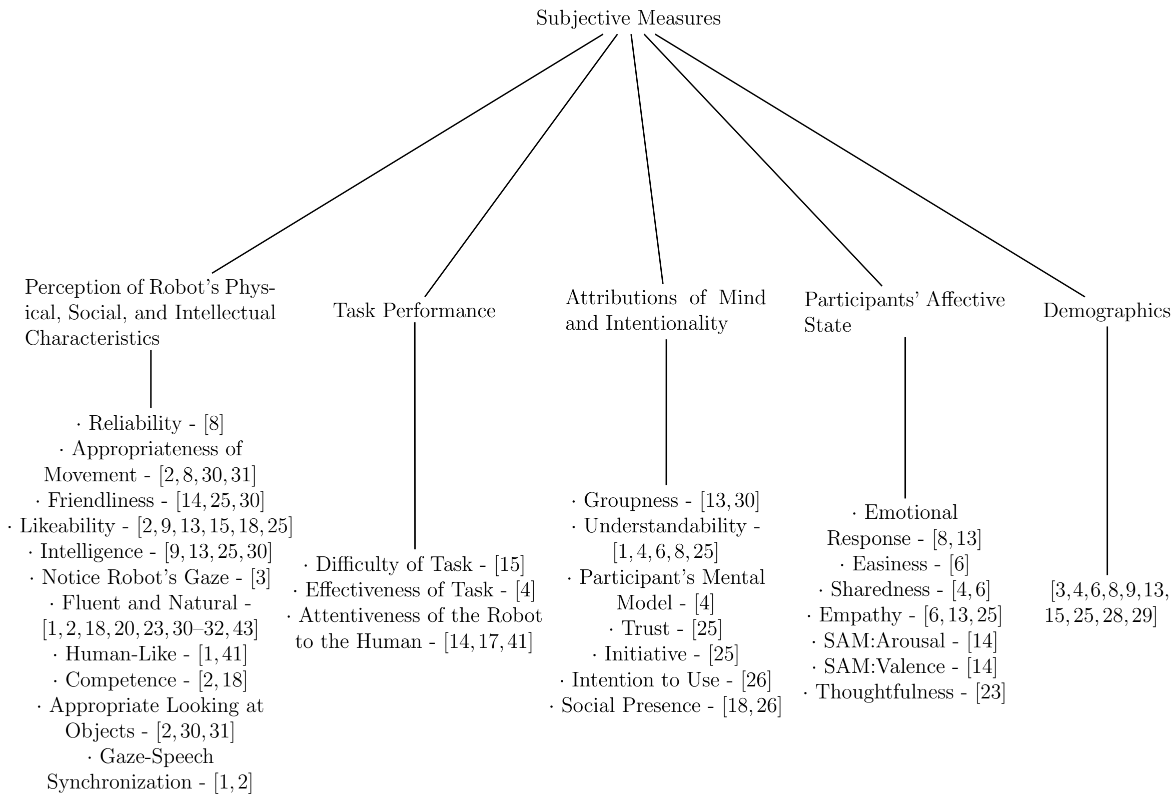


Figure 2.6: Subjective Measures

of Perception of the Robot's Physical, Social, and Intellectual characteristics, Task Performance, Attribution of Mind and Intentionality to the Robot, Participants' Affective State, and Demographics [1–4, 6, 8, 9, 13–18, 20, 23, 25, 26, 28–32, 41]. Perception of the Robot's Physical, Social, and Intellectual characteristics comprised of 11 measures – Reliability, Appropriateness of movement, Friendliness, Likeability, Intelligence, Notice robot's gaze, Fluent and Natural, Human-like, Competence, Appropriate looking at objects, and Gaze-speech synchronization. Task Performance included three measures - Difficulty of Task, Effectiveness of Task, Attentiveness of the Robot to the Human. The measures for Attribution of Mind and Intentionality were - Groupness, Understandability, Participants' Mental Model, Trust, Initiative, Intention to Use, and Social Presence. The Participants' Affective State was measured by eight measures - Emotional Response, Easiness, Sharedness, Empathy, SAM:Arousal, SAM:Valence, and Thoughtfulness. Finally, the Demographics instrument was used for obtaining information on participants' age, gender, ethnicity, experience with robots, video gaming experience, and pet ownership. The typical methods used in subjective measurement were questionnaires and interviews. While each of the 23 studies used questionnaires, only seven studies [15, 18, 23, 25, 30, 42, 43] used the interview method.

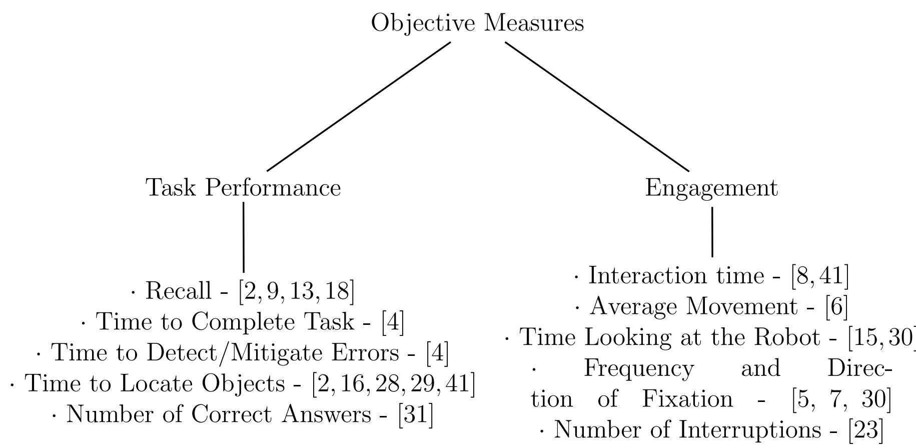


Figure 2.7: Objective Measures

Objective measures directly assess social gaze by comparing an individual's perception of the interaction to some ground truth. This is a rating based on a comparison of the data between the participant's perceptions and what is currently happening in the interaction. An objective measurement typically defies interpretation; it does not require the operator or the observer to make judgments. Objective measures were used in 19 of the 32 studies [2, 4–9, 11, 13, 15, 16, 18, 23, 25, 28–31, 41]. The objective measures for social head gaze are categorized as either Task Performance or Engagement (Fig. 2.7). The five measures used for Task Performance were - Recall, Time to Complete Task, Time to Detect/Mitigate Errors, Time to Locate Objects, and Number of Correct Answered. Engagement comprised of five measures - Interaction Time, Average Movement, Time Looking at the Robot, Frequency and Direction of Fixation, and Number of Interruptions. The objective video coding [56, 57] method was used in each of the 19 studies.

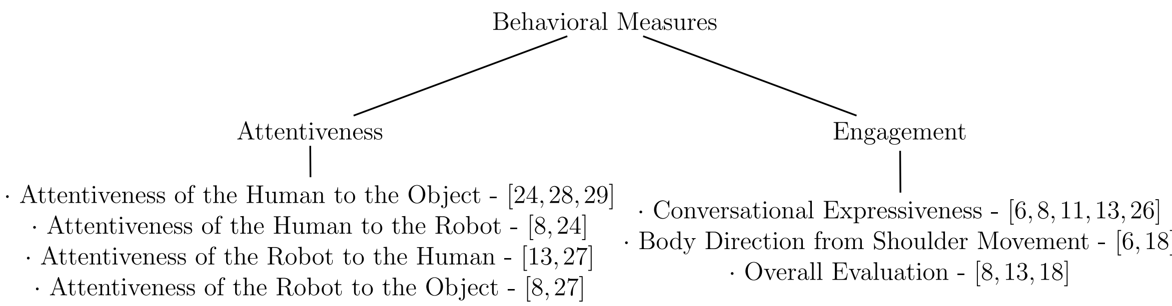


Figure 2.8: Behavioral Measures

Behavioral measures infer from the actions that participants choose, based on the assumption that good actions/response will follow from good social head gaze and vice-versa. Observers are often forced to rely more on the participants' observable actions and verbalizations in order to infer the effectiveness of social head gaze. Since behavioral measures rely primarily on observer ratings, they are somewhat subjective in nature. Behavioral measures were analyzed in 10 studies (Fig. 2.8), and can be categorized as measures of Attentiveness and Engagement [6, 8, 11, 13, 18, 24, 26–29]. Attentiveness is measured by four measures - Attentiveness of the Human to the Object, Attentiveness of the Human to the Robot, Attentiveness of the Robot to the Human, and Attentiveness of the Robot to the Object. Engagement of the participants' in the interaction is inferred from three measures - Conversational Expressiveness, Body Direction from Shoulder Movement, and Overall Evaluation.

In each of the 10 studies, the recorded video was examined by independent observers, and every overt action/reaction of the participant was coded (behavioral analysis) to infer Attentiveness and Engagement. In two studies [28, 29], Eye-tracker was used to evaluate one Attentiveness measure - Attentiveness of the Human to the Object (Fig. 2.9).

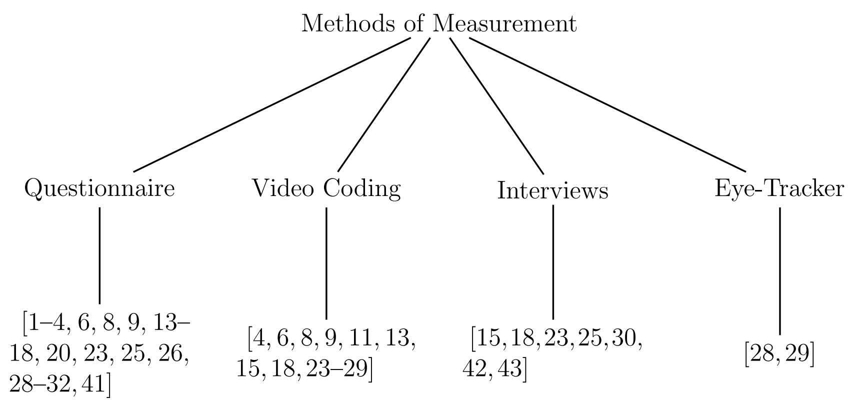


Figure 2.9: Methods of Evaluation

2.10 Summary

This section describes 32 distinct studies relating to social head gaze, and analyzes each study in terms of a novel taxonomy: *synchronization*, *percepts*, *head gaze acts*, *robots used*, *group configuration*, *implementation styles*, *tasks*, and *measures*. The current state-of-the-art TSHG-S was discussed and three limitations were presented. First, TSHG-S requires manual annotation and semantic content understanding. Second, if the robot uses a preset library to select appropriate head gaze behaviors, the head gaze cannot be generated in open-ended, interactive scenarios. Third, tight synchronization of gaze and speech that mimics human gaze may not be feasible due to limitations of the robot. Three percepts (*external*, *internal*, and *linguistic*) and six distinct head gaze acts (*fixate*, *avert*, *concurrence*, *scan*, *confusion*, and *short glance*) were defined. The robots used were categorized as one of four types: *android*, *humanoid*, *animal-like*, and *non-anthropomorphic*. Three group configurations (*dyad*, *triad*, and *multi-party*) used in human-robot interaction were identified. The implementation styles of head gaze were grouped as follows: *fully autonomous*, *partially autonomous*, *Wizard-of-Oz*, and *video*. The tasks used to evaluate social head gaze were categorized as: *presentation*, *learning*, *conversation*, *game*, and *fundamental research*. Three types of measures (*objective*, *subjective*, *behavioral*) were detailed along with a discussion on four different methods for evaluating social head gaze: *questionnaires*, *video coding*, *interview*, and *eye-tracker*.

The survey of the 32 distinct studies identifies at least four gaps:

1. The studies investigated the social head gaze phenomena using only TSHG-S, which are replications of how a human generates head gaze. TSHG-S uses semantic understanding, which requires significant human effort and time to manually annotate synchronization events in advance, restricts interactive dialog, and requires the operator to act as a puppeteer.

2. Implementations of social head gaze generation are at the manual Wizard-of-Oz stage, and little work has considered formal methods that transfer findings into autonomous implementations.

3. No programmable framework of social head gaze generation appears to exist in the social or computer sciences, resulting in robotic implementations using only partial understandings of head gaze generation, or creating ad hoc frameworks.

4. The studies were mostly limited to anthropomorphic robots, neglecting interactions with robots, which may be constrained by function to have a semi-

anthropomorphic or non-anthropomorphic form (e.g., a rescue robot).

This work addresses gaps (1) and by (2) by conceiving LSHS-S, a practical alternative to TSHG-S. In order to address gaps (3) and (4), this research constructs a behavioral robotics framework of social head gaze that provides a broad representation of robotics applications with wide applicability.

Social head gaze plays a central role in human-robot interaction. When head gaze is absent, breakdowns occur in conversations and can engender negative outcomes like ostracism [58]. Human-robot interaction studies provide strong evidence that robot gaze leads to increased task performance [2, 4–9, 13, 15, 25, 28, 41], increased engagement [5–8, 11, 15, 25, 41], improved perception of a robot's physical, social and intellectual characteristics [1–3, 6, 8, 9, 13, 15, 20, 25, 41, 44], increased attributions of mind and intentionality to the robot [1, 4, 6, 8, 25], increased positive affective state [6,8,12,45], improved attentiveness to the robot and task [8,13,27,28]. Therefore, it is imperative to capture a computational theory of social head gaze and autonomously generate head gaze-speech acts.

3. THEORY AND APPROACH

The existing approaches to generate head gaze-speech acts in human-robot interaction requires significant human effort and time to manually annotate synchronization events in advance. This is because the identification of these synchronization events depends on the semantics of dialog, which requires inference. Additionally, these approaches [2, 4, 7, 7, 8, 8, 16] typically used theory of mind systems to decide on actions to be performed based on the robot's beliefs about objects in the world. Theory of mind systems model cognitive processes such as Visual Attention, Working Memory, and Behavior Arbitration. They decide on actions to be performed based on the robot's beliefs about objects in the world. However, they do not describe what the system does in terms of specific mechanisms or terminologies. There is a lack of clarity on when to generate head gaze and how to generate it. The approach to capture a computational theory of social head gaze and generate autonomous head gaze-speech acts has two components: (1) the substitution of affordances for linguistic and internal percepts of head gaze to infer the state of an interaction (e.g., end of turn, beginning of turn, middle of turn theme, middle of turn rheme), and (2) the use of a behavioral robotics framework to map affordances onto head gaze acts, and enable a robotic implementation. The affordances for linguistic and internal percepts of head gaze are computationally trivial, are independent of semantics, support autonomous generation, and are useful for interactive, open-ended conversations. They can serve as a reasonable substitute until deeper methods of determining the state of an interaction can be developed. This section introduces affordances and translates the review of literature into the behavioral robotics framework. Section 4 will discuss the specific implementation choices made for the "Survivor Buddy" robot.

3.1 Affordances

Affordances are conditions or objects that are directly perceivable without any memory, inference, or interpretation, and have been widely used by behavioral roboticists because they simplify computation [37]. In robot soccer competitions, a goal area may be painted a specific color, and that color is only permitted to be used for the goal; thus if a robot perceives that color, it is the goal, and the robot does not have to remember where it is located.

The development of LSHG-S is made possible by the substitution of affordances for linguistic and internal percepts (turn-taking, semantics, and internal states). The affordances can be based on sentence structure, time delays, typing, prosody, or inflection of speech. For example, during a human-human conversation, a person always averts his or her gaze at the beginning of a turn, and at the start of the "theme" [22]. A limitation of using "rheme" and "theme," is that they are subjective, and vary with sentences. A word that is at the beginning of the "rheme" in one sentence, need not mark the beginning of the "rheme" in another sentence. The locations of the "rheme" and "theme" need to manually inferred by a human before the start of a interaction. However, in straightforward simple sentences, the theme is at the beginning of the sentence and the rheme is at the end [53]. Thus punctuation (*! ?, and carriage return at the end of a paragraph*) in text-to-speech or inflection in voice recognition are approximations of the location of the "rheme" or "theme," and act as affordances. Another example of an affordance is *elapsed time*. The *elapsed time* since the start of a turn affords back-channeling; the listener typically nods to show that they are still listening [8, 21].

The proposed approach for the realization of head gaze through the use of affordances supports autonomous coordination of head gaze-speech. They are unique, can be easily computed, and do not require manual annotation. It enables interactive, open-ended conversations, and can be tailored to the limitations of specific robots and applications. The specific affordances using the implementation of LSHG-S on the "Survivor Buddy" rescue robot are described in Section 4.

3.2 Mapping of Social Head Gaze Unto Behavioral Robotics Framework

A behavioral robotics framework, also called "programming by behavior," was selected to represent social head gaze, since this has been shown to express ethological, psychological, and robotic concepts, and it is consistent with good software engineering principles, such as modularity and extensibility [37,38]. The constructed behavioral robotics framework needs to capture the commonalities, essence, and experience of previous successful head gaze generation implementations, so that it provides a broad coverage and wide applicability. Hence, a two-step methodology commonly used to derive reference architectures [59, 60] was followed:

1. *Construction of Conceptual Architectures* - *Conceptual Architectures* were derived for each of the 32 previous implementations of social head gaze using behavioral robotics theory [37, 38] as the common framework. Conceptual architectures are abstract representations of subsystems and inter-subsystem relations, not specific procedures or variables [60].

2. *Commonality Analysis* - A *commonality analysis* [61] was employed to synthesize shared elements between the resulting 32 conceptual architectures to form a behavioral robotics framework.

3.2.1 Construction of Conceptual Architectures

The key construct in behavioral robotics is a behavior b, which maps a percept s onto an act r [38]. An agent may have multiple behaviors active at the same time

	Common Terms in the Ethology Community	Common Terms in the Robotics Community	Example from Social Head Gaze
b	behavior	behavior	communicating social attention
s	stimulus	percept	human shows initial interest
r	response	act	fixate
G	gain	gain	culture and gender
C	coordination function	coordination function	arbitration by prioritization
ρ	overall response	overall response	fixate

Table 3.1: Common Terminology in the Ethology/Psychology and Behavioral Robotics Communities Illustrated with a Social Head Gaze Example.

therefore, the combined observable response is given as $\rho = C(G \times B(S))$, where B is a vector of behaviors, S is a vector of sensed percepts, G is a vector of the gain functions, and C is the coordination function that determines the overall response ρ. The strength of act r may be modified by a gain G, which may amplify or reduce the contribution of an individual behavior to the overall behavior. Examples of gains in social head gaze are covariate factors such as culture and gender, which are identified to have a significant influence on head gaze [9, 13]. Table 3.1 illustrates a social head gaze example using behavioral robotics terminologies; the terminology from ethology/psychology is also included for readers familiar with those fields. For example, while communicating social attention to a human, a robot should consider gender when determining the amount of time it will fixate [9]. Here the behavior is $fixate = f(gender) \times$ COMMUNICATING SOCIAL ATTENTION$(human)$.

As the first step towards the construction of the behavioral robotics framework, this work derives conceptual architectures for each of the 32 previous implementations of head gaze using the behavioral robotics notations described above. The conceptual architectures are comprised of behaviors, percepts, acts, gains, and coordination functions. Figure 3.1 provides a diagrammatic representation of the conceptual architectures.

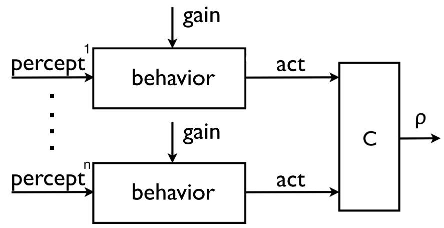


Figure 3.1: Example of an Conceptual Architecture for Two Individual Behaviors with Gains Passing Through a Coordination Function.

3.2.2 Commonality Analysis

A commonality analysis is an analytical technique used to determine the components of an architecture [61]. It helps identify the domain concepts that represent the common elements of the domain at its highest level of abstraction; it is also useful for normalizing the existing notations produced by previous implementations.

The 32 conceptual architectures derived previously were iteratively analyzed to identify and create useful abstractions common to all conceptual architecture components. The nomenclature was then standardized. For example, *Avert* [15,17], *Look Away* [1,9,13,22], and *Avoid Gaze* [11] all corresponded to the same head gaze act; hence, that head gaze act was standardized to *Avert*. This step ensured that implementations with different overall functionality, environments, and robot types were taken into consideration and supported. The commonality analysis identified three types of percepts, six head gaze acts, five behaviors, and one coordination function based on prioritization. Section 2 defines the three types of percepts – *external percepts*, *linguistic percepts*, and *internal percepts* – and the six head gaze acts – *fixate*, *avert*, *concurrence*, *short glance*, *confusion*, and *scan*.

Five social head gaze behaviors were identified:

1. Communicating Social Attention is a behavior where head gaze is used

by robots to look interested in human(s) [1–16, 24, 26, 40]. This behavior maps an *external* percept *Human Shows Initial Interest* on to the *fixate* head gaze act. This behavior is initiated for non-verbal communication, and typically occurs at the beginning of an interaction or if the robot is not capable of speech.

2. Regulating an Interaction is a behavior where head gaze is used for engaging in a verbal conversation with human(s) [1,2,6,9–11,13,16,20,22]. This behavior maps combinations of *linguistic* and *external* percepts on to the *fixate*, *avert*, or *concurrence* head gaze acts. The linguistic percepts facilitate turn-taking and are as follows: *Start of Turn*, *Middle of Turn*, *End of Turn*, *First Word in Theme*, and *First Word in Rheme*. The external percept *Listening to Human* and internal percept $Internal\ State_{acknowledge}$ are used to activate back-channeling.

3. Manifesting an Interaction is a behavior where head gaze is used to direct attention towards objects in the environment [1, 2, 4, 8, 16, 27–29]. The behavior maps *external* and *linguistic* percepts onto the *fixate* head gaze act. A combination of the *external* percept *Presence of Object* and the *linguistic* percept *Onset of Object Utterance* facilitates *referential gaze*.

4. Projecting Mental State is a behavior where head gaze is used for showing various mental states, such as emotions. This behavior maps an *internal* percept such as $Internal\ State_{mental\ state}$ onto the head gaze act for mental state such as confusion [4] or emotions such as happiness, sadness, surprise, etc [24]. The internal state of the robot can be set based on its beliefs about the people and objects in the world.

5. Establishing Agency is a behavior where a head gaze is used to convey

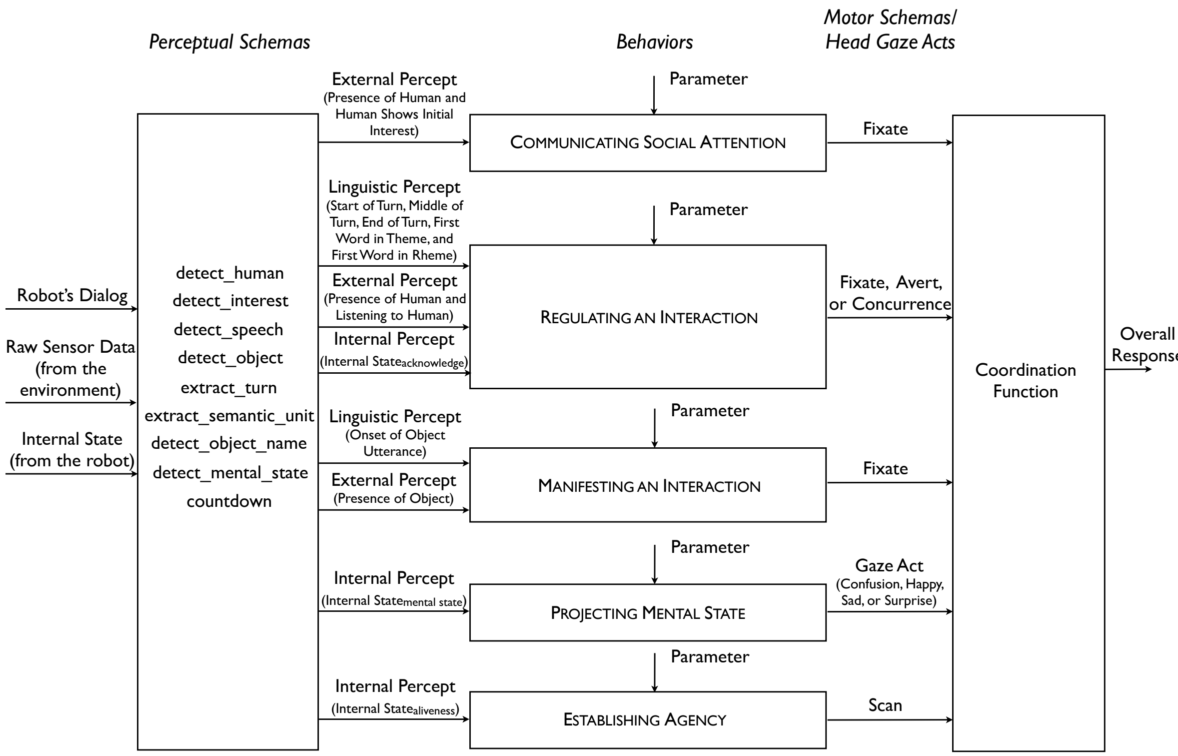


Figure 3.2: Notational View of the Possible Behaviors in the Behavioral Robotics Framework for Social Head Gaze. The Range of Possible Perceptual Schemas, Percepts, and Head Gaze Acts Have Been Enumerated.

general liveliness and awareness [4, 8, 26]. This behavior maps an *internal* percept such as $Internal\ State_{aliveness}$ onto the *scan* head gaze act.

A coordination function fuses the responses of multiple active behaviors [37, 38]. The existing social head gaze system architectures [1,2] use a competitive arbitration method based on the prioritization of behaviors to select a single overall response. None of the 32 studies reported implementations using gain parameters such as gender or culture to actively influence the generation of a head gaze.

3.2.3 Resulting Behavioral Robotics Framework

Figure 3.2 shows the behavioral robotics framework resulting from the commonality analysis. The components of the master framework are grouped into *Perceptual Schemas*, *Behaviors*, *Motor Schemas*, and *Coordination Function*. This grouping of the components of the master framework is as was suggested by Arkin [38].

Perceptual Schemas have at least one method that takes sensor input and transforms it into a data structure called a percept [37]. Perceptual schemas are used to generate the external, linguistic, and internal percepts or the affordances for these percepts. Figure 3.2 enumerates a list of nine possible perceptual schemas to generate the percepts. The perceptual schemas used to extract the affordances are discussed in Section 4. The *detect_human* perceptual schema is used to detect the *Presence of Human* percept. The *detect_interest* perceptual schema generates the *Human Shows Initial Interest* percept. The *detect_speech* perceptual schema computes the *Listening to Human* percept. The *Presence of Object* is indicated by the *detect_object* perceptual schema. The *extract_turn* perceptual schema returns *Start of Turn*, *Middle of Turn*, and *End of Turn* percepts. The *Start of Rheme* and *Start of Theme* semantic units are extracted by the *extract_semantic_unit* perceptual schema. The *detect_object_name* perceptual schema computes the *Onset of Speech Utterance* percept. This perceptual schema sets the $\textit{Internal State}_{\textit{mental state}}$ percept. Finally, the *countdown* perceptual schema provides two internal percepts $\textit{Internal State}_{\textit{aliveness}}$ and $\textit{Internal State}_{\textit{acknowledge}}$. The perceptual schemas can share the same sensors. For example, the *detect_human* perceptual schema shares the sensor data from the webcam with the *detect_object* perceptual schema. Additionally, the computational processes in the behaviors can share the percepts created by the perceptual schemas. The *Presence of Human* percept is shared between COMMUNICATING SOCIAL ATtention and Regulating an Interaction social head gaze behaviors.

The behavioral robotics framework consists of five behaviors: Communicating Social Attention, Regulating an Interaction, Manifesting an Interaction, Projecting Mental State, and Establishing Agency. These behaviors become active when their corresponding percept is perceived. The behaviors are transformation units; they map the percept to an appropriate head gaze act. The behaviors can employ different computational mechanisms (models based on probability or learning). The gains are shown as parameters that apply to each of the individual behaviors. The gain parameters can be used to modify duration, speed, and range of head gaze acts based on covariate factors that influence head gaze, such as the culture and gender [9, 13] of the interaction partner. However, at the present time there is no literature informing the implementation values for these factors.

The motor schema represents the template for physical activity and are connected to actuators [37]. The motor schema of the behavioral robotics framework currently supports six head gaze acts: *Fixate, Avert, Concurrence, Short Glance, Confusion,* and *Scan.*

The coordination function is used to coordinate the responses of multiple active behaviors [37,38]. The coordination function ensures that the robot is sensitive to the current context and conveys the appropriate meaning. The prioritization rules used by existing architectures [1, 2] are ad hoc [1], lack implementation details [2], and are limited to two or three behaviors. Hence, a coordination scheme based on timestamps (highest priority for the most recent behavior) and the nature of the behavior (atomic or non-atomic) is proposed in Section 4. While any other appropriate coordination action selection methods, fuzzy logic, or voting can be used, this was the first scheme to be implemented for the five behaviors.

The behavioral robotics framework has been constructed from six categories of robotics applications that implement head gaze: healthcare [24, 26], victim management [14, 40], robot guides [8, 10, 11, 13, 41], entertainment [1, 2, 4, 5, 9, 16], telepresence [42, 43], and fundamental research [3, 6, 7, 15, 17, 20, 27–29], and captures their commonalities, essence, and experience. Therefore, it is expected that the behavioral robotics framework for head gaze can be instantiated for each of these original applications, in addition to new ones.

The proposed behavioral robotics framework's broad coverage and wide applicability provides a foundation for further discussion. It identifies many of the key aspects and components that will be present in any well-designed head gaze generation system. However, it is not a comprehensive representation of all understandings of social head gaze. The behaviors, percepts, and head gaze acts were selected from implementations of head gazes within the robotics domain, rather than directly from human-human communication literature. Eye gaze is excluded from the master framework, as it includes additional complexities that are the subject of future work. Head gaze alone has significant value and can be used without eye gaze. Additionally, it is widely applicable in the robotics domain (very few robots have movable eyes). If additional competencies are discovered to be important for robotic head gaze, the behavioral robotics framework can be extended to accommodate these changes.

3.3 Summary

This section discussed the theory and approach to a computational theory of social head gaze for social agents, and captured it as a programmable framework, so that head gaze-speech acts can be autonomously generated. The approach is comprised of two components: 1) the substitution of affordances for linguistic and internal percepts of head gaze (e.g., end of turn, beginning of turn, middle of turn theme, middle of turn rheme), and (2) the use of a behavioral robotics framework to map affordances onto head gaze acts, and enable a robotic implementation.

Affordances were introduced as conditions or objects that are directly perceivable without any memory, inference, or interpretation [37]. The generation of LSHG-S using affordances from sentence structure, time delays, typing, prosody, or inflection of speech instead of linguistic and internal percepts was examined. The three advantages of using affordances were detailed: (1) support for autonomous generation, (2) unique and can be easily computed, (3) independent of semantics and manual annotation, and (3) enables interactive open-ended conversations.

A two-step procedure by [59, 60] was followed to derive the behavioral robotics framework from 32 existing implementations of head gaze. First, conceptual architectures for each of 32 previous implementations were constructed. Second, a commonality analysis was employed to synthesize shared elements between the resulting conceptual architectures and construct the behavioral robotics framework. The constructed behavioral robotics framework consisted of *Perceptual Schemas*, *Behaviors*, *Motor Schemas*, and *Coordination Function*. Nine perceptual schemas were used: *detect_human*, *detect_interest*, *detect_speech*, *detect_object*, *extract_turn*, *extract_semantic_unit*, *detect_object_name*, *detect_mental_state*, and *countdown*. The perceptual schemas generated three types of percepts: *external*, *linguistic*, and *internal*. Five social head gaze behaviors were identified – COMMUNICATING SOCIAL ATTENTION, REGULATING AN INTERACTION, MANIFESTING AN INTERACTION, PROJECTING MENTAL STATE, and ESTABLISHING AGENCY – to map the percepts on to the head gaze acts. The motor schema comprised of six head gaze acts: *Fixate, Avert, Concurrence, Short Glance, Confusion,* and *Scan.* The behavioral robotics framework provides broad coverage, wide applicability, and identifies many of the key aspects and components that will be present in any well-designed head gaze generation system.

4. IMPLEMENTATION *

In order to answer the secondary research question *How can social head gaze be expressed as behaviors or schemas, which are common representations in both psychology and robotics?*, the important aspects of head gaze such as percepts, head gaze acts, behaviors, and coordination function described by the behavior robotics framework in Section 3 needs to be implemented. The implementation specific details for mechanisms to perceive the percepts, execute head gaze acts, map the head gaze act onto the percept, and coordinate multiple active behaviors are described in this section.

The implementation of LSHG-S on the "Survivor Buddy" robot is summarized in Fig. 4.1. Moving from left to right, the Fig. 4.1 shows perceptual schemas that extract the affordances, the affordances from sentence structure, time delays, and typing that approximate the linguistic and internal percepts for a social head gaze behavior, the gaze acts along with the probability of its occurrence, the coordination function, and the five social head gaze behaviors that serve as the mapping. The experiment conducted to evaluate this implementation using the Survivor Buddy robot in a victim management scenario is discussed in Section 5.

4.1 Robot Platform Description

The proposed head gaze generation method was implemented on Survivor Buddy, an affective multimedia head mounted on an Inuktun Extreme-VGTV robot (Fig. 4.2 and 5.1). Survivor Buddy has four degrees of freedom, and is capable of very agile movements. Survivor Buddy's head is a 7" touch screen monitor with a webcam and

*Reprinted with permission from "Evaluation of Head Gaze Loosely Synchronized with Synthetic Speech for Social Robots" by Srinivasan, V., Bethel, C.L., Murphy, R.R, 2014. IEEE Transactions on Human-Machine Systems, Accepted for Publication, Copyright [2014] by Vasant Srinivasan.

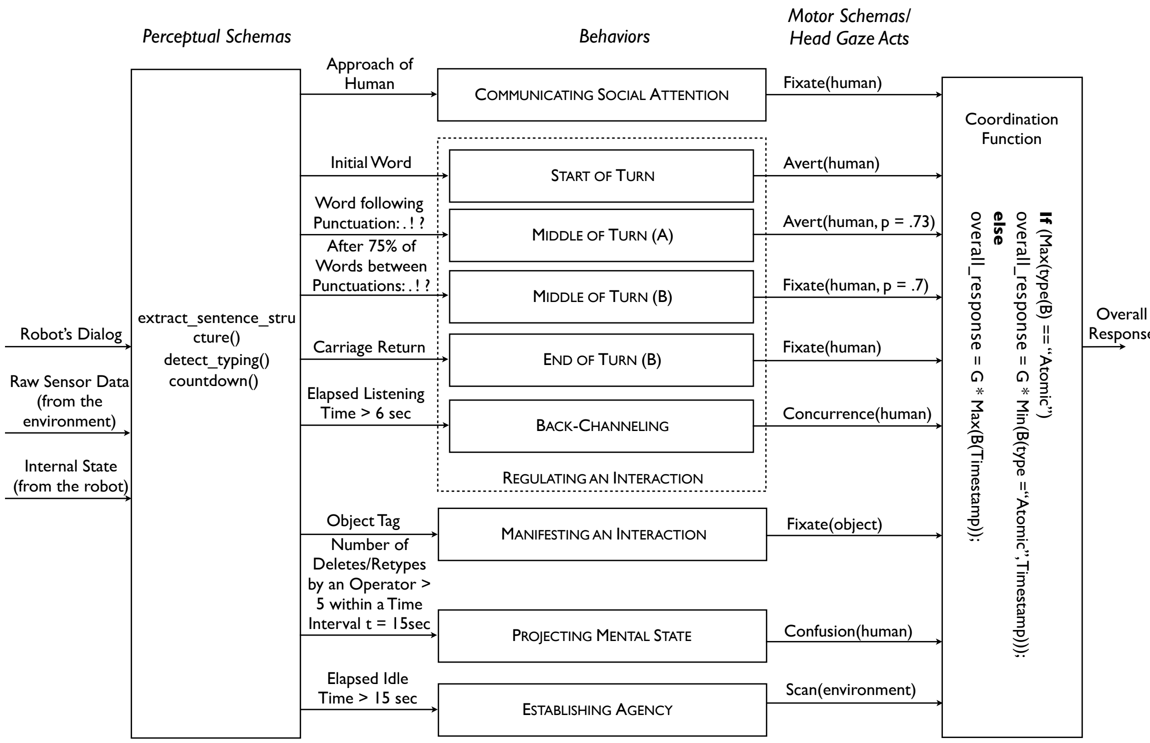


Figure 4.1: Implementation of LSHG-S

microphone manufactured by MIMO Monitors. The neck of the robot contains the speaker system. Note that the Survivor Buddy robot has low degree of anthropomorphism, with a mechanical appearance. The computation theory can be implemented independent of the choice of robot. However, factors such as (perceived) appearance of the robot or degrees of freedom may impact the user perception of head gaze. There is no existing literature informing this.

The software implementation of the LSHG-S system was in C# and used Microsoft text-to-speech with the Microsoft Anna voice. While any other programming language can be used to implement LSHG-S, C# was chosen because of the availability of pre-existing libraries for motor control and support for object-oriented development. Microsoft Anna voice was chosen because of clarity of speech and clear

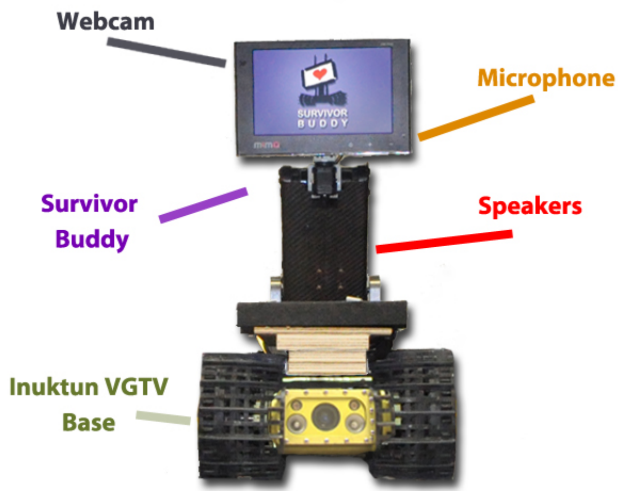


Figure 4.2: Survivor Buddy Robot

pronunciation present in this text-to-speech voice.

The implementation currently requires typed text and punctuation. This is available transparently to a robot that is operated autonomously, or can be provided by the operator if it is tele-operated. The TSHG-S implementation used pre-recorded audio files for each turn of the robot dialog, so that head gaze can be tightly synchronized with speech by manual inspection.

4.2 Affordances for Linguistic and Internal Percepts

The current LSHG-S implementation uses eight affordances from the sentence structure used in dialog, time delays, and typing to infer linguistic percepts and internal percepts respectively. The five linguistic percepts from the robot's dialog: *End of Turn*, *Start of Turn*, *Middle of Turn "Theme"*, and *Middle of Turn "Rheme"* are substituted with affordances from the sentence structure: *Initial Word*, *Word following Punctuation : . ! ?*, *After 75% of Words between Punctuation : . ! ?*, *Carriage Return*, and *The Object Name Tag*. Two internal states of the robot: *inter-*

	Percepts from Literature	**Substituted Affordances in Sentence Structure, Time Delays, and Typing**
1	First Word in Theme Start of Turn [9, 13, 22]	Initial Word
2	First Word in Theme Middle of Turn [9, 13, 22]	Word following Punctuation : . ! ?
3	First Word in Rheme Middle of Turn [9, 13, 22]	After 75% of Words between Punctuation : . ! ?
4	First Word in Rheme End of Turn [9, 13, 22]	Carriage Return
5	800 msec to 1 sec before Onset of Utterance of Object [2, 16, 28, 29]	The Object Name Tag
6	Internal State$_{acknowledge}$ [4, 8, 21]	Elapsed Listening Time > 6 sec
7	Internal State$_{aliveness}$ [4, 8]	Elapsed Idle Time > 15 sec
8	Internal State$_{confused}$ [4]	Number of Deletes/Retypes by an Operator > 5 within a Time Interval t = 15 sec

Table 4.1: Affordances from Sentence Structure, Time Delays, and Typing.

nal state$_{\text{acknowledge}}$ and *internal state*$_{\text{aliveness}}$ are approximated using affordances from time delays: *Elapsed Listening Time > 6 sec* and *Elapsed Idle Time > 15 sec*. One internal state of the robot – *internal state*$_{\text{confused}}$ is approximated with an affordance from typing – *Number of Deletes/Retypes by an Operator > 5 within a Time Interval t = 15 sec*. Table 4.1 lists the eight percepts in the literature for turn events, linguistic events, and internal states, that require semantic or speech understanding (column 1) and how this research substitutes an affordance of either sentence structure, time delay, or typing (column 2) for a linguistic or internal percept, which produces a social head gaze act.

4.2.1 Affordances for Linguistic Percepts

Linguistic percepts occur in the robot's dialog (text or audio). For example, such percepts include *Start of Turn*, *Middle of Turn*, *End of Turn*, *First Word in*

Theme, *First Word in Rheme*, and *Onset of Speech Utterance*. These percepts require inference and interpretation. For example, the theme specifies the topic of a sentence, (i.e., what the sentence is all about), while the rheme specifies what is new or interesting about the topic [53]. However, the limitation of using semantic units "theme" and "rheme" is that they are subjective, require sentence understanding, and vary with sentences. A word that is at the beginning of the "rheme" in one sentence need not mark the beginning of the "rheme" in another sentence. However, in straightforward simple sentences, the theme is at the beginning of the sentence and the rheme is at the end [53]. Thus punctuation (! ?, and carriage return at the end of a paragraph) in text-to-speech or inflection in voice recognition are approximations of the location of the "theme" or "rheme" and act as affordances.

The first five affordances (rows 1 - 5) in Table 4.1 – *Initial Word*, *Word following Punctuation : . ! ?*, *After 75% of Words between Punctuation : . ! ?*, *Carriage Return*, and *Object Name Tag* – are extracted by the *extract_sentence_structure* perceptual schema. The *Initial Word*, *Word following Punctuation : . ! ?*, *After 75% of Words between Punctuation : . ! ?*, *Carriage Return* affordances are the sentence structure approximations of linguistic percepts (*Start of Turn*, *Middle of Turn*, *End of Turn*, *First Word in Rheme* and *First Word in Theme*) from models of human-human interaction. Since the theme occurs at the the beginning of an independent clause or simple sentence [53], the occurrence of *First Word in Theme*) and *Start of Turn* can be substituted with *Initial Word* of a new turn in the robot's dialog. Similarly, the occurrence of *First Word in Theme*) and *Middle of Turn* can be substituted with the first word of a new sentence within the same turn of the robot's dialog or *Word following Punctuation : . ! ?*. The rheme occurs toward the end of the independent clause or simple sentence [53]. Therefore, the occurrence of *First Word in Rheme*) and *Middle of Turn* can be substituted by *After 75% of Words* *between Punctuation : . ! ?*. The occurrence of *First Word in Rheme* and *End of Turn* can be substituted with the last character of the current turn of the robot's dialog or *Carriage Return*. The affordance listed in Row 5 is a *Object Name Tag* inserted before an object name. This affordance approximates the gaze at an object in the environment from an *800 msec to 1 sec before Utterance of Object* [2,16,28,29] to gaze at the object at the onset of Utterance of Object. The *Object Name Tag* can be inserted either manually when there is an operator present and typing a sentence, or can be autonomously inserted by a reasoning system. The advantages of using these affordances over the "theme" or "rheme" are that they are unique, can be easily computed, and support autonomous annotation.

4.2.2 Affordances for Internal Percepts

Internal percepts are self-perception of an internal state of the robot. This is based on the robot's beliefs about the people and objects in the world. Internal states are typically determined by the Cognitive System of the robot. Three internal states of the robot $\textit{internal state}_{\text{acknowledge}}$, $\textit{internal state}_{\text{aliveness}}$, and $\textit{internal state}_{\text{confused}}$ [4] are substituted with three affordances (rows 6 - 8) from time delays and typing.

The *countdown* perceptual schema provides the two affordances based on time delays: *Elapsed Listening Time > 6 sec* and *Elapsed Idle Time > 15 sec*. The affordance listed on row 6 approximates the $\textit{internal state}_{\text{acknowledge}}$ by thresholding the time interval between the robot's responses to the human during dialog. If the *Elapsed Listening Time > 6 sec*, the internal state of the robot is set to acknowledge. The affordance for the $\textit{internal state}_{\text{aliveness}}$ from row 7 is a timeout based on the idleness of the robot. If the *Elapsed Idle Time > 15 sec*, the internal state of the robot is set to aliveness. Since the existing literature does not provide guidance on specific values for back-channels and acknowledgements [1, 8], or aliveness [4], a suitable timeout was estimated by the researchers to communicate the corresponding function of head gaze effectively.

The affordance listed in Row 8 is based on typing and used only when an operator is using the robot. This affordance is returned by the *detect_typing* perceptual schema. It approximates the *internal state*$_{\text{confused}}$ [4] of the robot by thresholding the number of deletes and retypes in a time interval t. If the *Number of Deletes/Retypes by an Operator* > 5 *within a Time Interval* $t = 15$ *sec*, the internal state reflects confusion.

4.3 Head Gaze Acts

The robot used five head gaze acts: *fixate, avert, concurrence, confusion, and scan.* The gaze acts are robot-dependent, as each robot would have its own implementation of acts based on its degrees of freedom and motor characteristics. The Survivor Buddy robot moved with an average velocity of $33°/sec$ for all gaze acts. Since the existing literature does not provide guidance on specific values for the velocity, a suitable velocity was chosen by the researchers to communicate the corresponding function of head gaze effectively. The implementation specifics of the gaze acts matched the known parameters used by earlier implementations of head gaze in the literature, and are described below.

1. *Fixate* moves the robot's head to a position facing the human directly at a velocity of $33°/sec$. Fixation occurs for an indefinite duration until another gaze act activates [1–13, 15, 16].

2. *Avert* is a +/- 7° simultaneous horizontal and vertical movement of the head, away from the fixation point [21] at a velocity of $33°/sec$. Aversion occurs for an indefinite duration until another gaze act activates [1,2,6,9–11,13,16,19–22].

3. *Concurrence* is a repetitive vertical head movement of +/- 10° [1, 8, 21] at a velocity of 33°/*sec*. Concurrence occurs once every 3 seconds.

4. *Confusion* is a head roll of +/- 20° at a velocity of 33°/*sec*. The head returns to the fixation point after 1 second [4] at a velocity of 33°/*sec*.

5. *Scan* is a fixation persisting between 0.77 sec to 1 sec to a series of three random points in space [4, 8, 13] at a velocity of 33°/*sec*.

4.4 Production Rules

The five social head gaze behaviors were implemented using production rules. The production rules are if-then statements; if an affordance (LSHG-S) or percept (TSHG-S) is perceived, the head gaze act is called. A total of nine production rules (Table 4.2) were implemented on the Survivor Buddy robot. The production rules directly correspond to the five behaviors of head gaze, and a single head gaze behavior may be comprised of one or more production rules.

Production rule 1 is used by the robot for *Communicating Social Attention* in a human. The robot fixates toward the human to indicate attention [1–16, 40]. Production Rules 2-6 are designed for *Regulating an Interaction* with the human. The robot uses rules of turn-taking to fixate and avert from the human [1, 2, 6, 9–11, 13,16,20–22]. *Establishing Agency* is implemented on the robot using production rule 7. The robot uses the *scan* gaze act to randomly look at different points in space, and to indicate it is alive and functioning properly [4, 8]. *Projecting Mental State*, such as confusion, is accomplished by production rule 8 [4]. Production Rule 9 is used to implement *Manifesting an Interaction*, where the robot fixates toward an object in the environment when it utters the object name in speech [1, 2, 4, 8, 16, 27–29]. These production rules were identified from the literature (Table 4.2) of head gaze

	Behavior	Production Rule (LSHG-S)	Production Rule (TSHG-S)
1	Communicating Social Attention [3–5] [6–11] [1, 2, 12, 13, 15, 16]	IF **Approach of Human**, THEN **Fixate** toward the human for an indefinite duration at a velocity of 33°/sec.	IF **Presence of Human** and **Human Shows Initial Interest**, THEN **Fixate** toward the human for an indefinite duration at a velocity of 33°/sec.
2		IF **Initial Word**, THEN **Avert** from the human with a +/ -7°/sec simultaneous horizontal and vertical movement for an indefinite duration at a velocity of 33°/sec.	IF **Presence of Human**, **First word in Theme**, and **Start of Turn**, THEN **Avert** from the human with a +/ -7°/sec simultaneous horizontal and vertical movement for an indefinite duration at a velocity of 33°/sec.
3		IF **Word following Punctuations . ? !**, THEN **Avert (p = .73)** from the human with a +/- 7°/sec simultaneous horizontal and vertical movement for an indefinite duration at a velocity of 33°/sec.	IF **Presence of Human**, **First word in Theme**, and **Middle of Turn**, THEN **Avert (p = .73)** from the human with a +/- 7°/sec simultaneous horizontal and vertical movement for an indefinite duration at a velocity of 33°/sec.
4	Regulating an Interaction [19], [6, 9–11, 13], [1, 2, 16, 20–22]	IF **After 75% of Words between Punctuation : . ! ?**, THEN **Fixate (p = .7)** toward the human for an indefinite duration at a velocity of 33°/sec.	IF **Presence of Human**, **First word in Rheme**, and **Middle of Turn**, THEN **Fixate (p = .7)** toward the human for an indefinite duration at a velocity of 33°/sec.
5		IF **Carriage Return**, THEN **Fixate** toward the human for an indefinite duration at a velocity of 33°/sec.	IF **Presence of Human**, **First word in Rheme**, and **End of Turn**, THEN **Fixate** toward the human for an indefinite duration at a velocity of 33°/sec.
6		IF **Elapsed Listening Time > 6 sec**, THEN **Concurrence** toward the human with repetitive vertical head movement of +/- 10° every 3 seconds at a velocity of 33°/sec.	IF **Listening to Human** and **Internal State**$_{acknowledge}$ THEN **Concurrence** toward the human with repetitive vertical head movement of +/- 10° every 3 seconds at a velocity of 33°/sec.
7	Establishing Agency [4, 8]	IF **Elapsed Idle Time > 15 sec**, THEN **Scan** three random points in the environment at a velocity of 33°/sec.	IF **Internal State**$_{aliveness}$, THEN **Scan** three random points in the environment at a velocity of 33°/sec.
8	Projecting Mental States [4]	IF **Number of Deletes/Retypes by an Operator > 5 within a Time Interval t = 15 sec**, THEN **Confusion** toward the human with a head roll of +/- 20° and return to the fixation point at a velocity of 33°/sec.	IF **Internal State**$_{\boldsymbol{confused}}$, THEN **Confusion** toward the human with a head roll of +/- 20° and return to the fixation point at a velocity of 33°/sec.
9	Manifesting an Interaction [28] [1, 2, 4, 8, 16]	IF **The Object Name Tag**, THEN **Fixate** toward the object in the environment at a velocity of 33°/sec.	IF **Presence of Object** and **Onset of Object Utterance**, THEN **Fixate** toward the object in the environment at a velocity of 33°/sec.

Table 4.2: Comparison of the Nine Production Rules for LSHG-S and TSHG-S

in human-robot interaction, and can be expanded if new studies identify new uses of head gaze.

The C# code snippet for the implementation of production rules 2-5 of Regu-

LATING AN INTERACTION behavior for LSHG-S is illustrated below:

```
using namespace Behavior
{
       StringBuilder Regulating_an_Interaction (string current_robot_turn)
       {
       static int location = 0;
       StringBuilder buffer = new StringBuilder();
       buffer.Append(current_robot_turn);
       For (int i = 0; i < GetWords.TotalNumber(buffer); i++)
       {
       if (PSchema.extract_sentence_structure(‘‘Initial Word’’, location))
           buffer.Insert( ‘‘<Avert>’’, location);
       if (PSchema.extract_sentence_structure(‘‘Word Following
           Punctuations’’, location))
           buffer.Insert(‘‘<Avert,.73>’’, location);
       if (PSchema.extract_sentence_structure(‘‘After 75% of Words
           between Punctuations’’, location))
           buffer.Insert(‘‘<Fixate,.7>’’, location);
       if (PSchema.extract_sentence_structure(‘‘Carriage Return’’,
           location))
           buffer.Insert(‘‘<Fixate>’’, location);
       }
       return buffer;
       }
 }
```

Figure 4.3: C# Code for REGULATING AN INTERACTION Behavior for LSHG-S.

4.5 Coordination of Multiple Head Gaze Behaviors

Coordination ensures that the robot is sensitive to the current social function and conveys the appropriate meaning. The implementation of the production rules occur in parallel and use separate threads; therefore multiple production rules may be active at any time t, and need to be coordinated. This work coordinates head movements simultaneously for five behaviors. Prior work [1,2] focusses on coordination of just two or three behaviors, not all five together. The coordination function is implemented using a prioritization scheme based on timestamps (highest priority for the most recent behavior) and the nature of the behavior (atomic or non-atomic). The Projecting a Mental State and Manifesting an Interaction behaviors are atomic and run to completion without interruption. All other behaviors are non-atomic and may be interrupted by the most recent behavior. The algorithm below describes this function.

```
if (Max(type(B) = = ‘‘atomic’’)
    overall_response = (G * Min(B(type=‘‘atomic’’,timestamp)));
else
    overall_response = (G * Max(timestamp(B)));
```

Projecting a Mental State and Manifesting an Interaction are classified as atomic because any interruptions from other head gaze acts may detract from the goals of the robot, perceived understanding of the robot actions, and social competence of the robot. These behaviors are typically uninterrupted in human-human communication. Projecting a Mental State, such as anger makes any detected interruption, such as an aversion of head for turn-taking, of secondary importance. Expressing the emotion fully is more important than a filler gaze act like an aversion of the head that can be used to build rapport. Referential head gaze in humans is used to draw the attention of the human to an object. The purpose of the Manifesting an Interaction behavior will not be accomplished if it is interrupted midway through the process. In terms of implementation, this translates to production rules 8 and 9, shown in Figure 4.2, running to completion without interruption. All other production rules (1-7) can be interrupted by the most recent production rule that is activated.

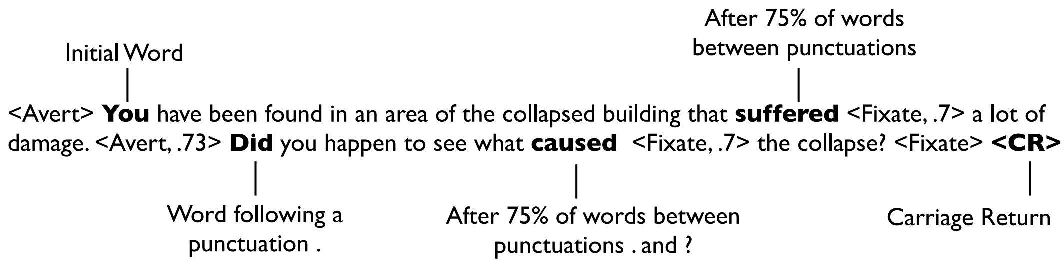


Figure 4.4: Buffer (LSHG-S)

The activation of production rules using the LSHG-S and TSHG-S implementations are described using an example. The input sentence from the script is as follows: "You have been found in an area of the collapsed building that suffered a lot of damage. Did you happen to see what caused the collapse?"

LSHG-S: This input fills the buffer variable (Fig. 4.3) that interleaves both speech and head gaze. The affordance *Initial word* "You" activates production rule 2 and inserts the <*Avert*> head gaze tag. Next, affordance *After 75% of words between punctuations* activates production rule 4 and inserts <*Fixate, .7*> after the word "suffered." An <*Avert, .73*> gaze tag is inserted when the affordance *Word following a punctuation .* is perceived, before the word "Did" as described by production rule 3. Using production rule 4 another middle of turn <*Fixate, .7*> is inserted after the word "caused" when the affordance *After 75% of words between punctuations . and ?* is perceived. Finally, the affordance *Carriage Return* activates production rule 5 to insert a <*Fixate*> to indicate the end of turn.

The contents of the buffer variable at the end of the robot's current turn is shown in Fig. 4.4. This passes through the Microsoft speech system, which converts the text-to-speech and triggers the gaze act when it encounters a gaze tag. The run time for this implementation is $O(n)$ where n is the length of the text that needs to be parsed. The affordances from sentence structure has low computational costs and makes it attractive to be used in real-time applications. However some of cues necessary for implementation of autonomous head gaze like object recognition or understanding of natural language may not run in linear time and act as limiting factors.

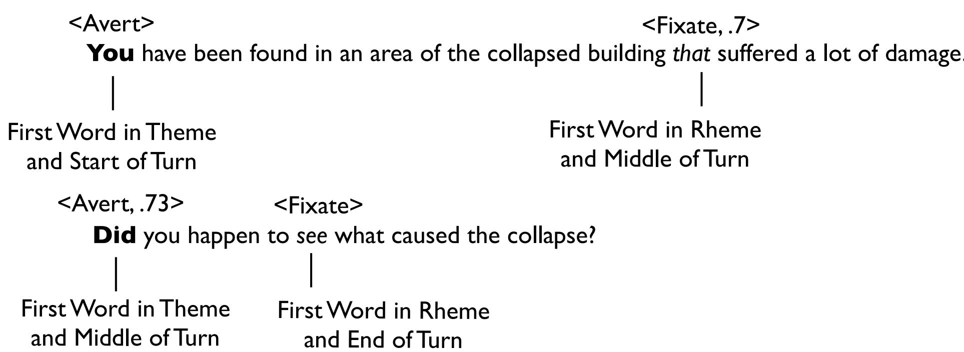


Figure 4.5: Marked Pre-Recorded Audio File(TSHG-S)

TSHG-S: The linguistic percepts and turn events (for example, First Word in Theme and Start of Turn) are identified by manual inspection, and marked on a pre-recorded audio file (Fig. 4.5). The percept *First Word in Theme and Start of Turn* “You” activates production rule 2 and the <*Avert*> head gaze act is called. Next, percept *First Word in Rheme and Middle of Turn* “that” activates production rule 4 and the <*Fixate, .7*> head gaze act is triggered. This is followed by an <*Avert, .73*> head gaze act, which is generated when the percept *First Word in Theme and Middle of Turn* is perceived, at the onset of the word “Did”, as described by production rule 3. Finally, the percept *First Word in Rheme and End of Turn* “see” activates production rule 5 to call the <*Fixate*> head gaze act.

4.7 Summary

This section outlines the implementation of LSHG-S on "Survivor Buddy," an affective four degree of freedom robot. The software implementation of LSHG-S system was C# and used the Microsoft text-to-speech system with Microsoft Anna voice. The mechanisms for executing five head gaze acts *Fixate*, *Avert*, *Concurrence*, *Confusion*, and *Scan* are described in terms of range, velocity, and duration. The eight affordances from structure of sentences used in dialog, time delays, and typing used to substitute linguistic and internal percepts are provided: *Initial Word*, *Word following Punctuation : . ! ?*, *After 75% of Words between Punctuation : . ! ?*, *Carriage Return*, *Elapsed Listening Time > 6 sec*, *Elapsed Idle Time > 15 sec*, *Number of Deletes/Retypes by an Operator > 5 within a Time Interval t = 15 sec*, and *The Object Name Tag*. These affordances are unique, can be easily computed, and support autonomous annotation. The behaviors are implemented using production rules. The production rules are if-then statements where if a percept is perceived, the head gaze act is called. The production rules directly correspond to the five behaviors, and a single behavior may comprise one or more production rules. Detailed descriptions of nine production rules (Table 4.2) used to map the affordances on to head gaze act(s) are given. The coordination function is implemented using a prioritization scheme based on timestamps (highest priority for the most recent behavior) and the nature of the behavior (atomic or non-atomic). This work coordinates head movements simultaneously for five behaviors. Prior work [1, 2] focuses on coordination of just two or three behaviors, not all five together. The run time for LSHG-S system is $O(n)$ where n is the length of the text that needs to be parsed. Finally, the activation of production rules, using both TSHG-S and LSHG-S implementations, are described with an example.

5. EXPERIMENTS

This section describes the experimental methodology and design of a large-scale 93-participant experiment for assessing the social acceptance of LSHG-S. An experimental validation was necessary because there are two potential problems with using LSHG-S: (1) The activation of the head gaze act for LSHG-S can lead, lag, or occur at the onset of the speech utterance. Unlike the well validated TSHG-S, it is not precise and the head gaze acts may not precisely match the dialog presented. This may annoy or confuse the human. (2) Social head gaze for realistic scenarios comprises of five functions of head gaze: *look interested in human(s)*, *engage in a verbal conversation with human(s)*, *convey general liveliness and awareness*, *showing various mental states*, and *referential gaze* to objects in the environment. The change in synchronization for two functions such as *engaging in a verbal conversation with human(s)* and *referential gaze* to objects in the environment may impact the human's perception of the robot's overall head gaze during the interaction. Five hypotheses concerning the impact of the LSHG-S and TSHG-S were formulated to answer the primary research question described in Section 1. In this section, details of the experiment are described relating to the participants, equipment, personnel, and experimental methods. 23 measures were used to evaluate the overall social acceptance of the robot.

5.1 Hypotheses

Five hypotheses were tested in this experiment to evaluate secondary research question 3 – "*Is it possible to evaluate through sound experimental methods the effectiveness and appropriateness of the head gaze acts generated using the behavioral robotics framework?*"

- *Hypothesis 1 (H1)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

- *Hypothesis 2 (H2)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

- *Hypothesis 3 (H3)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

- *Hypothesis 4 (H4)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

- *Hypothesis 5 (H5)*: The LSHG-S condition improvements over the NHG-S condition will be comparable to those of the TSHG-S condition.

Hypotheses H1, H2, H3, and H4 tests the effectiveness of LSHG-S and TSHG-S in eliciting high levels of social acceptance, similar to that observed in the current head gaze literature. Hypotheses H1 and H2, specifically test whether the simplifying inferences used to enable LSHG-S are sufficient for social head gaze. Hypothesis H5 tests whether participants notice the difference in the quality of head gaze generated by the LSHG-S and TSHG-S conditions.

5.2 Experimental Design

A 1 x 3 between-subjects experiment was designed to evaluate three head gaze conditions: (1) LSHG-S, (2) TSHG-S, and (3) no head gaze-speech (NHG-S). The scenario was a simulated disaster scenario of a parking garage collapse (Fig. 5.1), wherein Survivor Buddy, the rescue robot shown in Fig. 5.1, has a dialog with a trapped victim based on a 911 dispatch and triage protocols (Appendix F). This setting was chosen because (1) it is an extension of [14], which illustrated that an extreme setting heightens affective responses from participants, (2) participants tend to naturally follow the dispatcher protocol and thus a pre-scripted dialog would not confound the experiments, and (3) the dialog captured all five social head gaze behaviors at least once (none of the 32 previous studies reported a dialog that captured all known social head gaze behaviors).

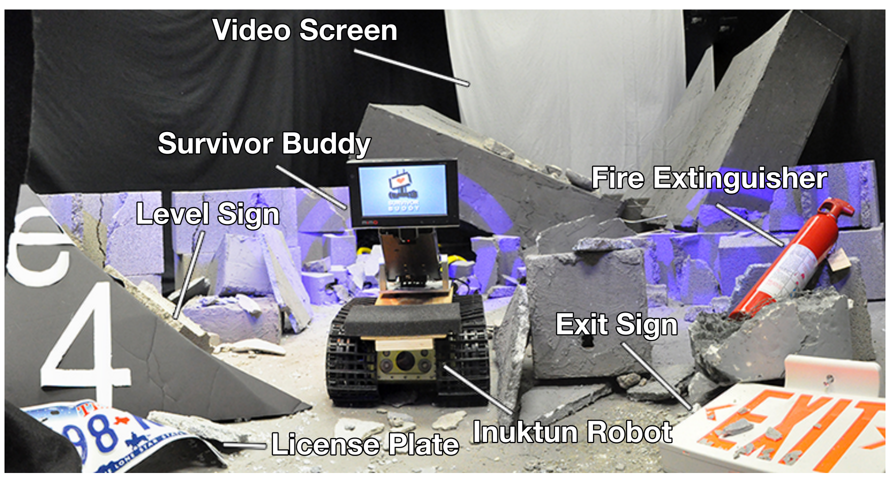


Figure 5.1: The Simulated Disaster Area from the Participant's Point of View

The environment was comprised of the trapped victim, prop concrete floors, columns with rebar, simulated glass pieces, and objects typically found in a parking garage, such as a fire extinguisher, license plate, parking level sign, and exit sign. The environment also had a full theatrical stage lighting system in order to provide optimum visibility without sacrificing a lifelike effect. The participant interacted with Survivor Buddy for approximately 15 minutes. As per the 911 dispatch protocol, the dialog focused on assessing the participant's physical health, and gaining information about the location and nature of the event. The dialog ensured the activation of each of the nine production rules at least once (none of the previous studies reported a dialog that captured all nine production rules) with a total of 162 possible gaze acts.

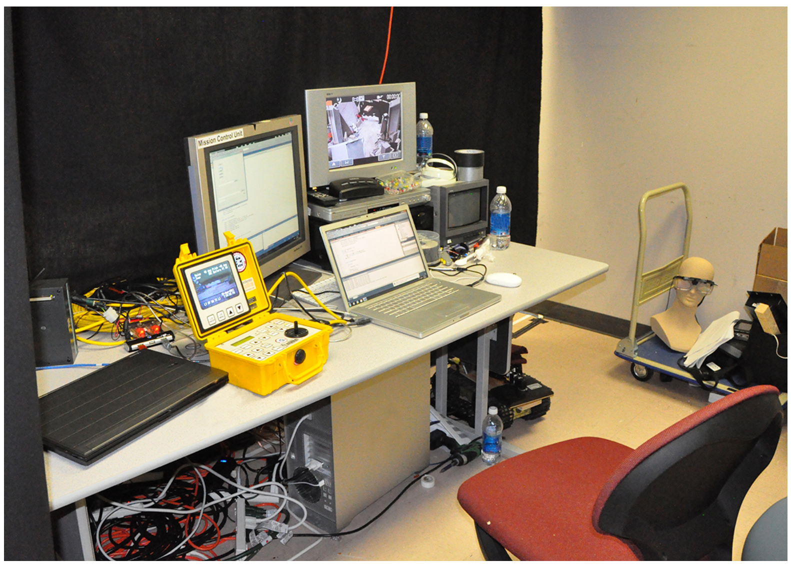


Figure 5.2: Robot Control Area.

The experiment utilized *a priori* sensor data for object locations, participant head locations, and *internal state*$_{\text{confused}}$ of the robot. A hidden operator ("wizard") [47] controlled the robot present in the simulation area from the robot control area (Fig. 5.2 and 5.5). The hidden operator received video feeds of three camera viewpoints: (1) overhead view, (2) robot point of view, and (3) participant point of view. The operator used this information to determine the state of the interaction and stepped through conversation turns, using predefined sentences and phrases rather than real-time typing. The hidden operator's interface is shown in Fig. 5.3. This approach overcame limitations in existing state-of-the-art object and speech recognition systems, and ensured repeatability and consistency across conditions. This led to the robot (via the hidden operator) controlling the direction of the conversation.

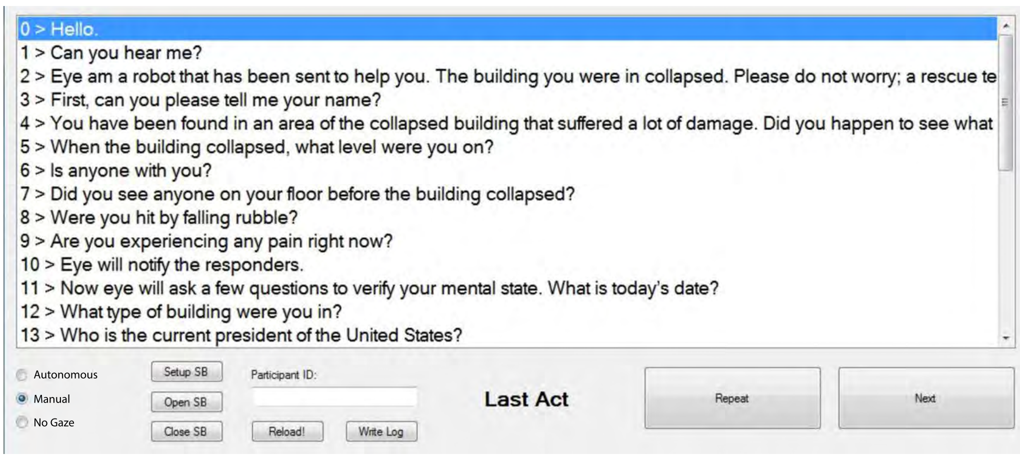


Figure 5.3: Robot Operator's Interface

In the LSHG-S condition, Survivor Buddy displayed head gaze behaviors, using the proposed method for the generation of head gaze based on sentence structure and time delays. The TSHG-S condition displayed gaze behaviors based on the semantic content of the dialog, which was similar to gaze behaviors exhibited in human-human conversation. The semantic units of "theme" and "rheme" were manually annotated by inspection using definitions described by Halliday [53]. The theme refers to the part of an utterance that sets the tone of the utterance and connects the previous utterance to the next one. The rheme contains the new information that the utterance intends to communicate. This same procedure was used by [9] and [22]. In the

NHG-S condition, Survivor Buddy looked directly at the participant throughout the interaction, without displaying any head gaze acts, and used only speech to interact.

5.3 Participants

An *a priori* power analysis projected a minimum of 42 participants based on having three groups, 80% power, a medium effect size of 0.25, and an $\alpha = 0.05$. Significance level α is the probability of making a wrong decision when the null hypothesis is true [62]. A total of 93 participants completed the experiment. The participants included 53 males and 40 females within an age range of 18-67 ($M = 32.38, SD = 14.84$). People with diverse backgrounds including students, engineers, administrators, firefighters, technicians, and doctors participated in the experiment. The ethnic backgrounds of the participants consisted of 68.8% Caucasian, 7.5% Asian, 14% Hispanic, 3.2% African American, and 6.5% Middle Eastern. Of the 93 participants, 64 reported owning a pet, and 8 reported owning a robot. The participants' familiarity with robots was low ($M = 2.09$, $SD = 1.67$ on a scale of 1 to 7). Video gaming experience was moderate ($M = 3.23$, $SD = 5.7$ on a scale of 1 to 7). Compensation for the participants included a chance to win a door prize, and they were not required to complete the study to be included in the drawing.

5.4 Equipment and Personnel

Seven pieces of equipment were used for this study. The Survivor Buddy [63] platform discussed in Section 4 was used for the robot role. Eye trackers (Figure 5.4) were used to capture where the participants were looking during their interaction with the robot. The robot was controlled using a Desktop Macbook Pro computer using a tether. Participants used a laptop to take pre- and post-assessments (Appendix G and H) used in the study. Finally, three video and audio feeds from overhead view, robot point of view, and participant point of view, were recorded by three cameras.

Figure 5.4: iScan Eye Tracker.

Additional equipment (e.g., power, cables) was included as needed.

The investigator of this work was primarily responsible for setting up experimental protocols, running all of the trials, and for analysis of the results. Two graduate students helped conduct this study. The first graduate student was responsible for recording the eye tracking data and the calibration process. This student was also responsible for leading the participants into the physical simulated confined space. Another graduate student helped place the robot in the starting position for each interaction, and ensured that the robots were powered on, and that the batteries were charged. They were also responsible for all video operations and maintenance, including: charging of batteries, syncing recordings, and labeling and backing up interactions.

5.5 Experimental Method

The experiment was conducted in three phases: 1) pre-interaction phase - participant check-in, consent (Appendix C), and pre-interaction questionnaires (Ap-

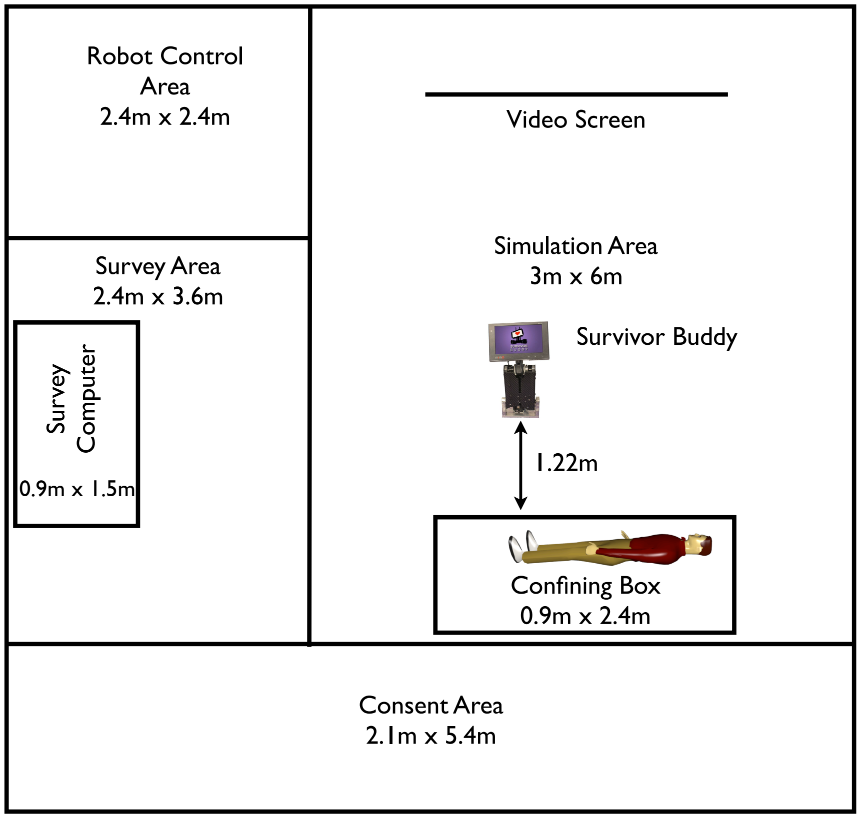


Figure 5.5: The Layout of the Study Site.

pendix G), 2) interaction phase - interaction with the robot, and 3) post-interaction phase - post-interaction questionnaires (Appendix H) and debriefing (Appendix J). The entire study took approximately 60 minutes per participant to conduct, with data collection occurring over a period of 5 weeks. The experiment protocol is summarized in Appendix E.

5.5.1 Pre-Interaction Phase

In pre-interaction phase, the participants were first shown a video explaining the IRB consent process, after which each participant read and signed the consent forms (Appendix C), as required by the Texas A&M University Institutional Review Board. The participants were then read instructions regarding what they could expect while participating in the study (Appendix D).

Each participant was then assigned to one of three conditions (between-subjects factor): LSHG-S, TSHG-S, and NHG-S. The participants were then outfitted with the eye tracker, and after the eye tracking calibration process, the participants were requested to step into the simulated confined space (wooden structure) (0.91 m x 0.91 m x 2.44 m) and play the role of a victim. The lights in the study site were turned off, and the participants were asked to lie down on their right side in the confined space.

5.5.2 Interaction Phase

During the interaction phase, the participant covered themselves with a sleeping bag, and the lights illuminating the pathway to the wooden structure in the experiment site were turned off. The participants saw a brief dramatic video, which showed a first-person view of a parking garage collapse. The video ended with flashes of light, followed by lighting sufficient to see the robot. At this point in the experiment, the simulation of a collapsed parking structure disaster environment, and the robot system were made visible to the participant (Fig. 5.1). The Survivor Buddy robot system was at a distance of 1.22 meters, within the participant's personal zone [14], so as to increase the likelihood of social interaction. The screen displayed only a Survivor Buddy logo so that the only social cues were voice and head gaze acts. The interaction with the participant lasted approximately 15 minutes, and the robot followed a predefined script consisting, of questions and simple directions (Appendix F). The robot could also repeat portions of the script upon request. The robot supervisor (hidden from view) would activate the text for the robot's turns in the dialog.

The robot activated the COMMUNICATING SOCIAL ATTENTION behavior to gain the participant's attention and convey that it was interested and ready for an interaction. The robot then used the REGULATING AN INTERACTION behavior for effective human-like turn-taking and back-channeling, during a purposeful dialogue with the victim based on 911 dispatch and triage protocols. The dialogue focused on assessing the participant's physical health and gaining information about the location and nature of the event. The robot posed questions like: *"Are you experiencing any pain right now?"* and *"Did you happen to see what cause the collapse?"* Then the dialog shifted to questions that assessed mental function, for example, "Paper is used commonly for writing. Can you name as many alternative uses as you can?" The robot also monitored the area surrounding the participant and used the MANIFESTING AN INTERACTION behavior to point toward objects of interest such as fire extinguishers or hazardous objects. The robot activated the PROJECTING MENTAL STATE behavior to indicate confusion and provide feedback to the participant about its internal state. The robot used the ESTABLISHING AGENCY behavior to convey aliveness and let the participant know that it was functioning properly.

The five behaviors COMMUNICATING SOCIAL ATTENTION, REGULATING AN INTERACTION, MANIFESTING AN INTERACTION, PROJECTING MENTAL STATE, and ESTABLISHING AGENCY were manifested at least once during the study. The interaction concluded with the participant being informed that rescuers had arrived, and Survivor Buddy's head closing to signify no further engagement. After the completion of the interaction, the participants were removed from the confined space (wooden box) and the eye tracker system was removed.

5.5.3 Post-Interaction Phase

In post-interaction phase, the participants were asked to complete a post-interaction assessment (Appendix H), after [9, 14, 64], in order to access participants' affective state and perceptions of the robot. Following the completion of the post-interaction questionnaire, the participants were debriefed about the study (Appendix J), after which they were free to leave.

5.6 Experimental Measures

The experiment used pre-interaction and post-interaction questionnaires, eye tracking measures, and video observations to evaluate the social acceptance of the robot.

1. *Pre-Interaction Questionnaire* - The pre-interaction questionnaire consisted of nine standard attributes regarding the participant's age, occupation, gender, education level, prior robot experience, ethnicity, prior video gaming experience, robot ownership, and pet ownership [14, 64]. Appendix G reports the pre-interaction questionnaire items used in the study.

2. *Post-Interaction Questionnaires* - The post-interaction questionnaires consisted of 21 measures. These measures comprised of 16 standardized attributes used to determine participants' affective state and participants perception of the robot. They are: SAM: Valence [55], SAM: Arousal [55], Chance of Rescue [64], Robot Engagement [1, 8, 65], Robot Likeability [64], Human-Like Behavior [9], Robot Intelligence [9], Robot Detachment [64], Robot Confidence [64], Robot Competence [64], Robot Unpleasantness [64], Robot Extraversion [64], Understanding Robot Behavior [65], Gaze-Speech Synchronization [65], Looking at Objects at Appropriate Times [65], and Natural Movement [9, 65].

   Five *new* measures were developed for victim management: Person at Ease, Robot Empathy, Robot Loyalty, Robot Integrity, and Robot Caring. These measures were developed to understand if the participants' perceive the robot as a “friendly companion” [66] during a victim management scenario. These measures have not been reported before for head gaze.

Two questions from the Self Assessment Manikin (SAM) [55] used a nine point Semantic Differential scale. The remainder of the questions used a seven point Likert scale, with one indicating strong disagreement or strongly negative, and seven indicating strong agreement or strongly positive. Measures with multiple items were checked for internal consistency, using a Cronbach’s Alpha statistic [67] (Cronbach’s $\alpha > .70$ is considered to be reliable). Several questions were reverse coded to prevent participants from uniformly selecting a single rating. Appendix I reports the list of items, and their reliability with regard to the 23 measures used to evaluate the performance of LSHG-S.

3. *Eye Tracking Measures* - For this experiment ISCAN eye trackers were utilized. Eye trackers can be useful in determining where exactly the participant looks during the course of the interaction with robot. Other measures like the time taken to look at an object or total time the participant spent looking at the robot can also be calculated.

4. *Video Observations* - Infrared night vision cameras were utilized because the robot interactions occurred in low light conditions. Images were obtained from three camera perspectives: face view, participant view, and overhead view. This data was analyzed to evaluate two objective measures, Memory [68] and Creativity [69], captured during the interaction. Creativity and Memory were measured as part of this study because research has demonstrated that these measures inversely correlate with stress [70, 71]. Stress indicates arousal, which is a dimension of affect [72]. These measures have not been reported before for head gaze.

5.7 Summary

This section began with two motivations for conducting an experimental validation - (1) For LSHG-S, the head gaze acts may not precisely match the dialog. This may annoy or confuse the human. (2) The change in synchronization for two functions *engaging in a verbal conversation with human(s)* and *referential gaze* to objects in the environment may impact the human's perception of the robot's overall head gaze during the interaction. This was followed by a description of five hypotheses used to evaluate research question R3. The experimental design and three justifications on the choice of setting and dialog was presented. The simulated parking garage collapse scenario was presented with detailed descriptions on the objects in environment, lighting effects, and the 911 dispatcher protocol dialog. The experiment employed a hidden operator ("wizard") [47] to control the robot. The role of the hidden operator and the rationale for using this approach was explained. A total of 93 diverse participants between ages 18-67 participated in the study. The details on power analysis used for participant recruitment, demographics, and sources of recruitment were presented. In addition, a description of the equipment used in the study and responsibilities of personnel involved in the study were provided. The study was conducted in three phases with a duration of approximately 60 minutes per participant occurring over a period of 5 weeks. Details on all three phases (1) participant check-in, consent, and pre-interaction questionnaires, (2) interaction with the robot, and 3) post-interaction questionnaires and debriefing were provided. Lastly, the experimental measures used to evaluate the effectiveness of LSHG-S were presented. Three types of measurement tools – (1) pre-interaction and post-interaction questionnaires, (2) eye tracking measures, and (3) video observations – comprising a total of 23 measures were used to evaluate the social acceptance of the LSHG-S.

Out of the 23 measures, 16 measures use standardized questionnaires. Five *new* measures were developed for victim management: Person at Ease, Robot Empathy, Robot Loyalty, Robot Integrity, and Robot Caring. Additionally, two objective measures Creativity and Memory were measured. These seven measures have not been reported before for head gaze. The details in this section provide adequate information for reproducibility of the study by other investigators.

6. DATA ANALYSIS AND RESULTS *

This section presents the details of the data analysis conducted on data collected from three measurement tools – (1) pre-interaction and post-interaction questionnaires, (2) eye tracking measures, and (3) video observations – during the large-scale 93 participant human-robot interaction study. Each of the five hypotheses presented in Section 5 were evaluated:

1. *Hypothesis 1 (H1)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

2. *Hypothesis 2 (H2)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

3. *Hypothesis 3 (H3)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

4. *Hypothesis 4 (H4)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

5. *Hypothesis 5 (H5)*: The LSHG-S condition improvements over the NHG-S condition will be comparable to those of the TSHG-S condition.

*Reprinted with permission from "Evaluation of Head Gaze Loosely Synchronized with Synthetic Speech for Social Robots" by Srinivasan, V., Bethel, C.L., Murphy, R.R, 2014. IEEE Transactions on Human-Machine Systems, Accepted for Publication, Copyright [2014] by Vasant Srinivasan.

A Principal Component Analysis (PCA) was performed to group the 21 reliable measures into three independent categories: *Participants' Positive Affective State, Participants' Perception of the Robot*, and the *Consistency and Appropriateness of the Robot's Head Movements*. The three categories helped to better understand and interpret the results from such a large set of measures (Fig. 6.1). This section presents an results from an analysis of seven covariates – gender, age, video gaming experience, robot ownership, pet ownership, past experience with robots, and ethnicity – on the measured attributes. The interaction log files from the experiment were examined to determine the average number of head gaze acts activated in the LSHG-S and TSHG-S conditions.

6.1 Data Analysis

The data collected from three measurement tools – (1) pre-interaction and post-interaction questionnaires, (2) eye tracking measures, and (3) video observations – was analyzed. The data analysis for pre-interaction and post-interaction questionnaires, and video observations was straightforward. It consisted of an univariate ANOVA for each measure and a post-hoc analysis using Tukey's HSD. An ANCOVA was performed to further account for other potential sources of variance such as gender, age, video gaming experience, robot ownership, pet ownership, past experience with robots, and ethnicity.

The eye tracker data was analyzed, but was found to be unreliable. This was because of three reasons: 1) The eye tracker shifted from its original position as people laid down on the ground resulting in a change to the initial calibration; 2) The environment was very dark causing the eye tracker positions to be inaccurate; and 3) The eye lashes of people interfered with the ability of the tracking algorithm to perform accurately.

Both the Bonferroni correction [73] and FDR (Benjamini-Hochberg [74]) correction for multiple testing was applied. The Bonferroni correction method is the simplest and most conservative method used to correct for multiple testing. To perform the Bonferroni correction, the critical p-value (α) is divided by the number of comparisons being made. Since 21 dependent variables were analyzed and 21 statistical tests were performed, the corrected significance level after the Bonferroni correction [73] is $p < 0.0024$ (0.05/21).

The False Discovery Rate (FDR) Benjamini-Hochberg procedure is a standard iterative stepwise algorithm that controls the false discovery rate at level α [74] to correct for multiple testing.

FDR is defined as -

$$FDR = E\left[\frac{V}{R}\right]$$

where V is the total number of false discoveries (type I errors) and R is the total number of rejected null hypotheses (known as discoveries). Additionally, $\frac{V}{R}$ is defined to be 0 if R is 0. The advantage of FDR over the Bonferroni correction is that FDR has greater power at the cost of higher type I errors than the Bonferroni [74]. This is because the Bonferroni correction is designed to reduce the probability of even one false discovery (incorrectly rejected null hypotheses) from occurring, but FDR is designed to reduce the proportion of false discoveries, and is considered less conservative. Algorithm 1 (Appendix K) describes the procedure for 21 tests and $\alpha = .05$. The corrected significance level after the FDR (Benjamini-Hochberg) correction [74] is $p < 0.033$. An implementation of the procedure in R was also used [75] to cross-check the calculations.

	Measure	Main Effect	Post-Hoc Results			Mean (M), Standard Deviation (SD)					
			LSHG-S vs NHG-S	TSHG-S vs NHG-S	LSHG-S vs TSHG-S	LSHG-S		TSHG-S		NHG-S	
Participants' Affective State	SAM: Valence	F[2, 90] = 3.97, p = .02, η^2 = .08	t(90) = -2.63, p = .03, d = -.55	p = .08	p = .90	1.09	1.49	1.25	1.36	2.06	1.45
	Creativity	F[2, 90] = 3.30, p = .02, η^2 = .07	t(90) = 2.52, p = .04, d = .53	t(90) = 2.51, p = .04,, d = .53	p = .99	14	4.49	13.96	4.43	11.17	3.50
	Memory	F[2, 90] = .766, p = .47, η^2 = .02	p = .49	p = .60	p = .98	4.50	2.20	4.40	1.25	3.93	1.96
	Person at Ease	F[2, 90] = 3.8, p = .03, η^2 = .08	t(90) = 2.48, p = .04, d = .52	p = .06	p = .98	6.23	1.63	6.19	.91	5.61	1.18
	SAM: Arousal	F[2, 90] = 8.43, p < .001, η^2 = .16	t(90) = 3.63, p = .001, d = .77	t(90) = 3.48, p = .002, d = .73	p = .99	7.68	1.42	7.61	1.70	6.09	1.97
	Chances of Rescue	F[2, 90] = 1.22, p = .3, η^2 = .03	p = .53	p = .29	p = .90	3.97	.50	4.03	.57	3.83	.40
Participants' Perception of the Robot	Robot Empathy	F[2, 90] = 4.54, p = .01, η^2 = .09	t(90) = 2.49, p = .04, , d = .53	t(90) = 2.72, p = .02, d = .57	p = .97	5.59	.93	5.65	.89	5.02	.86
	Robot Loyalty	F[2, 90] = 4.08, p = .02, η^2 = .08	t(90) = 2.73, p = .02, d = .56	p = .09	p = .79	5.43	.85	5.27	.81	4.75	.10
	Robot Integrity	F[2, 90] = 3.79, p = .03, η^2 = .08	t(90) = 2.55, p = .03, d = .54	p = .08	p = .93	6.02	.911	5.95	.80	5.49	.76
	Robot Caring	F[2, 90] = 4.20, p = .02, η^2 = .09	t(90) = 2.41, p = .047, d = .51	t(90) = 2.60, p = .03, d = .55	p = .98	5.18	.83	5.23	.89	4.56	1.23
	Robot Engagement	F[2, 90] = 3.30, p = .04, η^2 = .07	t(90) = 2.46, p = .04, , d = .52	p = .15	p = .82	6.06	.89	5.87	1.17	5.26	1.65
	Robot Likeability	F[2, 90] = 6.75, p = .002, η^2 = .13	t(90) = 3.05, p = .01, d = .64	t(90) = 3.33, p = .004, d = .70	p = .967	4.93	.60	4.97	.63	4.36	.89
	Human-Like Behavior	F[2, 90] = 8.9, p < .001, η^2 = .17	t(90) = 4.03, p < .001, d = .85	t(90) = 3.10, p = .01, d = .65	p = .627	5.10	1.42	4.74	1.34	3.54	1.74
	Robot Intelligence	F[2, 90] = 1.05, p = .35, η^2 = .02	p = .67	p = .33	p = .84	6.29	.75	6.40	.57	6.14	.77
	Robot Competence	F[2, 90] = .68, p = .51, η^2 = .02	p = .52	p = .65	p = .97	5.99	.77	5.94	.77	5.76	.89
	Robot Unpleasantness	F[2, 90] = .76, p = .47, η^2 = .02	p = .99	p = .50	p = .57	1.55	.79	1.37	.49	1.57	.70
	Robot Extraversion	F[2, 90] = .29, p = .75, η^2 = .006	p = .9	p = .95	p = .73	5.20	1.30	4.97	1.15	5.06	1.20
Consistency and Appropriateness	Understanding Robot Behavior	F[2, 90] = 18.087, p < .001, η^2 = .56	t(90) = 4.63, p < .001, d = .98	t(90) = 5.29, p < .001, d = 1.12	p = .99	5.57	.95	5.63	1.59	2.83	1.15
	Gaze-Speech Synchronization	F[2,90] = 47.9, p < .001, η^2 = .52	t(90) = 8.66, p < .001, d = 1.83	t(90) = 8.28, p < .001, d = 1.75	p = .93	5.90	1.27	5.78	1.08	2.93	1.63
	Looking at Objects at Appropriate Times	F[2,90] = 14.6, p < .001, η^2 = .54	t(90) = 4.82, p < .001, d = 1.02	t(90) = 5.12, p < .001, d = 1.08	p = .98	5.72	1.67	6.10	.99	3.28	1.43
	Natural Movement	F[2,90] = 16.7, p < .001, η^2 = .27	t(90) = 4.79, p < .001, d = 1.01	t(90) = 5.19, p < .001, d = 1.09	p = .92	4.70	1.62	4.83	1.60	2.77	1.38
	1	2	3	4	5	6	7	8	9	10	11

Figure 6.1: Summary of the Results

6.2 Evaluation of the Proposed Hypotheses

Fig. 6.1 lists the main effects (column 2) from the ANOVA, Tukeys' post-hoc results between the conditions (columns 3-5) for the 21 measures (column 1), and the mean and standard deviation for each of the measures (columns 6-11). The eta-squared (η^2) and Cohen's d effect sizes can be interpreted using the scale [76] shown in Table 6.1. The results for original ANOVA (Fig. 6.1), Bonferroni Correction (Fig.(6.2), and FDR (Benjamini-Hochberg) correction (Fig. 6.3) are described for each hypothesis.

Effect Size	**Eta-Squared (η^2)**	**Cohen's (d)**
Small	0.01 - 0.06	0.20 - 0.49
Medium	0.06 - 0.14	0.50 - 0.79
Large	0.14+	0.80+

Table 6.1: Interpretation of Effect Size

6.2.1 Results for Hypothesis 1 (H1)

Hypothesis 1 (H1) states that "Participants who interact with a robot exhibiting the LSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition." Consistent with H1, LSHG-S was preferred to NHG-S.

1. Original Results ($p < .5$): Support for four measures: SAM: Valence, Creativity, Person at Ease, and SAM: Arousal.

2. Bonferroni Correction ($p < .0024$) - Support for one measure: SAM:Arousal.

3. FDR (Benjamini-Hochberg) Correction ($p < .033$) - Support for four measures: SAM: Valence, Creativity, Person at Ease, and SAM: Arousal.

6.2.2 Results for Hypothesis 2 (H2)

Hypothesis 2 (H2) states that "Participants who interact with a robot exhibiting the LSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition." As predicted by H2, LSHG-S was preferred to NHG-S.

1. Original Results ($p < .5$): Support for 11 measures: Robot Empathy, Robot Loyalty, Robot Integrity, Robot Caring, Robot Engagement, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement.

2. Bonferroni Correction ($p < .0024$) - Support for six measures: Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement.

3. FDR (Benjamini-Hochberg) Correction ($p < .033$) - Support for 10 measures: Robot Empathy, Robot Loyalty, Robot Integrity, Robot Caring, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement.

6.2.3 Results for Hypothesis 3 (H3)

Hypothesis 3 (H3) states that "Participants who interact with a robot exhibiting the TSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition." As revealed by H3, TSHG-S was preferred to NHG-S.

	Measure	Main Effect	Post-Hoc Results			Mean (M), Standard Deviation (SD)					
			LSHG-S vs NHG-S	TSHG-S vs NHG-S	LSHG-S vs TSHG-S	LSHG-S		TSHG-S		NHG-S	
Participants' Affective State	SAM: Valence	$F[2, 90] = 3.97,$ $p = .02, \eta^2 = .08$	$t(90) = -2.63,$ $p = .03, d = -.55$	$p = .08$	$p = .90$	1.09	1.49	1.25	1.36	2.06	1.45
	Creativity	$F[2, 90] = 3.30,$ $p = .02, \eta^2 = .07$	$t(90) = 2.52,$ $p = .04, d = .53$	$t(90) = 2.51,$ $p = .04,, d = .53$	$p = .99$	14	4.49	13.96	4.43	11.17	3.50
	Memory	$F[2, 90] = .766,$ $p = .47, \eta^2 = .02$	$p = .49$	$p = .60$	$p = .98$	4.50	2.20	4.40	1.25	3.93	1.96
	Person at Ease	$F[2, 90] = 3.8,$ $p = .03, \eta^2 = .08$	$t(90) = 2.48,$ $p = .04, d = .52$	$p = .06$	$p = .98$	6.23	1.63	6.19	.91	5.61	1.18
	SAM: Arousal	$F[2, 90] = 8.43,$ $p < .001, \eta^2 = .16$	$t(90) = 3.63,$ $p = .001, d = .77$	$t(90) = 3.48,$ $p = .002, d = .73$	$p = .99$	7.68	1.42	7.61	1.70	6.09	1.97
	Chances of Rescue	$F[2, 90] = 1.22,$ $p = .3, \eta^2 = .03$	$p = .53$	$p = .29$	$p = .90$	3.97	.50	4.03	.57	3.83	.40
Participants' Perception of the Robot	Robot Empathy	$F[2, 90] = 4.54,$ $p = .01, \eta^2 = .09$	$t(90) = 2.49,$ $p = .04, , d = .53$	$t(90) = 2.72,$ $p = .02, d = .57$	$p = .97$	5.59	.93	5.65	.89	5.02	.86
	Robot Loyalty	$F[2, 90] = 4.08,$ $p = .02, \eta^2 = .08$	$t(90) = 2.73,$ $p = .02, d = .56$	$p = .09$	$p = .79$	5.43	.85	5.27	.81	4.75	.10
	Robot Integrity	$F[2, 90] = 3.79,$ $p = .03, \eta^2 = .08$	$t(90) = 2.55,$ $p = .03, d = .54$	$p = .08$	$p = .93$	6.02	.911	5.95	.80	5.49	.76
	Robot Caring	$F[2, 90] = 4.20,$ $p = .02, \eta^2 = .09$	$t(90) = 2.41,$ $p = .047, d = .51$	$t(90) = 2.60,$ $p = .03, d = .55$	$p = .98$	5.18	.83	5.23	.89	4.56	1.23
	Robot Engagement	$F[2, 90] = 3.30,$ $p = .04, \eta^2 = .07$	$t(90) = 2.46,$ $p = .04, , d = .52$	$p = .15$	$p = .82$	6.06	.89	5.87	1.17	5.26	1.65
	Robot Likeability	$F[2, 90] = 6.75,$ $p = .002, \eta^2 = .13$	$t(90) = 3.05,$ $p = .01, d = .64$	$t(90) = 3.33,$ $p = .004, d = .70$	$p = .967$	4.93	.60	4.97	.63	4.36	.89
	Human-Like Behavior	$F[2, 90] = 8.9,$ $p < .001, \eta^2 = .17$	$t(90) = 4.03,$ $p < .001, d = .85$	$t(90) = 3.10,$ $p = .01, d = .65$	$p = .627$	5.10	1.42	4.74	1.34	3.54	1.74
	Robot Intelligence	$F[2, 90] = 1.05,$ $p = .35, \eta^2 = .02$	$p = .67$	$p = .33$	$p = .84$	6.29	.75	6.40	.57	6.14	.77
	Robot Competence	$F[2, 90] = .68,$ $p = .51, \eta^2 = .02$	$p = .52$	$p = .65$	$p = .97$	5.99	.77	5.94	.77	5.76	.89
	Robot Unpleasantness	$F[2, 90] = .76,$ $p = .47, \eta^2 = .02$	$p = .99$	$p = .50$	$p = .57$	1.55	.79	1.37	.49	1.57	.70
	Robot Extraversion	$F[2, 90] = .29,$ $p = .75, \eta^2 = .006$	$p = .9$	$p = .95$	$p = .73$	5.20	1.30	4.97	1.15	5.06	1.20
Consistency and Appropriateness	Understanding Robot Behavior	$F[2, 90] = 18.087,$ $p < .001, \eta^2 = .56$	$t(90) = 4.63,$ $p < .001, d = .98$	$t(90) = 5.29,$ $p < .001, d = 1.12$	$p = .99$	5.57	.95	5.63	1.59	2.83	1.15
	Gaze-Speech Synchronization	$F[2,90] = 47.9,$ $p < .001, \eta^2 = .52$	$t(90) = 8.66,$ $p < .001, d = 1.83$	$t(90) = 8.28,$ $p < .001, d = 1.75$	$p = .93$	5.90	1.27	5.78	1.08	2.93	1.63
	Looking at Objects at Appropriate Times	$F[2,90] = 14.6,$ $p < .001, \eta^2 = .54$	$t(90) = 4.82,$ $p < .001, d = 1.02$	$t(90) = 5.12,$ $p < .001, d = 1.08$	$p = .98$	5.72	1.67	6.10	.99	3.28	1.43
	Natural Movement	$F[2,90] = 16.7,$ $p < .001, \eta^2 = .27$	$t(90) = 4.79,$ $p < .001, d = 1.01$	$t(90) = 5.19,$ $p < .001, d = 1.09$	$p = .92$	4.70	1.62	4.83	1.60	2.77	1.38
	1	2	3	4	5	6	7	8	9	10	11

Figure 6.2: Summary of the Results Using Bonferroni Correction. The Highlighted Cells Indicate Statistically Significant Results At $p < .0024$

1. Original Results ($p < .5$): Support for two measures: Creativity and SAM: Arousal.

2. Bonferroni Correction ($p < .0024$) - Support for one measure: SAM:Arousal.

3. FDR (Benjamini-Hochberg) Correction ($p < .033$) - Support for two measures: Creativity and SAM: Arousal.

6.2.4 Results for Hypothesis 4 (H4)

Hypothesis 4 (H4) states that "Participants who interact with a robot exhibiting the TSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition." Consistent with H4, TSHG-S was preferred to NHG-S.

1. Original Results ($p < .5$): Support for eight measures: Robot Empathy, Robot Caring, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement.

2. Bonferroni Correction ($p < .0024$) - Support for six measures: Robot Likeability, Human-Like Behavior, and Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement.

3. FDR (Benjamini-Hochberg) Correction ($p < .033$) - Support for eight measures: Robot Empathy, Robot Caring, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement.

	Measure	Main Effect	Post-Hoc Results			Mean (M), Standard Deviation (SD)					
			LSHG-S vs NHG-S	TSHG-S vs NHG-S	LSHG-S vs TSHG-S	LSHG-S		TSHG-S		NHG-S	
Participants' Affective State	SAM: Valence	$F[2, 90] = 3.97$, $p = .02$, $\eta^2 = .08$	$t(90) = -2.63$, $p = .03$, $d = -.55$	$p = .08$	$p = .90$	1.09	1.49	1.25	1.36	2.06	1.45
	Creativity	$F[2, 90] = 3.30$, $p = .02$, $\eta^2 = .07$	$t(90) = 2.52$, $p = .04$, $d = .53$	$t(90) = 2.51$, $p = .04,$, $d = .53$	$p = .99$	14	4.49	13.96	4.43	11.17	3.50
	Memory	$F[2, 90] = .766$, $p = .47$, $\eta^2 = .02$	$p = .49$	$p = .60$	$p = .98$	4.50	2.20	4.40	1.25	3.93	1.96
	Person at Ease	$F[2, 90] = 3.8$, $p = .03$, $\eta^2 = .08$	$t(90) = 2.48$, $p = .04$, $d = .52$	$p = .06$	$p = .98$	6.23	1.63	6.19	.91	5.61	1.18
	SAM: Arousal	$F[2, 90] = 8.43$, $p < .001$, $\eta^2 = .16$	$t(90) = 3.63$, $p = .001$, $d = .77$	$t(90) = 3.48$, $p = .002$, $d = .73$	$p = .99$	7.68	1.42	7.61	1.70	6.09	1.97
	Chances of Rescue	$F[2, 90] = 1.22$, $p = .3$, $\eta^2 = .03$	$p = .53$	$p = .29$	$p = .90$	3.97	.50	4.03	.57	3.83	.40
Participants' Perception of the Robot	Robot Empathy	$F[2, 90] = 4.54$, $p = .01$, $\eta^2 = .09$	$t(90) = 2.49$, $p = .04$, , $d = .53$	$t(90) = 2.72$, $p = .02$, $d = .57$	$p = .97$	5.59	.93	5.65	.89	5.02	.86
	Robot Loyalty	$F[2, 90] = 4.08$, $p = .02$, $\eta^2 = .08$	$t(90) = 2.73$, $p = .02$, $d = .56$	$p = .09$	$p = .79$	5.43	.85	5.27	.81	4.75	.10
	Robot Integrity	$F[2, 90] = 3.79$, $p = .03$, $\eta^2 = .08$	$t(90) = 2.55$, $p = .03$, $d = .54$	$p = .08$	$p = .93$	6.02	.911	5.95	.80	5.49	.76
	Robot Caring	$F[2, 90] = 4.20$, $p = .02$, $\eta^2 = .09$	$t(90) = 2.41$, $p = .047$, $d = .51$	$t(90) = 2.60$, $p = .03$, $d = .55$	$p = .98$	5.18	.83	5.23	.89	4.56	1.23
	Robot Engagement	$F[2, 90] = 3.30$, $p = .04$, $\eta^2 = .07$	$t(90) = 2.46$, $p = .04$, , $d = .52$	$p = .15$	$p = .82$	6.06	.89	5.87	1.17	5.26	1.65
	Robot Likeability	$F[2, 90] = 6.75$, $p = .002$, $\eta^2 = .13$	$t(90) = 3.05$, $p = .01$, $d = .64$	$t(90) = 3.33$, $p = .004$, $d = .70$	$p = .967$	4.93	.60	4.97	.63	4.36	.89
	Human-Like Behavior	$F[2, 90] = 8.9$, $p < .001$, $\eta^2 = .17$	$t(90) = 4.03$, $p < .001$, $d = .85$	$t(90) = 3.10$, $p = .01$, $d = .65$	$p = .627$	5.10	1.42	4.74	1.34	3.54	1.74
	Robot Intelligence	$F[2, 90] = 1.05$, $p = .35$, $\eta^2 = .02$	$p = .67$	$p = .33$	$p = .84$	6.29	.75	6.40	.57	6.14	.77
	Robot Competence	$F[2, 90] = .68$, $p = .51$, $\eta^2 = .02$	$p = .52$	$p = .65$	$p = .97$	5.99	.77	5.94	.77	5.76	.89
	Robot Unpleasantness	$F[2, 90] = .76$, $p = .47$, $\eta^2 = .02$	$p = .99$	$p = .50$	$p = .57$	1.55	.79	1.37	.49	1.57	.70
	Robot Extraversion	$F[2, 90] = .29$, $p = .75$, $\eta^2 = .006$	$p = .9$	$p = .95$	$p = .73$	5.20	1.30	4.97	1.15	5.06	1.20
Consistency and Appropriateness	Understanding Robot Behavior	$F[2, 90] = 18.087$, $p < .001$, $\eta^2 = .56$	$t(90) = 4.63$, $p < .001$, $d = .98$	$t(90) = 5.29$, $p < .001$, $d = 1.12$	$p = .99$	5.57	.95	5.63	1.59	2.83	1.15
	Gaze-Speech Synchronization	$F[2,90] = 47.9$, $p < .001$, $\eta^2 = .52$	$t(90) = 8.66$, $p < .001$, $d = 1.83$	$t(90) = 8.28$, $p < .001$, $d = 1.75$	$p = .93$	5.90	1.27	5.78	1.08	2.93	1.63
	Looking at Objects at Appropriate Times	$F[2,90] = 14.6$, $p < .001$, $\eta^2 = .54$	$t(90) = 4.82$, $p < .001$, $d = 1.02$	$t(90) = 5.12$, $p < .001$, $d = 1.08$	$p = .98$	5.72	1.67	6.10	.99	3.28	1.43
	Natural Movement	$F[2,90] = 16.7$, $p < .001$, $\eta^2 = .27$	$t(90) = 4.79$, $p < .001$, $d = 1.01$	$t(90) = 5.19$, $p < .001$, $d = 1.09$	$p = .92$	4.70	1.62	4.83	1.60	2.77	1.38
	1	2	3	4	5	6	7	8	9	10	11

Figure 6.3: Summary of the Results Using FDR (Benjamini-Hochberg) Control. The Highlighted Cells Indicate Statistically Significant Results At $p < .033$

6.2.5 Results for Hypothesis 5 (H5)

Hypothesis 5 (H5) states that "The LSHG-S condition improvements over the NHG-S condition will be comparable to those of the TSHG-S condition." The results revealed that the number of measures (Bonferroni correction - 7, FDR correction - 14) in which LSHG-S showed significant improvements over NHG-S was comparable to the number of measures (Bonferroni correction - 7, FDR correction - 10) in which TSHG-S showed significant improvements over NHG-S.

1. Original Results ($p < .5$): LSHG-S showed significant improvements over NHG-S in 15 measures: SAM: Valence, Creativity, Person at Ease, SAM: Arousal, Robot Empathy, Robot Loyalty, Robot Integrity, Robot Caring, Robot Engagement, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. TSHG-S showed improvements over NHG-S in 10 measures: Creativity, SAM: Arousal, Robot Empathy, Robot Caring, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. These results support H5 as LSHG-S elicited positive responses in five additional measures than the TSHG-S, when compared to NHG-S.

2. Bonferroni Correction ($p < .0024$) - Both LSHG-S and TSHG-S showed significant improvements over NHG-S in seven measures: SAM: Arousal, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. These results fully support H5.

3. FDR (Benjamini-Hochberg) Correction ($p < .033$) - LSHG-S showed significant

improvements over NHG-S in 14 measures: SAM: Valence, Creativity, Person at Ease, SAM: Arousal, Robot Empathy, Robot Loyalty, Robot Integrity, Robot Caring, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. TSHG-S showed improvements over NHG-S in 10 measures: Creativity, SAM: Arousal, Robot Empathy, Robot Caring, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. These results support H5, since LSHG-S elicited positive responses in four additional measures than the TSHG-S, when compared to NHG-S.

6.3 Results for Social Acceptance

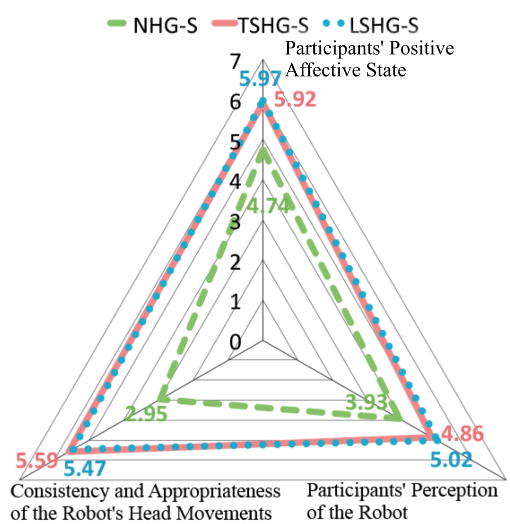


Figure 6.4: Radar Plot Comparing the Means of TSHG-S, LSHG-S, and NHG-S Conditions Using Bonferroni Correction

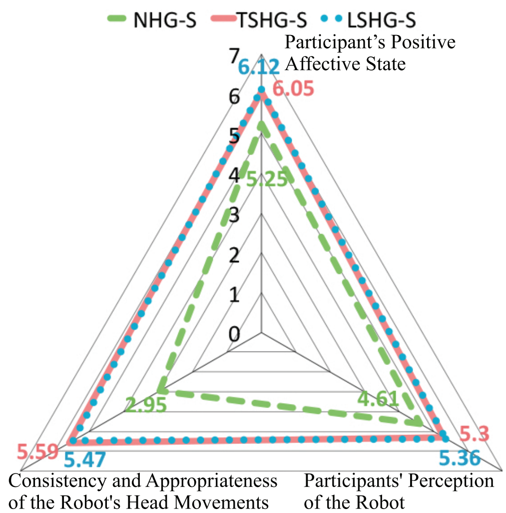


Figure 6.5: Radar Plot Comparing the Means of TSHG-S, LSHG-S, and NHG-S Conditions Using FDR Control

Social acceptance is a measure of how well the robot performs across three categories of measures: *Participants' Positive Affective State, Participants' Perception of the Robot,* and *Consistency and Appropriateness of the Robot's Head Movements.* In order to calculate social acceptance of each head gaze generation condition, the three independent categories for each condition were represented on a radar plot and the area under the radar plot was calculated.

The radar plot consists of three equiangular spokes, called radii, with each spoke representing one of three independent categories. The data values on each spoke are the mean corresponding to a head gaze condition and measure category. Since two questions – SAM:Valence and SAM:Arousal – from Self Assessment Manikin (SAM) [55] used a nine point Semantic Differential scale, their means were re-scaled to a seven point scale. For example, the mean for *Consistency and Appropriateness of the Robot's Head Movements* category in the LSHG-S condition using Bonferroni correction (Fig. 6.4) is $(5.57 + 5.9 + 5.72 + 4.7)/4 = 5.47$.

Condition	Bonferroni Correction (square units)	FDR Correction (square units)
LSHG-S	39	41.39
TSHG-S	38.55	41.36
NHG-S	19.15	23.07

Table 6.2: Area under the Radar Plot.

Fig. 6.4 illustrates the radar plot with Bonferroni correction, while Fig. 6.5 shows the radar plot using FDR correction. The calculated area for each condition and correction is shown in Table 6.2 and reflects a measure of overall social acceptance. Both LSHG-S and TSHG-S conditions have comparable areas, irrespective of the correction method applied, suggesting that the LSHG-S condition engendered high levels of social acceptance similar to the TSHG-S condition. In both plots, the TSHG-S and LSHG-S conditions have a larger area under the graph when compared to the NHG-S condition, suggesting that they achieved greater social acceptance than NHG-S.

6.4 Influence of Covariates on the Measured Attributes

Using an ANCOVA, the influence of gender, age, video gaming experience, robot ownership, pet ownership, past experience with robots, and ethnicity on the 21 measured attributes was analyzed.

1. Participants who had more experience with robots got more agitated with the robot ($F[1,84] = 4.82$, $p = .03$, $\eta^2 = .05$) and found the robot to be less Extraverted ($F[1,84] = 9.98$, $p < .001$, $\eta^2 = .11$).

2. Increased age led to a higher perception of a better Chance of Rescue ($F[1,84] = 8.97$, $p = .003$, $\eta^2 = .09$) by a robot.

Head Gaze Act	**Average number for TSHG-S**	**Average number for LSHG-S**
Fixate	64	58
Avert	56	61
Concurrence	22	17
Fixate(Object)	7	7
Confusion	1	1
Scan	1	1
Total	151	145

Table 6.3: Average Number of Head Gaze Acts for TSHG-S and LSHG-S from the Log Files Rounded to the Nearest Whole Number

3. Male participants rated the robot as being more unpleasant than female participants ($F[1,84] = 5.42$, $p = .02$, $\eta^2 = .06$).

4. Caucasian participants viewed the robot to be more extraverted than Asian participants ($F[1,84] = 2.58$, $p = .04$, $\eta^2 = .10$).

5. Increased video gaming experience resulted in increased perception of Natural Movement ($F[1,84] = 11.56$, $p = .03$, $\eta^2 = .04$) and Robot Synchronization ($F[1,84] = 11.02$, $p = .001$, $\eta^2 = .06$).

6.5 Log Analysis

The log files for each of the 62 interactions (31 TSHG-S and 31 LSHG-S) were analyzed post-hoc. The average number of overall head gaze acts for LSHG-S and TSHG-S conditions were similar at 151 and 145 respectively (Table 6.3), for an interaction of 15 minutes. However, there are small differences in the average number of *fixate*, *avert* and *concurrence* head gaze acts between the two conditions. The number of *fixate* and *avert* head gaze acts differ because of probability, and the total number of head gaze acts that are possible in the script. The reason for less number of *fixate* head gaze acts in the LSHG-S condition is because more number of *fixate* head gaze acts occur with probability 1 in the TSHG-S condition (40) than the LSHG-S condition (28). The reason for more number of *avert* head gaze acts in the LSHG-S condition is because the total number of *avert* head gaze acts possible in the script for the LSHG-S condition is 71, compared to 64 for the TSHG-S condition. In case of the *concurrence* head gaze act, the participant might have responded faster in the LSHG-S condition, and not allow the *concurrence* head gaze act to occur as often during the interaction.

6.6 Summary

This section presented the statistical analysis and results of data collected as a part of a complex, large-scale human-robot interaction study. The results were corrected for multiple testing using Bonferroni correction and FDR (Benjamini-Hochberg) correction. Three important results were observed irrespective of the correction used – (1) *Hypothesis H1 and H2* - The results for H1 and H2 showed that LSHG-S elicited higher levels of social acceptance than NHG-S across three categories of measures: *Participants' Positive Affective State*, *Participants' Perception of the Robot,* and the *Consistency and Appropriateness of the Robot's Head Movements*, (2) *Hypothesis H3 and H4* - The results for H3 and H4 showed that participants' in the TSHG-S condition rated the robot more positively than the participants' in the NHG-S condition across three categories of measures: *Participants' Positive Affective State*, *Participants' Perception of the Robot,* and the *Consistency and Appropriateness of the Robot's Head Movements*, and (3) *Hypothesis H5* -The results for H5 revealed that LSHG-S performs at least as well as the TSHG-S when compared to NHG-S, irrespective of the type of correction used. These results assume particular significance because the LSHG-S condition used loose head gaze-speech synchronization, and was not expected to perform as well as it did.

Using the Bonferroni correction, Hypothesis H1 was supported in one measure – SAM:Arousal, while the FDR correction revealed support for four measures – SAM: Valence, Creativity, Person at Ease, and SAM: Arousal. Hypothesis H2 was supported in six measures after Bonferroni correction – Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. However, FDR correction revealed support for Hypothesis H2 in four additional measures – Robot Loyalty, Robot Empathy, Robot Integrity, and Robot Caring. The Bonferroni correction showed support for Hypothesis H3 in one measure – SAM: Arousal, while the FDR correction supported Hypothesis H3 for two measures – Creativity and SAM: Arousal. Hypothesis H4 was supported in six measures after Bonferroni correction – Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. However, FDR correction revealed support for Hypothesis H4 in four additional measures – Robot Loyalty, Robot Empathy, Robot Integrity, and Robot Caring. Hypothesis H5 is supported because the number of measures (Bonferroni correction - 7, FDR correction - 14) in which LSHG-S showed significant improvements over NHG-S was comparable to the number of measures (Bonferroni correction - 7, FDR correction - 10) in which TSHG-S showed significant improvements over NHG-S.

7. DISCUSSION [*]

The results from the 93 participant human-robot interaction study revealed that LSHG-S was socially acceptable for human-robot interaction. LSHG-S was successful in eliciting positive responses from participants' for measures of *Participants' Affective State*, *Participants' Perception of the Robot*, and *Consistency and Appropriateness of the Robot's Head Movements*, when compared to NHG-S. Additionally, LSHG-S performs at least as well as the TSHG-S when compared to NHG-S irrespective of the type of correction used - Bonferroni or FDR (Benjamini-Hochberg). These results are surprising because LSHG-S did not use semantic understanding to precisely match head gaze with dialog. The findings from this experiment suggests that 1) affordances in the sentence structure of dialog and time delays that were developed as a part of this research effort are adequate, and 2) autonomously generated head gaze-speech coordination is both possible and acceptable. This section discusses and interprets the results of the experiment, analyzes four factors that may be relevant for user perception of head gaze-speech synchronization, and addresses five limitations of the experiment.

7.1 Interpretation and Discussion of the Results

The results from this experiment indicated that that LSHG-S performs better than NHG-S, and at least as well as TSHG-S when compared to NHG-S irrespective of the type of correction used - Bonferroni or FDR (Benjamini-Hochberg). The Bonferroni correction is the most conservative test for multiple testing. It is expected that the measures that are statistically significant after this correction will be highly

[*]Reprinted with permission from "Evaluation of Head Gaze Loosely Synchronized with Synthetic Speech for Social Robots" by Srinivasan, V., Bethel, C.L., Murphy, R.R, 2014. IEEE Transactions on Human-Machine Systems, Accepted for Publication, Copyright [2014] by Vasant Srinivasan.

replicable in other studies. The measures where LSHG-S was preferred to NHG-S are as follows: SAM:Arousal, Robot Likeability, Human-Like Behavior, Understanding Robot Behavior, Gaze-Speech Synchronization, Looking at Objects at Appropriate Times, and Natural Movement. The results for *Consistency and Appropriateness of the Robot's Head Movements* are particularly strong, and suggest that head gaze behaviors generated by the robot were consistent, appropriate, and sensitive to the context. The results reinforce recent findings from Kirchhof & Ruiter [54] that gesture comprehension is temporally more flexible than gesture production, and that participants tolerate loose synchronization of head gaze with speech.

FDR (Benjamini-Hochberg) provides a good balance between discovery of statistically significant measures and limitation of false discovery occurrences. It maintains good statistical power and attempts to ensure the accuracy of statistically significant results while controlling the proportion of false discoveries. Using the FDR correction, the LSHG-S demonstrated high levels of social acceptance when compared to NHG-S in seven *additional* measures across the three categories: SAM: Valence, Creativity, Person at Ease, Robot Empathy, Robot Loyalty, Robot Integrity, and Robot Caring. The high creativity level of the participants in the LSHG-S condition suggests that participants experienced less stress, and were more relaxed compared to the NHG-S condition [70].

The results for *Participants' Affective State* and *Participants' Perception of the Robot* suggest that the participants perceived the robot as a friendly companion [66]. While this result is significant for social robotics in general, it has strong applicability to eldercare and therapeutic robotics, in addition to the victim management application presented in this dissertation. A rescue operation might take up to 6-10 hours after a victim is located [40], during which it is psychologically helpful for the "victim" to perceive the robot to be social and following human interpersonal communication norms [14].

7.2 Interpretation and Discussion of the Covariates

The purpose of including covariates in ANOVA is two-fold:

1. To reduce error variance and explain unexplained variances in terms of the covariates, so as to improve the accuracy of the test and better assess the main effect.

2. To remove the bias of variables that confound the results (i.e., a variable that varies systematically with the experimental manipulation).

While covariates relationship does not imply causation, several interesting trends were observed:

1. People experienced with robots got more agitated with the robot and found the robot to be less Extraverted. This could be because of the type of robot these people owned. They might all be "Roomba" owners, who are used to robots that aren't social and are tools. They were confused and agitated by the sociableness of it. They also probably perceive robots as task completer's and so assumed it wouldn't be extraverted. The type of robot influences people's experience with the robot and therefore should be measured and controlled in future experiments.

2. Increased age led to a higher perception of a better Chance of Rescue. Old people seem to have a lot of faith in "the future" and "technology" making life better. This arguably makes sense because they have witnessed many life improvements due to other technology, so they are more likely predisposed to believing a rescue robot will be a life changing technology that actually increases Chance of Rescue.

3. Male participants rated the robot as being more unpleasant than female participants. Humans resonate more to communications delivered by a gendered voice that matches their own, rather than by an oppositely gendered voice, regardless of whether the voice is human or synthesized [77]. The male participants may not have related as well to the female voice used. Females may have related better or neutrally to a female voice.

4. Caucasian participants viewed the robot to be more extraverted than Asian participants. Caucasians are generally Western and individualistic, while Asians are generally Eastern and collectivistic [78, 79]. This would explain the attribution of extraversion as a projection of self onto the robot.

5. Increased video gaming experience resulted in increased perception of Natural Movement and Robot Synchronization. This could be because people with a lot of video gaming experience feel comfortable interacting with artificial intelligence characters and are used to movements and synchronization that are non-human, which may not be perfect.

7.3 Four Factors Relevant for User Perception of Head Gaze-Speech Synchronization

This work identifies four factors that are important for user perception of speech-gaze synchronization, which may also explain why the LSHG-S condition performed well.

1. *Gesture comprehension is temporally more flexible than gesture production:* The gesture-speech synchrony might be a consequence of the production system, but may not be essential for comprehension [54]. If people were very sensitive to TSHG-S, then the results would have indicated that the partici-

pants would have a preference for the TSHG-S over LSHG-S, which was not the case. The semantic and temporal flexibility of head gaze production with robots requires further investigation.

2. *Expectation of gaze in a semi-humanoid robot:* The expectations of gaze from a robot with a low degree of anthropomorphism and a more mechanical appearance may not be as high as those required for human interpersonal interactions. Survivor Buddy has a low degree of anthropomorphism, with a mechanical appearance. It does not have eyes, hands, or even the shape of a human head. The author considers that even though Survivor Buddy was able to have an intelligent conversation with the participants, they do not hold it to the same high standards required in human interpersonal communication, or even human-android communication.

3. *Importance of synchronization at the start and end of turns is greater than at the middle of turns:* Yamazaki *et al.* [11] discuss that timing is critical at the start and end of turns for conversation. Cassell *et al.* [22] emphasize the importance of middle of turn events for superior performance; however, the timing of these events has not been examined. The proposed method has precise synchronization at the start and end of turns, however, it approximates the semantic structure during the middle of turns to support autonomous generation. These approximations did not adversely impact the participants, and resulted in performance similar to that of TSHG-S. The timing of the middle of turn events may be of relatively low importance.

4. *Absence of lips:* Humans perceive speech-lip asynchrony as unnatural [80], and prior work suggests that speech needs to be tightly synchronized with lips, but not with gestures [54]. In the experiment, the Survivor Buddy robot did not

have lips. This could possibly explain why the participants did not perceive LSHG-S to be unnatural, but instead found it to be equivalent to that of TSHG-S.

7.4 Limitations of the Experiment

The study was limited in five aspects, listed below in no particular order.

1. *Implementation used a priori sensor data:* The robot did not have object recognition or speech recognition capabilities. This is because this research focused on the generation of appropriate head gaze behaviors, not recognition of objects or speech, which are challenging research areas in their own right. However, these limitations can be expected to be addressed in the future.

2. *Limited to head gaze, no eye gaze:* The implementation generated only head gaze acts and ignored eye gaze. However, the inference from the sentence structure, time delays, and typing could readily extend to eye gaze if the robot had mechanical or virtual eyes. Of the two body components, head gaze may be more immediately valuable as many robots do not have eyes capable of gaze. For example, robot animals often have fixed cameras or lights for eyes, and non-anthropomorphic robots do not have eyes at all.

3. *Single domain validation:* The experiment showed that gaze primitives validated in other domains (TSHG-S condition) had positive results for victim management, as shown by the high rating of the TSHG-S condition. Because the performance of the LSHG-S condition (which used approximations) was as good as the TSHG-S condition for victim management in a search and rescue domain, the expectation is that the results will transfer to other domains. The

next step is to validate the system in the field using a very realistic setting. Validation of the system in other domains can be explored in future work.

4. *Controlled conversational interaction:* The conversational interaction was directed by the robot (using a hidden operator). This method was adopted because speech recognition tools are unreliable, and affect the repeatability and consistency of the experiment across conditions. Future speech recognition systems will permit the participants to have a collaborative dynamic conversation with the robot.

5. *Purposeful dialog* - In the study, the robot followed a 911 protocol and its goal was to gather as much information as possible about the victim (participant) and the environment. The content of the dialog had a sense of purpose attached to it and the objective of the dialog was very goal-oriented, in a question/answer form. The LSHG-S generation mechanism needs to be evaluated for casual dialog, in which content is less goal-oriented and more of small talk (a discussion on what the weather is like etc).

7.5 Summary

This section has interpreted and discussed the results associated with a large-scale 93 participant human-robot interaction study designed to evaluate the effectiveness and appropriateness of LSHG-S. The results from the experiment demonstrated a high level of social acceptance of LSHG-S in the key areas of *Participants' Positive Affective State, Participants' Perception of the Robot,* and *Consistency and Appropriateness of the Robot's Head Movements*. LSHGS performs at least as well as the TSHG-S when compared to NHG-S, irrespective of the type of correction used - Bonferroni or FDR (Benjamini-Hochberg). Statistically significant results for all measures of the *Consistency and Appropriateness of the Robots Head Movements* category was found even with the more conservative Bonferroni correction, suggesting high replicability of these results. Since the participants tolerated LSHG-S, the affordances developed as a part of this research effort are adequate and socially acceptable for human-robot interaction. The LSHG-S is a more viable and practical method for the generation of head gaze than the commonly used TSHG-S. This is because LSHG-S does not require manual annotation, LSHG-S performs better than NHG-S, and at least as well as TSHG-S when compared to NHG-S. The research identified four factors that may have impacted the result outcomes: *gesture comprehension is temporally flexible than gesture production*, *expectation of gaze in a semi-humanoid robot*, *importance of synchronization in start and end of turns is more than middle of turn*, and *absence of lips*. This is followed by a discussion of five limitations of the study: *implementation used a priori sensor data*, *limited to head gaze, no eye gaze*, *single domain validation*, *controlled conversational interaction*, and *purposeful dialog*.

8. CONCLUSIONS AND FUTURE WORK

This work developed a computational theory for social head gaze as a programmable framework, so that head gaze-speech acts can be autonomously and consistently generated. The fundamental primary research question – *"What is a computational theory of social head gaze for social agents?"* was answered using four related secondary research questions as follows:

1. *What is the appropriate set of social head gaze behaviors required for a naturalistic human-robot interaction?*

   This research surveyed 32 human-robot interaction studies of social head gaze in Section 2 and identified through a process of generalization and mining, a set of six head gaze acts and three types of percepts that have been successful in making a robot comforting, more socially consistent, and predictable to the user(s). The six head gaze acts that have been identified were: *Fixate*, *Avert*, *Concurrence*, *Scan*, *Confusion*, and *Short Glance*. The three types of percepts reported were: *external*, *linguistic*, and *internal*.

   Sections 3 and 4 detail the substitution of affordances for linguistic and internal percepts of head gaze, so that the state of an interaction can be inferred autonomously. Affordances are conditions or objects that are directly perceivable without any memory, inference, or interpretation [37]. A set of eight affordances from the sentence structure, time delays, and typing (*Initial Word*, *Word following Punctuation : . ! ?*, *After 75% of Words between Punctuation : . ! ?*, *Carriage Return*, *Elapsed Listening Time* $>$ *6 sec*, *Elapsed Idle Time* $>$ *15 sec*, *Number of Deletes/Retypes by an Operator* $>$ *5 within a Time Interval* t *= 15 sec*, and *The Object Name Tag*) were proposed.

2. *How can social head gaze be expressed as behaviors or schemas, which are common representations in both psychology and robotics?*

   Section 3 maps social head gaze unto a behavioral robotics framework. The behavioral robotics framework expresses social head gaze as a set of five behaviors – COMMUNICATING SOCIAL ATTENTION, REGULATING AN INTERACTION, MANIFESTING AN INTERACTION, PROJECTING A MENTAL STATE, and ESTABLISHING AGENCY. Section 4 describes the implementation of the five behaviors as nine production rules. The production rules are if-then statements that uniquely map the affordances onto head gaze acts.

3. *Is it possible to evaluate through sound experimental methods the effectiveness and appropriateness of the head gaze acts generated using the behavioral robotics framework?*

   A large-scale 93-participant experiment was conducted to test the efficacy of LSHG-S. The details of the experiment were presented in Section 5. The data analysis and results were provided in Section 6. Five hypotheses were evaluated:

   (a) *Hypothesis 1 (H1)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

   (b) *Hypothesis 2 (H2)*: Participants who interact with a robot exhibiting the LSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

   (c) *Hypothesis 3 (H3)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate their experiences more positively than participants who interact with a robot exhibiting the NHG-S condition.

(d) *Hypothesis 4 (H4)*: Participants who interact with a robot exhibiting the TSHG-S condition will evaluate the robot more positively than participants who interact with a robot exhibiting the NHG-S condition.

(e) *Hypothesis 5 (H5)*: The LSHG-S condition improvements over the NHG-S condition will be comparable to those of the TSHG-S condition.

The results indicated strong support for each of the five hypotheses. The social acceptance of LSHG-S is similar to TSHG-S (manual annotation), and preferred to NHG-S. LSHG-S performs at least as well as the TSHG-S when compared to NHG-S in key areas of *Participants' Positive Affective State*, *Participants' Perception of the Robot*, and *Consistency and Appropriateness of the Robot's Head Movements*, irrespective of the type of correction used - Bonferroni or FDR (Benjamini-Hochberg). These results suggest that LSHG-S is adequate for human-robot interaction. The potential problem of imprecisely synchronized head gaze acts with speech affecting user perception of the robot did not arise. The lag between the robot's speech and gaze acts did not annoy or confuse the human.

4. *Does the level of synchronization between gaze acts and speech impact the naturalistic perception of the social interaction?*

   The analysis of the results in Section 7 suggests that LSHG-S elicits high levels of social acceptance when compared to NHG-S, and is adequate for human-robot interaction. The Survivor Buddy robot interacted with a human in a simulated victim management scenario using all five functions of head gaze (Section 5). The change in synchronization for two functions such as *engaging in a verbal conversation with human(s)* and *referential gaze* did not impact the human's perception of the robot's overall head gaze during the interaction.

This may have because of any of the four factors that could have impacted user perception of head gaze-speech synchronization identified in Section 7 – *Gesture comprehension is temporally more flexible than gesture production*, *Expectation of gaze in a semi-humanoid robot*, *Importance of synchronization at the start and end of turns is greater than at the middle of turns*, and *Absence of lips.*

8.1 Significant Contributions

This research provides seven contributions to the *social robotics community.* These contributions arranged in the order of abstraction (abstract to implementation-specific).

1. The findings from Sections 3, 4, 6, and 7 contribute to a fundamental understanding of the role of social head gaze in social acceptance (Section 6), particularly with regard to the question of when less competence is tolerable (Section 3), how social head gaze can be produced (Section 4), and the importance of the speech and gaze synchronization for the listener (Section 7).

2. It shows that autonomously generated head gaze-speech coordination is both possible and acceptable. Researchers and practitioners do not have to manually annotate every situation using the Wizard-of-Oz approach [47]. As seen in Section 6 a total of 137 out 145 head gaze acts that were manually annotated for TSHG-S were generated autonomously using LSHG-S. Note that this number reflects only *Fixate*, *Avert*, *Concurrence*, and *Scan* head gaze acts. Note that *Fixate(Object)* and *Confusion* head gaze were generated manually in both conditions.

   The results also indicate that synchronization of head gaze with speech is more flexible than initially thought. This finding promotes implementation on robots with lesser capabilities (for example, a toy "Keepon" with low velocity/joints limits).

3. Section 3 provides social robotics researchers and practitioners with a formal vocabulary for social head gaze comprising of behavioral robotics nomenclature, such as perceptual schemas, percepts (*external*, *linguistic*, and *internal*), behaviors (Communicating Social Attention, Regulating an Interaction, Manifesting an Interaction, Projecting a Mental State, and Establishing Agency), and motor schemas (*Fixate*, *Avert*, *Concurrence*, *Scan*, *Confusion*, and *Short Glance*). These provide a common lexicon and taxonomy that facilitates communication across diverse groups such as social scientists interested in understanding the fundamental aspects of the social head gaze phenomena, or robot behavior designers/practitioners who need to implement head gaze elements in a specific application that are autonomous, consistent, repeatable, and natural.

4. While the robot generated socially acceptable head gaze behaviors in real-time for a goal directed victim management scenario detailed in Section 5, it is expected that the robot can generate socially acceptable head gaze behaviors for very open-ended, interactive scenarios. This is because LSHG-S is independent of the content of dialog. However this needs to be validated in future human-robot interaction studies. The ability to generate real-time head gaze in open-ended interactive scenarios is very important in situations where robot responses cannot be anticipated *a priori* (e.g. personal robots for eldercare or museum tours).

5. This work contributes five new measures for victim management - Person at Ease, Robot Empathy, Robot Integrity, Robot Loyalty, and Robot Caring. The

list of items, and their reliability with regard to these five measures are reported in Appendix I. Since these measures were statistically significant only for the FDR correction and not the Bonferroni correction, these measures need to be validated in future human-robot interaction studies.

6. The economic impact relates primarily to the amount of labor involved and costs required for the modification of existing robots. Section 4 contributes a novel mechanism for inferring affordances from sentence structure, time delays, and typing that is independent of the semantics of dialog. This method reduces the amount of labor involved, since higher social acceptance can be generated with reduced manual effort and real-time operator workload for even unstructured dialog.

   The behavioral robotics framework was derived in Section 3 from 32 previous implementations of head gaze using a commonality analysis. Hence, it is applicable to a wide variety of robots (anthropomorphic, non-anthropomorphic).

7. The behavioral robotics framework simplifies creation, analysis, and comparison of social head gaze implementations. The instantiation of a social head gaze implementation is described in Section 4. The analysis and comparison of two existing head gaze architectures using the behavioral robotics framework and the Software Architecture Analysis Method (SAAM) is provided in Appendix B.

8.2 Open Research Questions

There were five main open research questions that were revealed while performing this research study. These following open research questions help inform Future Work in the area of social head gaze for social robotics.

1. What is the audio visual integration envelope for head gaze? The extent to which humans tolerate loose synchronization of head gaze and speech needs to be determined.

2. What is the impact of content conversation on head gaze acts? Currently, no study has investigated the role of content or what the robot is saying in head gaze.

3. What is the impact of head and eye gaze in a social interaction? Current work addresses head gaze, not both of them together. There are indications from psychology literature that eye gaze will have priority over head gaze [48], but this needs to be investigated.

4. How does head gaze extend to multi-party situations, and can this be modeled? Currently, head gaze is predominantly a feature of dyadic [1–11, 14–18, 20, 23–32,40,42,43,49] or triadic situations [13,41]. However, it is reasonable to assume that robots will encounter multi-party situations often in the real world. Thus, it is also worth asking whether a multi-party situation can be approximated to several dyadic encounters.

5. What is the impact of distance and culture on parameters of head gaze? Humans tend to have exaggerated movements when they are further away from a human [48, 50]. The effect of distance and culture on parameters of head gaze (such as duration and range) needs to be studied.

8.3 Future Work

Five directions for future work have been identified.

1. Head-gaze speech synchronization merits further investigation. Additional studies need to be conducted to answer critical questions such as: *What is the*

extent to which humans tolerate loose synchronization of head gaze and speech? Also, *What factors affect expected synchronization of head gaze-speech?* For example, what role does people's (perceived) appearance of the robot have on head gaze-speech synchronization? and What is the impact of content conversation on head gaze acts?

2. Validation testing needs to be conducted in other domains which have different content of dialogue. While the study showed that a set of gaze acts, validated in other domains, had positive results in the search and rescue domain, the collection of gaze acts need to be applied in each of the original domains as described in literature, as well as to new ones.

3. The behavioral robotics framework should be expanded to support both head and eye gaze. The Coordination Function needs to be upgraded to support both head and eye gaze at the same time.

4. The behavioral robotics framework should be extended to support multi-party interactions. This would involve updating the implementation of REGULATING AN INTERACTION behavior to support the short glance gaze act and upgrading the Coordination Function to resolve any conflicts resulting from interactions with multiple people.

5. The impact of time on human-robot interactions has not been investigated. In a search and rescue scenario, the victims are expected to interact with the robot for 4-6 hours. The effects of long-term interaction on social head gaze needs to be investigated.

8.4 Summary

This research was the first to propose LSHG-S and show that autonomous head gaze-speech coordination is possible and does not require semantic understanding. The fundamental primary research question – *"What is a computational theory of social head gaze for social agents?"* was answered using four related secondary research questions: (1) *What is the appropriate set of social head gaze behaviors required for a naturalistic human-robot interaction?*, (2) *How can social head gaze be expressed as behaviors or schemas, which are common representations in both psychology and robotics?*, (3) *Is it possible to evaluate through sound experimental methods the effectiveness and appropriateness of the head gaze acts generated using the behavioral robotics framework?*, and (4) *Does the level of synchronization between gaze acts and speech impact the naturalistic perception of the social interaction?*. LSHG-S was realized using eight novel affordances for turn-taking and semantics from the sentence structure, time delays, and typing: *Initial Word*, *Word following Punctuation : . ! ?*, *After 75% of Words between Punctuation : . ! ?*, *Carriage Return*, *Elapsed Listening Time* $>$ *6 sec*, *Elapsed Idle Time* $>$ *15 sec*, *Number of Deletes/Retypes by an Operator* $>$ *5 within a Time Interval* $t = 15$ *sec*, and *The Object Name Tag*. The results from a 93-participant experiment indicated that LSHG-S elicited high levels of social acceptance, performed as well as the TSHG-S condition when compared to the NHG-S condition, and the participants' were not annoyed or confused by the head gaze. This suggests that the affordances developed as a part of this research effort are adequate and socially acceptable for human-robot interaction. A behavioral robotics framework for social head gaze was developed to simplify creation, analysis, and comparison of implementations. Seven contributions of the research to the *social robotics community* were detailed followed by a discussion of five directions for future work.

REFERENCES

[1] A. Holroyd, C. Rich, C. L. Sidner, and B. Ponsler, “Generating connection events for human-robot collaboration,” in *20th IEEE International Workshop on Robot and Human Interactive Communication, RO-MAN*, 2011, pp. 241 – 246.

[2] C.-M. Huang and B. Mutlu, “Robot behavior toolkit: Generating effective social behaviors for robots,” in *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction.* New York, NY, USA: ACM, 2012, pp. 25–32.

[3] M. Imai, T. Ono, and H. Ishiguro, “Robot mediated round table: Analysis of the effect of robot’s gaze,” in *Proceedings of 11th IEEE International Workshop on Robot and Human Interactive Communication*, 2002, pp. 411–416.

[4] C. Breazeal, C. D. Kidd, T. Andrea L, G. Hoffman, and M. Berlin, “Effects of nonverbal communication on efficiency and robustness in human-robot teamwork,” in *in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2005, pp. 383–388.

[5] T. Minato, M. Shimada, H. Ishiguro, and S. Itakura, “Development of an android robot for studying human-robot interaction,” *Innovations in Applied Artificial Intelligence*, pp. 424–434, 2004.

[6] D. Sakamoto, T. Kanda, T. Ono, M. Kamashima, M. Imai, and H. Ishiguro, “Cooperative embodied communication emerged by interactive humanoid robots,” *International Journal of Human-Computer Studies*, vol. 62, no. 2, pp. 247 – 265, 2005.

[7] K. F. MacDorman, T. Minato, M. Shimada, S. Itakura, S. Cowley, and H. Ishiguro, "Assessing human likeness by eye contact in an android testbed," in *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society*, 2005, pp. 21–23.

[8] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1, pp. 140–164, 2005.

[9] B. Mutlu, J. Forlizzi, and J. Hodgins, "A storytelling robot: Modeling and evaluation of human-like gaze behavior," in *Proceedings of the International Conference on Humanoid Robots. IEEE*, 2006.

[10] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka, "Museum guide robot based on sociological interaction analysis," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2007, pp. 1191–1194.

[11] A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka, "Precision timing in human-robot interaction: Coordination of head movement and utterance," in *Proceeding of the Twenty-Sixth SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2008, pp. 131–140.

[12] N. Mitsunaga, C. Smith, T. Kanda, H. Ishiguro, and N. Hagita, "Adapting robot behavior for human–robot interaction," *IEEE Transactions on Robotics*, vol. 24, no. 4, pp. 911–916, 2008.

[13] B. Mutlu, T. Shiwa, T. K, H. Ishiguro, and N. Hagita, "Footing in human-robot conversations: How robots might shape participant roles using gaze cues," in

Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction. New York, NY, USA: ACM, 2009, pp. 61–68.

[14] C. Bethel and R. R. Murphy, “Non-facial and non-verbal affective expression for appearance-constrained robots used in victim management,” *Paladyn. Journal of Behavioral Robotics*, vol. 1, pp. 219–230, 2010.

[15] M. Shimada, Y. Yoshikawa, M. Asada, N. Saiwaki, and H. Ishiguro, “Effects of observing eye contact between a robot and another person,” *International Journal of Social Robotics*, vol. 3, pp. 143–154, 2011.

[16] C.-M. Huang and B. Mutlu, “The repertoire of robot behavior: Designing social behaviors to support human-robot joint activity,” *Journal of Human-Robot Interaction*, vol. 2, no. 2, pp. 80–102, 2013.

[17] H. Admoni, B. Hayes, D. Feil-Seifer, D. Ullman, and B. Scassellati, “Are you looking at me?: Perception of robot attention is mediated by gaze type and group size,” in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction.* New York, NY, USA: IEEE Press, 2013, pp. 389–396.

[18] C.-M. Huang and B. Mutlu, “Learning-based modeling of multimodal behaviors for humanlike robots,” in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction.* New York, NY, USA: ACM, 2014, pp. 57–64.

[19] Y. Matsusaka, S. Fujie, and T. Kobayashi, “Modeling of conversational strategy for the robot participating in the group conversation,” in *Interspeech’01*, 2001, pp. 2173–2176.

[20] C. T. Ishi, C. Liu, H. Ishiguro, and N. Hagita, "Head motions during dialogue speech and nod timing control in humanoid robots," in *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*. Piscataway, NJ, USA: IEEE Press, 2010, pp. 293–300.

[21] E. Gu and N. I. Badler, "Visual attention and eye gaze during multipartite conversations with distractions," in *Intelligent Virtual Agents (IVA'06), Marina del Rey, CA*, 2006, pp. 411 – 416.

[22] J. Cassell, O. E. Torres, and S. Prevost, "Turn taking vs. discourse structure: How best to model multimodal conversation," in *Machine Conversations*. Kluwer, 1998, pp. 143–154.

[23] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, 2014, pp. 25–32.

[24] H. Kozima, C. Nakagawa, and Y. Yasuda, "Interactive robots for communication-care: a case-study in autism therapy," in *14th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2005, pp. 341–346.

[25] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior," in *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, 2009, pp. 69–76.

[26] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Relating conversational expressiveness to social presence and acceptance of an assistive social robot," *Virtual Reality*, vol. 14, no. 1, pp. 77–84, 2010.

[27] M. Imai, T. Ono, and H. Ishiguro, "Physical relation and expression: Joint attention for human-robot interaction," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 4, pp. 636–643, 2003.

[28] M. Staudte and M. W. Crocker, "The effect of robot gaze on processing robot utterances," in *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, Citeseer. Cognitive Science, 2009, pp. 431–436.

[29] ——, "Visual attention in spoken human-robot interaction," in *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction.* New York, NY, USA: ACM, 2009, pp. 77–84.

[30] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction.* New York, NY, USA: ACM, 2014, pp. 49–56.

[31] A. Sauppé and B. Mutlu, "Robot deictics: How gesture and context shape referential communication," in *Proceedings of the 2014 ACM/IEEE International Conference on Human Robot Interaction.* New York, NY, USA: ACM, 2014, pp. 342–349.

[32] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zeng, B. A. Blumer, K. MacLean, and E. A. Croft, "Meet me where i'm gazing: How shared attention gaze affects human-robot handover timing," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction.* New York, NY, USA: ACM, 2014, pp. 334–341.

[33] S. Duncan, "Some signals and rules for taking speaking turns in conversations," *Journal of Personality and Social Psychology*, vol. 23, no. 2, pp. 283 – 292, 1972.

[34] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn-Taking for Conversation," *Language*, vol. 50, no. 4, pp. 696–735, 1974.

[35] A. S. Meyer, A. M. Sleiderink, and W. J. Levelt, "Viewing and naming objects: Eye movements during noun phrase production," *Cognition*, vol. 66, no. 2, pp. B25–B33, 1998.

[36] Z. M. Griffin, "Gaze durations during speech reflect word selection and phonological encoding," *Cognition*, vol. 82, no. 1, pp. B1–B14, 2001.

[37] R. R. Murphy, *Introduction to AI Robotics*. The MIT Press, 2000.

[38] R. C. Arkin, *Behavior-Based Robotics*. The MIT Press, May 1998.

[39] R. Cloutier, G. Muller, D. Verma, R. Nilchiani, E. Hole, and M. Bone, "The concept of reference architectures," *Systems Engineering*, vol. 13, no. 1, pp. 14–27, 2010.

[40] T. Fincannon, L. E. Barnes, R. R. Murphy, and D. L. Riddle, "Evidence of the need for social intelligence in rescue robots," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, vol. 2, september 2004, pp. 1089–1095.

[41] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke, "Towards a humanoid museum guide robot that interacts with multiple persons," in *Proceedings of the IEEE/RSJ International Conference on Humanoid Robots (Humanoids)*. IEEE, 2005, pp. 418–423.

[42] D. Sirkin, W. Ju, and M. Cutkosky, "Communicating meaning and role in distributed design collaboration: How crowdsourced users help inform the design

of telepresence robotics," in *Design Thinking Research*. Springer, 2012, pp. 173–187.

[43] C. Liu, C. T. Ishi, H. Ishiguro, and N. Hagita, "Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, 2012, pp. 285–292.

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

[45] V. Srinivasan, C. L. Bethel, R. R. Murphy, and C. I. Nass, "Validation of a behavioral robotics framework for head social gaze," in *Proceedings of the Workshop on Gaze in HRI: From Modeling to Communication*, 2012.

[46] M. Cherubini, R. de Oliveira, N. Oliver, and C. Ferran, "Gaze and gestures in telepresence: Multimodality, embodiment, and roles of collaboration," *CoRR*, vol. abs/1001.3150, 2010.

[47] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of oz studies: Why and how," in *Proceedings of the 1st International Conference on Intelligent User Interfaces*. New York, NY, USA: ACM, 1993, pp. 193–200.

[48] N. J. Emery, "The eyes have it: the neuroethology, function and evolution of social gaze," *Neuroscience and Biobehavioral Reviews*, vol. 24, no. 6, pp. 581 – 604, 2000.

[49] K. Pitsch, A.-L. Vollmer, and M. Muhlig, "Robot feedback shapes the tutor's presentation how a robot's online gaze strategies lead to micro-adaptation of

the human's conduct," *Interaction Studies*, vol. 14, no. 2, pp. 268–296, 2013.

[50] M. Argyle and M. Cook, *Gaze and Mutual Gaze*. Cambridge University Press, 1976.

[51] E. Goffman, "Footing," *Semiotica*, vol. 25, no. 1-2, pp. 1–30, 1979.

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

[53] M. A. Halliday, *Intonation and Grammar in British English*, ser. Janua linguarum: Series practica. Mouton, 1967.

[54] C. Kirchhof and J. d. Ruiter, "On the audiovisual integration of speech and gesture," in *Book of Abstracts of the 5th Conference of the International Society for Gesture Studies*, Lund, Switzerland, 2012, p. 62.

[55] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[56] J. C. Goodwin, *Research in Psychology: Methods and Design*. Wiley, 1995.

[57] B. Johnson and L. Christensen, *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. Sage Publications, 2010.

[58] B. Mutlu, "Designing gaze behavior for humanlike robots," Ph.D. dissertation, Carnegie Mellon University, 2009.

[59] R. Kazman, L. Bass, M. Webb, and G. Abowd, "Saam: a method for analyzing the properties of software architectures," in *Proceedings of the 16th International Conference on Software Engineering*. IEEE Computer Society Press, 1994, pp. 81–90.

[60] A. E. Hassan and R. C. Holt, “A reference architecture for web servers,” in *Proceedings of the Seventh Working Conference on Reverse Engineering.* IEEE, 2000, pp. 150–159.

[61] D. M. Weiss, “Commonality analysis: a systematic process for defining families,” in *Development and Evolution of Software Architectures for Product Families.* Springer, 1998, pp. 214–222.

[62] S. R. A. Fisher, *Statistical Methods for Research Workers.* Oliver and Boyd, 1970, vol. 14.

[63] R. R. Murphy, A. Rice, N. Rashidi, Z. Henkel, and V. Srinivasan, “A multi-disciplinary design process for affective robots: Case study of survivor buddy 2.0,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA).* IEEE, 2011, pp. 701–706.

[64] V. Groom, V. Srinivasan, C. L. Bethel, R. R. Murphy, L. Dole, and C. I. Nass, “Responses to robot social roles and social role framing,” in *Proceedings of International Conference on Collaboration Technologies and Systems (CTS)*, May 2011, pp. 194–203.

[65] C. Rich, B. Ponsleur, A. Holroyd, and C. L. Sidner, “Recognizing engagement in human-robot interaction,” in *Proceeding of the 5th ACM/IEEE International Conference on Human-Robot Interaction.* New York, NY, USA: ACM, 2010, pp. 375–382.

[66] B. A. Fehr, *Friendship Processes*, ser. Sage Series on Close Relationships. Sage Publications, 1996.

[67] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[68] H. Ebbinghaus, *Memory: a Contribution to Experimental Psychology.* Teachers College, Columbia University, 1913.

[69] J. P. Guilford, "Creativity: Its measurement and development," *A Source Book for Creative Thinking*, pp. 151–167, 1962.

[70] W. H. Teichner, E. Arees, and R. Reilly, "Noise and human performance, a psychophysiological approach," *Ergonomics*, vol. 6, no. 1, pp. 83–97, 1963.

[71] L. Schwabe and O. T. Wolf, "Learning under stress impairs memory formation," *Neurobiology of Learning and Memory*, vol. 93, no. 2, pp. 183–188, 2010.

[72] J. A. Russell, "Evidence of convergent validity on the dimensions of affect," *Journal of Personality and Social Psychology*, vol. 36, no. 10, p. 1152, 1978.

[73] J. P. Shaffer, "Multiple hypothesis testing," *Annual Review of Psychology*, vol. 46, no. 1, pp. 561–584, 1995.

[74] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.

[75] T. Hothorn, F. Bretz, and P. Westfall, "Simultaneous inference in general parametric models," *Biometrical Journal*, vol. 50, no. 3, pp. 346–363, 2008.

[76] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences.* Routledge, 1988.

[77] C. I. Nass and S. Brave, *Wired for Speech: How Voice Activates and Advances the Human-computer Relationship.* MIT Press Cambridge, 2005.

[78] R. C. Page and D. N. Berkow, "Concepts of the self: Western and eastern perspectives," *Journal of Multicultural Counseling and Development*, vol. 19, no. 2, pp. 83–93, 1991.

[79] C. Kagitcibasi, "Individualism and collectivism," *Handbook of Cross-cultural Psychology*, vol. 3, pp. 1–49, 1997.

[80] A. Vatakis, J. Navarra, S. Soto-Faraco, and C. Spence, "Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments," *Experimental Brain Research*, vol. 185, no. 3, pp. 521–529, 2008.

[81] P. C. Clements, "Software architecture in practice," Ph.D. dissertation, Carnegie Mellon University, 2002.

[82] C. Bethel and R. R. Murphy, "Review of human studies methods in hri and recommendations," *International Journal of Social Robotics*, vol. 2, pp. 347–359, 2010.

APPENDIX A

ROBOTS, SUBSYSTEMS, SOCIAL SCIENCE MODELS, AND SYNCHRONIZATION USED FOR THE 32 MAJOR STUDIES.

Studies	**Robot**	**Subsystems**	**Model**	**Sync**
Imai *et al.* [27]	Humanoid	Sensor System, Perceptual System, Dialogue Mechanism, Joint Attention Mechanism, and Action Executive.	-	-
Imai *et al.* [3]	Humanoid	-	-	-
Matsusaka *et al.* [19]	Humanoid	-	[34, 50]	Tight
Breazeal *et al.* [4]	Animal-Like	Sensor System, Perceptual System, Reasoning System, Cognitive System, Action Executive.	-	-
Fincannon *et al.* [40]	Non-Anthropomorphic	-	-	-
Minato *et al.* [5]	Android	-	-	-
Sakamoto *et al.* [6]	Humanoid	Sensor System, Communicative Units, and Action Executive.	-	-
Sidner *et al.* [8]	Animal-like	Sensor System, Perceptual System, Conversation Model, and Action Executive.	[33]	Tight

Continued on next page

Table A.1 – *Continued from previous page*

Studies	**Robot**	**Subsystems**	**Model**	**Sync**
Macdorman *et al.* [7]	Android	-	-	Tight
Kozima *et al.* [24]	Animal-Like	Perceptual System, Attention Map, Habituation Mechanism, and Emotion Expression.	-	-
Bennewitz *et al.* [41]	Humanoid	Sensor System and Behavior System	-	-
Mutlu *et al.* [9]	Humanoid	-	[22]	Tight
Kuno *et al.* [10]	Humanoid	-	[34]	Tight
Yamazaki *et al.* [11]	Humanoid	-	[34]	Tight
Staudte & Crocker [28, 29]	Humanoid	-	[35, 36]	Tight
Mutlu *et al.* [13]	Humanoid	-	Experiment, [33, 51, 52]	Tight
Mutlu *et al.* [25]	Humanoid, Android	-	Experiment	Tight
Ishi *et al.* [20]	Humanoid, Android	-	Experiment	Tight
Bethel & Murphy [14]	Non-Anthro-pomorphic	-	-	-
Heerink *et al.* [26]	Animal-Like	-	-	-

Continued on next page

Table A.1 – *Continued from previous page*

Studies	**Robot**	**Subsystems**	**Model**	**Sync**
Shimada *et al.* [15]	Android	-	-	-
Holroyd *et al.* [1]	Humanoid	Collaboration Manager, Engagement Recognition, Turn Policy, Reference Policy, Response Policy, Maintenance Policy, and BML Realizer.	[33, 34] [65]	Tight
Sirkin *et al.* [42]	Humanoid	-	-	-
Liu *et al.* [43]	Humanoid, Android	-	Experiment	-
Huang & Mutlu [2]	Humanoid	Perceptual System, Cognitive System, Behavioral System, Behavior Coordination System, Behavior Generator, Activity Model, Memory, and Social Behavior Knowledge Base.	[33, 34] [35, 36]	Tight
Huang & Mutlu [16]	Humanoid	Perceptual System, Cognitive System, Behavior Selection System, Activity Model, Memory, and Social Behavior Knowledge Base.	[33, 34] [35, 36]	Tight
Pitsch *et al.* [49]	Humanoid	-	-	-

Continued on next page

Table A.1 – *Continued from previous page*

Studies	**Robot**	**Subsystems**	**Model**	**Sync**
Admoni *et al.* [17]	Animal-Like	-	-	-
Andrist & Mutlu [23]	Humanoid	Gaze Controller, Head Controller, Speech Recognition, and Dialogue Manager.	-	-
Huang & Mutlu [18]	Humanoid	-	Experiment	Tight
Sauppe & Mutlu [31]	Humanoid	-	[35, 36]	Tight
Admoni *et al.* [30]	Humanoid	-	-	Tight
Moon *et al.* [32]	Humanoid	-	[35, 36]	Tight

APPENDIX B

ANALYSIS OF EXISTING ARCHITECTURES

The Software Architecture Analysis Method (SAAM) is a five step process used for evaluating existing architectures [81]:

1. *Characterize a reference architecture of the domain.* The behavioral robotics framework was synthesized from 32 previous implementations using the two step methodology for deriving reference architectures outlined in [59,60]. Therefore, for this analyses we used the behavioral robotics framework described in Section 3.

2. *Describe the existing architecture in terms of the reference architecture.* The structural decomposition of the two systems architectures are mapped on to the behavioral robotics framework, followed with an allocation of functionality to the structure.

3. *Choose a set of quality attributes with which to assess the architecture.* The two system architectures are evaluated for *overall functionality*. While any other quality attributes – such as the modifiability to new environments, extension of capabilities, and portability to different robot types [81] – can be used, these attributes are not considered in the current evaluation because the existing architectures are still in development and not mature.

4. *Choose a set of concrete tasks which test the desired quality attributes. Overall functionality* is the number of head gaze behaviors supported by the architecture.

5. *Evaluate the degree to which each architecture provides support for each task.* To *architecturally support overall functionality*, a subsystem to support the behavior must be present. Additionally, best practices for architectural design [59] require that the subsystems responsible for supporting behaviors should be a) isolated in architectural description, that is the subsystem should be isolated from the rest of the architecture, and b) non-monolithic. There should be support for subdivision of functionality within the subsystem.

B.1 Human-Robot Collaboration Architecture

This section details the architectural description and analysis of the Human-Robot Collaboration architecture.

B.1.1 Architectural Description

The re-characterization of the Human-Robot Collaboration architecture is shown in Table B.1. The Raw Sensor Data component of the Human-Robot Collaboration architecture is assigned to the Sensor Processing Module. Two subsystems, Collaboration Manager and Behavior Recognition are allocated to the Perception Module. The Collaboration Manager contains dialogue annotated with turn status. The functionality of the behavior Recognition subsystem is to perceive behavior indicators such as when a human initiates a connection. Three subsystems – Response Policy, Turn Policy, and Reference Policy – are assigned to the Behavior Module. The Turn Policy subsystem generates the head gaze required for engaging in a conversation. This subsystem performs the function of the REGULATING AN INTERACTION subsystem. The Response Policy subsystem generates the head gaze necessary for looking interested in humans, which is the function of the COMMUNICATING SOCIAL ATTENTION subsystem. The Reference Policy subsystem generates referential head gazes for looking at objects in the environment. The subsystem captures the functionality of the Manifesting an Interaction subsystem. Two subsystems, Maintenance Policy and Collaboration Manager, are allocated to the Action Arbitration Module. The role of the Maintenance Policy subsystem is to prioritize the head gaze policy. The Collaboration Manager described above has one additional function. This subsystem is responsible for inhibiting turn or point gestures. Both of these subsystems perform the function of the Coordination Function of the behavioral robotics framework. The BML Realizer subsystem is allocated to the Action Execution Module. This subsystem executes the overall output of the robot.

There are four points of interest to note in this re-characterization of the Human-Robot Collaboration architecture:

1. The description of the Collaboration Manager subsystem is monolithic; hence, it does not lend itself to a subdivision of functionality. This is because there is limited structural separation between the perception of turn status, content of dialogue, and behavior arbitration. The Collaboration Manager must provide the dialogue, identify the turn events, and provide conflict resolution.

2. The coordination mechanisms exist in both the Collaboration Manager and Maintenance Policy subsystems and their interactions are not fully defined and isolated. For example, what happens when two rules have the same priority has not been addressed.

3. In its current form, the architecture doesn't include mechanisms for Projecting Mental State and Establishing Agency, which have been shown to be important components of head gaze in other systems [4, 8, 25].

Behavioral Robotics Framework		Human Robot Collaboration Architecture	Robot Behavior Toolkit Architecture
Module	**Component**	**Component**	**Component**
Sensor Processing Module	Raw Sensor Data	Raw Sensor Data	Raw Sensor Data
	Internal State		
Perceptual Module	Perceptual System	Collaboration Manager	Perceptual System
		Behavior Recognition	Cognitive System
Behavior Module	Communicating Social Attention	Response Policy	Behavior Selection System and Knowledge Base
	Regulating an Interaction	Turn Policy	
	Manifesting an Interaction	Reference Policy	
	Projecting Mental State	-	-
	Establishing Agency	-	-
Action Arbitration Module	Coordination Function	Maintenance Policy	Behavior Coordination System
		Collaboration Manager	
Action Execution Module	Overall Response	BML Realizer	Behavior Generator

Table B.1: Allocation of the Components of Human-robot Collaboration Architecture [1] and Robot Behavior Toolkit Architecture [2] on the Behavioral Robotics Framework Based on Functionality.

B.1.2 Architecture Analyses

The *overall functionality* of the Human-Robot Collaboration architecture is three, since the architecture is only capable of generating head gaze in three out of the five behaviors: Communicating Social Attention, Regulating an Interaction, and Manifesting an Interaction. This because it consists of only those subsystems (see Table B.1).

B.2 Robot Behavior Toolkit Architecture

This section details the architectural description and analysis of the Robot Behavior Toolkit architecture.

B.2.1 Architectural Description

The re-characterization of the Robot Behavior Toolkit architecture is shown in Table B.1. The Raw Sensor Data component of the Robot Behavior Toolkit architecture is assigned to the Sensor Processing Module. Two subsystems, the Perceptual System and the Cognitive System, are allocated to the Perception Module. The Perceptual System transforms stimuli into a percept. The Cognitive System provides internal and external percepts based on the information from the Perceptual System and the current action prescribed by the Activity Model. Two subsystems, the Knowledge Base and the Behavior Selection System, are assigned to the Behavior Module. The Knowledge Base is a collection of behavioral specifications in XML. The Behavior Selection System queries the Knowledge Base for an appropriate behavior based on the percept. Both these subsystems are responsible for the generation of head gaze and perform the function of three behavioral robotics framework subsystems: COMMUNICATING SOCIAL ATTENTION, REGULATING AN INTERACTION, and MANIFESTING AN INTERACTION. The Behavior Coordination subsystem is assigned to the Action Arbitration Module. The role of the Behavior Coordination subsystem is to resolve conflicts and overlaps among behaviors by prioritization. This subsystem performs the function of the Coordination Function of the behavioral robotics framework. However, as was explicitly mentioned by Huang et al. [16], this subsystem has not been implemented. The Behavior Generator subsystem is allocated to the Action Execution Module. This subsystem organizes the coordinated behavior in XML for execution.

There are two points of interest to note in this re-characterization of the Robot Behavior Toolkit architecture:

1. The Knowledge Base subsystem is a collection of behavioral specifications in

XML. The description of this subsystem is monolithic and not isolated. As seen in Table B.1, each of the behaviors use the same subsystem.

2. The architecture does not support the following two components: Projecting Mental State and Establishing Agency. These components have been shown to be important components of head gaze in other systems [4, 8, 25].

B.2.2 Architecture Analyses

The *overall functionality* of the Robot Behavior Toolkit architecture is three. The architecture is only capable of generating head gaze in three out of five behaviors: Communicating Social Attention, Regulating an Interaction, and Manifesting an Interaction. This is because only these three behaviors have been implemented (see Table B.1).

APPENDIX C

EXPERIMENTAL STUDY CONSENT FORM

Version: 04/28/2011

CONSENT FORM
Evaluation of a search and rescue robot in a confined space simulated disaster site

Introduction

The purpose of this form is to provide you (as a prospective research study participant) information that may affect your decision as to whether or not to participate in this research.

You have been asked to participate in a research study to evaluate robot technology used in search and rescue operations. The purpose of this study is to a) determine how participants evaluate the robots, b) assess how participants feel during their interactions with robots c) record these interactions for video coding and assessment of the interactions between the participants and robots. You were selected to be a possible participant because you have little or no experience interacting with search and rescue robots and can provide an evaluation of the robot in a confined space simulated disaster environment. These robots may be used in future Urban Search and Rescue operations. This study is being sponsored/funded by NSF grant No. 0905485.

What will I be asked to do?

Your participation in this effort is as follows: you will be requested to complete several assessments prior to starting the actual study. A demographics questionnaire and two different brief psychological assessments will be provided for completion. You will be asked to wear eye tracking goggles to determine where you look while interacting with the robot. You will be asked to lie down on a padded surface in a simulated disaster setting. This setting is a somewhat confined space. Video recordings will be made of the interactions between you and the robot in one scenario. A brief interview of your experiences will occur following the interactions and you will be given one final assessment before being released from the study. These activities will take about 60 minutes to complete.

What are the risks involved in this study?

You may experience some discomfort from laying on a padded surface in a somewhat confined space, but this should not be any greater than discomfort you may experience in your daily life.

What are the possible benefits of this study?

There is no direct benefit to you from taking part in this study; however, you will contribute to a better understanding of human-robot interactions and aid in better understanding of how victims may interact with these robots in real search and rescue situations.

Do I have to participate?

No. Your participation is voluntary. You may decide not to participate or to withdraw at any time without your current or future relations with Texas A&M University being affected.

Texas A&M University IRB Approval	**From:** 6-22-11	**To:** 6-21-12
IRB Protocol # 2011-0430	**Authorized by:** Jean Wuy...	

Page 1 of 3

Will I be compensated?

You will not be paid for your participation. You will be provided a door prize ticket at the start of the study for your participation in the study. Completion of the study is not required to receive the door prize ticket. The door prize winning ticket(s) will be drawn at the conclusion of the research study and you do not need to present at the time of the drawing to win.

Who will know about my participation in this research study?

Your privacy and research records will be kept confidential to the extent of the law. Authorized research personnel, employees of the Department of Health and Human Services, and the TAMU Institutional Review Board may inspect the records from this research project.

Data files will be available upon demand to human-robot interaction researchers. Results extracted from the datasets acquired in this study may be published. However, the data obtained from you will be combined with data from others in the publication. The published results will not include your name or any other information that would personally identify you in any way.

Data files will not be identified with, or contain, a participant's name or social security number. In some cases, the affiliation of a participant with a university or employer may be inferred from logos on clothing; we will not edit the video to disguise these logos. Data files will only be identified by the arbitrary identification number created for this study, and the types of interactions. Each participant may have a copy of their own data. Data files released to members of the Human-Robot Interaction community will be referenced only by the arbitrary identification number, and type of interaction.

Whom do I contact with questions about the research?

This research study has been reviewed by the Human Subjects' Protection Program and/or the Institutional Review Board at Texas A&M University. For research-related problems or questions regarding your rights as a research participant, you can contact these offices at (979)458-4067 or irb@tamu.edu.

Whom do I contact about my rights as a research participant?

This research study has been reviewed by the Human Subjects' Protection Program and/or the Institutional Review Board at Texas A&M University. For research-related problems or questions regarding your rights as a research participant, you can contact these offices at (979)458-4067 or irb@tamu.edu.

Texas A&M University IRB Approval	**From:** 6-22-11 **To:** 6-21-12
IRB Protocol # 2011-0430	**Authorized by:** [signature]

Version: 04/28/2011

Signature

Please be sure you have read the above information, asked questions and received answers to your satisfaction. You will be given a copy of the consent form for your records. By signing this document, you consent to participate in this study.

______ I agree to be Video recorded.

______ I do not want to be Video recorded.

Signature of Participant: ______________________________ **Date:** ____________

Printed Name: ____

Signature of Person Obtaining Consent: ____________________ **Date:** ____________

Printed Name: ____

Texas A&M University IRB Approval	**From:** 6-22-11	**To:** 6-21-12
IRB Protocol # 2011-0430	**Authorized by:** Jan Wuegs	

APPENDIX D

EXPERIMENTAL STUDY INFORMATION SHEET

Version: 04/28/2011

INFORMATION SHEET

Evaluation of a search and rescue robot in a confined space simulated disaster site

Introduction

The purpose of this form is to provide you (as a prospective research study participant) information that may affect your decision as to whether or not to participate in this research.

You have been asked to participate in a research study to evaluate robot technology used in search and rescue operations. The purpose of this study is to a) determine how participants evaluate the robots, b) assess how participants feel during their interactions with robots c) record these interactions for video coding and assessment of the interactions between the participants and robots. You were selected to be a possible participant because you have little or no experience interacting with search and rescue robots and can provide an evaluation of the robot in a confined space simulated disaster environment. These robots may be used in future Urban Search and Rescue operations.

This study is being sponsored/funded by NSF grant IIS No. 0905485.

What will I be asked to do?

Your participation in this effort is as follows: you will be requested to complete several assessments prior to starting the actual study. A demographics questionnaire, a health questionnaire, and two different brief psychological assessments will be provided for completion. You will be asked to wear eye tracking goggles to determine where you look while interacting with the robot. You will be asked to lie down on a padded surface in a simulated disaster setting. This setting is a somewhat confined space. Video recordings will be made of the interactions between you and the robots in one scenario. A brief interview of your experiences will occur following the interactions and you will be given one final assessment before being released from the study. These activities will take about 60 minutes to complete.

What are the risks involved in this study?

You may experience some discomfort from laying on a padded surface in a somewhat confined space, but this should not be any greater than discomfort you may experience in your daily life.

What are the possible benefits of this study?

There is no direct benefit to you from taking part in this study; however, you will contribute to a better understanding of human-robot interactions and aid in better understanding of how victims may interact with these robots in real search and rescue situations.

Do I have to participate?

No. Your participation is voluntary. You may decide not to participate or to withdraw at any time without your current or future relations with Texas A&M University being affected.

Will I be compensated?

Texas A&M University IRB Approval	**From:** 6-22-11	**To:** 6-21-12
IRB Protocol # 2011-0430	**Authorized by:** [signature]	

Page 1 of 2

You will not be paid for your participation. You will be provided a door prize ticket at the start of the study for your participation in the study. Completion of the study is not required to receive the door prize ticket. The door prize winning ticket(s) will be drawn at the conclusion of the research study and you do not need to present at the time of the drawing to win.

Who will know about my participation in this research study?

Your privacy and research records will be kept confidential to the extent of the law. Authorized research personnel, employees of the Department of Health and Human Services, and the TAMU Institutional Review Board may inspect the records from this research project.

Data files will be available upon demand to human-robot interaction researchers. Results extracted from the datasets acquired in this study may be published. However, the data obtained from you will be combined with data from others in the publication. The published results will not include your name or any other information that would personally identify you in any way.

Data files will not be identified with, or contain, a participant's name or social security number. In some cases, the affiliation of a participant with a university or employer may be inferred from logos on clothing; we will not edit the video to disguise these logos. Data files will only be identified by the arbitrary identification number created for this study, and the types of interactions. Each participant may have a copy of their own data. Data files released to members of the Human-Robot Interaction community will be referenced only by the arbitrary identification number, and type of interaction.

Whom do I contact with questions about the research?

If you have questions regarding this study, you may contact Vasant Srinivasan (vasant_s18@tamu.edu, ph 979-324-4720) or Zack Henkel (zmhenkel@neo.tamu.edu, ph 409-781-9846)

Whom do I contact about my rights as a research participant?

This research study has been reviewed by the Human Subjects' Protection Program and/or the Institutional Review Board at Texas A&M University. For research-related problems or questions regarding your rights as a research participant, you can contact these offices at (979)458-4067 or irb@tamu.edu.

Participation

Please be sure you have read the above information, asked questions and received answers to your satisfaction. If you would like to be in the study, contact Vasant Srinivasan (vasant_s18@tamu.edu, ph 979-324-4720) or Zack Henkel (zmhenkel@neo.tamu.edu, ph 409-781-9846).

Texas A&M University IRB Approval	From: 6-22-11	To: 6-21-12
IRB Protocol # 2011-0430	Authorized by: [signature]	

APPENDIX E

EXPERIMENT PROTOCOL

The name of this study is: “Evaluation of a Search and Rescue Robot in a Confined Space Simulated Disaster Site.” The purpose of this study is to evaluate a robot that may be used in urban search and rescue operations. We will ask you to evaluate your feelings and reactions to the robots with which you are interacting. Here is how the study will go:

1. We will have you read and complete the informed and videotaping consent forms.

2. You will be assigned a unique identifier or participant ID that you will use throughout the experiment.

3. You will be given your door prize ticket.

4. If at any time you feel as though you are not able to continue with the experiment just let us know and we will assist you in getting out of the confined space box as quickly as possible.

5. You will be asked to complete a demographics questionnaire, so that we can gather some basic information about you.

6. Once you complete this, you will be asked to wear an eye tracking goggle and calibrate the eye-tracking system.

7. Next you will be asked to view a 3 minute video from an actual disaster to set the scene.

8. After the video is completed minutes are completed we will take you into the simulated disaster site and place you in a moderately confined space so you will have the sensation of being in a disaster setting.

9. You will be lying down on your right side during the robot interactions.

10. Next you will interact with the robot in one randomly assigned scenario. The interaction will take approximately 7 minutes. During this scenario your interaction will be videotaped and we will be obtaining eye tracking information.

11. Once the interaction is complete we will assist you with sitting up, remove you from the confined space simulated disaster site and remove the eye tracking goggle.

12. You will be taken back to the desk area where you began the research study and we will have you complete a post interaction questionnaire.

13. If at any point you have questions or do not understand any item(s) on the assessments, please feel free to ask questions.

14. We will debrief you on the goals of the study. Then you will be free to leave the study area. We ask that you do not discuss the details of your experiences with others so that the study will not be impacted by participants having prior knowledge of the study.

APPENDIX F

INTERACTION SCRIPT

First word in Theme - **Bold**

First word in Rheme - *Italics*

Object Name - SMALL CAPS

Hello. [Pause]

Can you hear me? [Pause]

I am a robot that *has* been sent to help you. **The** building you were in *collapsed.*

Please do not worry; **a** rescue team *is* aware that you are trapped and knows where you are. **They** are currently working *to* free you.

I will *stay* with you and **remain** in contact with *the* rescuers as they work to reach you.

In the meantime, I will be here *to* help you and to keep you company. **I** will also *be* assessing your health and mental state periodically as we wait.

I'm going to start *by* asking you a few questions. [Pause]

First, can you please tell me *your* name? [Pause]

You have been found in an area of the collapsed building *that* suffered a lot of damage. **Did** you happen to *see* what caused the collapse? [Pause]

When the building collapsed, *what* level were you on? [Pause]

Is anyone with you? [Pause]

Did you see anyone *on* your floor **before** the building collapsed? [Pause]

Were you *hit* by falling rubble? [Pause]

Are you *experiencing* any pain right now? [Pause]

I will *notify* the responders. [Pause]

Now I will ask a few questions *to* verify your mental state.

What is *today's* date? [Pause]

What *type* of building were you in? [Pause]

Who is *the* current president of the United States? [Pause]

Thank you. Now, to *keep* you alert as we wait, **I** will *lead* you through a memory exercise.

I'm going to *list* some objects and I will later ask you to repeat *as* many of them as you can. Are you ready to begin? [Pause]

Here we *go*: vacuum, cat, doorknob, ladder, turkey, planet, pillow, fountain, chocolate, wire, stone, lemon, concrete, vase, boat, candy cane, speaker, tape, steering wheel, sock.

The rescuers are *working* very hard to rescue you. **They** are *getting* closer now.

I'm going to *examine* you for injuries. **Please** *follow* my instructions.

I will first be checking for neck injury. Can you comfortably *move* your head towards the direction of the EXIT sign? [Pause]

In order to check for spinal injury, *could* you try to wiggle the toes on your RIGHT leg.

Now your LEFT leg.

Did you have any trouble with **either** of those tasks? [Pause]

There is a FIRE EXTINGUISHER over there. **Can** you *point* your free arm toward it? [Pause]

Did moving your arm *cause* you any discomfort? [Pause]

I am passing your answers *to* the responders.

I will now *test* how many words you can remember from **the** list I gave you earlier.

When I say "go," you will have up to 30 seconds to **list** as many of the words as you can *remember*. Go.

[Wait 30 seconds]

All right, Good job.

I am now going to assess *the* area surrounding you and will appreciate your help.

Can you tell me if there is anything *hanging* above you? [Pause]

What are you *resting* on? [Pause]

I can see *that* there is a level sign over there, but I can't read the level number. **Can** you *tell* me what it is? [Pause]

I didn't hear you clearly. **Could** you *repeat* that? [Pause]

The rescuers are *almost* here. *Hang* in there.

I will now *conduct* another alertness test.

I will *name* a common object and **your** goal is to come up with *as* many uses for the object as you can think of that do not include its common use. **So**, for example, if I say "pencil," *you* won't say "writing" because a pencil is typically used for writing, **but** you might say you could use it *as* a "chopstick," or as a "dagger."

Once I state the word, *you* will then have 30 seconds **to** state as many uses for that object as you can think of. Don't worry about being correct, *just* try to be creative.

Are you *ready* to start? [Pause for answer].

OK.

The first object *is* "shoe." **What** can you use a shoe for apart from *wearing* it to walk?

[Wait for 30 seconds]

Time's up.

That's great!

The next object *is* "a sheet of paper." **What** can you use a sheet of paper for *apart* from writing on it? [Wait for 30 seconds]

Time's up.

For the last object, think of uses for that LICENSE PLATE lying over there. **What** can you use it for *apart* from identifying a car?

[Wait for 30 seconds]

Time's up.

That is all the information I *need* right now. *Thank* you for your help.

The rescuers are *now* approaching, so our interaction is complete. **Please** *lay* still and await further instruction from the rescuers.

APPENDIX G

PRE-INTERACTION QUESTIONNAIRE ITEMS

1. What is your gender?

   Male ○ Female ○

2. What is your age (in years)?

3. How many hours a week do you spend playing video games?

4. Do you own a robot?

   Yes ○ No ○

5. Do you have a pet dog or cat?

   Yes ○ No ○

6. Think about your previous experience interacting with robots.

	No Experience						A lot of Experience
	1	2	3	4	5	6	7
How much experience have you had interacting with robots?							

7. Please check one or more of the circles below that best describes your race/ethnicity.

Hispanic or Latino.	○	American Indian or Alaska Native.	○
Asian.	○	Black or African American.	○
Caucasian (White).	○	Native Hawaiian or Other Pacific Islander.	○
Middle Eastern.	○		

APPENDIX H

POST-INTERACTION QUESTIONNAIRE ITEMS

1. Think of the robot you interacted with during the simulation. How well do these words describe the robot?

	Describes Very Poorly 1	2	3	4	5	6	Describes Very Well 7
Enthusiastic	○	○	○	○	○	○	○
Frustrated	○	○	○	○	○	○	○
Feminine	○	○	○	○	○	○	○
Happy	○	○	○	○	○	○	○
Inefficient	○	○	○	○	○	○	○
Confident	○	○	○	○	○	○	○
Funny	○	○	○	○	○	○	○
Arrogant	○	○	○	○	○	○	○
Cheerful	○	○	○	○	○	○	○
Honest	○	○	○	○	○	○	○
Helpful	○	○	○	○	○	○	○
Kind	○	○	○	○	○	○	○
In Control	○	○	○	○	○	○	○
Humorless	○	○	○	○	○	○	○
Jovial	○	○	○	○	○	○	○
Extroverted	○	○	○	○	○	○	○

Continued on next page

Table H.1 – *Continued from previous page*

	Describes Very Poorly						Describes Very Well
	1	2	3	4	5	6	7
Introverted	○	○	○	○	○	○	○
Warm	○	○	○	○	○	○	○
Masculine	○	○	○	○	○	○	○
Trustworthy	○	○	○	○	○	○	○
Cold	○	○	○	○	○	○	○
Confident	○	○	○	○	○	○	○
Reliable	○	○	○	○	○	○	○
Sympathetic	○	○	○	○	○	○	○
Outgoing	○	○	○	○	○	○	○
Likeable	○	○	○	○	○	○	○
Sly	○	○	○	○	○	○	○
Sincere	○	○	○	○	○	○	○
Concerned about me.	○	○	○	○	○	○	○
Unemotional	○	○	○	○	○	○	○
Empathetic	○	○	○	○	○	○	○
Shy	○	○	○	○	○	○	○

2. Indicate your agreement with the following statements.

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
The robot's primary purpose was to help me.	○	○	○	○	○	○	○
The robot's primary purpose was to help the rescuers.	○	○	○	○	○	○	○
The robot would only do things that were in my best interest.	○	○	○	○	○	○	○
The robot would follow the rescuers' orders, even if it caused me harm.	○	○	○	○	○	○	○
The robot was more loyal to me than the rescuers.	○	○	○	○	○	○	○
The robot was on my side.	○	○	○	○	○	○	○

3. Indicate your agreement with the following statements.

	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
The robot was engaging.	○	○	○	○	○	○	○
I liked the robot.	○	○	○	○	○	○	○
The robot annoyed me.	○	○	○	○	○	○	○
The robot was friendly.	○	○	○	○	○	○	○
The robot made me feel relaxed.	○	○	○	○	○	○	○
The robot was shy.	○	○	○	○	○	○	○
The robot made me nervous.	○	○	○	○	○	○	○
I trusted the robot.	○	○	○	○	○	○	○
The robot made me feel safe.	○	○	○	○	○	○	○
The robot liked you.	○	○	○	○	○	○	○

4. Indicate your agreement with the following statements.

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
The robot saw the situation from my perspective.	○	○	○	○	○	○	○
The robot was concerned about me.	○	○	○	○	○	○	○
The robot was oblivious to my emotional state.	○	○	○	○	○	○	○
The robot wanted me to be rescued.	○	○	○	○	○	○	○
The robot was empathetic.	○	○	○	○	○	○	○
I felt better with the robot than I would have felt if I were alone.	○	○	○	○	○	○	○
If I were ever trapped, I would prefer to wait for rescue by myself than with the robot.	○	○	○	○	○	○	○
If I were ever trapped, the robot would help me pass the time.	○	○	○	○	○	○	○

5. How well do these words describe the robot?

	Describes Very Poorly 1	2	3	4	5	6	Describes Very Well 7
Intelligent	○	○	○	○	○	○	○
Harsh	○	○	○	○	○	○	○
Fair	○	○	○	○	○	○	○
Friendly	○	○	○	○	○	○	○
Competent	○	○	○	○	○	○	○
Incompetent	○	○	○	○	○	○	○
Qualified	○	○	○	○	○	○	○
Unpleasant	○	○	○	○	○	○	○
Experienced	○	○	○	○	○	○	○
Rude	○	○	○	○	○	○	○
Cooperative	○	○	○	○	○	○	○
Skilled	○	○	○	○	○	○	○
Motivated	○	○	○	○	○	○	○
Informed	○	○	○	○	○	○	○
Unkind	○	○	○	○	○	○	○
Quick Learner.	○	○	○	○	○	○	○
Committed to the task.	○	○	○	○	○	○	○
Trained	○	○	○	○	○	○	○

Continued on next page

Table H.5 – *Continued from previous page*

	Describes Very Poorly 1	2	3	4	5	6	Describes Very Well 7
Assertive	○	○	○	○	○	○	○
Difficult to Use	○	○	○	○	○	○	○
Dishonest	○	○	○	○	○	○	○
Understandable	○	○	○	○	○	○	○
Adaptive	○	○	○	○	○	○	○
Aggressive	○	○	○	○	○	○	○
Unhelpful	○	○	○	○	○	○	○

6. Think about your feelings while participating in the simulation. Indicate your agreement with the following statements.

	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
I remained focused.	○	○	○	○	○	○	○
I felt stressed.	○	○	○	○	○	○	○
I felt claustrophobic.	○	○	○	○	○	○	○
I was bored.	○	○	○	○	○	○	○
I felt like another person was physically close to me.	○	○	○	○	○	○	○
I believed that rescuers were on their way.	○	○	○	○	○	○	○
I felt crowded.	○	○	○	○	○	○	○
I felt optimistic.	○	○	○	○	○	○	○
I felt lonely.	○	○	○	○	○	○	○
I felt like I was part of a team.	○	○	○	○	○	○	○
I felt frustrated.	○	○	○	○	○	○	○
I was scared.	○	○	○	○	○	○	○
I was confident the rescuers would find me.	○	○	○	○	○	○	○
I felt like I was all alone.	○	○	○	○	○	○	○
I felt trapped.	○	○	○	○	○	○	○

7. Think back to your interaction with the robot and indicate your agreement with the following statements.

	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
The robot was looking at me.	○	○	○	○	○	○	○
The robot was looking away from me.	○	○	○	○	○	○	○
The robot was attentive to me.	○	○	○	○	○	○	○
I was interested in the information presented to me.	○	○	○	○	○	○	○
The information presented to me was enjoyable.	○	○	○	○	○	○	○
I could understand the robot well.	○	○	○	○	○	○	○
I was attentive to the robot.	○	○	○	○	○	○	○
The robot behaved human-like.	○	○	○	○	○	○	○
The robot was attractive.	○	○	○	○	○	○	○

Continued on next page

Table H.7 – *Continued from previous page*

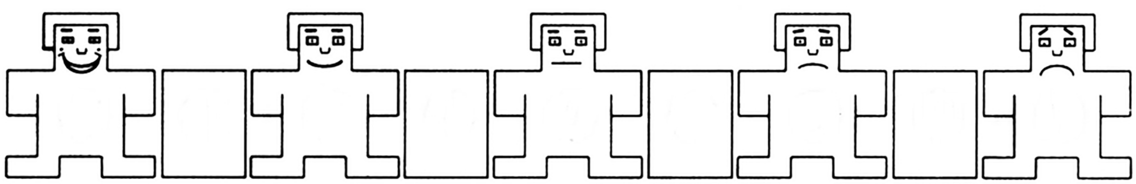
	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
The robot was friendly.	○	○	○	○	○	○	○
The robot was optimistic.	○	○	○	○	○	○	○
The robot was happy.	○	○	○	○	○	○	○
The robot was knowledgeable.	○	○	○	○	○	○	○
The robot was irresponsible.	○	○	○	○	○	○	○
The robot was intelligent.	○	○	○	○	○	○	○
The robot was foolish.	○	○	○	○	○	○	○
The robot was ignorant.	○	○	○	○	○	○	○
The robot was sensible.	○	○	○	○	○	○	○
The robot felt like a stranger.	○	○	○	○	○	○	○
The robot was aware of its surroundings.	○	○	○	○	○	○	○
The robot was focused on me.	○	○	○	○	○	○	○
The robot had a personality.	○	○	○	○	○	○	○

Continued on next page

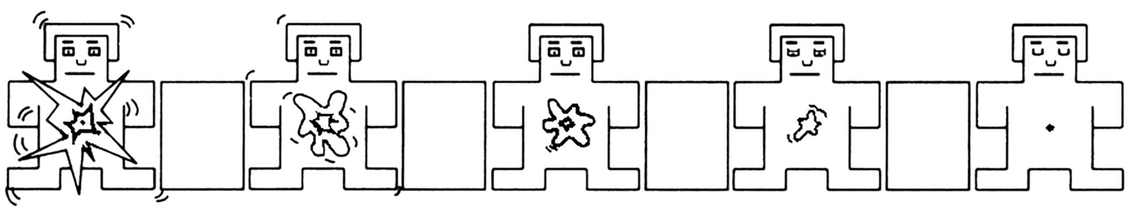
Table H.7 – *Continued from previous page*

	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
The robot moved naturally.	○	○	○	○	○	○	○
The robot looked at the objects at appropriate times.	○	○	○	○	○	○	○

8. How positive/negative did you feel about your interaction with the robot presented?



9. How agitated/comforted did you feel about your interaction with the robot presented?



APPENDIX I

LIST OF ITEMS AND RELIABILITY FOR THE 23 MEASURES

Measure	**Item**	**Cronbach's α**
SAM: Valence [55]	"How positive/negative did you feel about your interaction with the robot presented?"	-
Creativity [69]	Summation of the the total number of alternate uses participants generated within 30 seconds for three items: "shoe," "sheet of paper," and "license plate" during the interaction.	-
Memory [68]	Summation of the the total number of memorized items recalled by the participant during the interaction. The robot read off twenty different memory items ("vacuum," "cat," "doorknob," etc) then the robot diverted the participants attention for 30 seconds. The robot then allowed 30 seconds for participants to state as many of the items as they could remember.	-

Continued on next page

Table I.1 – *Continued from previous page*

Measure	Item	Cronbach's α
Person at Ease [64]	Index of four items: "I was scared," "I felt stressed," "I felt frustrated," and "I felt trapped."	.71
SAM: Arousal [55]	"How agitated/comforted did you feel about your interaction with the robot presented?"	-
Chance of Rescue [64]	Index of three items: "I was confident the rescuers would find me," "I believed that rescuers were on their way," and "I felt optimistic."	.78
Robot Empathy [64]	Index of five items: "kind," "sincere," "empathetic," "sympathetic," and "concerned about me."	.76
Robot Loyalty [64]	Index of four items: "the robot's primary purpose was to help me," "the robot would only do things that were in my best interest," "the robot was more loyal to me than the rescuers," and "the robot was on my side."	.76
Robot Integrity [64]	Index of five items: "likeable," "trustworthy," "helpful," "honest," and "reliable."	.77

Continued on next page

Table I.1 – *Continued from previous page*

Measure	**Item**	**Cronbach's α**
Robot Caring [64]	Index of five items: "the robot liked me," "the robot saw the situation from my perspective," "the robot was concerned about me," "the robot was empathetic," and "the robot wanted me to be rescued."	.75
Robot Engagement [1, 8, 65]	"The robot was engaging."	-
Robot Likeability [64]	Index of five items: "I liked the robot," "the robot was friendly," "the robot made me feel relaxed," "I trusted the robot," and "the robot made me feel safe."	.87
Human-Like Behavior [9]	"The robot behaved human-like."	-
Robot Intelligence [9]	"Intelligent."	-
Robot Detachment [64]	Index of three items: "humorless," "unemotional," and "cold."	.53 (unreliable)
Robot Confidence [64]	Index of three items: "confident," "in control," and"masculine."	.26 (unreliable)

Continued on next page

Table I.1 – *Continued from previous page*

Measure	**Item**	**Cronbach's α**
Robot Competence [64]	Index of eight items: "committed to the task," "competent," "experienced," "informed," "intelligent," "qualified," "skilled," and "trained."	.84
Robot Unpleasantness [64]	Index of seven items: "difficult to use," "dishonest," "incompetent," "rude," "unhelpful," "unkind," and "unpleasant."	.79
Robot Extraversion [64]	Index of seven items: "outgoing," "extraverted," "vivacious," "jovial," "enthusiastic," "cheerful," and "perky."	.71
Understandability of Robot Behaviors [65]	Index of three items: "I always knew what object the robot looked at," "I could easily tell which objects the robot looked at," and "I could understand the robot."	.86
Gaze-Speech Synchronization [65]	"The robot synched its movements with what it was saying."	-
Looking at Objects at Appropriate Times [65]	"The robot looked at the objects at appropriate times."	-
Natural Movement [9, 65]	"The robot movements were natural."	-

APPENDIX J

QUESTIONS FOR DEBRIEFING

Participants were requested to answer the following questions following Bethel and Murphy [82]:

1. What were you feeling during the interaction?
2. Were there any feelings that arose during the interaction that impacted you in a positive way?
3. Were there any feelings that arose during the interaction that impacted you in a negative way?
4. Was there anything that occurred during the interaction that was problematic for you in any way?
5. Do you have any suggestions for improving the experimental process?
6. Do yo have any other comments or suggestions about this experiment?

APPENDIX K

CALCULATION OF FDR CORRECTED SIGNIFICANCE LEVEL

1: Create a Vector A by sorting observed p-values
2: Create the vector B by computing $j * \frac{\alpha}{21}$.
3: Subtract vector A from vector B; call this vector C.
4: Find the largest index, d, (from 1 to 21) for which the corresponding number in vector C is negative.
5: Reject all null hypotheses whose p-values are less than or equal to p_d (d indexes vector A). The null hypotheses for the other tests are not rejected.

The FDR control algorithm is applied to the original results as shown in the Table below. The largest index for which the corresponding number in vector C is negative is 14. Therefore, the corrected significance level after the FDR (Benjamini-Hochberg) correction [74] is $p < 0.033$.

	Original p**-values**	**Vector A (Sorted** p**-values)**	**Vector B** $(j * \frac{\alpha}{21})$	**Vector C (Vector B - Vector A)**
1	.02	.001	.0024	-.0014
2	.02	.001	.0048	-.0038
3	.47	.001	.0071	-.0061
4	.03	.001	.0095	-.0085
5	.001	.001	.0119	-.011
6	.3	.001	.0142	-.0133
7	.01	0.002	.0167	-.0147
8	.02	0.01	.019	-.009
9	.03	0.02	.0214	-.0014
10	.02	0.02	.0238	-.0038
11	.04	0.02	.0262	-.0062
12	.002	0.02	.0286	-.0086
13	.001	.03	.031	-.0095
14	.35	.03	.0333	-.0033
15	.51	.3	.0357	.2643
16	.47	.33	.0380	.2919
17	.75	.35	.0405	.3095
18	.001	.47	0.0429	.4271
19	.001	.47	.0452	.4248
20	.001	.51	0.0476	.4624
21	.001	.75	0.05	.7