Interaction Design of Sociable Driving Agents as an Effective In-Vehicle Interface

ドライバーと自動車との効果的なインタフェースに向けたドライビングエージェントのインタラクションデザイン

January, 2019

Doctor of Philosophy (Engineering)

Nihan Karatas
ニハン　カラタス

Toyohashi University of Technology

Acknowledgement

I would like to express my sincere gratitude to my advisor Prof. Michio Okada for his continuous support of my Ph.D study, also the motivation and immense knowledge he provided. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study. His inspirational concepts and ideas on robots, commonly known as "weak robots" brought me another insight to designing the interactions between the humans and the social robots. Thanks to him, I could work on very unique and promising projects in human-robot interaction research.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Jun Muira and Prof. Shigeru Kuriyama for their insightful comments and encouragement, but also for the hard questions which pushed me to think about my research from various perspectives. My sincere thanks also goes to Prof. Michio Kitazaki for his valuable feedback and encouragement. Without their precious support it would not be possible to conduct this research.

I thank my fellow lab mates in for the stimulating discussions, great help, support, co-working and the great enjoyment brought to my life. I feel very lucky to have such great friendships with all of my friends in this research atmosphere.

Last but not the least, I would like to thank my family: my parents and my brother for supporting me spiritually throughout writing this thesis and for encouraging me to follow my dreams. Without their love and support I won't be here today.

別紙４－１（課程博士（英文））

Date of Submission（January 7, 2019）:

Department Computer Science and Engineering	Student ID Number	D135311	Supervisors	OKADA Michio
Applicant's name	KARATAS Nihan			

Abstract（Doctor）

Title of Thesis	Interaction Design of Sociable Driving Agents as an Effective In-Vehicle Interface

Approx. 800 words

With the advancement of the technology, the interaction design of the dashboard of the cars have been changed a lot in the last years. The amount of buttons have been becoming increasingly confusing and sensory overload. On the other hand, a great deal of research has been conducted on highly autonomous vehicles which make their own driving decisions that minimise human interventions with the vision of decreasing human errors and achieving a safer, more energy efficient and more comfortable mode of transportation. The autonomous cars hold much more functionalities compared to the cars on the road today. Therefore, the design of the dashboard for the expected self-driverless cars should be created in a way that intuitively understandable by the wide range of users considering their naivety on the technology and cultural background, etc. It has been claimed that the human brain has evolved to be highly adaptive in social interactions therefore, people tend to anthropomorphise the technology. From the drivers' perspective, we believe that it is crucial to interact with an in-vehicle interface system in such a social, natural and familiar manner to reduce mental workload and create a more sociable environment inside a car. Since human brain has been evolved to be an expert in social interactions, social robots are envisioned as having the ability to interact with humans (and others) socially in order to achieve their designated goals. With this respect, a social robot platform what would mediate the interactions between a car and a driver can be effective in terms of obtaining environmental information and understanding the vehicle's intentions while interacting with the driver socially. In this thesis, we address several problems regarding to the interaction between such a social interface and a driver.

Firstly, we propose a social interface named NAMIDA that incorporates three conversational robots that can decrease the number of directed utterances towards a driver through a turn-taking process among the robots. First, we evaluated this model by employing virtually embodied social agents. Through this model, we show that drivers can gain necessary location-based information without joining the conversation among the robots. The results of our pilot study revealed that the proposed multi-party conversation based interaction model is more effective in alleviating certain workload factors for drivers compared to a conventional one-to-one communication based approach that directly addresses the driver. Then, we built and used our robotic driving agents to conduct an

experiment to investigate the lifelikeness and distractedness of the interaction of the multi-party conversation of three robotic agents and the one-to-one conversation between one robot with a driver. In this study, we show that overhearing information from the physically embodied multi-party conversation based driving agent system is perceived as possessing more lifelike characteristics compared to a conventional, one-to-one communication based driving agent that directly addresses the driver. Also, the proposed approach reduced the distraction level and increased the enjoyment of the drivers. Through these two studies, we demonstrated that an interaction design with the multi-party conversation of the driving agents can be more efficient and enjoyable when the driver's attention is required on driving (manual driving).

Considering the current researches on the autonomy Level 3 of the autonomous vehicles (limitedly autonomous), depending on the circumstances along the road (i.e. bad road conditions), the driver should take-over the control from the autonomous mode as smooth and rapid as possible. Therefore, our next study investigated a paradigm for keeping the drivers' situation-awareness active during the autonomous driving by utilising our social robot system, NAMIDA. We analysed the effectiveness of NAMIDA on maintaining the drivers' attention to the road, by evaluating the response time of the drivers to a critical situation on the road. An experiment consisting of a take over scenario showed that existence of NAMIDA significantly reduced the response time of the drivers with eye gaze behaviours of the robots. In addition, we inferred that the robots facilitated the drivers to put them in social confirmity where the drivers' attention was on the road more often when the robots were always watching the road. However, in this study, the effects of eye gazing behaviours of the robots on the perceptions of the drivers in terms of comprehending the intention of the robots and feeling the autonomous driving safer remained unknown.

In order to achieve a reliable interaction with the autonomous cars, intersubjectivity should be built between the autonomous car and the human operator where the human will believe that the car possess the same intentions with the human. One critical social cue for human to understand the intentions of others is eye gaze behaviours. In our next study, we demonstrated that when the robots followed the eye gaze behaviours of the driver, the perception of the intersubjectivity and the autonomous car as a social entity were increased. The results of this study also revealed that the autonomous system was perceived safer and more enjoyable compared to the condition with not using the robots and the condition with using random gazing behaviours of the robots, respectively.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we introduce the future interface of the automobiles by addressing first a brief history of the in-car interfaces, and then how they have changed in time. Then, we discuss the human-centric design principles for these interfaces that should adapt to drivers' needs. Autonomous cars are seen as the future of the transportation with the vision of decreasing human errors and achieving a safer, more energy efficient and more comfortable mode of transportation. However, eliminating human involvement from driving might threaten the trust and perceived safety, and suppress drivers' joy of driving and the desire to control the vehicle which in turn lead to a refusal to use autonomous cars. Herein, it is very important to consider an in-car interface in order to mediate the interaction between the autonomous vehicle and the human operator. In this chapter, we will also mention the important factors in communication with an autonomous car, and how a social robot platform can be useful to mediate the interaction between the autonomous car and the driver with the aspects of the intersection of cognitive science as well as the ecological, developmental and social physiology.

1.1 User-Centered Design for an In-Car Interface

Over time, the change in user interface (UI) of cars has been dramatic. In 1908, the UI model of the cars mainly direct controlled, where the driver had only one dedicated leaver and gauge, for each function. It was a simple and purposeful design that did not distract the driver from operating the car in a safe way. Since, it was not as technologically advanced compared to today's cars, as many dials, buttons and gauges were not necessary. In 1930, when airplanes were becoming more popular, many luxury car companies started to design the UI of their cars inspired by these complex airplane cockpit UI designs. It was not because of necessarily to increase the usability of the product, but rather boast the technology under the hood. Dashboard complexity was driven by style instead of function. In this sense, these features became a definition of luxury. This was a time in which designers were imagining the future to be more about technology and less about people. During the 1990's car interiors were beginning to feel the burn from feature overload. The amount of buttons was becoming increasingly confusing and sensory overload. Technology adoption in cars today is hitting an inflection point, and the UI model we have grown accustomed to cannot handle it.

Driving is a dangerous activity and all these new features are beginning to become increasingly distracting to the driver. Furthermore, it is important for the naive users to address these technologies intuitively. Intuitive design is defined as "the design that a user is able to understand and use a design immediately that is, without consciously thinking about how to do it". With the proliferation of workplace computers in the early 1990s, user experience started to become a concern for designers. Norman [89] discussed that a system's design model should be identical to the user's mental model, so that a user should not give much effort to be able to use the technology. The user-centered design tries to optimize the product around how users can, want, or need to use the product, rather than forcing the users to change their behavior to accommodate the product [89]. The users thus stand in the center of two concentric circles: 1)The inner circle includes the context of the product, objectives of developing it and the environment it would run in. 2)The outer circle involves more granular details of task detail, task organization, and task flow [1]. With the introduction of further elements to car UI, such as: smartphone integration, Bluetooth and Wi-Fi

[1]https://www.w3.org/WAI/redesign/ucd

connectivity and the touchscreen displays; the automobile manufacturers, technology companies and researchers in the field focused on how to make the interaction intuitive for the users. However, the combined effect of long automotive design cycles that cannot keep pace with digital technology and the legacy of direct control make for clumsy dashboard design.

1.1.1 Driver Behavior in Autonomous Car

Autonomous car concept has received a great degree of attention in recent years. Considerable outreach activities undertaken by companies such as Google, and also following legislation in favour of the operation of 'autonomous cars' by the States of Nevada (March 2012), Florida (April 2012) and California (September, 2012). In the UK, the government has recently pledged the testing of autonomous cars on UK roads by 2013. The notion is also much favoured by the automotive industry who are currently quoted in the media on an almost daily basis, for instance, with Nissan (Wall Street Journal, 2013), General Motors (USA today, 2013) and Mercedes (Daily Mail, 2013) all committing the sale of 'self-driving cars' by 2020.

However, from the policy and research perspective, activities in this domain have been a little more gradual, and at least an understanding of the impact of such vehicles on overall road traffic management are not yet well understood [2]. Whilst the technology to allow the realisation of such cars is perhaps relatively advanced and more readily available, the challenge for human factors professionals and researchers is to ensure that the operators of such vehicles: i.e. the drivers – are able to comprehend the capabilities and limitations of the systems in place for automated driving [79].

The National Highway Traffic Safety Administration (NHTSA) defines four levels of car autonomy: Level 1: function-specific automation, Level 2: combined function automation, Level 3: limited self-driving and Level 4: full self-driving [1]. Among these, Level 3 has significant importance due to its being that of the expected next generation of the vehicles [13]. In this stage of driving, the driver is not essentially

[2]https://www.theengineer.co.uk/issues/august-2013-online/autos-on-autopilot-the-evolution-of-the-driverless-car/#ixzz2hJHF31Q3

required to monitor the road all the time; they can enjoy driving by engaging with non-driving-related activities. However, studies have demonstrated the effects of automation such as a loss of situational-awareness and overreliance on the increased level of automation [105], [34], [95]. Since in Level 3, there might be situations that the vehicle cannot handle (e.g. bad road conditions, increased traffic density), the driver should be available to take over the control within a sufficient transition time [1], [44]. In order to provide a quick and smooth handover, maintaining the drivers' attention on the road is crucial.

Recent studies in human factors focus on modalities such as visual, audio, speech and tactile to take the driver's attention efficiently and inform them about the handover process [84], [101]. However, each of these modalities or their combinations have been reported as more or less annoying for drivers [97], [72]. Due to the takeover request alerts the driver to react urgently in order to take an immediate reaction, in the case of a false alarm (e.g. a request is suggested although it is not necessary) the effectiveness and reliability of the system decreases and causes the rejection of usage of automated vehicle applications [14], [31]. In addition, announcing a take over request through these modalities is unilateral and is not perceived as considering the driver's stance.

1.2 A Social Robot Platform as an In-Car Interface

An in-car system that would interact with the driver should be persuasive so that it can influence them in their actions or beliefs. Fogg remarked that one persuasive element in technology is the role of social actors [39]. Considering people tend to treat computer systems as if they are real people [103], and the tendency of human brain in anthropomorphizing the technology [10], it is not surprising that the automated vehicles becoming persuasive when they perform more anthropomorphic features [129], [53].

Studies investigating the influences of passengers on driving behavior showed that people tend to drive in a riskier manner when they are alone [36], while collaboration between a driver and a co-driver leads to increase in safety [46]. In this respect, it can be thought that a social entity would make a driver more alert. We believe that, a socially interactive robot could be useful in terms of the increased awareness of a driver.

Researchers have focused on developing robotic interfaces as in-car companions to deliver the necessary information and monitor the driver's state of alertness while interacting with the driver socially [99, 77, 131]. At the same time, the potential benefits of conversational social robots as personal driving agents have been recognized by researchers and car manufacturers [131, 86, 93]. It has been demonstrated that a robotic driving agent is more noticeable, familiar, and acceptable compared to voice-only and display-based driving agents [118]; and also creates a stronger social bond with the driver while transmitting necessary information to them [131].

1.3 Minimal Design Method

Minimal Design Policy is first proposed by Matsumoto et al., who conclude that the robot's appearance should be minimized in its use of anthropomorphic features so that the humans do not overestimate or underestimate the robot's skills [78]. By minimal design, we mean eliminating the non-essential components and keep only the most fundamental functions. We expect that in the future minimally designed robots will be affordable. People will use such minimally-designed robots for many tasks such as cleaning, and here we may mention Roomba the robot [41] or to engage more with autistic children through therapeutic sessions of interaction while cooperating with Keepon the robot [71], etc. Minimal design policy is applied to developmental other robots such as Muu [92], CULOT [64], etc. The simple nature of minimally designed robots allows humans to interact easily with such robots on a daily basis. On the other hand, we must pay attention to sociability and adaptation factors. In fact, interacting with an affordable minimally designed robot may represent the first experience of a human interacting with a robot. This, leads us to assume that people will possibly have high expectations about the robot's adaptive capabilities.

1.4 Developmental Psychology and Epigenetic Robotics

Theories of child cognitive development, such as Vygotsky's "child in society" [73], can offer a framework for constructing robot architecture and social interaction design [27, 26]. Attention is perhaps the most studied aspect of human behavior in developmental robotic systems. Particularly in case of caregiver and robot relationships, focus of attention and all of its secondary aspects form core functionalities for social interaction and, eventually, learning. The robot Kismet was built primarily for studying models of human attention and visual search [16, 15]. This research proposes a minimal functionality research query: what is the minimal interaction functionality required for a robot to be capable of normal social interaction with its caregiver? Using a behavior-based approach with activation thresholds varied over time based on state parameters, Kismet responds to cues while tending towards a homeostatic middle-ground. Thus the same user input, which triggers surprise at first may soon trigger annoyance when repeated.

1.4.1 Attention Manipulation

Shared attention models are inspired directly by those of Baron-Cohen [8]. Separate modules enable theory-of-mind, intentionality, shared attention and eye-direction control. In implementing these modules with a physical robot, the challenges are two-fold: how will the primitive sensors of the robot enable the requisite perception (eye gaze direction, etc.); and how will the modules literally combine to result in an emergent behavior that is significantly richer than the composite parts (e.g imperative pointing).

An epigenetic bridge between embodiment and situatedness based on the conceptual framework of "situated embodiment" [134]. The term Epigenesis refers to individual development through incremental change between levels of competence of increasing complexity, achieved through the interaction with the physical and the socio-cultural environment. The idea of incremental adaptation is a simple, but developmentally and evolutionally appropriate way to ground the human-robot communication in knowledge deriving from our bodies and from our social environment. Major prerequisite of human communication is to be able to understand and manipulate each other's focus of attention—cf.the notion of manifestness in Relevance Theory [112]. The reason for this is that one's focus of attention determines, or at least well predicts, one's future behavior. Therefore one can safely presume that human beings have evolutionally acquired the skill to exchange clues for their attentional focuses in order to be able to cooperate and compete with each other efficiently. The fundamental skill to exchange the focus of attention is the ability to participate in acts of joint attention, whose ontogeny and phylogeny have been studied in developmental psychology [21, 8] and comparative psychology [60]. Joint attention is the activity of sharing each other's attentional focus and the mutual acknowledgment of this fact. Usually one's focus of attention can be derived from one's gaze, face, or pointing direction. Once the infant or robot established joint attention with an agent, they begin to share some information in the environment. We think this is the most primordial form of communication.

1.5 Thesis Contributions

In this thesis, we focus on designing interactions to mediate the relationship between the car and the driver by considering the different level of automation mode of the car. Considering the Level 2 or under (manual driving), the focus and attention of the driver is important. The information should be given to the driver without distracting the them, at the same time, without the deprivation of the pleasantness of the system. Considering the Level 3 (partially automated), the driver's awareness should be assured in order to make the handover the control from autonomous system should be smooth and as fast as possible. At the same time, it is important to assure the perceived safety and enjoyability of the driver in the autonomous system. Finally, the attention of the driver should be manipulated by considering the engagement with the individuals and keep the their situational-awareness towards the environment. In addressing these questions above, this thesis yields these core contributions:

1.5.1 Designing a Social Robot Platform for In-Car Usage

We developed a social interface named NAMIDA that incorporates three conversational robots that can decrease the number of directed utterances towards a driver through a turn-taking process among the robots. Through this model, the driver can gain necessary location-based information without joining the conversation. The results of our pilot study revealed that the proposed multi-party conversation based interaction model is more effective in alleviating certain workload factors for drivers compared to a conventional one-to-one communication based approach that directly addresses the driver. Moreover, an analysis of the attention behaviors of drivers showed that the proposed approach could encourage drivers to focus on the road better than that of a one-to-one communication based system. Finally, the results of a subjective impression showed that the multi-party conversation based system seemed more autonomous and was more animated; it also demonstrated more natural conversation.

1.5.2 Lifelikeness of the System and the Enjoyability and Distraction of Drivers

The applications of conversational robots are gaining popularity due to their potential in providing information while engaging the user in a conversation. However, when the user already is focused attention on a task, engaging them in conversation may be difficult or even risky. Human-robot interaction (HRI) field should consider interaction methods where a conversational robot can keep the human informed but without the obligation of engagement in the conversation and the deprivation of lifelikeness of the interaction. In this study, we discuss this approach within a driving scenario by utilizing a multi-party social robot platform. In our previous study, we showed that a virtually embodied multi-party conversational agents had the ability to decrease the number of directed utterances toward a driver through a turn-taking process among the robots which helped to reduce certain workload of a driver while the agents providing environmental information compared to a conventional one-to-one conversational agent. In our current study, we employed our robotic driving agents and conducted an experiment to investigate the lifelikeness and distractedness of the interaction of the multi-party conversation of three robotic agents and the one-to-one conversation between one robot with a driver. The results of this study revealed that overhearing information from the multi-party conversation of driving agents is perceived as possessing more lifelike characteristics compared to a conventional, one-to-one communication based approach that directly addresses the driver. Also, the proposed approach reduced the distraction level and increased the enjoyment of the drivers.

1.5.3 Response Time of Drivers

The increased automation level creates room for the drivers to shift their attention to non-driving related activities. However, there are cases that cannot be handled by automation where a driver should take over the control. This pilot study investigates a paradigm for keeping the drivers' situation-awareness active during autonomous driving by utilizing a social robot system, NAMIDA. NAMIDA is an interface consisting of three sociable driving agents that can interact with the driver through eye-gaze behaviors. We analyzed the effectiveness of NAMIDA on maintaining the drivers' attention to the road, by evaluating the response time of the drivers to a critical situation on the road. An experiment consisting of a take over scenario was conducted in a dynamic driving simulator. The results showed that existence of NAMIDA significantly reduced the response time of the drivers. However, surprisingly, NAMIDA without eye-gaze behaviors was more effective in reducing the response time than NAMIDA with eye-gaze behaviors. Additionally, the results revealed better subjective impressions for NAMIDA with eye-gaze behaviors.

1.5.4 Perceived Trust and Safety in Autonomous Car

In order to achieve a reliable interaction with the autonomous cars, intersubjectivity should be built between the autonomous car and the human operator where the human will believe that the car possess the same intentions with the human. One critical social cue for human to understand the intentions of others is eye gaze behaviors. This paper proposes an interaction method by utilizing the eye gazing behaviors of a robotic in-car driving agent platform with the purpose of enabling to perceive the autonomous car as a social entity. The results of this study revealed an increase in perception of the intersubjectivity in the case of the usage of the gaze following behaviors of the robots. Also, the proposed interaction method demonstrated that the autonomous system was perceived safer and more enjoyable compared to the condition with not using the robots and the condition with using random gazing behaviors of the robots, respectively. Moreover, a positive correlation has been found between comprehending the intentional stance of the agents and the social presence of the autonomous car.

1.5.5 Adaptive Attention Manipulation

An interaction method with joint attention of a social robot platform for autonomous cars and their human operators is discussed. We developed an in-car driving agent platform that incorporates three robots that can perform eye gaze behaviors to build a joint attention with the human operator to increase the driver's situational-awareness. We built an on-line adaptive mechanism for this purpose. We conducted an experiment to observe the real-time environment engagement of the driver.

Chapter 2

How Multi-Party Conversation Can Become an Effective Interface While Driving

2.1 Introduction

Significant amount of drivers rely on the need of location-based information resources during long hours behind the wheel. Mobile devices are available as an option for many users, yet these devices tend to easily divert a driver's attention and can increase the risk of accident. In-vehicle infotainment (IVI) systems are designed to meet a driver's needs inside a car, but these systems also require attention to initiate and frequent attention to monitor the system. Klauer et al. suggests that the risk of a traffic accident increases exponentially the longer a driver takes their eyes off the road [70].

Some newer generation IVI systems facilitate the driver's ability to pay attention to the road by utilizing Bluetooth and windshield projection as well as gesture and speech recognition technologies [85], [30], [119]. Nevertheless, these systems are still very reactive and the interactions with these technologies are not intuitive enough to alleviate mental workload, therefore distraction is inevitable. Herein, researchers

Figure 2-1: Multi party conversation between NAMIDA robots.

claim that these technologies are not any less dangerous than the devices that require a driver's eyes or hands.

A driver's mental process plays a very critical role during a driving maneuver. Hart et al. have explained that once the mental workload reaches an unacceptable level, driving safety may suffer [48]. As information receiving technology becomes more intelligent and complex, engineers face the challenge of developing a communication channel between drivers and IVI systems that takes into consideration a design with a method of interaction that is more natural and intuitive. Barton et al claimed that the human brain has evolved to be highly adaptive in social interactions therefore, people tend to anthropomorphize the technology [10]. From the drivers' perspective, we believe that it is crucial to interact with an IVI system in such a social, natural and familiar manner to reduce mental workload and create a more sociable environment inside a car.

Recently, some car manufacturers have focused on developing robotic interfaces as in-car companions to deliver the necessary information while interacting with the driver in a social, natural and familiar manner. Pivo is a robotic agent that is a co-pilot that directs the driver and monitors the driver's state of alertness [87]. Carnaby has been envisioned as a driving assistant to make driving safe and fun [99]. Moreover, Quin has been developed to handle location services and driver fatigue detection [77]. Further, with the collaborative work of MIT and Audi, AIDA (Affective Intelligent Driving Agent) can leverage a driver's mobile device and deliver personal, vehicle and city information by speech coupled with expressive body movements [131]. In this research, it was determined that AIDA, as an expressive robot and a static-mounted agent, could decrease the mental workload of a driver and prevent distractions as opposed to a mobile phone. This shows that using an affective driving agent is useful for the purpose of decreasing the mental workload. Nevertheless, in the type of communication that occurs between two parties (AIDA and driver), the only interlocutor AIDA system setup is the driver. In other words, the driver needs to maintain constant interaction with AIDA and take on the burden of managing and sustaining the conversation/interaction (e.g. asking questions of and responding to the system) for the purpose of acquiring the requested information. We believe that a one-to-one conversation approach cannot diminish the mental workload of the driver sufficiently for the above reasons.

In this study, we propose NAMIDA as an in-car social interface. NAMIDA designed to be located on the dashboard of a car within the peripheral vision of the driver (Fig 4-1). NAMIDA system consists of three sociable robots that can perform multi-party conversation which can help to alleviate the driver's hearership burden by allocating the conversational overload among the robots. Thence, the instances of directed utterances toward the driver can be diminished, and they can obtain the necessary information exclusively by listening to the conversation with less distraction. NAMIDA conducts context-aware interaction to provide location-based information within the conversation. During the conversation, the robots perform a persuasive utterance by employing turn initials and hedges using an informal polite language. This leads the system being intuitively comprehensible by providing a causal conversational ambiance. In our study, we examined the effects of multi-party conversation on the workload and attentional behaviors of drivers as well as the subjective assessments to evaluate the proposed communication method. We initially aimed to evaluate this model through a pilot study by employing virtually embodied social agents as NAMIDA robots.

The present work explains the concept of NAMIDA in detail in the next section "Concept of the NAMIDA." In the part "Design of the System", we explain the appearance, utterance generation and conversation structure of NAMIDA. We explain our experimental design in the "Experimental Protocol" section. In "Results", we evaluate our results and in the section "Discussion", we provide brief discussion about the results from different aspects. Finally, in the "Conclusion & Future Work" part, we summarize our research and propose plans for our future study.

2.2 Concept of the NAMIDA

The overall processing architecture of NAMIDA consists of context-aware interaction, multi-party conversation that includes role changing, utterance components (TCU, TRP, turn initials and hedges) and the non-verbal behaviours (eye gaze and head movements) in order to achieve a persuasive interaction with a driver (Fig. 2-2).

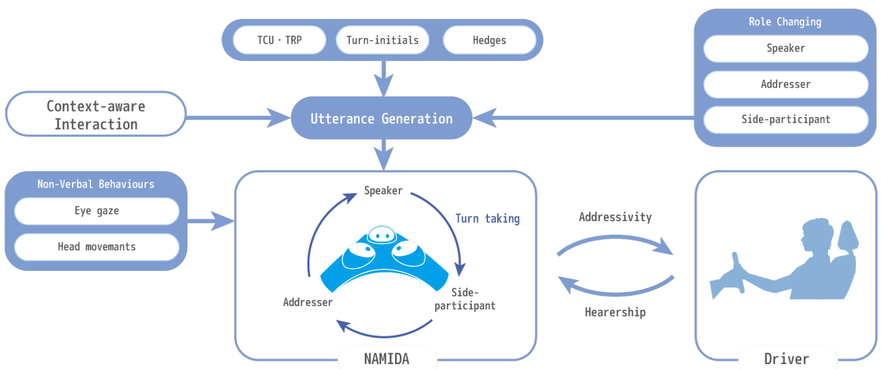

Figure 2-2: Depiction of the overall processing architecture for NAMIDA. Utterance generation utilizes a turn-taking process, using turn-initial elements and hedges and taking into account hearership and addressivity status for a multi-party conversation.

2.2.1 Context-Aware Interaction

In order to provide efficient and safe driving, possessing the right information at the right time is very important. It is also crucial to improve the driving behaviors that emerge from the intricate interaction between the driver, vehicle and the environment for the driver's decision-making process. Rakotonirainy et al. claims that context-aware systems can improve the driver's handling of a car by augmenting the awareness of the cars' state (e.g. following distance), the environment (e.g. location based information) and the physiological and psychological states of the driver (e.g. current attention level) [102]. Context- aware systems use Information Communication Technologies (ICT) to provide a greater awareness of relevant information about the physical world in order to assist the information recipient in the decision making process. However, this information requires the allocation of attention for it to register, and registering information cognitively is not an effortless task. Therefore, a system should be designed by considering the mental resources (short and long-term memory) of driver as well as having the capability of satisfying the driver's requirements through a social and enjoyable environment.

2.2.2 Multi-Party Conversation

The conversational structure of a driving agent system should be very smooth, well designed and aimed at reducing workload. Bakhtin discussed a persuasive conversation structure through analyzing the relationship between the hearer and addresser in the state of hearership and addressivity [6]. In a one-to-one communication modality, when the system directs individual words towards to the user (addressivity), the user is compelled to react to the addresser through a verbal or non-verbal channel (hearership), which creates a conversational burden and mental workload for the user.

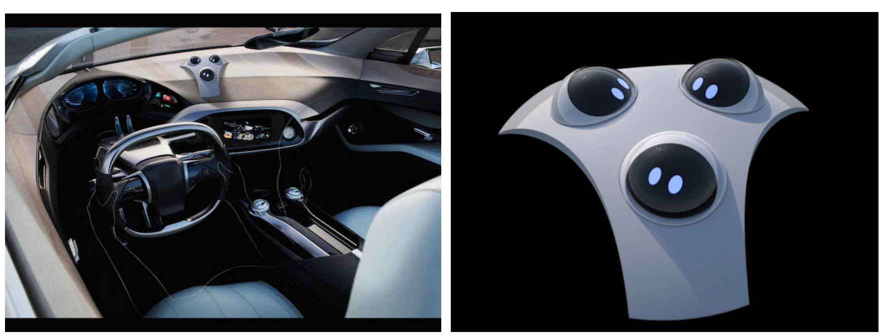

Figure 2-3: Base unit of NAMIDA, designed to be secured on the dashboard of the car within the peripheral vision of the driver *(left)*. With movable heads and eyes, NAMIDA projects a life-like appearance *(right)*.

Yoshiike et al. studied a system called MAWARI, presents in three social robots who can conduct multi-party conversations as an interactive social medium [132]. With the socially-designed interface of MAWARI, the user just listens to the conversation among the robots to obtain broadcasted news. The results of this study showed that when the user is mentally busy, three MAWARI robots (multi-party conversation based) could reduce the workload on the user through their communication modality while performing friendly and sociable conversation, compared to a system using only one MAWARI robot (one-to-one communication based). Another benefit of multi-party conversation has been claimed by Suzuki et al. which is one that includes different personalities, giving the user an opportunity to obtain information from different aspects [116]. On the other hand, the one-to-one communication method not only requires active involvement of the subject in the conversation, but also only allows one-sided individual information, which limits the scope of the conversation.

Moreover, Todo et al. asserts that the multi-party conversation setup is superior and leads to an improvement in user satisfaction with the following benefits of: (1) the conversation becoming more lively, (2) various interactive controls are made possible (all information can be shared among agents), and (3) more applications of a speech dialog system can be considered [122]. Furthermore, Ishizaki et al. discussed that multi-party conversation not only presents new topics/details, but also lessens the stress of conversation initiation [59].

The proposed multi-party conversation model was envisioned to alleviate the conversational burden. This approach not only helps reduce the mental workload for the driver, but it also provides for the different point of views of the participants about a location. For example, a discourse between agents concerning nearby restaurants allows the driver to gain certain details (e.g., the type of restaurants, menus, ratings, etc.) and expedites their decision making process without conversational effort. In addition, when the driver needs to acquire more details or make a different request, they can join and lead the discourse by assuming an active role (e.g., speaker). However, in this study, the driver has been preserved as a bystander, which enables us to evaluate the multi-party conversation utilities at a basic level from the driver's perspective.

Role Changing

In such a multi-party conversation, there are certain roles for the participants and these roles shift according to the verbal or non-verbal behaviors/cues of the participants. Goffman discussed the concept of footing, which explains the participant roles in a conversation [43]. In a more than two-party conversation (multi-party), we can define the main roles as: speaker, addressee and side-participant. The role of a participant when they do not contribute to the conversation becomes that of bystander. Research claimed that when a subject is in a dyadic interaction with a single entity (between a speaker and an addressee), they are heavily forced to interact with the other party [116]. Even the existence of two parties in the same environment (one with an intention and one without an intention to have conversation) loads the conversational role to the other individual who does not have an intention to talk. That is why even when the entity performs a monologue, since the only interlocutor in the environment is the subject, they will be under a conversational burden. Likewise, the presence of two entities will still bring the conversational load to the subject at least as a side-participant. However, the presence of three entities and their conversation within each other yields a situation where the subject can escape the conversational burden as a bystander.

2.2.3 Utterance Components and Non-Verbal Behaviours

Considerable evidence suggests that more spontaneous and natural utterances in a conversation help the agents to be more persuasive which contributes to engagement between the agent and the subject. Research showed that direct commands should be avoided in order to avoid a negative impression of the agent by the subject; instead, usage of hedges and discourse markers can soften the conversation and will be perceived as more natural and polite [126]. Non-verbal behaviors of an agent are also crucial to achieve comprehensive communication. Goodwin asserts that eye gaze behaviors are very important in human-human communication in conveying the message and determining conversational roles [45]. Therefore, in the area of HRI, it has become an essential task to develop eye gaze cues to enhance human-robot interaction [83]. In this sense, we employed hedges and turn-initials as components of verbal behaviors and eye gaze movements as a component of non-verbal behaviors of NAMIDA to make multi-party conversation more natural and cohesive.

2.3 Design of NAMIDA

We implemented NAMIDA as virtual, embodied social agents contained in a small display. NAMIDA consists of one base unit that attaches to the dashboard of a car, containing three movable robots with one degree of freedom each (Fig 4-1). We assumed that this design could reduce the number of modalities involved in a conversation and thereby decrease the potential of shifting the attention of the driver that might be caused by various movements of the system.

The NAMIDA system is located in the driver's peripheral vision. In this manner, the minimal design of NAMIDA attempts to minimize the appearance of the agents' competence being overestimated or underestimated by the human user ([78], [9]). The round-shape display of NAMIDA allows for the positioning of their eyes. We used three different discernible colors (red, green and blue) for composing the eyes, and used varied voices of the three to imply that each NAMIDA character is a different personality.

2.3.1 Utterance Generation Mechanism

Recently, research in human-robot interaction (HRI) has tackled not only the conveyance of content to users, but also the style of the transmission of information [2], [126]. Psychological studies suggest that the appearance and behavior of machines have the potential to influence human perception and behavior toward machines [81], [20], [52], [80]. According to Leech, the behaviors that allow humans to engage in social interactions in a relatively harmonious atmosphere can be defined as politeness [74]. A more thorough concept of politeness is described in the work of Brown et al. as; "regressive action taken to counter-balance the disruptive effect of face-

TCU/TRP	Non-verbal behaviours (NVB)	Turn-initials	Hedges
TCU	NVB1: Eye gaze towards addressee and side participant.	TI1:"a-a", TI2:"ano-", TI3:"anone", TI4:"anosa", TI5:"e-tto", TI6:"e-ttone", TI7:"etto", TI8:"etto-", TI9:"ne-ne", TI10:"ntto", TI11:"nttone"	
TRP	NVB2: Eye gaze towards addressee		H1:"ne", H2:"kedo", H3:"tte", H4:"ka"

Table 2.1: NAMIDA utterances coupled with nonverbal behaviors (turn-initial elements, hedges and nonverbal behaviors indicated).

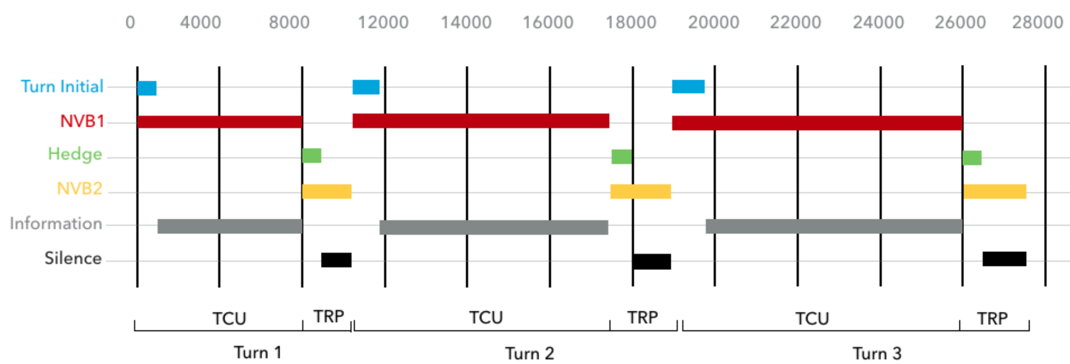

Figure 2-4: Figure depicts the utterance generation and non-verbal interaction design of the multi-party conversational agents in a time-scale base. The utterance of the speaking agent starts with a turn initial and ends with a hedge within a one turn.

threatening acts" [19]. According to this concept, FTAs are, "the acts that infringe on the hearers' need to maintain their self-esteem, and be respected." For example, an alerting utterance of a driving agent like "Turn your lights on!" does not minimize the chance that the hearer will take offense. Since driving can be a stressful act that can put people on edge, using a model that minimizes offensive statements and relieves stress by using polite utterances, even though not directed at the driver, has significant importance. In this regard, the communication design of a driving agent system should be designed around a politeness approach to elicit positive behaviors from the driver. In order to apply a polite-utterance model, we followed the linguistic cues described by Itani [61]. Another important point is to maintain the transitions between the utterances. Ford and Thompson suggest that humans employ turn-initials for changing direction, error handling and enhancement to maintain the liveliness of a conversation [40]. In our study, the utterances of NAMIDA emerged from informal, polite Japanese language, utilizing the turn-initials and hedges (shown in Table 2.1) which are used randomly within each utterance. In order to generate utterances, we employed Wizard Voice (ATR-Promotions) as a voice synthesis engine.

The persuasive behaviors of agents may function as a tool to induce changes in human behavior [47]. Researches showed that merging the utterance mechanism with bodily movements (head rotating and eye gazing) leads to a persuasive impression (as an example [91]). Since NAMIDA is located in the driver's peripheral vision, the persuasive acts of the system could imitate a more natural and sociable communication that would facilitate to understand the conversation flow, therefore the given information. To ensure that kind of communication design, we have integrated some non-verbal social cues such as eye gazing and orienting the head towards the speaker or addressee based on the conversation phase. While the speaker generates an utterance, it directs its eye gaze towards the hearers (both of the agents, which are addressee and side-participant, or just one agent, which is the addressee). Accordingly, the hearers (addresser and side-participant) incline their heads towards the speaker.

2.3.2 Conversation Structure

During everyday informal conversations, the role of speaker, addresser and other participants intuitively alternates on turn-taking bases. [106] introduced the components in a turn-taking system: 1) a Turn Construction Unit (TCU) defines an utterance as a whole turn, 2) a Transition Relevance Place (TRP) corresponds to the end of a TCU where the turn could legitimately pass from one speaker to another, and 3) a Turn Allocation Component (TAC) describes how the next turn is allocated among the participants (by the current speaker's selection or self-selection).

We built a conversational structure based on the above pattern. In order to emphasize the change of direction, dummy error-handling of the conversation and the lessening of FTAs, the speaker conducts turn-initials within the utterances. Also, for the purpose of indicating the TRPs at the end of each TCU and softening the utterance, the speaker chooses a hedge, which is shown in Table 2.1. In this way, it becomes easy for the driver to recognize when the speaker will be able to start or end the turn in each TCU. We followed such a strategy to make the conversation turns perceived as natural, and also to take into consideration the driver joining the conversation based on the turn-taking system for future implementations.

During a TCU, the speaking agent starts its utterance with a turn initial, then continues with giving information (Fig. 2-4). Meanwhile it directs its eye gaze mainly at the addresser while giving a short glance to the side-participant. When the speaking agent finishes uttering the information, it ends its utterance with a hedge and directes its eye gaze towards the addresee (TRP). Following a TRP, TAC occurs in two ways: the current speaker may select the next speaker by directing its eye gaze or self-selection occurs. For example, when there is silence in a conversation, the side-participant selects itself as the next speaker, takes the conversational burden and sustains the conversation. In a one-to-one conversation model, this silence duration can emerge as a conversational burden for a subject.

2.4 Experimental Protocol

This pilot study focuses mainly on exploring the effectiveness of our multi-party conversation approach on mental workload and the attention behavior of drivers as well as their subjective impressions.

We set up an experiment with two conditions, one-to-one communication based NAMIDA (OOCN) and multi-party conversation based NAMIDA (MPCN). Each participant performed mock-driving routines with a projected driving simulation while communicating with each NAMIDA setting. In the simulation, the participants were in unfamiliar streets containing restaurants, skyscrapers, exhibition halls, shopping malls, and ordinary houses. In each condition, NAMIDA agents introduce the environment to the driver. We employed the same verbal and non-verbal patterns for both NAMIDA settings (Table. 2.1). However, in the OOCN case, there is no side-participant, so that during a TCU, the speaking agent directs its eye-gaze only the addressee. The turn-taking occurres only in the MPCN condition among the three agents by allocating the script among them which creates their animated behaviors. In the OOCN condition, the location-based information content was uttered with less utterances through one agent whose head moves randomly to create the animated behaviors. The symbol (. . .) in the extracts below represents a pause of about 1.0-1.5 seconds.

2.4.1 Condition1: One-to-one Conversation Based NAMIDA (OOCN)

In this condition, a participant is always the "addressee" and receives the relevant, location-based information from OOCN while the NAMIDA agent is always the "speaker". Under this condition, the speaker express animated behaviors by directing its eye gaze and head (non-verbal behaviors), also utterances (verbal behaviors) towards the participant. In order to create an animated behaviour for OOCN, we implemented a series of movements as directing eye gaze towards slight right, slight left and front (facing to driver) synchronized with the utterances (Fig. 2-5). The conversational turn does not change in terms of the participants' roles. Below is an extract from a conversation under the OOCN condition:

```
1 Turn 1   N1:  [TCU I think, this is
2    a very nice street.]  [TRP](...)
3 Turn 2   N1:  [TCU There
4    must be
5    an old temple
6    around here.]  [TRP]
7 Turn 3   N1:[TCU This is amazing
8    isn't it?][TRP] (...)
```

2.4.2 Condition2: Multi-party Conversation Based NAMIDA (MPCN)

In this condition, a participant is always a "bystander" and receives the location-based information from the MPCN as an overhearer. During a turn changes in the conversation, the participant's roles and non-verbal behaviors (eye gaze and head directions) change animatedly as well.

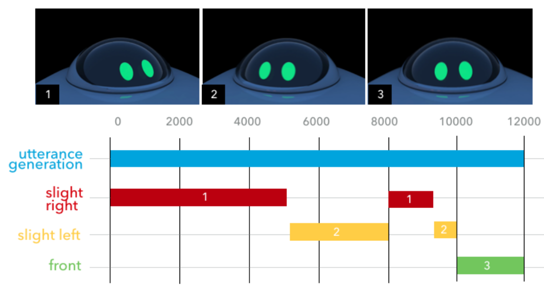

Figure 2-5: Figure depicts one turn of the conversation in the Condition 1.

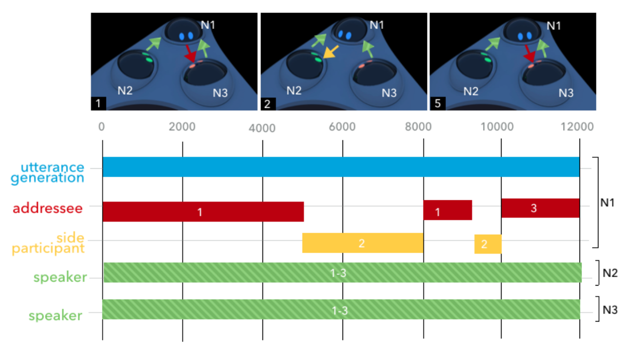

Figure 2-6: Figure depicts one turn of the conversation in Condition 2.

In order to create persuasive animated behaviours for MPCN, we implemented a series of movements for all of the agents. In a turn, when the speaking agent generates its utterances, it directs its eye gaze towards the addressee and the side-participant, meanwhile these two agents direct their eye gaze towards the speaking agent (Fig. 2-6). Below is an extract from a conversation under the MPCN condition with four turns:

```
1 Turn1   N1:  [TCU I think, this is
```

```
2    a very nice street.]  [TRP](...)
3 Turn2   N3:  [TCU Oh, really?]  [TRP]
4 Turn3   N2:  [TCU Yes.][TRP]
5    [TCU There,
6    must be
7    an old temple
8    around here.]  [TRP]
9 Turn4   N3:[TCU This is amazing
10    isn't it?][TRP](...)
```

In this example, in ***Turn1***, ***N1*** is the speaker, ***N2*** is the addresser and ***N3*** is side-participant however; in ***Turn2***, after the silence ***N3*** becomes the speaker, ***N1*** is the addresser and ***N2*** is a side-participant.

2.4.3 Experiment Setup

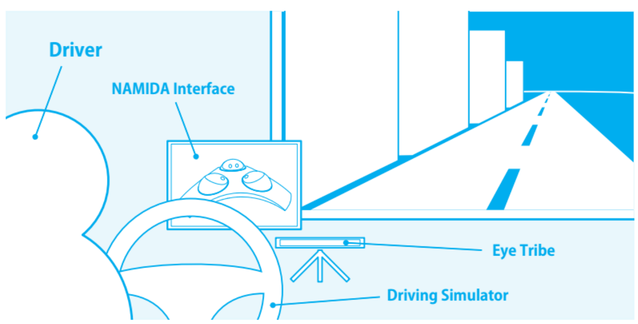

Figure 2-7: The setup of the experiment. The driver goes along the road while listening to the conversation of NAMIDA.

We set a mock driving environment as it is shown in Fig.2-7. We developed a simulated road environment (in Unity 3D [version 5.2]) with many buildings (e.g. shopping mall, restaurants, mansions, etc.). In the experiment, each participant performed a mock-driving routine by watching a projected driving simulation on a big wall. The NAMIDA interface was displayed as virtual, embodied agents, which were placed on the left side of the dashboard. On behalf of a context-aware system, we prepared scripts for each agent corresponding to the road rotation. The agents started to utter the scripted lines one minute before arriving at each destination spot. We set Eye Tribe gaze tracker in front of the driver in order to track the eye gaze of the each participants to be able understand their attentional focus during the experimental sessions.

In total, 14 Japanese participants (3 female and 11 male) of ages varying from 20 to 35 years old (average age of 23.15) took part in the experiment. Since the interaction between the participants and the system was limited, we kept each session to approximately 5 minutes in length in order to maintain a high level of concentration of the participants. All participants had a driving license. We divided the participants in half: one group from participants completed the experiment starting first with Condition 1 and then Condition 2; the other half completed the experiment starting with Condition 2 and then Condition 1. Such a strategy was useful in acquiring a counterbalance in the data, thereby reducing the effect of trial sequence in the results.

Upon arrival, each participant was given an orientation about the experiment and their task. The participants were asked to memorize the content of the conversation that involves information about nearby places while they are driving along the simulated road. The information provided by the agents' conversation was such as: "There is a nice Italian restaurant on the right side. Today their special menu is tomato sousse spaghetti.", "The building on the left is a big shopping mall. They also have an IMAX cinema inside.", etc. With this strategy, each participant had to pay attention to the conversation during the driving activity, which was effective in obtaining data on the attention variation of the participants. Also, in a real life driving case, drivers would like to remember the new places they have seen that might be interesting to visit afterwards. We expected that in the MPCN condition, the participants would recall the information better than the OOCN case. At the end of the each session, participants are given five questions about their recollection of the conversation under both conditions.

Dimention	Endpoints	Questions	Descriptions
Attention demand	Low/High	How do you rate the global attention required during the test with regard to what you usually feel while driving?	To evaluate the attention required by the activity to think about, to decide, to choose, to look for
Visual demand	Low/High	How do you rate the visual demand required during the test with regard to what you usually feel while driving?	To evaluate the visual demand necessary for the activity
Auditory demand	Low/High	How do you rate the auditory demand required during the test with regard to what you usually feel while driving?	To evaluate the auditory demand necessary for the activity
Temporal demand	Low/High	How do you rate the pressure related to the time available to run the whole activity during the test with regard to what you usually feel while driving?	To evaluate the specific constraint owing to timing demand when running the activity
Situation Stress	Low/High	How do you rate the stress required during the test with regard to what you usually feel while driving?	To evaluate the level of constraints/stress while conducting the activity such as fatigue, insecure feeling, irritation, discouragement and so on

Table 2.2: Factors of DALI, based on the context and the associated questionnaire with their description.

2.5 Results

We measured the workload of the participants objectively and subjectively. As an objective approach, we recorded the driver's eye-gaze behaviors with the Eye Tribe Tracker in order to measure the attention behavior of each participant. As a subjective approach, we employed a Driving Activity Load Index (DALI) in which the participants were required to answer five questions related to the five demands of mental workload and six questions related to their impression on the each NAMIDA system [98]. Each question had scale of 1 to 5 to rank participant opinion. We also evaluated the subjective impressions on both systems through a questionnaire includes six questions regarding the system's human-likeness, likability of the interaction, animacy, friendliness, persuasiveness and the sense of spontaneity of each NAMIDA system.

2.5.1 Workload Factors

In order to evaluate the mental workload, after each experimental session, the participants received the DALI questionnaire consisting of five questions, corresponding to the five DALI factors. Then we applied a paired t-test to determine if there was a statistical difference between the MPCN and OOCN cases.

Driving Activity Load Index (DALI)

In order to evaluate/compare the mental workload of the subjects, we employed a Subjective Workload Assessment Technique (SWAT). This kind of method consists of evaluating the driver's own judgment about the workload they experienced. DALI (Driving Activity Load Index) is a SWAT technique, which was proposed by [98] as a revised version of the NASA-TLX ([48]) and adapted to driving tasks. Since workload is multidimensional and depends on the type of loading task, there is a scale rating procedure for six predefined factors in terms of perceptive (attention, interference and stress demands), cognitive (visual and auditory demands) and temporal components (temporal demand) followed by a weighting procedure in order to combine the six individual scales into a global score (see Table. 2.2). However, in our study, we used five factors, excluding the interference factor because this factor is most suitable only when it is used in a real driving environment. One of the main advantages of DALI is the possibility to identify the origins of the driver's workload and allow for improvement of the proposed system at this identified level.

Each DALI factor has been calculated based on the subjects' ratings according to the work of [98].

Attention Demand: 88% and 85% of the participants answered the memorization questions correctly for the MPCN and OOCN, respectively. The participants' recall of the information showed no significant difference across the two conditions ($p=0.409>0.05$). However, the t-test for attention demand of the DALI revealed a significant difference ($t(13)=2.10$, $P<0.05$, significant) (Fig. 2-8). According to these results, the OOCN required more attention than the MPCN with regard to remem-

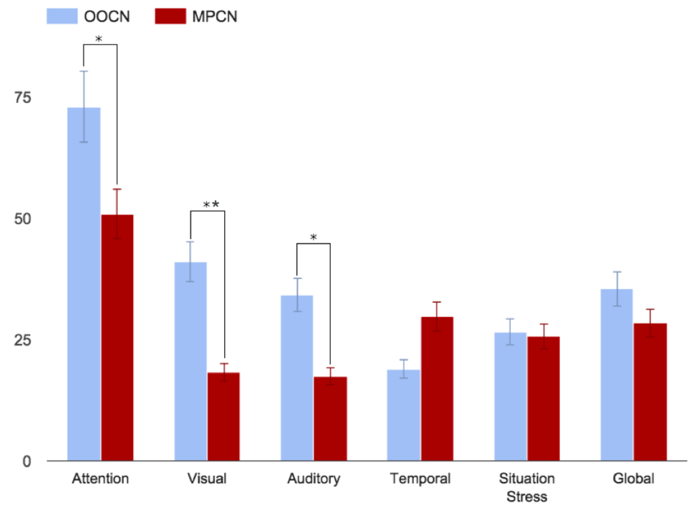

Figure 2-8: The figure depicts the results of the DALI factors under the OOCN and MPCN cases. (*:p<.05, **:p<.01)

bering the presented information.

Visual Demand: There was a significant difference in visual demand (t(13)=2.86, P=0.009<0.01, highly significant) (Fig. 2-8). This reveals that the participants had to exert more visual effort for the OOCN as compared with the MPCN because of the directed utterances by the interface.

Auditory Demand: There was a significant difference in auditory demand (t(13)=1.83, P=0.0449<0.05, significant) (Fig. 2-8). These results indicate that the participants allocated less auditory effort when listening to the MPCN as compared with the OOCN. This was because the driver was excluded from the conversation, yet could still listen and discern the presented information.

Temporal Demand: We found a relatively high, yet non-significant difference for this demand (t(13)=-1.10, P=0.145>0.05, non-significant). This may be because of more utterance generated by the MPCN compared with the OOCN during the same period of time (Fig. 2-8).

Situation Stress: This factor also does not show a significant difference (t(13), P=0.45992>0.05, non-significant) (Fig. 2-8). This may be because the experiments were conducted in a mock driving environment rather than a more realistic driving simulation or in a real-world environment.

Global Value: Overall, the global value didn't show significant difference between the MPCN and the OOCN (t(13) = 0.81, P=0.224>0.05, non-significant) (Fig. 2-8). We can claim that this is because of the non-significant results of temporal and situation stress demands, and the relatively higher rate of temporal demand for the MPCN case.

From Fig. 2-8, we can see that DALI's attention, visual and auditory demands showed significant differences while the differences were non-significant when considering temporal and situational stress. We can infer that the MPCN required significantly less attentional, visual and auditory efforts from participants than with the OOCN. A mock driving environment and higher utterance generation of the MPCN may have had an effect on the temporal and situation stress demands. This circumstance can be the reason for observing the non-significance found in the global value.

2.5.2 Attention Behavior of Driver

The eye gaze movements of a person provide significant cues about their attentional behaviors ([42], [38], [58]). During a driving activity, the eye gaze of a driver should be on the road as much as possible. However, the eye gaze behaviors can be easily altered with a one-to-one communication-based driving agent system whose utterances directly address the driver and put them under a conversational burden. We hypoth-

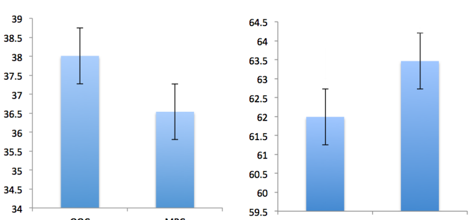

Figure 2-9: Results show the percentage of the collected eye gaze data on the NAMIDA system (*left*) and on the simulated road (*right*).

esized that a multi-party conversation-based driving agent system would require less attention from the driver, consequently the eye-gaze movements of the driver would mostly focus on the road. In order to analyze visual demand allocation, we tracked the participants' eye-gaze movements between the driving simulation screen and the

NAMIDA system during the each session of the experiment.

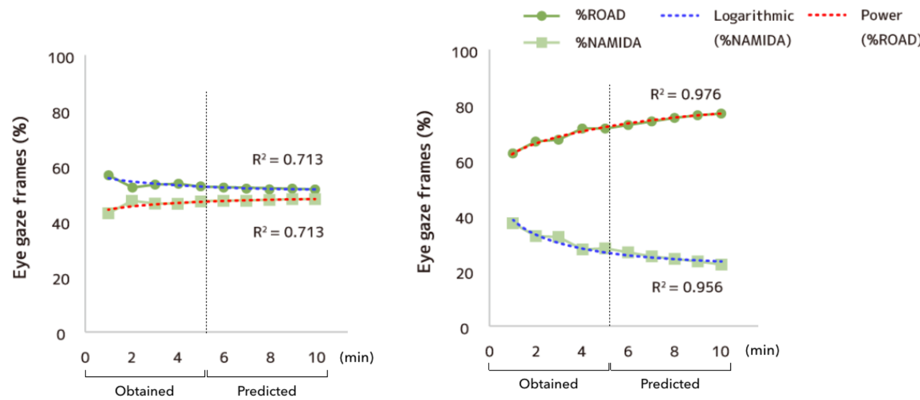

Figure 2-10: Figure shows the trendline of the eye gaze behaviours in OOCN *(left)* and MPCN *(right)* conditions. The part with first five minutes shows the eye gaze data collected during the experiment *(Obtained)*. The second part shows the predicted eye gaze behaviours for another five minutes *(Predicted)* by using the Obtained data.

Eye-gaze tracking enabled us to observe and evaluate participants' attention during the experiments, objectively and non-intrusively. We gathered dynamic interaction data via the Eye Tribe tracking tool. In the experiment room, because the lights were off and only the NAMIDA screen and the simulation screen were emitting light; we divided the attention region into two different areas and then counted the number of the driver's attention frames for the NAMIDA interface and also on the simulated road. The eye gaze tracking system allowed us to capture approximately 30 frames/second. Each frame was represented by a pair of (x, y) coordinates given on the simulation screen with instant-time information (hour/minute/second). Since the NAMIDA interface and the driving simulation were located on different screens, while the participant's attention was on the NAMIDA, the (x, y) data values are represented as (0, 0). For each session (OOCN and MPCN), we acquired a significant amount of eye gaze data.

We calculated the rate of participants' attention (eye gaze position) on the simulated road and on the NAMIDA, separately, for each session by utilizing the collected eye-gaze-position data (Fig. 2-9). According to the results, participants could pay relatively more attention to the simulated road (63.46%) during the MPCN session rather than the OOCN condition (61.98%). The eye gaze data also showed that, the participants exert comparatively more attention to the NAMIDA under the OOCN condition (38.01%) rather than the MPCN condition (36.53%).

According to these results, however, we couldn't observe significant difference on participants' attention between two conditions. We expected that the directed utterances towards the driver in the one-to-one communication system would cause the driver to lose attention to the road significantly more often than in a multi-party conversation-based system that would divert the utterances and allocate the conversation burden among the other participants inside the conversation. However, the results couldn't validate our hypothesis on this point.

Trend line Analysis

Research in HRI suffers from the limited time of interactions during experiments. It may not be possible to predict relatively longer interactions from shorter-timed experiment extrapolation. However, a trend line analysis can help to identify the trends of user behaviors and forecast future interactions. For the purpose of interpreting the attention behaviors of the participants in our study, we applied a trend line analysis and then fit the user data in a mathematical model to estimate the future tendency of behaviors, using a regression model.

Within the trend line analysis, with the OOCN case, we obtained a negative linear regression with a slope of $m=-0.751$ and a coefficient of determination of $R^2=0.424$, on the road, and a positive regression with a slope of $m=0.751$ and a coefficient of determination of $R^2=0.424$, on NAMIDA. The first five minutes shown in the Fig. 2-10*(left)* depicts the decline and incline of eye gaze behavior during five minutes of the OOCN experiment. The coefficient of determination, equal to *0.424*, on both the road and NAMIDA, indicates that about *48%* of the variation in eye gaze data can be explained by the participants reducing their attention on the road while increasing their attention on NAMIDA, in the OOCN case, over time. This would be considered a good fit to the data in the sense that it would substantially improve the ability to predict the eye gaze behavior of the participants.

On the other hand, with the MPCN case, we obtained a negative linear regression with a slope of *m=-2.238* and a coefficient of determination of R^2=*0.897* , on NAMIDA, and a positive regression with a slope of *m=2.238* and a coefficient of determination of R^2=*0.897*, on the road. The first five minutes shown in the, Fig. 2-10*(right)* depicts the decline and incline of eye gaze behavior during five minutes of the MPCN experiment. The coefficient of determination, equal to *0.897*, on both the road and NAMIDA, indicates that about *89%* of the variation in eye gaze data can be explained by the participants reducing their attention on NAMIDA while increasing their attention on the road, in the MPCN case, over time. This result can be considered a very good fit to our data in the sense that it would substantially improve the ability to predict the eye gaze behavior of the participants.

In order to predict eye gaze behavior, we fit our results into a statistical model. For the OOCN case, we obtained a power regression, for the road, with a coefficient of determination of R^2=*0.713* and a logarithmic regression, on NAMIDA, with a coefficient of determination of R^2=*0.713*. The last five minutes shown in the Fig. 2-10*(left)* depicts the declining and inclining attention behavior trend for the road and the NAMIDA system over a five-minute period. Since both the power regression and logarithmic regression models show high correlation coefficients (R^2), these models can be taken as good predictors in explaining the future trend of variation in eye gaze behaviors.

In the MPCN case, we obtained a logarithmic regression with a coefficient of determination of R^2=*0.972*, for the road, and a Power Regression with a coefficient of determination of R^2=*0.956* for NAMIDA. The last five minutes shown in the Fig. 2-10*(right)* depicts the inclining and declining attention behavior trend on the road and the NAMIDA system over a five-minute period. Since both the power regression and logarithmic regression models show high correlation coefficients, these models can be taken as good predictors in explaining the future trend of variation in eye gaze behaviors.

These results revealed that the subjects tend to exert more attentional behavior on road rather than NAMIDA in the MPCN case; while in OOCN case, they tend to devote more attentional behavior on NAMIDA rather than the road. This is an im-

Code	Question	Pair-wise t-test t(d.f.) = t-value P <0.05 (significant)	Result
Q1	Did you feel human-likeness from the conversation?	t(13) = -3.775 P = 0.0008 <0.05	Significant
Q2	How often did you want to interact with the robot(s)?	t(13) = 1.467 P = 0.082 >0.05	Non-significant
Q3	Do you feel that the robot(s) exhibited animacy?	t(13) = -3.964 P = 0.0009 <0.05	Significant
Q4	Did you feel the robot(s) as friend(s)?	t(13) = -0.718 P = 0.2414 >0.05	Non-significant
Q5	Did you feel the robot(s) is/are persuasive?	t(13) = -0.510 P = 0.308 >0.05	Non-significant
Q6	Did you feel the robot(s) conversation was spontaneous?	t(13) = -3.142 P = 0.003 <0.05	Significant

Table 2.3: Questionnaire and t-test results of subjective impression on MPCN and OOCN.

portant finding in determining how the eye gaze behavior of drivers can change during a conversation/interaction with a driving-agent system, especially when distinguishing between cases of directed utterances towards to driver and cases of allocating the utterances among other participants (agents) in a conversation/interaction. These results are also important in revealing how drivers' attention behaviors exhibit a tendency to change in the future.

2.5.3 Measurements of the Subjective Impression

This section focuses on exploring the subjective evaluations of the questionnaire which was designed to evaluate the effectiveness of the proposed communication method. Participants rated the persuasiveness of the utterance, sense of behavior autonomy, authenticity of the conversation, the sense of communication spontaneity, etc. for two NAMIDA systems (MPCN and OOCN). The questionnaire was designed using a 5-point, Likert scale (1=strongly disagree to 5=strongly agree). The analysis considered paired comparisons of each question through a t-test.

The NAMIDA system was initially designed as virtual embodied agents so that persuasiveness was of critical importance. Persuasive agents may function as a tool to induce changes in human behavior. Also, research has shown that the human likeness of such agents influences their effectiveness ([47]). Human likeness often corresponds to anthropomorphism, the attribution of human form, human characteristics or human behavior to non-human objects such as robots, computers, and animals. Thus, we measured and compared the human-likeness rate of our two NAMIDA systems.

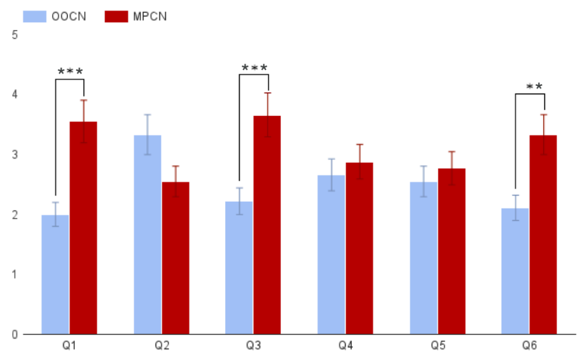

Figure 2-11: Figure shows the mean comparison of subjective impression questionnaire on MPCN and OOCN. (*:p<.05, **:p<.01, ***:p<.001).

Research reveals that lifelike creatures can deeply involve users emotionally and that this involvement can be used to influence users [47]. Heider and Simmel have devoted research to the perceived animacy and intentions of geometric shapes on computer screens [51]. This gradient of "aliveness" is a critical benchmark when comparing robotic systems. It has been reported that the way in which people form positive impressions of robotic systems depends on the visual and vocal behavior of the systems ([23]). That is why we decided to compare the animated behavior of the two systems.

Since computers and thereby robots, in particular, are to some degree treated as social actors ([103]), it can be assumed that people are able to judge robots in a similar way. Moreover, the perceived intelligence of a robotic system will depend on its competence. This affects the desire of a user to interact with a robotic system. In order to expose the differences of two systems we created, we also compared the desire of the participants to use the systems.

Friendliness has an important role in the early development of the relationship ([121]). In order to facilitate the development of the relationships between humans and social robots, these robots should be perceived as being friendly to humans in their interaction. Because of this, it was decided to explore how friendly the systems were perceived by participants by including this aspect in the subjective ratings.

Finally, we wanted to measure how natural the conversations produced by the NAMIDA systems were. Even though we prepared the conversation from the same script in the driving simulation for all participants, we wanted to see which communicative approach would elicit more spontaneous and natural conversation behaviors.

Results of Subjective Assessment

Table 2.3 and Fig. 2-11 shows the significant and non-significant differences for the comparison of the MPCN and OOCN. The MPCN showed a high significant difference for Q1, Q3 and Q6 with p-values of $t(13)=-3.775$, $p=0.0008<0.01$; $t(13)=-3.964$, $p=0.0009<0.01$ and $t(13)=-3.142$, $p=0.003<0.01$ respectively. Due to the MPCN's non-verbal behaviors (e.g., eye gaze and head movement) possessing different agent characteristics and a lively turn-taking mechanism (to lessen the conversational burden and sustain the conversation), the MPCN approach performed better than the OOCN in the areas of sense of behavior autonomy, conversation authenticity and communication spontaneity.

However, we observed non-significant differences for Q2, Q4 and Q5, p-values of $t(13)=1.467$, $p=0.082>0.05$; $t(13)=-0.718$, $p=0.241>0.05$ and $t(13)=-0.510$, $p=0.357>0.05$ respectively. Moreover, the OOCN showed non-significant but slightly high results for Q2 (Table 2.3 and Fig. 2-11). The reason for these results could be that the participants might have felt excluded from the conversation in the MPCN since the participants did not receive eye contact from the agents during the experiment and not involved in the interaction. It seems that the participants would prefer to interact with the system in a way that includes them in the conversation directly.

The results above show that even if the system does not direct utterances towards a driver, the MPCN has certain important aspects that have the potential to elicit positive social effects from the driver.

2.6 Discussion

In the current research, we argued the effects of a multi-party conversation based robotic agent and a one-to-one conversation based robotic agent on drivers' mental workload, attention behavior and subjective impression of the driver. As explained in sections 2.3.1 and 2.3.2, the proposed design of the multi-party conversation system employs well-established turn-taking and role-changing techniques.

2.6.1 Mental Effort of Drivers

We hypothesized that coupled with utterance generation, eye gazing and agent body movements, MPCN will create a more enjoyable, natural and intuitive environment which is easier to follow, thus, reduce certain mental workload demands. Although, we could not observe a significant difference on the global value of DALI, we can infer that in the OOCN case, the participants felt the responsibility of the conversation by the directed utterances from the system, which requires more mental resources, resulting in more attentional behaviors that indicate distraction (e.g., staring at the NAMIDA longer). Fig. 2-8 show the significant and non-significant differences between the mental workload factors observed in the MPCN and OOCN.

The results from the memorization test showed that recalling the given information during the conversation in both experimental cases had a non-significant effect. As in the experiment of [83], the roles of a participant in the conversation did not affect information recall. That is, the significantly higher rate of the attention demand of DALI in the OOCN implies that the participants (as addressers) exerted more attentional effort when memorizing the information. On the other hand, overhearing the same information via the MPCN, required less attentional demand as a cognitive component of the workload.

We claim that the reason for the high rates of attention demand for the DALI, in both cases, can be found in the fact that the NAMIDA interfaces are in the form of embodied social agents rather than physical robotic agents. Since the research proves that a social robotic agent provides more natural and intuitive interaction with humans over that of an embodied agent ([131]), we were not able to obtain as less attention demand value as we would have using a physically developed social robot. This will be involved in the next challenge for this project.

We also observed significant differences between the perceptive components (visual and auditory factors) of the workload. Considering the visual and auditory factors of each session, we observed very low values of these demands that were displayed in the situation where the driver has to memorize the presented information from the MPCN. Taking into account the fact that in both situations, the driver relied on the auditory information coming from the system, and due to having the conversational burden as an addresser in OOCN, the driver was obligated to be in direct interaction with the system which emerged as the visual and auditory efforts of the driver. Since direct interaction by the one-to-one communication required more workload than overhearing by a multi-party conversation as the nature of the communication model, it was relevant to find the highly significant difference for these workload factors as we expected.

The non-significant yet higher rate of temporal demand in the MPCN (Fig. 2-8) can be the reason for more utterance generation occurring in this case than the OOCN in a same period of time. However, we believe that in an unscripted, real-time interaction case, both cases would demonstrate less temporal demand, with MPCN requiring the least. It is because, in a real-time interaction scenario, the participants would take interactive roles: in the MPCN, a participant would take on one of the roles of speaker, addresser, side-participant or by-stander, while in the OOCN case, the role would be only speaker or addresser. Due to the higher conversation responsibility of the OOCN as the research [116] mentioned, participants would feel less temporal demand in the MPCN. We also observed non-significant and relatively low ratings in situational stress demand section (Fig. 2-8). This might be because the experiments were performed in a mock driving environment and the participants were relaxed during the experiments. In a realistic driving simulation, the results might change.

2.6.2 Attentional Focus While Driving

The trend-line analysis results supported the subjective findings for visual demand in the DALI by showing decreasing eye-gaze instances on the NAMIDA system during the MPCN case, and incremental eye-gaze behaviors on the road, unlike with the OOCN (Fig. 2-10). The coefficients of determination for visual attention on the road and the NAMIDA system revealed a good fit to the eye-gaze data such that we could predict the eye-gaze behaviors for the next five minutes.

Therefore, we fit our results into a statistical model to predict the tendency of the interaction based on the user's attention towards to the NAMIDA system. For the OOCN case, we obtained a power regression for the road and a logarithmic regression for NAMIDA. Also, for the MPCN case, we obtained a logarithmic regression for the road and a power regression on NAMIDA (Fig. 2-10). With the high correlation coefficients, these results provide reliable data in predicting the next five minutes of conversation/interaction between the users and the NAMIDA system. We can infer that the MPCN exhibits considerable potential in reducing eye gaze behaviors on an in-car agent system, whilst enhancing attentional focus on the road.

This study has been done by using virtually embodied NAMIDA agents. It has been demonstrated that a robotic driving agent is more noticeable, familiar, and acceptable [118], and also creates a stronger social bond with the driver while transmitting necessary information to them [131] compared to voice-only and display-based driving agents. In this sense, it can be expected that a physically developed robotic driving agent system would draw attention from the drivers in a different level than virtually embodied NAMIDA agents. Moreover, because of the mock driving environment in this study did not reflect the difficulties in real driving environment, the drivers often could find a room to shift their attention towards NAMIDA. We believe that in a more realistic driving environment, the drivers would give much more attention towards the road, correspondingly, we could observe a different result on the attention shifts of the drivers between OOCN and MPCN conditions.

2.6.3 Subjective Impression Towards NAMIDA

According to the subjective impression questionnaire, the MPCN presents significantly more human-like communication, animacy and spontaneous conversation. We can infer that it is because of the multi-party conversation based turn-taking mechanism of MPCN can sustain the conversation without the driver's participation and exhibit a theatrical performance that also entertains the driver. Since human-likeness encourages empathy in a system, with a multi-party conversation approach, subjects are more likely to exhibit positive feelings and implicitly feel more familiar with the system. Moreover, the highly significant difference of animacy in the MPCN reveals that the subjects were able to adapt more to the system. Also, the conversation with the turn-taking mechanism, coupled with the unique, utterance-generation mechanism contributed to the MPCN having a more natural conversation ability. Further, the significantly high rating for spontaneous conversation corresponds to the better, stress-free quality of the MPCN system (Fig. 2-11).

The non-significant differences on Q2, Q4 and Q5 correspond to the degree of cooperation, friendliness and sympathy perceived of the systems, respectively (see Fig. 2-11). These results show that the turn-taking mechanism, the different characteristics of the agents, and the lively, sustainable conversation aspects had no effect on the system's cooperation, friendliness and sympathy aspect. Moreover, the OOCN elicited slightly more cooperation. The reason for this may be that the participants felt the system to be more of a dynamic interaction due to the directed utterances of the OOCN, unlike with the MPCN.

The MPCN system has the potential to be considerably important in eliciting positive social behaviors. Specifically, because our study proposes a novel, human-robot, interaction method, we wanted to replicate the components for the attachment bond. Since we achieved highly significant differences on human-likeness, animacy and natural conversation, our proposed MPCN model shows potential in building a satisfying social bond with the driver.

2.6.4 Limitations

The low number of participants and the recruitment of mostly male subjects limited our results and our ability to make broader generalizations. Ideally, a study with more participants, across a wider age range with greater gender balance, using/not using the conventional in-car navigation system would produce more reliable results in terms of the effects of both systems on the drivers. Because we conducted our experiment with Japanese participants, the cultural context of our study constitutes another limitation: the fact that the Japanese participants are more accustomed to robotic interfaces.

2.7 Conclusion & Future Work

The proposed multi-party conversation based interface of NAMIDA presents a unique interaction between the car and driver. As a social interface, it has been designed to assist drivers by conducting a context-aware interaction during driving. We believe that this conversation approach is enjoyable as it requires less attention in obtaining necessary information. We designed an experiment to verify our hypothesis by comparing two different cases, MPCN and OOCN, in a mock driving environment. In the current research, we examined the mental workload, attention behaviors and subjective impression of drivers by comparing a multi-party conversation-based system with a one-to-one conversation-based system.

We evaluated our proposed system using a DALI questionnaire, a trend analysis of the eye-gaze data gathered during the experiments and a subjective impression questionnaire. The results of DALI revealed that even though the MPCN cannot fulfill all the workload factors, it induced less cognitive and perceptive components of workload. That is, overhearing the location-based information via a conversation between the sociable agents required significantly less attentional, visual and auditory efforts. It has been also shown that, MPCN required less eye gaze behavior during the experimental conditions. The trend analysis demonstrated that our proposed multi-party conversation-based system is promising in reducing the attention behavior on the system over use. Through the turn-taking based lively conversation, the MPCN system exhibited an enjoyable performance such that according to the subjective impression ratings, it had significantly more human-like and animated behaviors, and natural conversation aspects than the OOCN system. For a more enjoyable and sociable environment inside a car, our next study will involve the driver in the multi-party conversation by considering the real-time condition (in behavioural and workload aspect) of the driver. Our future study is also required to generalize the results and investigate the different aspects of the multi-party conversation on drivers during a real-time interaction with physically developed robotic agents.

Chapter 3

Let's Get Ready to Turn!: The Effects of Multi-party Conversation of Driving Agents on Perceived Lifelikeness and Distraction

3.1 Introduction

Conversational social robots have received broad attention in various contexts such as education [24], health care [113], entertainment [35], etc, due to their sociable interaction capabilities. In recent years, the potential benefits of conversational social robots as personal driving agents have been recognized by researchers and car manufacturers [127, 131, 86, 93]. It has been demonstrated that a robotic driving agent is more noticeable, familiar, and acceptable compared to voice-only and display-based driving agents [118]; and also creates a stronger social bond with the driver while transmitting necessary information to them [131]. However, in these studies, the one-to-one interaction between the robot and the driver is based on transmitting information unilaterally (always from robot to driver). Thrun [120] discussed that this kind of unilateral interaction creates a "master-slave" relationship and can be seen in industrial or professional service robots that lack of social interaction abilities. Norman [89] argues that humans are much better at interactions when on an "equal-footing" compared to a "master-slave" relationship. Therefore, when systems get smarter in the personal, sociable robot domain, it is expected that these robots should be designed based on an interaction in which the robot and the person can transmit information bi-directionally, where they can maintain a conversation together.

At the point of incorporating the driver's involvement within a real-time interaction with a robot in a driving environment, there are some essential issues to consider. First, even though the robot expects a response from the driver, because of the low speech-recognition accuracy and insufficient response from the robot in a real-time driving environment, the robot will encounter difficulties in maintaining a conversation. In order to overcome this problem, the conversation of the robot can be scripted in advance with proper utterances. However, it would be difficult for one robot to avoid unexpected responses from the interlocutor. Another problem is that because in a one-to-one communication that occurs between the two parties (the only interlocutor of the robot being the driver), the driver needs to maintain constant interaction with the robot and has to take on the burden of managing and sustaining the conversation (e.g., asking questions or backchanneling) as a result of a natural interaction. Bakhtin ([6]) discussed the conversation structure through analyzing the relationship between the hearer and addresser by considering the state of hearership and addressivity. According to this analysis, in a one-to-one conversation, when the robot directs individual utterances toward the driver (addressivity), they are compelled to react to the addresser through a verbal or non-verbal channel (hearership), which creates a conversational burden for the driver. This conversational burden may cause an increase on the mental workload, therefore, it creates a distraction for the driver. In some cases, in order to avoid any visual distraction, the driver, consciously or unconsciously, may not make an effort (verbal or non-verbal) as the hearer (hearership) while their attention is focused on driving. When the driver stops contributing to the conversation verbally or non-verbally, the interaction between the driver and the robot will lose its naturalness, and the robot will be perceived as a machine-like agent rather than a life-like agent. This situation may not only reduce the conversational engaging capability of the robot, but may also raise the negative feelings toward the system.

Increasing the number of robots and decreasing human involvement in conversation has been demonstrated as a natural and socially acceptable approach to overcome the issues mentioned above. Research has shown that the conversation among multiple robots, based on scripted utterances, could not only avoid problems related to recognition difficulties and an insufficient response repertoire of the system, but also enable users to feel that the conversation is more coherent, enjoyable, lively and natural compared to a single robot conversation ([122, 3, 56, 57, 4]). Hence, the multi-party conversation of robots has been utilized in different concepts by approaching human users through indirect interactions such as those found in on-stage entertainment applications ([49]), and broadcasting information as a passive-social medium in public places ([107, 94]). Observing such persuasive inter-robot conversations does not only endow the system with a more life-like sense, but also relieves the stress of initiating or maintaining a conversation, lessening the conversational burden on the person, helping them keep their attention on their ongoing tasks. The research has indicated that overhearing information from the conversation of multiple, passive-social robots helped users to obtain necessary information while exerting less mental effort in sustaining the conversation when mentally distracted with another task [132]. In a driving context, it is necessary to consider the conversational burden of a driver during an interaction with a driving agent without concession to the lifelikeness of the system. In this sense, we believe that the multi-party conversation of driving agents could help a driver to obtain necessary information (such as navigational directions, road conditions, etc.) with the feeling of less conversational burden and distraction, with the agents possessing a greater sense of lifelikeness, making driving more enjoyable and engaging.

In ([65]), we showed through a pilot study that the virtually embodied driving agents were more effective in encouraging drivers to focus on the road and possessing more autonomous, animated and natural characteristics when they are in a form of

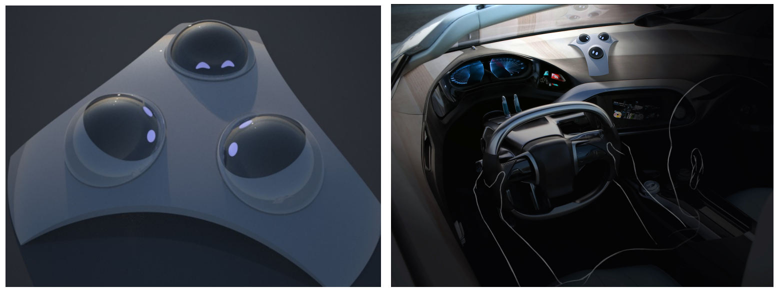

Figure 3-1: The conceptual figures of NAMIDA platform. It consists of one base unit containing three movable heads with one degree of freedom each (*left*). The base unit of NAMIDA that attaches to the dashboard of a car is within the peripheral vision of the driver (*right*).

multi-party conversational agents rather than one-to-one conversational agent. In the current study, we employed our robotic driving agent platform, NAMIDA (Fig. 3-1), within a more realistic driving environment to investigate the effectiveness of the multi-party conversation of driving agents in terms of the lifelikeness, enjoyment, subjective and objective evaluations of the conversational burden of the driver while receiving necessary information by overhearing the conversation; To accomplish these objectives, we conducted an experiment to evaluate the factors above by comparing two forms of the robots' conversation: one-to-one conversation and multi-party conversation. The NAMIDA platform is an in-car social interface consisting of three conversational robots that can perform a multi-party conversation, providing some necessary information to the driver indirectly. The present work explains the concept of the conversation of the robots in the next section: "The Concept of Multi-party Conversation of Driving Agents." In the part, "Design of NAMIDA Platform", we explain the appearance, utterance generation and conversation structure of NAMIDA. We explain our experimental design in the "Method" section. In "Results", we evaluate our results and in the section "Discussion", we provide a brief discussion about the results from different aspects. Finally, in the "Conclusion & Future Work" part, we summarize our research and plans for future study.

3.2 The Concept of Multi-Party Conversation of Driving Agents

Overhearing is defined as an indirect communication where an interlocutor receives information from another when they are not the addressee. Therefore, overhearing does not impose an obligation to maintain the conversation for an individual. Berelson remarked that overhearing information can be more effective in changing the opinion of a listener than deliberately directing the information to them [12]. This argument has been expanded in human-agent interaction studies in which it has been demonstrated that overhearing the conversation of persuasive agents could manipulate the attitudes of users [115, 68], and promote better learning [33]. Likewise, we believe that overhearing information from multi-party conversation-based driving agents could be an effective interaction method in improving driving skills and adopting safer driving behaviors in traffic. However, as an initial step of this study, rather than focusing on driving behaviors, we focused on how individuals would react to and perceive this kind of inter-robot conversation while their attention was on driving. We took this situation in hand in terms of the perceived lifelikeness, enjoyment, and distraction of the system by comparing two forms of the conversational agents: the one-to-one conversation based NAMIDA and the multi-party conversation based NAMIDA.

In order to make the robots' conversation persuasive, the non-verbal behaviors, such as turn-taking comprising the eye gaze behaviors, have significant importances. With regard to one-to-one conversation, when the speaking robot directs its utterances and eye gaze toward the driver, the addressee feels a conversational burden and is compelled to respond to the system verbally and/or non-verbally (Fig 3-2). Under this condition, one possible reaction of the driver can be to exhibit hearership behaviors as such turning their head toward the robot and providing a backchanneling response (Fig 3-2*(left)*). In this condition, they will be distracted from driving visually. Another possible reaction of the driver can be to keep focusing on the road without exhibiting hearership behaviors (they do not respond to the system verbally and/or non-verbally). In this condition, the nature of the one-to-one conversation will

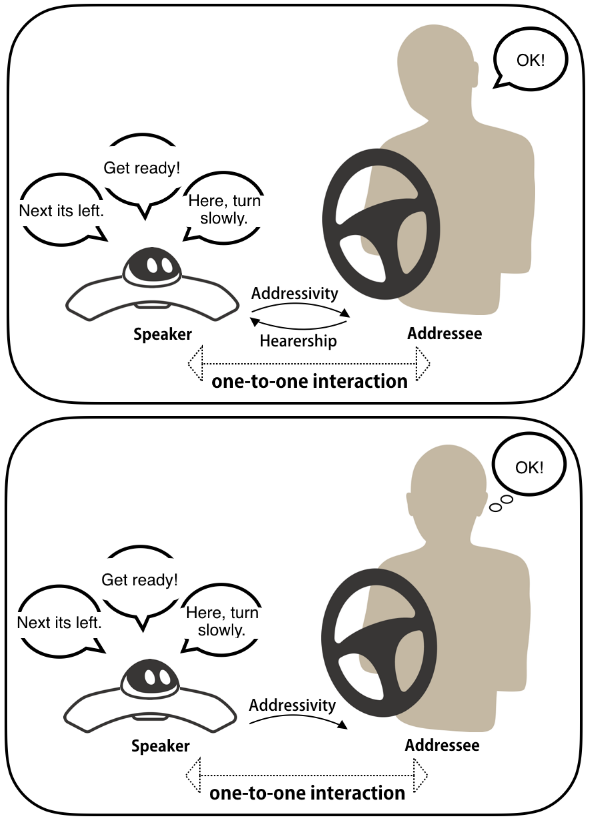

Figure 3-2: The figure depicts two scenarios during one-to-one conversation in which the robot directs its utterances toward the driver (addressivity) that makes them compelled to give a response (hearership). In the first scenario, the driver gives a backchanneling response to the robot *left*. In this case, the driver is visually distracted from driving. In the second condition, the driver keeps focusing on driving without responding to the robot *right*. In this case, the lifelikeness of the system is decreased.

be disrupted; therefore, the robot will lose its persuasiveness and the robot's sense of lifelikeness will be affected negatively (Fig 3-2*(right)*). On the other hand, in the case of the multi-party conversation of robots, the driver will not be imposed upon to contribute to the conversation; therefore, the overhearer can continue to focus on the road while listening to the robots' conversation (Fig. 3-3). Even though the driver will not give a backchanneling response, the robots' interaction will not disrupt the driver's attention in a behavioral manner. Thence, the robots can sustain the conversation within a logical flow, and the driver can obtain the necessary information from their conversation with little effort.

In the multi-party conversation, the robots should exhibit persuasive behaviors to

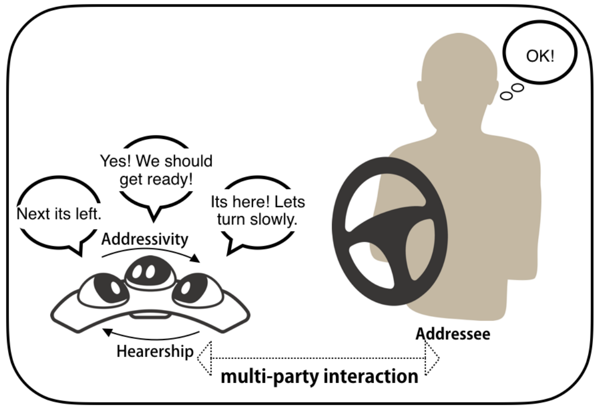

Figure 3-3: In multi-party conversation, the driver is not imposed upon to contribute to the conversation; therefore, they (the overhearer) can continue to focus on the road while listening to the robots' conversation. The robots can sustain the conversation in a life-like manner, and the driver can obtain necessary information from their conversation with little effort.

reflect the nature of the multi-party conversation in a way that the driver will not find strange or unfamiliar when compared to everyday conversation between people. Accordingly, the verbal and non-verbal behaviors of the robots should be consistent, coherent and familiar to the driver. The following section will describe the design factors for verbal and non-verbal behaviors of NAMIDA to create such a perception.

3.3 Design of NAMIDA Platform

3.3.1 Appearance

The NAMIDA platform was built as designed in the previous study [65]. NAMIDA consists of one base unit that attaches to the dashboard of a car, containing three movable heads with one degree of freedom each in the driver's peripheral vision. The round shaped head of each NAMIDA allows for the positioning of their eyes with full color LED light emission. The movement of the robots is enabled by three servo-motors inside each head attached to the main board. The robots has a conversation mechanism that allows them to generate verbal and non-verbal behaviors within each turn-taking.

3.3.2 Conversation Mechanism

In a multi-party conversation, there are certain roles for the individuals in the conversation, and these roles shift according to the verbal or non-verbal behaviors/cues of the individuals. Goffman discussed the concept of footing, which explains the individuals' roles in a conversation [43]. In a conversation involving two or more participants (multi-party), the main roles of the individuals are: speaker, addressee and side-participant. The role of an individual when they do not contribute to the conversation becomes that of overhearer. Research has claimed that when an individual is in a one-to-one interaction with a single entity (between a speaker and an addressee), they are heavily forced to interact with the other party [116]. On the other hand, as one of the key points of the multi-party conversation of robots in being persuasive is smooth turn-taking whereby the driver can be persuaded that the robots are having a conversation among themselves, thus they will not feel a conversational burden and can assume the role as an overhearer in the conversation environment.

We adopted the multi-party conversation approach that consists of nonverbal behaviors, turn-initials and hedges by utilizing the turn-taking components [106] as it has been used in our previous study [66]. Ford and Thompson suggest that humans employ turn-initials for changing direction, error handling and enhancement to maintain the liveliness of a conversation [40]. In our study, the utterances of the robots emerged from informal, polite Japanese language, utilizing the turn-initials and hedges, which are used randomly within each utterance. In order to generate utterances, we employed Wizard Voice (ATR-Promotions) as a voice synthesis engine. To ensure this kind of communication design, we integrated some non-verbal social cues such as eye gazing and orienting the head toward the speaker or addresser based on the conversation phase seen in casual human conversations [66]. In the NAMIDA system, when the speaker robot generates an utterance, it directs its eye gaze toward the hearers (both of the robots, which are addresser and side-participant, or just one robot, which is the addresser). Accordingly, the hearers (addresser and side-participant) incline their heads toward the speaker. Fig. 3-4 represents the structure

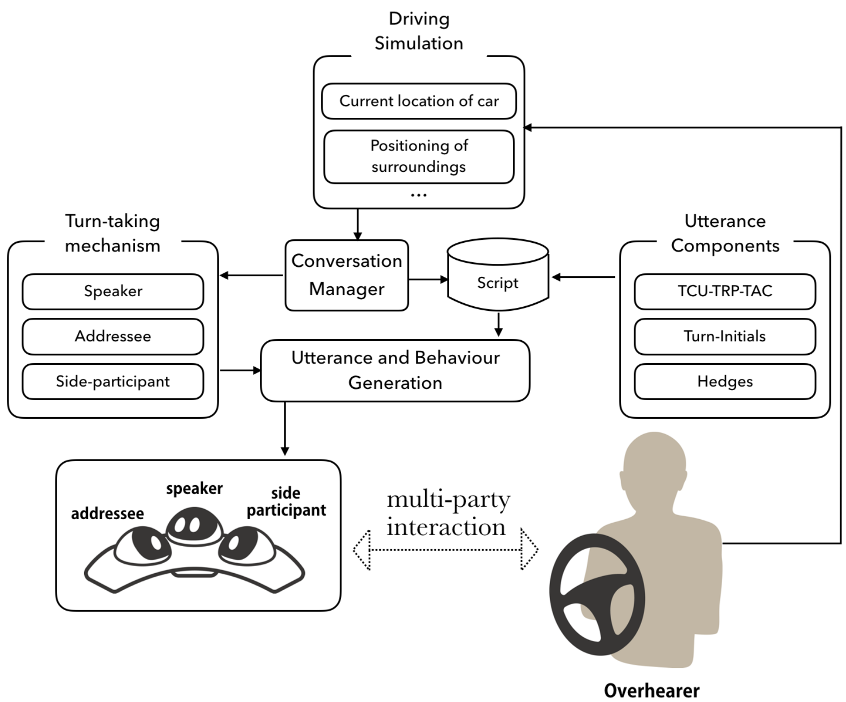

Figure 3-4: The internal structure of NAMIDA and its communication protocol.

of NAMIDA system and the communication protocol among the driving simulation, conversation manager, script, turn-taking mechanism, the utterance and behavior generation of the robots. The script was prepared with the utterance components and is triggered by the conversation manager that is activated when the current location of the car getting closer to the designated situations on the simulated road. The conversation manager controls the turn-taking by assigning conversational roles of the robots (speaker, addressee and side-participant) which change after each conversational turn.

3.4 Method

The multi-party conversational driving agent system that possesses the design aspects mentioned above, is expected to be perceived more positively compared to a one-to-one conversation based driving agent system due to its ability to maintain a conversation without imposing on the driver the burden of contributing to the conversation. In order to achieve this kind of interaction, the robots should exhibit persuasive behaviors, and these behaviors should be familiar and expectable for drivers. The persuasiveness and familiarity of the social robots are correlated with their life-like features. For this reason, in this study, as a first step in evaluating our robotic driving agent platform, we investigated the effects of the multi-party conversation of the driving agents on perceived lifelikeness, distraction and enjoyment from the driver's perspective.

In our previous study [66], we showed that virtually embodied driving agents could help drivers to reduce attentional, visual and auditorial workload factors. In our current study, we also examined the perceived workload factors to explore if we will find similar results with our previous study with our robotic driving agents in a more realistic driving environment.

We conducted the experiment by incorporating a driving simulation with the inclusion of two conditions, employing the NAMIDA platform: one-to-one communication-based NAMIDA (OOCN) and multi-party conversation-based NAMIDA (MPCN). We analyzed our experiment via the evaluation of subjective assessment questionnaires and analyzing the eye gaze behaviors of the participants.

3.4.1 Experimental Conditions

In each condition, the participants drove a simulated car and received information from the OOCN or MPCN platforms. In each condition, the driving agents gave navigation advice to the driver, suggested speed change and gave information about the car's surroundings. We employed the same verbal and non-verbal patterns for both NAMIDA settings. However, in the OOCN case, there was no side-participant, so that the speaking robot directs its eye-gaze only to the addressee. The turn-taking occurred only in the MPCN condition between the three robots by allocating the script among them.

Figure 3-5: The figure depicts the OOCN condition from the driver's point of view. In the OOCN condition, the robot directs its eye gaze and utterances toward the driver. The driver is the *addressee*.

Figure 3-6: The figure depicts the MPCN condition from the driver's point of view. In the MPCN condition, the robots conduct a multi-party conversation through turn-taking within each other. The driver is the *overhearer*.

Condition1

In this condition, there was no turn-taking process, therefore, the conversational roles did not change: the participant was always the "addressee" and the driving agent was always the "speaker" (Fig. 3-5). Under this condition, the speaker robot expressed verbal and non-verbal behaviors by directing its eye gaze and utterances toward the participant. The robot expressed animated behaviors by directing its eye gaze as well as utterances toward the participant. In order to create animated behaviors for the OOCN, we implemented a series of movements that included directing eye gaze toward slightly to the right, slightly to the left and front (facing the driver) which were synchronized with the utterances. The conversational turn did not change in terms of the participants roles.

Condition2

In this condition, the turn-taking process occurred and conversational roles changed only among the robots- the participant was always the "overhearer" (Fig. 3-6). During a turn changes in the conversation, the participant's roles and non-verbal behaviors (eye gaze directions) change animatedly as well. In order to create persuasive animated behaviors for the MPCN, we implemented a series of movements for all of the agents similar to the OOCN condition. During a turn, when the speaking robot generates its utterances, it directs its eye gaze toward the addressee and the side-participant (slightly to the right (e.g., addressee), slightly to the left (e.g., side-participant) which were synchronized with the utterances), meanwhile these two agents direct their eye gaze toward the speaking robot.

By considering these two conversational conditions of NAMIDA, we derived the following hypotheses:

- (H1) The MPCN will exhibit higher life-like qualifications compared to the OOCN.
- (H2) With the MPCN the drivers' attentional focus will not shift toward the robot as frequently compared the OOCN.
- (H3) The MPCN will be perceived as more positively (i.e. more enjoyable, less annoying and distractive, etc.) compared to the OOCN.
- (H4) The MPCN will be evaluated as requiring less mental workload.

3.4.2 Participants

In total, 22 Japanese university students (3 female and 19 male) of ages varying from 20 to 28 years old (M=23.064, SD=2.23) took part in this experiment. All participants had a driving license. The experiment was set up as a counterbalanced, within-subject study in such one group exposed first OOCN and then MPCN, other group exposed first MPCN and then OOCN condition. Upon arrival, the participants were introduced with the NAMIDA platform as a creation of our laboratory. Then, they were given an orientation about the experiment and their task. The participants were only told that, for both conditions, they had to drive starting from a parking area until arriving a specified train station in the simulated town. The task was to listen to the instructions and information from NAMIDA, and drive accordingly. The experimenters were careful about not to mentioning about the participants' expected responses toward the system in both conditions in order to avoid unconscious changes of the participants' behaviors. They were then given a demographic questionnaire. Before starting the experiment, the participants were given enough time to practice on the driving simulation until they feel comfortable. After each session, there was an approximately 10 - 15 minutes break, during which, the participants were asked to complete the subjective assessment questionnaires.

3.4.3 Experiment Setup

We set up our experiment in a driving simulator environment, using the UC-win Road Ver.13[1]. The simulator system consisted of three monitors placed on the dashboard, an adjustable driver seat, a steering wheel, a brake and an accelerator. The light in the experiment room was dimmed to enhance the reality of the driving task. Moreover, one professional camera was placed in front of the participants to record the participant's behaviors. The road had 17 checkpoints (the situations designed on the

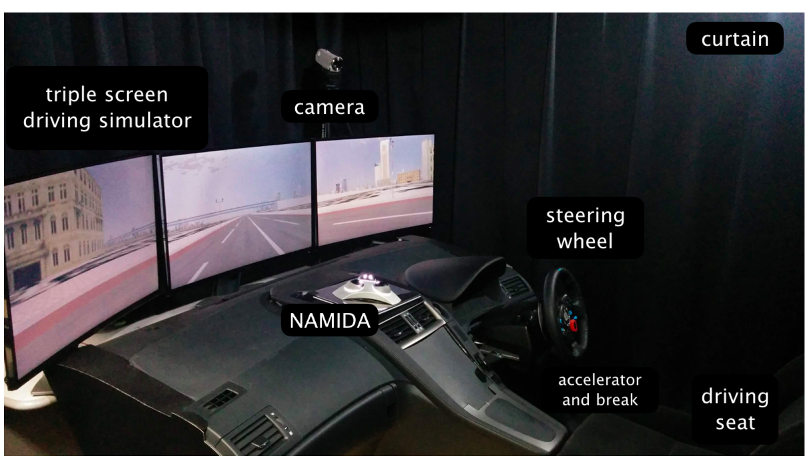

Figure 3-7: The experiment room and experimental setup.

[1]http://www.forum8.co.jp/product/ucwin/road/ucwin-road-1.htm

simulated road): starting from a parking area, continuing through six intersections (left and right), five sections with abnormal road conditions (e.g., car accident on the road, asphalt wet from rain, etc.), one different road situation (an underpass) and two interesting structures along the road (an old and majestic shrine and a building), one checkpoint reporting news (about weather condition on that day) and finally the goal/destination, which was in front of a central train station. The robot(s), under each condition, provided instructions or information depending on the checkpoints along the road. During the trip, NAMIDA gave instructions for navigation, suggestions for changing the speed and brief information about the surroundings directly (in the OOCN condition) or indirectly (in the MPCN condition) 30-60 seconds before arriving at the checkpoints. The silence duration of the robot(s) between two checkpoints depended on each participant's individual speed, yet the approximate time of the robot(s) being silent was 1.5 - 2 minutes. The maximum speed in the simulation was set at 50km/h in the city environment, and the total route was approximately 10km long. The approximate time for arriving at the destination was 10 - 13 minutes.

3.4.4 Measurements

For the subjective evaluation part, in order to analyze the lifelikeness of the system, the participants were given Godspeed questionnaires consisting of five dimensions [9]. To analyze their subjective impressions, they were given a set of questionnaire that included 10 questions regarding the interaction's perceived annoyance, enjoyment, distraction and conversational burden (Table 4.3). We also analyzed the mental workload of the participants through a subjective assessment questionnaire consisting of 10 questions based on the research [98]. For the objective evaluation part, we analyzed the attentional focus of the participants by tracking their eye gaze movements, using Tobii Pro Glasses 2[2] tracker.

[2]https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

						OOCN		MPCN	
		F	p	η^2		mean	std. dev.	mean	std. dev.
Anthrop.	Order	9.584	0.005**	0.324	G1	2.727	0.658	3.544	0.805
	Condition	9.620	0.005**	0.324	G2	2.327	0.588	2.854	0.537
	Interaction	0.449	0.51	0.022					
Animacy	Order	15.542	<.0.001***	0.437	G1	3.06	0.632	3.654	0.494
	Condition	6.990	.0.015**	0.259	G2	3.06	0.875	3.654	0.621
	Interaction	0.005	0.940	0.000					
Likability	Order	1.145	0.297	0.054	G1	4.181	0.821	4.545	0.042
	Condition	3.318	0.083+	0.1423	G2	4.036	741	4.218	0.404
	Interaction	0.368	0.55	0.018					

Table 3.1: Table shows the two-way mixed ANOVA results of Godspeed dimensions (F, p and η^2 values) and descriptive statistics of order groups (G1: Group 1, G2: Group 2). (+p<0.10, *p<0.05, **p<0.01, ***p<0.001)

3.5 Results

3.5.1 Analysis of Godspeed Factors

To confirm the validity of the Godspeed questionnaire, an internal reliability analysis was conducted on the items of each dimension of the questionnaire. The results showed that the Cronbach's alpha was greater than 0.68 in both OOCN and MPCN conditions for the dimensions of anthropomorphism ($\alpha=0.771$ and $\alpha=0.726$); animacy ($\alpha=0.854$ and $\alpha=0.717$); likability ($\alpha=0.876$ and $\alpha=0.842$); perceived intelligence ($\alpha=0.912$ and $\alpha=0.919$); and perceived safety ($\alpha=0.736$ and $\alpha=0.778$). The data from each dimension passed the Shapiro-Wilk normality test at $p=0.05$ level for both conditions, therefore a paired t-test was conducted to compare each dimension. The results revealed a significant difference in anthropomorphism *(t(21)=-3.914, ***p<.001)*, animacy *(t(21)=-3.271, **p=0.003)* and likability *(t(21)=-2.569, *p=0.017)* No significant difference was observed on perceived intelligence *(t(21)=0.24, p=0.812)* or perceived safety *(t(21)=1.311, p=0.203)*.

In order to check whether the dimensions that have been found statistically significant depended on the order of exposing the experiment, we conducted a mixed two-way ANOVA by including the two groups of the participants (Group 1: exposed

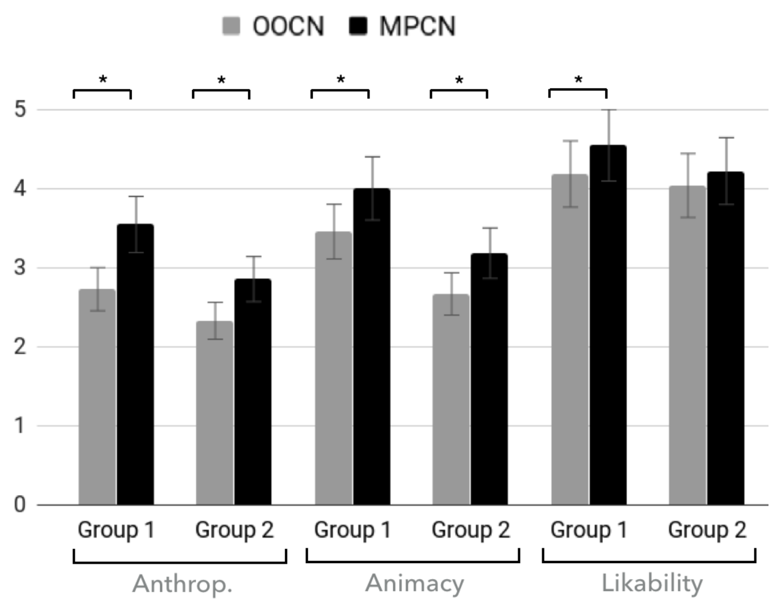

Figure 3-8: Graph indicates pair-wise comparisons of conditions (OOCN and MPCN) with orders (Group 1 and Group 2) of Godspeed dimensions (*p<0.05, **p<0.01, ***p<0.001).

first OOCN then MPCN, and Group 2: exposed first MPCN then OOCN; between groups) and the experiment conditions (OOCN and MPCN; within groups) (Table 3.1 and Fig. 3-8). The results showed that the interaction of order and condition was not significant for the three dimensions (anthropomorphism: $F(1,20) = 0.449$, $p = 0.51$, $\eta^2=0.022$; animacy: $F(1,20)=0.005$, $p=0.94$, $\eta^2=0.000$; likability: $F(1, 20)=0.368$, $p=0.550$, $\eta^2=0.018$). The main effect of condition on anthropomorphism and animacy was significant such that the MPCN was rated higher than the OOCN condition irrespective of the order ($F(1, 20)=9.62$, $p=$**0.005, $\eta^2=0.324$; $F(1, 20)=6.99$, *$p=0.015$, $\eta^2=0.259$; respectively) which was consistent with our expectations. On the other hand, the results also showed that the main effect of the order on anthropomorphism and animacy was significant ($F(1, 20)=9.584$, **$p=0.005$, $\eta^2=0.324$, $F(1, 20)=15.542$, ***$p<0.001$, $\eta^2=0.437$, respectively). This result showed that the ratings of the Group 1 was higher than the ratings of Group 2 which may point to a carry-over effect. As for the likability, though there was not significant difference on the main effect of condition ($F(1, 20)=3.318$ $p=0.083$, $\eta^2=0.142$), the MPCN was rated higher than the OOCN in likability irrespective of the order.

3.5.2 Attention Behavior of Driver

We analyzed the overall ratio of the eye gaze data of the participants on the simulation screen and driving agents by comparing the OOCN and MPCN conditions. The gathered data couldn't passed the Shapiro-Wilk normality test in both categories and conditions at $p=0.05$ level, therefore as a non-parametric test, we used the Wilcoxon signed-rank test (Table. 3.2). We observed that in the MPCN condition, the participants' eye gaze was on the road significantly more than in the OOCN condition (*Z=-3.392, ***p<.001*) (Fig. 3-9). Furthermore, in the OOCN condition, the participants looked at the NAMIDA platform significantly more than in the MPCN condition (*Z=-3.392, ***p<.001*). These results indicated that the MPCN system could help a driver to focus on the road more than the OOCN system.

We observed three specific eye gaze behaviors of the participants within the time duration when the robots start to talk: 1) gaze duration (the time [in seconds] when the participant maintains eye contact with NAMIDA during the robots' conversation), 2) gaze response (the gap time [in seconds] for the participant's first gaze to NAMIDA after the robots start to talk), and 3) response rate (the ratio of the times [from 17] the participant looked at NAMIDA when the robots started to talk). The gathered data couldn't passed the Shapiro-Wilk normality test in these three categories for both conditions at p=0.05 level, therefore as a non-parametric test, we used the Wilcoxon signed-rank test. The results indicated that there was no significant difference in the ranks of the gaze response time rates *Z=-1.328, p=0.183)* and durations *(Z=-0.957, p=0.168)* of the participants when the robots started to talk (Fig. 3-10). However, we found significant difference on the timing of participants' first gaze response after the driving agents start to talk *Z=-4.106, ***p=4E-05* (Table 3.3). It has to be noted that for the results of gaze response, there were times that the driver was already looking at the robots before they started to talk. These times were recorded as 0 (in second). The results revealed that the multi-party conversation gave the impression to the participants that they did not feel a response burden toward this conversation due to the turn-taking approach between the robots.

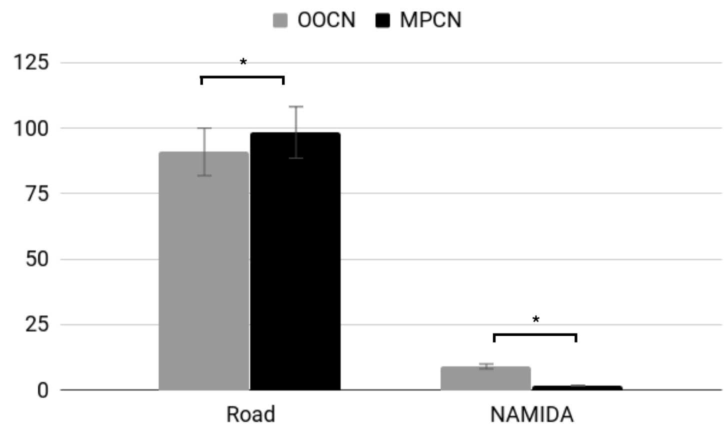

Figure 3-9: Graph indicates the mean values (%) and results of the rates of eye gaze positions. (*p<0.05, **p<0.01, ***p<.001).

				OOCN		MPCN	
	N	Z	p-value	mean	std. dev.	mean	std. dev.
NAMIDA	22	-3.392	<.001***	9.072	10.916	1.66	1.07
Road	22	-3.392	<.001***	90.919	10.915	98.339	1.07

Table 3.2: The Wilcoxon signed-rank test results of the eye gaze data rates observed on the NAMIDA and the road and the descriptive statistics.

				OOCN		MPCN	
	N	Z	p-value	mean	std. dev.	mean	std. dev.
Gaze duration	22	-0.957	0.168	38.425	33.138	49.915	33.593
Gaze response	22	-4.106	<.001***	2609.37	1676.89	8735.394	5169.136
Response rate	18	-1.328	0.183	21.598	16.867	20.588	12.393

Table 3.3: The Wilcoxon signed-rank test results of gazing behaviors of the participants after the robot(s) start conversation and the descriptive statistics related to each behavior.

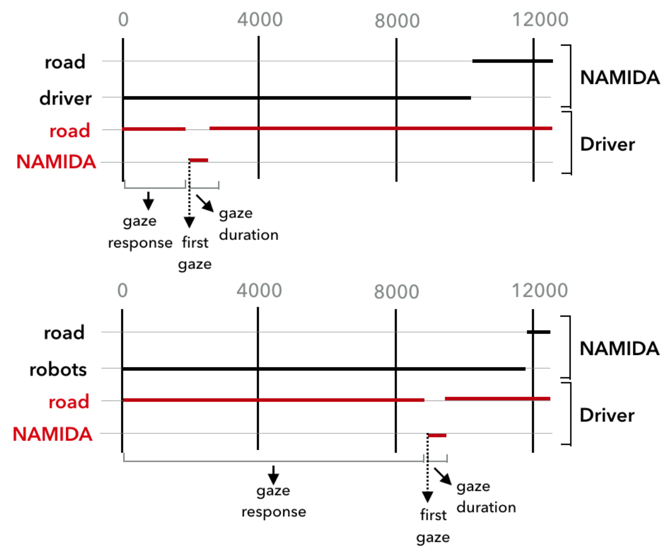

Figure 3-10: The time chart shows an example of gazing behaviors of "NAMIDA" on the "road" and "driver"/"robots"; and the gazing behaviors of "Driver" on the "road" and "NAMIDA". The gaze response and gaze duration of the "Driver" in OOCN *left* and MPCN *right* conditions are based on the average results (Table 3.3.

Group	Question
Enjoyability	I think the driving was enjoyable.
	I think the driving was fascinating.
Annoyance	I thought that NAMIDA's conversation was noisy.
	I annoyed with NAMIDA's conversation
Conversational burden	While NAMIDA talking, I felt like I had to give an answer.
	While NAMIDA talking, I felt like I needed to look at them.
Distraction	I was distracted from driving.

Table 3.4: The content of the subjective impression questionnaire.

Additionally, we analyzed our results in order to clarify if there is a carry-over effect by the subjects' gaze responses. The two-way mixed ANOVA showed non-significant result on the interaction of order and condition ($F(1, 20)=0.014,\ p=.907,\ \eta^2=0.000$). On the other hand, we found significant result on the main effect of condition ($F(1, 20)=26.097,\ p<.001,\ \eta^2=0.566$) that was consistent with our expectations.

3.5.3 Evaluation of the Subjective Impression

In order to evaluate the subjective impressions, the participants were given a total of eight questions about perceived annoyance, enjoyability, conversational burden and distraction that they felt from each NAMIDA condition (Table 4.3). The question-

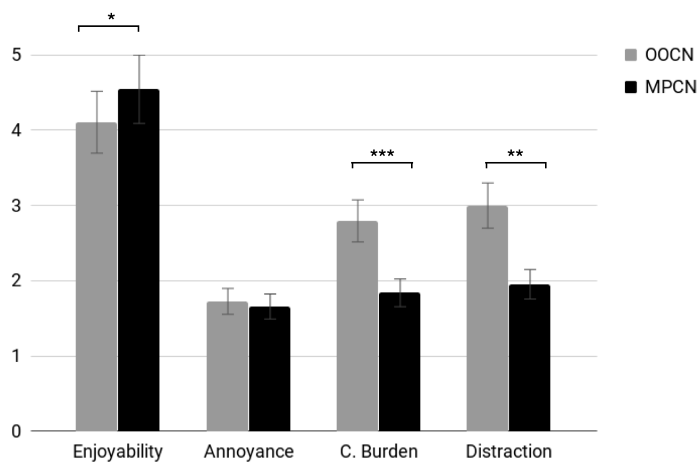

Figure 3-11: Graph indicates the mean value and results of Wilcoxon signed-rank test of subjective impression questionnaire. (*p<0.05, **p<0.01, ***p<0.001).

				OOCN		MPCN	
	N	Z	p-value	mean	std. dev.	mean	std. dev.
Enjoyability	15	-2.271	0.01*	4.106	0.744	4.454	0.604
Annoyance	16	-0.31	0.378	1.727	0.869	1.659	0.917
C. Burden	19	-3.219	<.001***	2.795	1.191	1.84	0.917
Distraction	16	-2.74	0.003**	2.59	1.402	1.954	1.174

Table 3.5: The results of Wilcoxon signed-rank test of subjective impression questionnaire and the descriptive statistics. SQ1: Enjoyability, SQ2: Annoyance SQ3: Conversational Burden, SQ4: Distraction

	N	Z	p-value	OOCN		MPCN	
				mean	std. dev.	mean	std. dev.
Attention	17	-0.047	0.48	1.681	1.427	1.636	1.176
Auditory	9	⋆W=22	5	0.636	0.902	0.636	0.847
Visual	9	⋆W=14	5	1	1.309	0.772	0.812
Temporal	17	-2.579	0.004**	2.681	0.716	2.068	0.876
Situation Stress	18	-2.395	0.0082	1.159	0.829	1.556	0.896
Global	21	-0.521	0.603	7.159	3.472	6.67	2.885

Table 3.6: The results of Wilcoxon signed-rank test of subjective impression questionnaire and the descriptive statistics. ⋆ represents W value for the situations when the size of N is below *10*.

naire was designed using a 5-point, Likert scale (1=strongly disagree to 5=strongly agree). First, we conducted an internal reliability analysis for the questionnaire. For the OOCN condition, the results revealed $\alpha=0.63$, $\alpha=0.935$ and $\alpha=0.837$ for annoyance, enjoyability and conversational burden, respectively. For the MPCN condition, the results revealed $\alpha=0.822$, $\alpha=0.898$ and $\alpha=0.817$ for annoyance, enjoyability and conversational burden, respectively. We applied the Wilcoxon signed-rank test to compare the conditions for each factor within (Table 3.5). We found no significant difference for annoyance ($Z=-0.3103$, $p=.378$). On the other hand, we found significant differences for enjoyability ($Z=-2.2718$, $^{**}p=0.011$), conversational burden ($Z=-3.219$, $^{***}p<.001$) and distraction ($Z=-2.74$, $^{*}p=0.003$) (Fig.3-11).

We analyzed our results in order to clarify if there is a carry-over effect by the subjects' subjective impressions. The two way ANOVA with the order (Group 1 and Group 2) and the conditions (MPCN and OOCN) showed non-significant result on the enjoyability ($F(1, 20)=0.688$, $p=.416$, $\eta^2=0.033$), conversational burden ($F(1, 20)=7.953$, $p=0.010$, $\eta^2=0.284$) and distraction ($F(1, 20)=2.525$, $p=.127$, $\eta^2=0.112$) On the other hand, significant results were found on the main effect of condition on these items ($F(1, 20)=7.429$, $^{*}p=0.013$, $\eta^2=0.27$, $F(1, 20)=18.413$, $^{***}p<.001$, $\eta^2=0.479$, $F(1, 20)=16.261$, $^{**}p=.002$, $\eta^2=0.619$ respectively) that was consistent with our expectations.

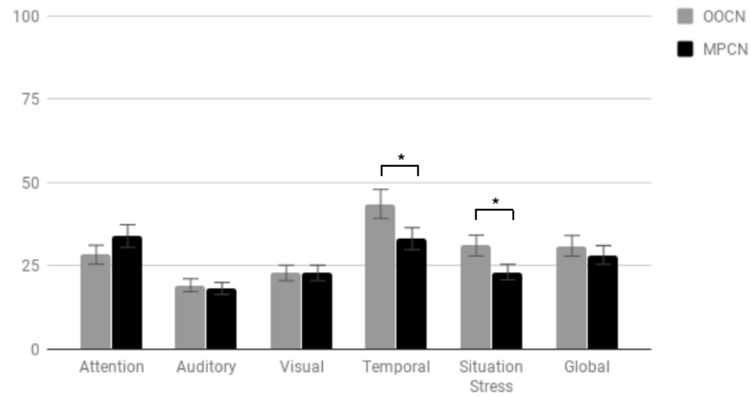

Figure 3-12: Graph indicates the mean values and results of Wilcoxon signed-rank test for DALI questionnaire. (*p<0.05, **p<0.01, ***p<0.001).

3.5.4 Analysis of Mental Workload Factors

A Wilcoxon signed-rank test showed that unlike our previous study, there was no significant difference on neither cognitive components nor the perceptive components (Table 3.6, Fig.3-12).

On the other hand, we observed significant difference on temporal demand *Z=-2.579, **p=0.004* and the stress demand *Z=-2.395, **p=0.008* (Fig. 3-12). We can infer from these results that the participants felt the time pressure and stress aspect under the OOCN condition significantly more than in the MPCN condition, due to the conversation style of the each condition.

Additionally, we analyzed our results in order to clarify if there was a carry-over effect of the questionnaire responses by conducting a two-way ANOVA with the session order as an independent variable. The analysis showed non-significant results on the interaction of the order and condition in situational stress ($F(1, 20)=0.105$, $p=.749$, $\eta^2=0.005$) and temporal demands ($F(1, 20)=0.588$, $p=.451$, $\eta^2=0.028$). On the other hand, significant results were found on the main effect of condition on these demands ($F(1, 20) = 26.097$, $***p<.001$, $\eta^2=0.566$, $F(1, 20)=8.762$, $**p=0.007$, $\eta^2=0.304$, respectively) that was consistent with our expectations.

3.6 Discussion

In this paper, based on the results of this comparative study, we discuss the effects of robotic driving agents within a multi-party conversation platform with that of a one-to-one conversation based platform. We investigated drivers' perceptions about the robots' lifelike qualities, attention manipulation, subjective impressions and perceived mental workload. The design of the multi-party conversation model employs turn-taking actions during a conversation in which we hypothesized that, coupled with verbal (utterance generation) and non-verbal (eye gazing) behaviors, the robots would create a more lifelike, enjoyable and less distractive environment compared to a one-to-one conversation model where the robot's utterances were solely directed toward the driver.

3.6.1 Godspeed Dimensions

The analysis of two-way mixed ANOVA on Godspeed dimensions revealed that MPCN was rated significantly higher than OOCN in anthropomorphism and animacy regardless of the order that was consistent with our prediction. On the other hand, we found a main effect of the order that led us to infer that there might be a carry-over effect between the conditions depending on the exposing order. In Group 1, participants first exposed the OOCN condition, in which one robot was trying to have an interaction with the driver. The expressive conversation of the robot might bring out a certain level of anthropomorphism and animacy. After the OOCN, exposing the MPCN condition showed the participants that the one robot they exposed in the first condition had an ability to have multi-party conversation which creates a more lively interaction. Nevertheless, these features of multi-party conversational robots were realized with the knowledge of the interactions of one-to-one conversational robot which indicates a contrast effect for these dimensions. The contrast effect is a concept in social psychology that when a situation is compared with another one that is enhanced or diminished, relative to the normal situation, this situation will be evaluated lesser or greater [110]. Since anthropomorphism refers to the attribution of a human form and human characteristics to non-human entities, we can infer that in the MPCN condition, the robots were perceived as possessing a more natural, human-like conversation with their coherent utterances and turn-taking actions when they compare it with OOCN that incorporates with addressing directly to the participants. Likewise, compared to the previous interaction experience, the participants rated MPCN better as they exhibit the verbal and non-verbal behaviors that induced the perception of the robots as having more animacy, which refers to being that of lifelike creatures. Piaget emphasized the major factors of lifelikeness consists of movements and intentional behaviors, and that perceiving something as lifelike allows humans to distinguish humans from machines [96]. Therefore, the coherent non-verbal behaviors of the multiple robots contributed in the MPCN led to a perception of the robots as more animated.

As for the likability dimension, we did not observed a statistical potential of a carry-over effect. Even though the results did not mark a statistically significant main effect of condition, the ratings of likability on MPCN condition was higher than the OOCN regardless of the order. It is described that likability is the first positive impression of people by others. Since a human-likeness component encourages empathy in a system, in the MPCN condition, the participants were more likely to exhibit positive feelings and implicitly feel more familiar with the system. We believe that these results indicate the system's capability of achieving a human-like nature to a certain extent.

3.6.2 Subjective Assessment of Enjoyability, Annoyance, Conversational Burden and Distraction

The results of subjective assessment showed that in the MPCN condition, the robots were perceived as more enjoyable than in the the OOCN condition. In the MPCN condition, the informative conversation of the robots was presented as a theatrical performance, and this might entertain the participants. As the research demonstrated[49], when people attribute human-like behaviors to robots, in our case, having a lively multi-party conversation, it makes the robots be perceived as more enjoyable. Moreover, as for the placement of NAMIDA within the peripheral vision of the drivers, the participants could enjoy the companionship of the MPCN more so than with the OOCN. These results can also indicate that the indirect interactions within the peripheral vision of the drivers have persuasive effects on the users.

The results also revealed that even though, in the MPCN condition, there were slightly more lines in the conversation script than in the OOCN condition, that might be evaluated as annoying, the multi-party conversation was perceived as significantly more enjoyable. The MPCN condition was also perceived as significantly less distractive and giving a feeling of less conversational burden, due to its persuasive turn-taking nature of the system.

3.6.3 Objective Assessment of Distraction

It is expected that when the driving agent starts to talk in the OOCN condition, the driver is tempted to look at the driving agent as one would in a natural one-to-one conversation. However, in the MPCN condition, because the utterances were not directed at the driver, he/she would not attempt to look at the driving agents. With this indirect communication method, the participants could get the necessary information within the lively conversation occurring between the robots.

The results of the subjective assessment of conversational burden and distraction, and the objective analysis of the eye gaze behaviors of the participants were closely related. Considering the conversational burden, the results of the subjective analysis and objective analysis were consistent. The eye gaze data analysis demonstrated that in both conditions the participants focused on the road more than NAMIDA. However, we found a significant difference on the rate of eye gazing on NAMIDA and the road between the two conditions. In the OOCN condition, the participants' eye gazes were diverted toward NAMIDA significantly more than in the MPCN condition. Likewise, in the MPCN condition, where the robots could sustain the conversation without the driver's participation, the subjects' eye gazes were on the road significantly more than in the OOCN condition. In addition the results of gaze response analysis indicated that the directed utterances and eye gaze of NAMIDA in the OOCN condition imposed upon the participants the role of addressee causing an increase the conversational burden. Even though, in both situations, the driver relied on the auditory information coming from the robots, yet for the OOCN condition, however, due to having the conversational burden as an addresser, the driver was obligated to be in direct interaction with the system which emerged as visual and auditory efforts required for the driver. Since the direct interaction by the one-to-one communication platform requires more effort than overhearing a multi-party conversation, it was relevant to find, as we expected, a significant difference on perceived conversational burden.

Moreover, the results indicated that the participants perceived that they were distracted in the OOCN condition significantly more than in the MPCN condition even though during a conversation term, the eye gaze rate was not statistically significant. We regarded the perceived distraction as being aware of shifting attention rather than performing reflexive behaviors. In physiology, reflexive behaviors are defined as the behaviors that the individual is not aware of while they are doing them. With these results, it has been shown that the participants were aware of the situation they evaluated OOCN system as being that of more distractive.

3.6.4 Workload Factors

As for the workload factors, interestingly, we did not observe the similar results with our previous study [65]; the results did not conform to the results we obtained from our previous study where we used embodied virtual agents [65]. One reason for this might be the fact that the drivers focused on driving more than in the previous study due to the fact that the current driving simulator was much closer to a real-world scenario, and it required much more focus to control the car and manage the traffic. Consequently, the drivers concentrated more on the road rather than the robots in both conditions and, thus, a ceiling effect might have occurred that prevented the participants to distinguish on their perceptions on the visual and auditorial demands of the mental workload. The objective results from 3.5.2 which showed that the participants' eye gaze were on the NAMIDA platform significantly more often in OOCN condition than in MPCN condition, and the results from 3.5.3 which showed that the participants felt distracted more often in the OOCN condition support the possibility of ceiling effect that might caused the participants to evaluate the cognitive and perceptive demands of the workload.

For the temporal and situational stress demands, we can infer from these results that the participants felt that the time pressure aspect under the OOCN condition was significantly more than in the MPCN condition, due to the conversation style of the each condition. The participants might have felt that the conversation in the OOCN condition was open and a response was expected unlike the MPCN condition. Hence, the utterance of the robot might be perceived as a command to be obeyed, which created a time pressure aspect. Likewise, the significant result of stress demand may point to a "master-slave" interaction in which the driver should comply with whatever the robot says, which would lead to an increase in stress demand. However, these results have not been observed on the previous research [65]. Since the physically presence of others is physiologically arousing and causing "social facilitation [133], the presence of the robots rather than virtual embodiment could make the robots more salient and closer to a human-like manner in their interactions [69]. Hence, we can infer that the persuasiveness of the robotic agents in this study was more effective and distinguishable in both conditions of the OOCN and the MPCN than the virtually embodied driving agents in the previous study.

3.6.5 Limitations

The recruitment of mostly male subjects limited our results and our ability to make broader generalizations. Ideally, a study across a wider age range with greater gender balance and conducting the study in a real car environment would produce more reliable results in terms of the effects of both systems on the drivers. Because we conducted our experiment with Japanese participants, the cultural context of our study constitutes another limitation: the fact that the Japanese participants are more accustomed to robotic interfaces. Also, when we compare the high usage of polite language in Japan with other countries, the utterance structure of a driving agent may need to be varied, considering the differences in the casual everyday language of each country.

3.7 Conclusion & Future Work

The multi-party conversation-based interface of NAMIDA presents a unique interaction between the car and driver. As a social interface, it has been designed to assist drivers by conducting a multi-party conversation This study showed that this conversational approach has the potential to be perceived as more lifelike and enjoyable as it possess a lively conversation of robots that requires less attention from the driver in obtaining necessary information. We designed an experiment to verify our hypothesis with our NAMIDA platform by comparing two different cases, the MPCN and OOCN, in a realistic driving simulation environment.

We evaluated the interactions between the driver and the NAMIDA platform in both conditions using the subjective questionnaires that examined the system's lifelikeness, enjoyability, annoyance, perceived conversational burden and perceived distraction; the objective analysis that examined the eye-gaze data gathered during the experiments. The results of the subjective questionnaires revealed that in the MPCN condition, the robots created more anthropomorphic feelings and were perceived as more likable and having more animacy with their coherent utterance and turn-taking actions. However, the order of the experimental conditions might have an influence on the anthropomorphism and the animacy ratings where the participants rated higher the MPCN when they interacted with the OOCN before. Moreover, the objective eye gaze analysis revealed that overhearing information via a conversation among the sociable agents required significantly less attentional efforts. The subjective analysis also supports this result by showing that in the MPCN condition, the perceived conversational burden and perceived distraction were observed as significantly less than in the OOCN condition. Through the turn-taking based lively conversation of the robots where the driver, as an overhearer, is not compelled to contribute to the conversation, the MPCN system exhibited a more stress-free and enjoyable driving environment.

For a more sociable environment inside a car, our next study will focus on the driver in the multi-party conversation by considering an adaptive approach toward the driver's needs. The future study is also required to further generalize the results and investigate the different aspects of the multi-party conversation platform on drivers during a real-world vehicle experiment.

Chapter 4

The Effects of Driving Agent Gaze Following Behaviors on Human-Autonomous Car Interaction

4.1 Introduction

Recently, a great deal of research has been conducted on highly autonomous vehicles which make their own driving decisions that minimize human interventions with the vision of decreasing human errors and achieving a safer, more energy efficient and more comfortable mode of transportation [128, 100]. However, eliminating human involvement from driving might threaten the trust and perceived safety, and suppress drivers' joy of driving and the desire to control the vehicle which in turn lead to a refusal to use autonomous cars.

Researches has demonstrated that an increase in perceived anthropomorphism affects positively the perceived trust in autonomous vehicles [129], [53]. However, in these studies, the interaction between the system and the human is still not natural nor intuitive due to the system lacks the exhibiting of continuous sociability. A human operator should be able to understand the automation system, fully and intuitively. Norman [88] discussed that a system's design model should be identical to the user's mental model. He also suggests that the communication can be more effective in a form of an appropriate metaphor. Flemisch et al. [37] introduced H-Metaphor as an interaction concept between an autonomous car and a human operator based on understanding the situational-intentions of each other based on the idea of continuous haptic interactions between a horse and a rider. We believe that through building a reciprocal interaction between a human operator and an autonomous car where the parties can perceive each other as social entities and understand each other's intentions (emerging intersubjectivity), and build intersubjectivity, in which they can find motivation to engage with each other [123] (i.e. situational-awareness towards the shared environment to react when an action is needed), a reliable interaction can be established.

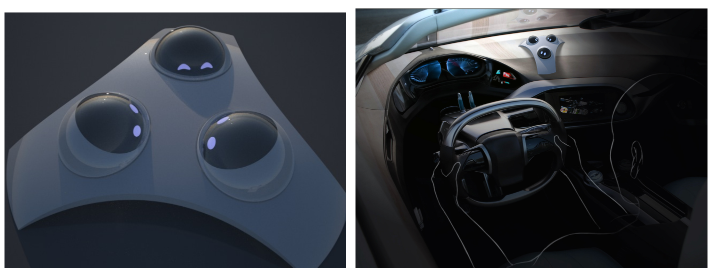

Figure 4-1: A conceptual diagram of the NAMIDA platform. It consists of one base unit containing three movable heads with one degree of freedom each (*left*). The base unit of NAMIDA, which attaches to the dashboard of a car is within the peripheral vision of the driver (*right*).

Intentional stance is a strategy for understanding an entity's behaviour by treating it as a rational entity whose actions are governed by its beliefs and desires [29]. Intentional stance is closely related to social presence, which defines the degree of awareness of the other entity in an interaction and the sense of access to the other's mind [108]. Intersubjectivity emerges as humans feel that others feel or act on as if they have the same intentions [11]. This intersubjective sharing is critical, because it creates a shared space of common psychological ground that enables everything from collaborative activities with shared goals to human-like cooperative communication through comprehending each other's intentions. In this respect, when the autonomous system persuades the human operator that the system possesses the same intentions as the human operator as a social entity, the relationship between them will gain a more organic shape, which will enable the establishment of a constant reliable interaction.

Humans engage in a wide variety of social interactions using their ability to reason about others' intentional stance. One social interaction for humans is to adopt the intentional stance of others using the ability to interpret the eye gazing of others and then interpreting their actions [29, 8]. Researchers in cognitive science and developmental psychology consider gaze following to be one of the essential components in social interaction and learning [18]. It also contributes to understanding of what the others think, feel and intend to do [123], [7].

A number of studies in Human-Robot Interaction and Human-Agent Interaction have shown that with eye gazing behaviours, robots can gain the ability to give information to their interlocutors [67], [83, 25]. In situated human-machine interaction, the robot's or agent's gaze could be used as a cue to facilitate the user's comprehension of the robot's instructions [111]. Expressive eye gaze is one behaviour (among many drawn from animation principles) that can make intentions and desires more explicit [117]. Even when users are unaware of the intended communication, robots can reveal their intentions implicitly through eye gazes and influence human behaviours [82]. Researches has demonstrated that robots can use gazes to establish joint attention in the attempts of learning from demonstrations, as well as in soliciting feedback when there is uncertainty [76]. When a robot student responds to joint attention by following the human teacher's gaze, it better conveys the robot's internal states and knowledge, which leads to more efficient teaching: fewer errors, faster recovery from errors, and less repetition of learned information [55]. People also rate a robot as more natural and competent when it builds joint attention while performing a task [55].

In this study, we employ a robotic, driving agent platform, NAMIDA (Fig. 4-1), to utilize eye gazing cues to reveal the intentions of the robots, correspondingly the social presence of an autonomous car. We analyzed whether the gaze following behaviours of the robots can facilitate the interaction between the human operator and the autonomous car.

4.1.1 Perceived Agency

Using a robotic driving agent as an interface for an autonomous car might create ambiguity for the humans' perceptions. The interface can either be perceived as a companion for the driver that is independent from the autonomous system (e.g., passenger), or as an authority who is directly connected to the system and responsible for the autonomous driving. When humans feel an engagement with a social entity, they tend to feel safer in their interactions with that entity [109]. We expect that when the robots' intentions (e.g., watching the road and being aware of) synchronize with the autonomous car's actions (take an action according to the situations on the road), the human operator will infer the existence of dependency (authority) of the driving agents to the autonomous system, which we will define it as the "perceived agency" of the robots. We also expect that the perceived agency would lead to an increase in the perception of safety.

4.2 Method

In this study, we investigated whether the gaze following behaviors of a robotic driving agent platform was effective on increasing the perceived agency and intentional stance of the robots, social presence of the autonomous car, perceived safety and enjoyability of a human operator that they could feel during a simulated autonomous driving. We also expected that these two factors will be depended on the intentional stance of and the perceived social presence of the system as the correlation between the intentional stance of the autonomous car and the driving agent platform. In addition, we wanted to explore any relation between the perceived agency with the other factors we analyzed.

4.2.1 System Design

For this study, we employed NAMIDA platform that is consisting of three robots [66]. The NAMIDA platform involves one base unit that attaches to the dashboard of a car, containing three movable heads with one degree of freedom each in the driver's peripheral vision. The round shaped head of each robot allows for the positioning of their eyes with full color LED light emission. The movement of the robots is enabled by three servomotors linked to the Arduino platform inside each head that is attached to the main board. NAMIDA has an utterance generation mechanism that allows the robots to generate verbal and non-verbal behaviors within each turn-taking. However, in the current study, we focused on the effectiveness of NAMIDA system's particular non-verbal behaviors (eye gaze behaviors), thus we did not employ the utterance generation mechanism.

4.2.2 Gaze Following Behaviors of NAMIDA

Joint attention is an active bilateral process which involves attention manipulation, but it can only be fully realized when the parties are aware of each other's attention [62]. Even though response and feedback behaviours are necessary to realize a joint attention that makes robots more competent and socially interactive within a human-robot interaction, however, in this study, we only focused on the one aspect of the joint attention which is the gaze following behaviours that is the active unilateral process of simultaneous looking.

In order for the robots to realize the gaze following of the human, we used Tobii Pro Glasses 2[1] tracker. From the tracker, two parameters of eye gazing data (i.e., points on x and y axes) were used to implement the gazing behaviours of the robots. In order to amplify the gaze following movements of the robots (to increase the human sense on the gaze following behaviors of the robots), the eye gaze data obtained from the tracker was multiplied by *3.5* on the x axis. We also put a 350 ms delay between the human gaze and the robots' gaze to make the gaze following more sensible by the human operator. For this study, when the human moves his/her eyes, all the three robots perform gaze following behaviours simultaneously within a cohesive manner.

[1]https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

4.2.3 Conditions

We conducted an experiment by employing our NAMIDA platform with three conditions in an autonomous driving environment:

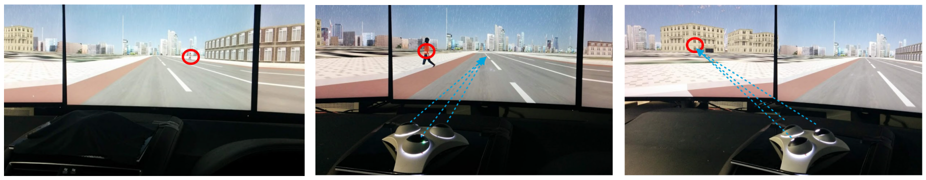

Figure 4-2: The NN (*left*), NRB (*middle*) and NFB (*right*) conditions are shown. The *red* circle represents the gaze point of the human operator. The *blue arrows* represent the gaze direction of the robots.

1. No NAMIDA (NN): the robots were not used. They were covered with a black piece of cloth (Fig. 4-2 (*left*)).

2. NAMIDA with Random gaze Behaviors (NRB): The robots were placed on the dashboard and were set to watch the front side of the road (as passengers) with the head movement of normally distributed from -15 to +15 degrees (Fig. 4-2 (*middle*)).

3. NAMIDA with Following Behaviors (NFB): The robots were constantly following the eye gaze of the participants (Fig. 4-2 (*right*)).

4.2.4 Experimental Setup

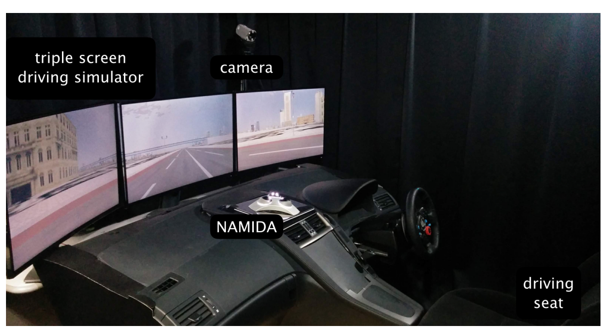

Figure 4-3: The experiment room and experimental setup.

Our aim was to investigate the effects of the gaze following behaviors of NAMIDA on the perceptions of the agency, intentional stance, social presence, perceived safety and enjoyment of the human operators. We set up our experiment in a simulated autonomous driving environment that consists of three monitors placed on the dashboard, an adjustable driver seat and a steering wheel (Fig.4-3). We used UC-win Road Ver.13[2] as the simulation software. The experiment room was dimmed to enhance the reality of the driving task.

In the simulated road, we placed five situations (checkpoints) along the way where each one triggered an action for the autonomous system to take. Three of the checkpoints indicate turning actions (signaling, slowing down, changing lanes and turning right or left), one checkpoint to indicate an underground passageway (signaling, slowing down, changing lanes) and one to indicate an automobile accident in the underground passageway (signaling, slowing down and passing by the automobile accident carefully). The maximum speed of the car was set to 40km/h. The robots were enabled to track the human operator's eye gaze using the Tobii Pro Glasses 2 tracker.

4.2.5 Procedure

22 participants (3 female, 19 male) from 19 to 40 years old (M=24.82, SD=6.31)) took part in our experiment. We conducted a counterbalanced within-subject-study. All participants had a current driving license. Upon arrival, each participant was given an explanation about the experiment. Before starting the experiment, the participants were asked to fill out a demographic questionnaire. After each session, they were asked to fill out the subjective assessment questionnaire; the three questions for Perceived Agency (PA) of NAMIDA, three questions for Perceived Enjoyment (PE), four questions for Perceived Safety (PS), five questions for Social Presence of the autonomous car (SP), 6 questions for Intentional Stance of Namida (ISN) (Table 4.1). Except for PA, the questionnaire items were prepared based on the work in [50, 5, 114].

In the simulation, the participants were told that they had to go to a train station, and that the autonomous car would take them to there using the best route, so their task was to carefully watch the environment and interact the robots during the

[2] http://www.forum8.co.jp/product/ucwin/road/ucwin-road-1.htm

autonomous driving. Before starting the experiment, the participants undertook a trial session for a few minutes in order for them to adapt to the environment. After the trial, the real sessions started. Each session took approximately five minutes.

4.3 Results

Perceived Agency (PA) and Intentional Stance of NAMIDA (ISN) questions were given only after the NRB and NFB conditions. For PA questions, the validity of the questions were analyzed through conducting an internal reliability analysis. The results showed that the Cronbach's alpha was greater than 0.68 under both NRB and NFB conditions (α=*0.855* and α=*0.688*, respectively). A subsequent paired t-test revealed significant difference between two conditions *(t(21)=-1.734, *p<.048)*. The NFB was rated significantly higher (M=3.189, SD=0.91) than the NRB condition (M=2.295, SD=1.032) (Fig. 4-5, *right*).

The results for the validity of ISN questionnaire showed that the Cronbach's alpha was α=*0.937*, α=*0.907* for NRB and NFB conditions, respectively. Then, we conducted a paired t-test to investigate whether the gaze behaviors of the robots had an effect on their perceived intentional stance. The results showed that under the NFB (M=3.114, SD=0.915) condition the participants rated the related questions significantly higher than under the NRB condition (M=2.205, SD=0.999), *(t(21)=-4.252 , ***p<.001)*.

The results of the validity analysis of the Social Presence (SP) questions showed that the Cronbach's alpha was α=*0.658*, α=*0.785*, and α=*0.811* for NN, NRB and NFB conditions, respectively. Then we conducted a one-way within subject ANOVA to investigate whether the gaze following behaviors of NAMIDA affect the perceived social presence. The results showed a significant difference among the conditions *(F(2,42)=9.872, ***p<.001)*. The Bonferroni correction revealed that the main score for the NFB (M=3.372, SD=0.458) was significantly higher than the NN (M=3, SD=0.436) condition; *t(21)=-3.552, **p=.006* and NRB (M=2.827, SD=0.701) condition; *t(21)=-3.355, **p=.009* (Fig 4-5).

Code	Questions
PA1	I felt that Namida and the car were connected.
PA2	I felt that Namida was independent from the car.
PA3	I felt that Namida reflected the car's mind.
PE1	I think the driving was enjoyable.
PE2	I think the driving was fascinating.
PE3	I think,the driving was boring.
PS1	I think the driving was safe.
PS2	I think the driving was relaxing.
PS3	I think the driving was calm.
PS4	I think the driving was surprising.
SP1	I perceived that I was in the presence of the car.
SP2	I felt that the car was watching me and was aware of my presence.
SP3	The thought that the car is not a real person crosses my mind often.
SP4	The car appeared to be sentient (conscious and alive) to me.
SP5	I perceived the car as being only machine, not as living creature.
ISN1	I felt that the autonomous car/robots could understand my intention.
ISN2	I thought that the autonomous car/robots shared my feelings.
ISN3	The autonomous car/robots seemed to care about me.
ISN4	The autonomous car/robots was trying to get involved with me.
ISN5	I thought the attention of the autonomous car/robots depended on my attention.
ISN6	I felt connection between me and the autonomous car/robots.

Table 4.1: Subjective assessment questionnaire consists of Perceived Agency (PA), Perceived Enjoyment (PE), Perceived Safety (PS), Social Presence (SP), Intentional Stance of NAMIDA (ISN).

		F/t value	p-value
t-test	PA	-1.734	<.048*
	ISN	-4.252	<.001***
ANOVA	SP	9.872	<.001***
	PS	4.858	.013*
	PE	4.27	.021*

Table 4.2: The table on the *left* shows the t values for a paired t-test analysis (PA and ISN, F values for a one-way repeated ANOVA (PA, SP, ISN, PS and PE factors), and p values for the corresponding analysis for each factor. The table on the *right* shows the Pearson's correlation analysis. The values on the correlation table represents the *r values* (*p<0.05, **p<0.01, ***p<0.001).

The Cronbach's alpha for the Perceived Safety questionnaire (SP) was under 0.68 in three conditions ($\alpha=0.664$, $\alpha=0.57$ and $\alpha=0.667$ for NN, NRB and NFB conditions, respectively). However, when we excluded the PS4 question, the Cronbach's alpha for each condition became $\alpha=0.774$, $\alpha=0.693$ and $\alpha=0.757$ for NN, NRB and NFB conditions, respectively. After excluding the PS4, we conducted a one-way within subject ANOVA to investigate whether the gazing behaviors of the robots affected perceived safety. The results showed a significant difference among the conditions $(F(2,42)=4.858,\ {}^{*}p=.013)$. The Bonferroni correction revealed that the main score for the NFB (M=3.727, SD=0.462) condition was significantly higher from the NN (M=3.364, SD=0.601) condition $(t(21)=-2.629,\ {}^{*}p=.047)$.

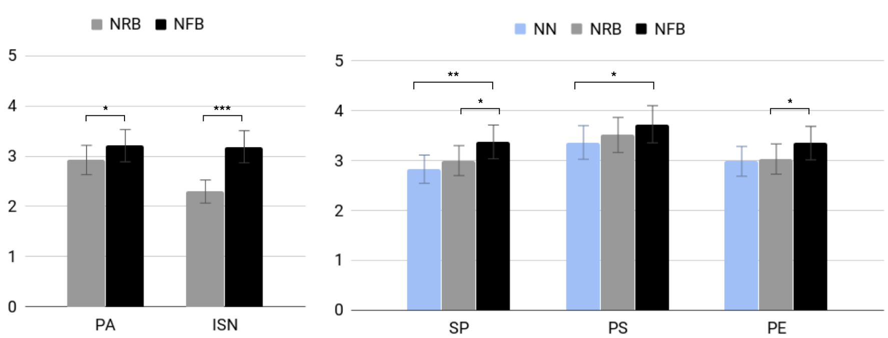

Figure 4-4: The graph indicates the mean values of each subjective factor and the conditions where the factors are significantly different (*p<0.05, **p<0.01, ***p<0.001).

The validity test of the Perceived Enjoyment (PE) questionnaire showed that the Cronbach's alpha was $\alpha=0.841$, $\alpha=0.844$ and $\alpha=0.811$ for NN, NRB and NFB conditions, respectively. We then conducted a one-way within subject ANOVA to inves-

	PA	SP	ISN	PS	PE
PA	–				
SP	0.208	–			
ISN	0.344*	0.569***	–		
PS	0.449**	0.449***	0.337*	–	
PE	0.218	0.066	0.270	0.196	–

Table 4.3: The table on the *left* shows the t values for a paired t-test analysis (PA and ISN, F values for a one-way repeated ANOVA (PA, SP, ISN, PS and PE factors), and p values for the corresponding analysis for each factor. The table on the *right* shows the Pearson's correlation analysis. The values on the correlation table represents the r *values* (*p<0.05, **p<0.01, ***p<0.001).

tigate whether the gaze following behavior of NAMIDA affected perceived enjoyment. The results showed a significant difference among the conditions *(F(2,42)=4.27, *p=.021)*. The Bonferroni correction revealed that the main score for the NFB (M=3.348, SD=0.43) condition was significantly higher from the NRB (M=3.03, SD=0.435) condition *(t(21)=-2.672, *p=.043)*.

The Pearson correlation coefficients results among five subjective assessment factors are shown in Table 4.3 (*left*). The results showed that there was a moderate positive correlation between ISN and PA ($r = 0.569, p = .022$), PS and PA ($r = 0.449, p < .022$), PS and SP ($r = 0.449, p < .001$), PS and ISN ($r = 0.337, p = .025$). On the other hand, there was a strong positive correlation between the ISN and SP ($r = 0.569, p < .001$) indicating that comprehending the intentions of the robots induces the social presence of the autonomous car.

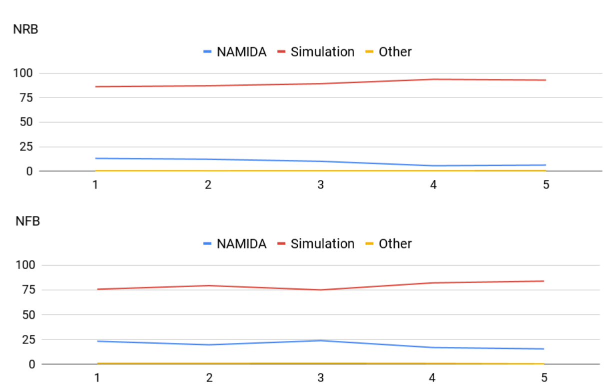

Figure 4-5: NRB time scale

	NRB			NFB			NRB-NFB
Minute	Namida Mean (STD)	Simulation Mean (STD)	Other Mean (STD)	Namida Mean (STD)	Simulation Mean (STD)	Other Mean (STD)	Wilcoxson signed-rank test
1	13.172	86.25	0.5806	23.244	75.79	0.9615	W=27.00
	(15.10)	(15.87)	(1.245)	(21.75)	(21.97)	(1.704)	*p=0.02
2	12.301	87.16	0.5428	19.660	79.40	0.9396	W=34.00
	(17.27)	(17.64)	(0.7691)	(22.55)	(22.92)	(1.569)	*p=0.047
3	10.171	89.27	0.5557	23.860	75.10	1.043	W=37.00
	(14.42)	(15.33)	(1.046)	(25.31)	(25.37)	(1.154)	p=0.065
4	5.664	93.74	0.5982	16.926	82.19	0.8838	W=10.00
	(14.59)	(14.99)	(1.469)	(23.62)	(23.77)	(1.652)	*p=0.014
5	6.356	92.96	0.6867	15.523	83.95	0.5232	W=28.00
	(13.16)	(14.44)	(2.057)	(25.59)	(26.08)	(1.476)	p=0.132

Table 4.4: The table shows the descriptive analysis of the NRB (*left*) and NRB (*right*) conditions. shows the Pearson's correlation analysis. The values on the correlation table represents the *r values* (*p<0.05, **p<0.01, ***p<0.001).

4.3.1 Eye Gazing Behaviours

We investigated the gazing behaviors of the robots and the participants by using the gathered gaze data from the tracker, in order to understand whether there is a statistical difference between the conditions in terms of gaze fixation counts on the robots for the NRB and NFB conditions. Three participants were omitted from the data set due to lack of sufficient amount of data. We conducted a Wilcoxon Signed-Rank test with the gaze data from 19 participants. The results revealed that in the NFB (M=*33.304*, SD=*16.895*) condition, the participants looked at the robots significantly more times compared to the NRB (M=*23.334*, SD=*16.419*) condition (*t(19)=1.844, p=0.036*).

According to the feedback from the participants, in the NFB condition, some of them felt that the robots were responsive to their eye gaze. Others noted that they felt that they were actively involved and felt that they were not alone, but they could not understand to what the robots' behaviors were responsive. They reported that the robots followed their eye gaze. When they realized this responsiveness of the robots, they played with the robots.

We divided the session period to each minute during the experiment. We found that in the NFB condition, in the first, second and fourth minutes, the participants looked at the robots significantly more than the NRB condition (Table 4.4). The participants looked at the robots in the beginning to reason the robots' behaviors. In NRB condition, the participants looked at Namida. On the other hand, in the NFB condition, most of the participants looked at the robots in beginning, understood that the robots are responsive towards them and they kept their interaction until the end of the session. This result also supports the PE results implying that the participants enjoyed with the robots' responsiveness.

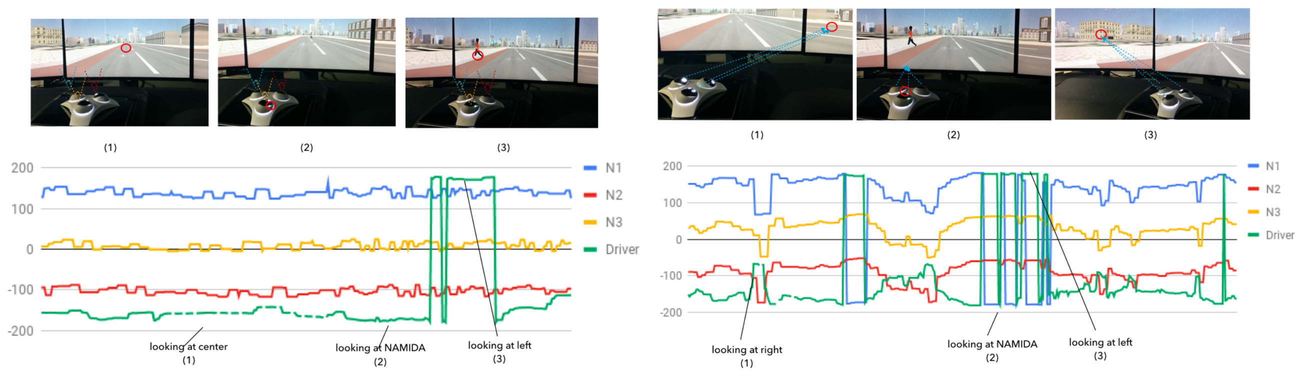

Figure 4-6: The scale graph shows the NRB condition where the robots moved left and right with a random Gaussian distribution of +-15 degrees as it was precoded.

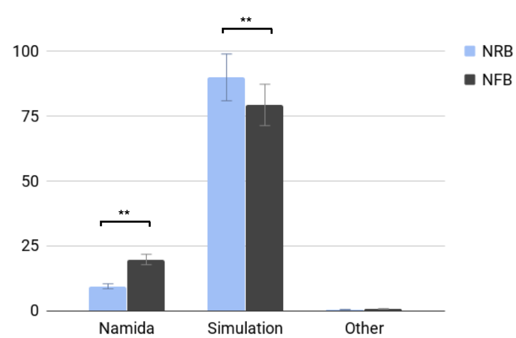

Figure 4-7: The graph shows the result of paired t-test between NRB and NFB conditions on the gaze fixation count (*p<0.05, **p<0.01, ***p<0.001).

4.4 Discussion

In this study, we investigated the effectiveness of gaze following behaviors of a robotic driving agent platform to enhance human-autonomous car interaction. The proposed interaction method was expected to increase the perceived agency and comprehending the intentional stance of the robots; and social presence of the autonomous car. In addition, we expected there to be an increase in perceived safety and enjoyability with the autonomous driving system.

The results verified that under the NFB condition, the participants attributed the robots a higher level of agency which means the gaze following behaviors were associated better with the actions of the simulated autonomous car by the participants. Consequently, the gaze following behaviors were an influence in the robots being perceived as the authority of the car rather than a passenger. We also found that under the NFB condition, the participants perceived the intentional stance of the robots significantly better than under the NRB condition which led us to infer that the gaze following behaviors of the robots hold the potential of building intersubjectivity with the human operator. Moreover, the results showed that under the NFB condition, the participants' sense of social presence of the autonomous system was better compared to the NN and NRB conditions. The strong correlation between the intentional stance of the robots and the social presence of the autonomous car indicated that realizing the intentions of robots contributed to perceiving the autonomous car as being a social entity more than under the other conditions.

Perceived safety was found significantly higher under the NFB condition compared to the NN condition as it is expected; however, it was not significant compared to the NRB condition. It can be inferred that the robots' front looking behaviors might have similar safety effects with the following gaze behaviors. On the other hand, perceived enjoyment was found significantly higher under the NFB condition compared to the NRB condition as it is expected; however, it was not significantly higher compared to the NN condition. It can be said that, without the robots, the participants could also enjoy the autonomous driving by observing and reasoning the car's actions. However, in the case of employing the robots, it was significantly more enjoyable when the robots were responsive to gazing of the participants.

Lichtenthäler et al. [75] have shown that when a robot's behavior was legible, perceived safety of humans increased. In this respect, we expected that the intentional stance of the robots will positively correlate with the perceived safety. We also expected that a positive correlation between the social presence of the autonomous car and the perceive safety. The moderate correlations between PS and SP, PS and ISN pointed out our expectations above. Also, the moderate correlations between PS and PA, PS and SP indicated that perceiving the robots as an authority has a potential to ameliorate the perception of the system as a social entity that might affect the perceived safety.

4.5 Conclusion

In this study, we proposed an autonomous car-human operator interaction paradigm in order to achieve a reliable interaction with an autonomous car, such that the autonomous system and the human could sense each other's intentions and be aware of each other's presence. The results of this pilot study showed that perceiving an autonomous car as a social entity through the gaze following behaviors of a driving agent platform was possible and has the potential to improve the perceived safety and enjoyment of the autonomous driving system. Future studies will investigate methods to improve the relationship between an autonomous car and human operator in terms of increasing the perceived safety, trust and the pleasure of autonomous driving.

Chapter 5

Sociable Driving Agents to Maintain Driver's Attention in Automatic Driving

5.1 Introduction

Autonomous cars are seen as the key to a safer, more comfortable, more energy efficient and more convenient method of transportation. In recent years, many automotive manufacturers have been testing autonomous car systems to be able to use the fully autonomous cars in the near future [128], [100]. The National Highway Traffic Safety Administration (NHTSA) defines four levels of car autonomy: Level 1: function-specific automation, Level 2: combined function automation, Level 3: limited self-driving and Level 4: full self-driving [1]. Among these, Level 3 has significant importance due to its being that of the expected next generation of the vehicles [13]. In this stage of driving, the driver is not essentially required to monitor the road all the time; they can enjoy driving by engaging with non-driving-related activities. However, studies have demonstrated the effects of automation such as a loss of situational-awareness and overreliance on the increased level of automation [105], [34], [95]. Since in Level 3, there might be situations that the vehicle cannot handle (e.g. bad road conditions, increased traffic density), the driver should be available to take over the control within a sufficient transition time [1], [44]. In order to provide a quick and smooth handover, maintaining the drivers' attention on the road is crucial.

Recent studies in human factors focus on modalities such as visual, audio, speech and tactile to take the driver's attention efficiently and inform them about the handover process [84], [101]. However, each of these modalities or their combinations have been reported as more or less annoying for drivers [97], [72]. Due to the takeover request alerts the driver to react urgently in order to take an immediate reaction, in the case of a false alarm (e.g. a request is suggested although it is not necessary) the effectiveness and reliability of the system decreases and causes the rejection of usage of automated vehicle applications [14], [31]. In addition, announcing a take over request through these modalities is unilateral and is not perceived as considering the driver's stance. In contrast, our research focus is to maintain the driver's attention seamlessly and create a sufficient time to evaluate the current situation if it is necessary to take over the control.

An in-car system that would interact with the driver should be persuasive so that it can influence them in their actions or beliefs. Fogg remarked that one persuasive element in technology is the role of social actors [39]. Considering people tend to treat computer systems as if they are real people [103], and the tendency of human brain in anthropomorphizing the technology [10], it is not surprising that the automated vehicles becoming persuasive when they perform more anthropomorphic features [129], [53].

Studies investigating the influences of passengers on driving behavior showed that people tend to drive in a riskier manner when they are alone [36], while collaboration between a driver and a co-driver leads to increase in safety [46]. In this respect, it can be thought that a social entity would make a driver more alert. We believe that, a socially interactive robot could be useful in terms of the increased awareness of a driver.

Researchers have focused on developing robotic interfaces as in-car companions to deliver the necessary information and monitor the driver's state of alertness while interacting with the driver socially [131], [65]. These studies mainly focused on how to assist drivers without distracting them while driving. In contrast, in this current study, we investigate how to draw the driver's attention constantly within a social manner while they engage with non-driving related activities. The utterance of a robot can be an effective solution, however, when the subtask involves auditorial component, using the same mental source can impose the workload of the driver [130]. We believe that a visual stimuli that is more social and familiar with humans can seamlessly draw attention and keep awake the driver's situational-awareness.

Attention is defined as an increased awareness [17] and it has critical importance for goal-directed behaviors [124]. Joint attention attributes on attentional processes and is an essential skill in communication and interaction. Deák et al. defined the joint attention as simultaneously allocating attention to a target as a consequence of attending to each other's attentional states [28]. Responding to joint attention refers to the ability to follow the directions of gaze and gestures of others in order to share a reference.

Eye gaze is a crucial component of typical social interactions, in that humans use gaze to indicate attention to an object of mutual interest. There are researches to investigate the assistive driving systems draw on observing eye gaze behaviors of drivers [38], [32]. However, according to our knowledge, eye gaze has not been used as an influential component to effect the driver's attentional behaviors.

A number of studies in HRI showed that with head and eye gaze behaviors, robots can gain the ability to give information to their interlocutors. Research suggests that with visual attention cues, robots were able to define conversational roles (addressee, bystander, or eavesdropper) [83]. Another study demonstrated that the eye-gaze cues of robots influence people's decisions during a game [82]. Moreover, Das et al. showed that the head movements of a social robot, that is in the peripheral vision of a human, is effective and socially acceptable to attract the partner's attention [25]. Their study also showed that through these behaviors by the robots, the participants' attention could be drawn in a high success rate to initiate and establish an interaction. In light of these findings, we believe that shifting the attention of a driver who focuses on a secondary-task (non-driving activities) to primary-target (road) via eye gaze and head movements of a robotic interface could be seamless, nonirritating and socially acceptable.

In this study, our purpose was to understand if a social agent that can use its head movements and eye gazing can be an effective interface to maintain the drivers' attention seamlessly within a social manner that would make their handover process from automatic to manual mode of the vehicle efficient. We used NAMIDA, the sociable driving agents to maintain the driver's attention during the automation mode of the vehicle. While a driver is engaged with a non-driving related activity, NAMIDA robots periodically move their heads and direct their eye-gaze towards the driver to shift their attention constantly to the road. We expected that by this approach, the driver's attention would be attracted with the robots directing their eyes to the driver and this will induce to shift the driver's attention to the road. With this modality, drivers would be on alert that would make their handover process efficiently.

5.2 System Design

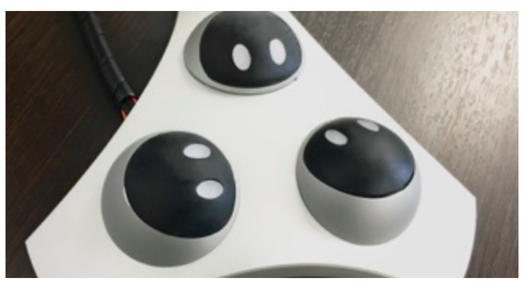

Figure 5-1: NAMIDA consists of three robots that is designed to be placed on dashboard on a car.

We built NAMIDA as it had been designed in our previous study ([65]). NAMIDA consists of one base unit that attaches to the dashboard of a car, containing three movable heads with one degree of freedom each in the driver's peripheral vision (Fig. 5-1). The round shape head of each NAMIDA allows for the positioning of their eyes with the full color LED light emission. The movement of the robots is enabled by three servomotors inside each head attached to the main board. NAMIDA has an utterance generation mechanism that allows the robots to generate verbal and non-verbal behaviors within each turn-taking [66]. However, in the current study, we focused on the effectiveness of NAMIDA system's particular non-verbal behaviors on the drivers' attention, thus we did not employ the utterance generation mechanism.

5.2.1 Attention Shifting of Driver

One definition of attention is intentionally perceiving the things that are relevant to the specific desires that correlate with the current goal [125]. The only way of understanding the intention of an agent is observing its behavior. Our aim was shifting the driver's attention to the NAMIDA's focus constantly. Establishing a joint attention between the driver and NAMIDA could help us to realize this aim. According to Kaplan and Hafner, reaching joint attention implies at least four kinds of prerequisites: attention detection, attention manipulation, social coordination and intentional stance [63]. For this study, NAMIDA has been designed as watching the road leading the driver to deduce that NAMIDA's attention focuses on the road (Fig. 2-7). When the driver doesn't look at the road, according to the rule in Algorithm 1, the robots turn their heads and gaze toward the driver. By this method, it was expected that a driver could detect NAMIDA's attentional behavior, and their attention could be manipulated by NAMIDA's gazing behaviors and would direct their gaze to NAMIDA's initial focus (road). The robots turn their heads in three

Algorithm 1: NAMIDA checks eye-gaze of the participant to generate its gazing behaviors. If the participant's eye gaze focuses on the simulation road, the *attention_counter* decreases; otherwise it increases. The *threshold* refers to approximately one minute that indicates the time duration of driver's attention not on the simulation screen.

```
procedure CHECK ATTENTION
    while attention_counter<threshold do
        if (attention == true) then
            attention_counter decrease
        else
            attention_counter increase
    namida_behavior.generate()
```

seconds, fix their eye gaze towards the driver for two and a half seconds and back to the initial position in three seconds. Even though response and feedback behaviors are necessary for joint attention [124] in a human robot interaction [54] to make robots more competent and socially interactive, however, in this study, we only focus on one component of joint attention which is recognizing the interlocutor's attention focus and shifting the driver's attention to that direction. The shifting attention process has been set as: (1) driver detects the attention focus of NAMIDA, (2) with the eye gaze behavior of NAMIDA, the driver believes that NAMIDA wants the driver's attention to shift towards its target (3) driver adopts the goal of joint the attention and performs a response-act (shifting attention to the road). This process has been repeated whenever the driver's attention stays on the non-driving related task according to the rule in Algorithm 1.

5.3 Method

Response time to the environmental changes is important in driving in order to create time to get the necessary action in the case of an abnormal situation. And response time is related to the mindset of drivers that correlates with their focus attention. If the driver's mindset is maintained to the road, then they will keep their alertness, otherwise, they cannot be aware of an emergency situation while they focus on non-driving related activities. NAMIDA has been designed as an interactive interface between the driver and the vehicle. The interface is expected to assist and support the driver based to the changing environmental conditions. In this study, the role of NAMIDA is to maintain the driver's attention to the road during the automation mode on the driving simulation (Fig. 5-2). With this experiment, we aimed to analyze the effectiveness of NAMIDA when an unexpected situation occurred on the road that the automation is unable to handle and performing a successful handover accordingly. Listed below are our hypotheses:

- H1: Drivers tend to drive more carefully when they collaborate with a co-driver [46]. In this respect, we hypothesize that the existence of NAMIDA interface, as a social actor, will maintain the driver's attention and lead the driver to be on alert such that in the case of encountering a critical situation on the road, they can notice the situation quickly (shorter response time) and take an action to avoid it. The response time of the driver will shorten by employing NAMIDA compared to not employing NAMIDA.

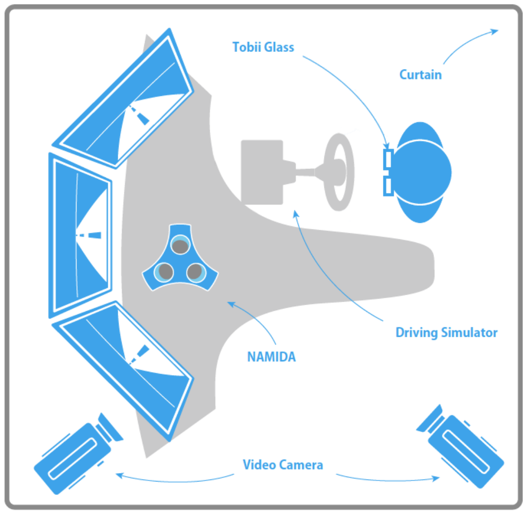

Figure 5-2: Figure depicts the experimental environment.

- H2: The head movements and eye gaze behaviors of robots are social stimulates that can draw attention from human and influence people's decisions respectively [25], [82]. We expected that NAMIDA with periodical head movements and gazing behaviors will draw attention constantly that it will ensure to maintain the driver's attention and lead to a shorter response time compared to NAMIDA without head movements and eye-gazing behaviors.

- H3: The head movements and eye-gazing behaviors of NAMIDA contribute its being of a social actor consequently, we expected that NAMIDA with gazing behaviors will be evaluated significantly more positively in terms of subjective impression and self-evaluation of the participants compared to NAMIDA without gazing behaviors.

We set an experiment to investigate the hypotheses above by comparing three conditions of driving: with no NAMIDA (NN), with NAMIDA but no non-verbal behaviors (NNB) and NAMIDA with non-verbal behaviors (NWB). We analyzed the attention states of the participants in each condition by analyzing the eye gaze behaviors of them utilizing by Tobii Pro Glasses 2[1] tracker. Moreover, we measured the participants subjective impressions about NAMIDA after the experiment for the NNB and NWB conditions.

5.3.1 Experimental Setup

Our aim was investigating the effectiveness of the NAMIDA system on maintaining the driver's attention on the road while they engaged with non-driving related actions (subtask) during the automated mode of a car. We set our experiment on a driving simulation environment (Fig. 5-3). The simulation consists of three monitors placed on the dashboard, an adjustable driver seat, a steering wheel, a break and an accelerator. The speed of the car was set to 70km/h. The experiment room was dimmed out to enhance the reality of the task. Participants' eye gaze behaviors were tracked by the Tobii Pro Glasses during the experiment. Moreover, two professional cameras were placed in the room to record all sessions.

5.3.2 Procedure

Figure 5-3: In the NWB condition, NAMIDA conducts eye gaze behaviors to draw attention from the driver.

24 participants attended our experiment (4 females, 20 males) from 19 to 40 years old (M=24.82, SD=6.31). We conducted a between-subject-study in order to avoid

[1]https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

influences of learning effects. All participants held a current driving license. The participants were divided into three groups randomly such that each condition had 8 participants. Before the experiment, the participants were given a demographic questionnaire and then they were told that if they see any abnormal situation on the road, they should take over the control from the automation and manage the car manually. Below explains the three conditions of the experiment:

- No NAMIDA (NN): NAMIDA was not placed on the dashboard and the participants did not receive any support to maintain their attention.
- NAMIDA with no behavior (NNB): NAMIDA was placed on the dashboard and the robots were always watching the front. They did not expose non-verbal behavior to take the driver's attention.
- NAMIDA with behavior (NWB): NAMIDA was placed on the dashboard and the robots were watching the front. When the driver's attention was not on the road according to the rule mentioned in Algorithm 1, the three robots gazed at the driver.

Before starting each session, the participants undertook trial driving for five minutes without having a subtask. The aim of this was for adaption of the participants to the driving environment. The driver's primary task was being aware of the environment and to decide if it is necessary to take over control from the automation, and, meanwhile, as a subtask they were asked to watch the movie being played on the tablet PC placed in the front (Fig. 5-3). The movie was showed from the beginning for each participant. With this way, it is aimed to show the same points of the movie for all participants while they exposed the same parts of the simulation. After the trial, the 20 minute sessions started.

A collision of two cars with a fire effect was set into the driving simulation on the last minute of the session. The collision was visible 11.5 seconds before arrival at the incident area. In NWB condition, the Algorithm 1 was terminated before coming to this point in order to avoid from unfair response time results among the

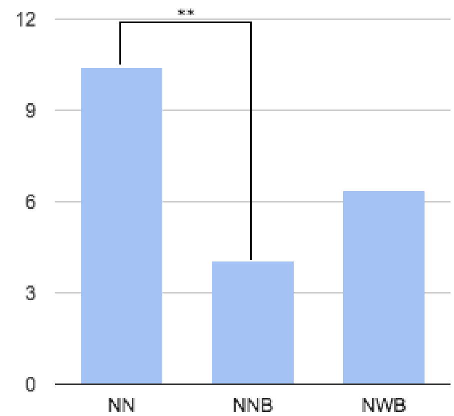

Figure 5-4: The results from our objective measures. The existence of NAMIDA effected participant's response time. ** indicates p-value smaller than .01.

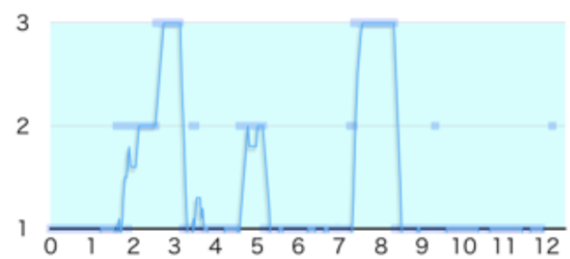

Figure 5-5: Figure shows a section of the gazing behavior of a participant (1: subtask, 2: NAMIDA, 3: simulation road).

Participant No	NAMIDA Gazing Number	Participant's Response	Response rate
1	17	17	100%
2	18	17	94.44%
3	9	9	100%
4	18	18	100%
5	17	16	94.12%
Average	**15.8**	**15.4**	**97.712%**

Figure 5-6: Eye gaze behaviors of NAMIDA could divert attention from the driver with a 97% success rate.

participants. There was no sign about the collision's existence throughout the road. At this point, the automation level automatically changed into the manual mode without any notification. If the participant was not aware of the collision, they could not avoid it, in result, the car would crash. On the contrary, if the driver could notice the existence of the collision on the road beforehand, they could avoid from the collision by controlling the car. All participants were asked to wear the eye tracking glasses, thus we could collect their eye-gaze behaviors and analyze their attention

Code	Questions	Mean value (standard error)		Pairwise t-test t(d.f.) = t-value
		NNB	NWB	
Q1	Did you think that NAMIDA was moving with its own intention?	2.25 (1.35)	3.89 (1.12)	**P = 0.00817<0.01
Q2	Did NAMIDA seem to care about you?	1.75 (0.92)	4.11 (1.39)	**P = 0.00001 <0.01
Q3	Was NAMIDA trying to get involve with you?	1.75 (096)	3.22 (1.23)	**P = 0.000854<0.01
Q4	Did you want to actively involve with NAMIDA?	2.13 (1.25)	3.00(1.00)	P = 0.12927 >0.05
Q5	Did you feel that you communicated with NAMIDA?	1.63 (0.74)	2.67 (0.88)	*P = 0.01848 <0.05
Q6	Did you want to interact more with NAMIDA?	2.25 (1.40)	1.89 (0.99)	P = 0.27714 <0.05

Table 5.1: Table shows the subjective impression and self-evaluation questions and results.

status in real-time.

5.4 Results

5.4.1 Response time to accident

In this experiment, our aim was to understand if NAMIDA could be an effective interface to maintain the driver's mindset during the automatic mode of a vehicle. The main goal for this section was how quickly the participants could notice the accident on the road and avoid crashing into it. We analyzed the eye gaze fixation by utilizing Tobii Pro Glasses of the participants to understand when they could notice the accident. We designated an invisible point on the simulated road for the measurements. The collision became visible at that point, thus if the participant was already looking at the screen at that time, the response time was recorded as 0s. Likewise, the response time was recorded as 11.5s if the participant collided. Fig. 5-4 shows the result of participants' recognition time of the collision.

There were three subjects whose eye-gaze data was damaged, thus the results were calculated for 21 participants out of 24. A one-way between subjects ANOVA was conducted to compare the response time for NN, NNB and NWB conditions. There was a significant effect on response time at the $p<.05$ level for the three conditions

[F(2,21)=4.47, p=0.024]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the NN condition (M = 10.43, SD = 3.87) was significantly different than the NNB condition (M = 4.03, SD = 4.69). However, the pytNWB condition (M = 6.37, SD = 4.92) did not significantly differ from NN and NNB conditions. According to these results, our first hypothesis were verified. (Fig. 5-4) shows that the longest response time to the accident was found in the NN condition. We confirmed from the movie, which was recorded from the participant's point of view through eye-gaze tracker, that two participants in the NN condition were not aware of the accident until they crashed. Without having NAMIDA, they concentrated on the subtasks and it turned out that it became too late to notice the accident. Among the NNB and NWB conditions, the reaction time was shorter in the NNB unlike our prediction; our hypothesis for this case was not verified. The results showed that four out of eight participants in NNB condition were already focused on the road when the collision became visible on the screen which means the participants focused on the road more than the subtask. According to these results, we could infer that the existence of NAMIDA contributes to maintaining the driver's attention.

5.4.2 Eye-gaze Data Analysis

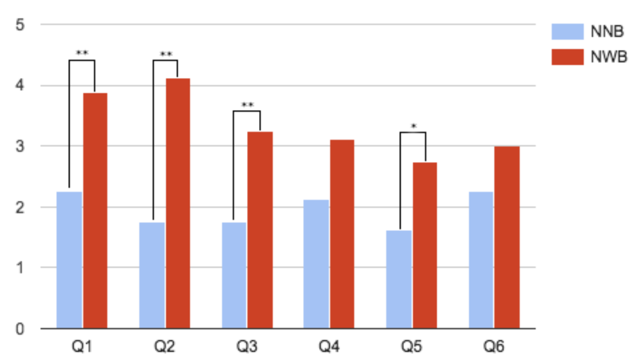

Figure 5-7: The results from our subjective measures. Gazing behavior of NAMIDA affected participant's impression and self-evaluation. ** and * indicates p-values smaller than .01 and .05, respectively.

In order to understand the interaction between the participants and NAMIDA in the NWB case, we analyzed the five participants' eye gaze behaviors in detail. An example of a movement of the eye-gaze data of a participant is shown in Fig. 5-5. NAMIDA's gazing behavior from start to end (turning heads to the driver, fixation and turning back to the initial position) takes 8.5 seconds. Fig. 5-5 shows that after NAMIDA starts to move, the participant's eye gaze shifted to NAMIDA and then the simulation road and then the subtask, NAMIDA and simulated road, and then road again. Fig. 5-6 shows how many times NAMIDA's gazing behaviors activated and how many times the participants responded to NAMIDA's gazing behaviors by shifting their eye gaze. According to these results, NAMIDA could shift the participant's attention with 97% success rate as seen in Fig. 5-6.

5.4.3 Subjective Impression and Self-Evaluation Analysis

The impression evaluation questionnaire was given to the participants who attended to NNB and NWB conditions. Questionnaire responses were scored on a 5-point Likert Scale: 1 (Strongly Disagree), 2 (Disagree), 3 (Neutral), 4 (Agree), 5 (Strongly Agree). The questions from Q1 to Q3 were evaluating the participants' impression about NAMIDA whilst those from Q4 to Q6 were about the participant's individual evaluation towards the NAMIDA system. We applied t-test to analyze this part. The results showed that there were significant differences in the first three question items (**$p=0.00817<0.01$, **$p=0.0001<0.01$, **$p=0.000854<0.01$) (Fig. 5.1). With the eye gaze behaviors of the robots, participants could get the robots' intention. Also, these gazing behaviors could lead the participants to feel of been cared and considered. Since the Cronbach's Alpha for these three items is 0.919 which is above the suggested 0.7 threshold, [90], we can say that the participants' impression on the NAMIDA questionnaire are internally consistent. We conducted a t-test to compare the participant's impression for the NNB and NWB conditions. There was a significant effect at the $p<.001$ level ($t(23)$, $P=0.0000000208<0.001$; $M=1.91$, $SD=1.05$; $M=3.75$, $SD=0.84$ respectively).

The results of Q4, Q5 and Q6 were relatively lower than the first three questions. As it can be seen from Fig. 5.1, significant differences were not found in Q4 and Q6 ($p=0.12927>0.05$, $p=0.27714>0.05$, respectively). One of the major reasons for this might be that we adopted minimal design of modality to express the NAMIDA behaviors. Using only eye-gaze behaviors might be sufficient to give a positive impression to a driver, however it might not be enough to improve their relationship with NAMIDA. However, the significant difference in Q5 (* $p = 0.01848 < 0.05$) suggests that it was more possible to initiate a connection between a driver and NAMIDA in the NWB condition compared to the NNB condition. The Cronbach's Alpha for these three items is 0.896 in which means that the internal consistency is reliable among the questions. We conducted a t-test to compare the participant's self-evaluation. There was a significant effect the $p<.001$ level for the NNB and NWB conditions ($t(23)$, $P=0.0028<0.001$; $M=2$, $SD=1.17$; $M=2.79$, $SD=0.88$ respectively).

5.5 Discussion

In this research, we investigated the effectiveness of gazing behaviors of our social robot platform NAMIDA to maintain the driver's attention on the road during the automation mode of a vehicle. The proposed design of the system employs the gazing behaviors according to the driver's gazing behaviors. We hypothesized that a driver could maintain their attention on the road better with NAMIDA, thus they can keep their situational-awareness and take quicker action in the case of an incident that cannot be handled by the automation mode of the vehicle. According to the results in 5.4.1, our hypothesis was verified. Research claimed that existence of another would help drivers to be more careful [46]. Considering the tendency of treating technology as if they were real people [104], likewise, we can infer that, having NAMIDA system in the car could help drivers to be more careful and keep their situation-awareness more than in a system that does not have NAMIDA.

Our second hypothesis was that NWB would lead a shorter reaction time of drivers than NNB. However, the results showed that participants paid more attention to the road in the NNB condition. In the NNB case, participants focused on the road more than the subtask. This situation made the driver aware of the collision on the road earlier. A reason for this can be that as being aware of the existence of NAMIDA as three social entities, the participants were effected by the influences of social conformity in the environment [22]. Nonetheless, this kind behavior does not comply with the automation mode of a vehicle. Automated vehicles are supposed to provide more time for drivers to enjoy their ride without having so much constant obligation to exert attention. In the NWB case, the drivers could pay attention to the subtask while focusing on the road with the NAMIDA's eye-gazing behaviors which created more time to enjoy their ride. Fig.5-6 shows the responses of five participants to the eye gaze behaviors of NAMIDA. With a high success rate, NAMIDA's movements could divert the participants' attention and made it shift to the road from the subtask.

We also expected that NWB would have a better subjective impression and self-evaluation than NNB. Our hypothesis was verified. When we examine each question one by one, the results showed that NAMIDA's behaviors in the NWN condition have been perceived as having significantly more comprehendible in their representation of intention, caring, involving and communication compared to NNB. On the other hand, the participants' desire to be involved in an interaction with NAMIDA did not show significant difference. One reason for this can be the limited non-verbal behaviors of the robots. We employed only one modality (eye-gazing) for NAMIDA to understand the participant's reactions within a basic level. Non-verbal behaviors coupled with verbal behaviors could make the system be perceived as more communicable and interactable.

5.6 Conclusion

In this study we proposed an interaction modality for our social robot platform NAMIDA to maintain the driver's attention during the semi-automation mode (Level-3) of a vehicle. Without having NAMIDA, participants could not allocate attention to the road and it turned out that it became too late when they noticed a collision on the road. On the other hand, NAMIDA with watching the road behavior without having the gazing behaviors (NNB) led the participants to be more alert on the road, so that most of participants in the NNB condition were already focused on the road when the collision became visible on the screen. Maybe it was a result of a social conformity in the ambient which we will investigate it in our future study in detail. This might lead the reaction time to become shorter in the NNB compared to the NBW condition unlike our prediction. On the other hand, the gazing behaviors of NAMIDA in the NWB condition created more time for a driver to enjoy their ride. Furthermore, NAMIDA could shift the participant's attention with a 97% success rate which demonstrated the potential effectiveness of the gazing behaviors of NAMIDA. According to these results, we could infer that the existence of NAMIDA contributes to maintaining the driver's attention.

Subjective results also showed that the participants' impressions of and their own individual evaluation towards NWB evaluated significantly better than NNB. However, maybe, one modality (eye-gazing behavior) for NAMIDA induced the perception of being less communicable. Our future study will also involve the inclusion of other modalities such as utterance generation and turn-taking modules to increase the sociability and enjoyment of automated driving.

Chapter 6

Conclusion

With the advancement of the technology, the interaction design of the dashboard of the cars have been changed a lot in the last years. The amount of buttons have been becoming increasingly confusing and sensory overload. On the other hand, a great deal of research has been conducted on highly autonomous vehicles which make their own driving decisions that minimise human interventions with the vision of decreasing human errors and achieving a safer, more energy efficient and more comfortable mode of transportation. The autonomous cars hold much more functionalities compared to the cars on the road today. Therefore, the design of the dashboard for the expected self-driverless cars should be created in a way that intuitively understandable by the wide range of users considering their naivety on the technology and cultural background, etc. It has been claimed that the human brain has evolved to be highly adaptive in social interactions therefore, people tend to anthropomorphise the technology. From the drivers' perspective, we believe that it is crucial to interact with an in-vehicle interface system in such a social, natural and familiar manner to reduce mental workload and create a more sociable environment inside a car.

Human brain has been evolved to be an expert in social interactions to keep their existence. Social robots are envisioned as having the ability to interact with others socially in order to achieve their goals. With this respect, a social robot platform what would mediate the interactions between a car and a driver can be effective in terms of obtaining environmental information and understanding the vehicle's intentions while interacting with the driver socially. In this thesis, we address several problems regarding to the interaction between such a social interface and a driver.

The proposed multi-party conversation based interface of NAMIDA (virtually embodied) presents a unique interaction between the car and driver. As a social interface, it has been designed to assist drivers by conducting a context-aware interaction during driving. We believe that this conversation approach is enjoyable as it requires less attention in obtaining necessary information. We designed an experiment to verify our hypothesis by comparing two different cases, MPCN and OOCN, in a mock driving environment. In the current research, we examined the mental workload, attention behaviors and subjective impression of drivers by comparing a multi-party conversation-based system with a one-to-one conversation-based system.

We evaluated our proposed system using a DALI questionnaire, a trend analysis of the eye-gaze data gathered during the experiments and a subjective impression questionnaire. The results of DALI revealed that even though the MPCN cannot fulfill all the workload factors, it induced less cognitive and perceptive components of workload. That is, overhearing the location-based information via a conversation between the sociable agents required significantly less attentional, visual and auditory efforts. It has been also shown that, MPCN required less eye gaze behavior during the experimental conditions. The trend analysis demonstrated that our proposed multi-party conversation-based system is promising in reducing the attention behavior on the system over use. Through the turn-taking based lively conversation, the MPCN system exhibited an enjoyable performance such that according to the subjective impression ratings, it had significantly more human-like and animated behaviors, and natural conversation aspects than the OOCN system. For a more enjoyable and sociable environment inside a car, our next study will involve the driver in the multi-party conversation by considering the real-time condition (in behavioural and workload aspect) of the driver. Our future study is also required to generalize the results and investigate the different aspects of the multi-party conversation on drivers during a real-time interaction with physically developed robotic agents.

The multi-party conversation-based interface of NAMIDA presents a unique interaction between the car and driver. As a social interface, it has been designed to assist drivers by conducting a multi-party conversation This study showed that this conversational approach has the potential to be perceived as more lifelike and enjoyable as it possess a lively conversation of robots that requires less attention from the driver in obtaining necessary information. We designed an experiment to verify our hypothesis with our NAMIDA platform by comparing two different cases, the MPCN and OOCN, in a realistic driving simulation environment.

We evaluated the interactions between the driver and the NAMIDA platform in both conditions using the subjective questionnaires that examined the system's lifelikeness, enjoyability, annoyance, perceived conversational burden and perceived distraction; the objective analysis that examined the eye-gaze data gathered during the experiments. The results of the subjective questionnaires revealed that in the MPCN condition, the robots created more anthropomorphic feelings and were perceived as more likable and having more animacy with their coherent utterance and turn-taking actions. However, the order of the experimental conditions might have an influence on the anthropomorphism and the animacy ratings where the participants rated higher the MPCN when they interacted with the OOCN before. Moreover, the objective eye gaze analysis revealed that overhearing information via a conversation among the sociable agents required significantly less attentional efforts. The subjective analysis also supports this result by showing that in the MPCN condition, the perceived conversational burden and perceived distraction were observed as significantly less than in the OOCN condition. Through the turn-taking based lively conversation of the robots where the driver, as an overhearer, is not compelled to contribute to the conversation, the MPCN system exhibited a more stress-free and enjoyable driving environment.

For a more sociable environment inside a car, our next study will focus on the driver in the multi-party conversation by considering an adaptive approach toward the driver's needs. The future study is also required to further generalize the results and investigate the different aspects of the multi-party conversation platform on drivers during a real-world vehicle experiment.

In our another study, we proposed an autonomous car-human operator interaction paradigm in order to achieve a reliable interaction with an autonomous car, such that the autonomous system and the human could sense each other's intentions and be aware of each other's presence. The results of this pilot study showed that perceiving an autonomous car as a social entity through the gaze following behaviors of a driving agent platform was possible and has the potential to improve the perceived safety and enjoyment of the autonomous driving system. Future studies will investigate methods to improve the relationship between an autonomous car and human operator in terms of increasing the perceived safety, trust and the pleasure of autonomous driving.

The interaction modality for our social robot platform NAMIDA to maintain the driver's attention during the semi-automation mode (Level-3) of a vehicle. Without having NAMIDA, participants could not allocate attention to the road and it turned out that it became too late when they noticed a collision on the road. On the other hand, NAMIDA with watching the road behavior without having the gazing behaviors (NNB) led the participants to be more alert on the road, so that most of participants in the NNB condition were already focused on the road when the collision became visible on the screen. Maybe it was a result of a social conformity in the ambient which we will investigate it in our future study in detail. This might lead the reaction time to become shorter in the NNB compared to the NBW condition unlike our prediction. On the other hand, the gazing behaviors of NAMIDA in the NWB condition created more time for a driver to enjoy their ride. Furthermore, NAMIDA could shift the participant's attention with a 97% success rate which demonstrated the potential effectiveness of the gazing behaviors of NAMIDA. According to these results, we could infer that the existence of NAMIDA contributes to maintaining the driver's attention.

Subjective results also showed that the participants' impressions of and their own individual evaluation towards NWB evaluated significantly better than NNB. However, maybe, one modality (eye-gazing behavior) for NAMIDA induced the perception of being less communicable. Our future study will also involve the inclusion of other modalities such as utterance generation and turn-taking modules to increase the sociability and enjoyment of automated driving.

Bibliography

[1] National Highway Traffic Safety Administration et al. Preliminary statement of policy concerning automated vehicles. *Washington, DC*, pages 1–14, 2013.

[2] Spannan E. Mutlu B. Andrist, S. Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 341–348. IEEE Press, 2013.

[3] T Arimoto, Y Yoshikawa, and H Ishiguro. Cooperative use of multiple robots for enhancing sense of conversation without voice recognition. *SIG-SLUD*, 5(2):76–77, 2015.

[4] Tsunehiro Arimoto, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Multiple-robot conversational patterns for concealing incoherent responses. *International Journal of Social Robotics*, pages 1–11, 2018.

[5] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments*, 10(6):583–598, 2001.

[6] Mikhail M Bakhtin. The problem of speech genres (vern w mcgee, övers.) i caryl emerson & michael holquist (red.), speech genres & other late essays (ss. 60-102), 1986.

[7] Simon Baron-Cohen. *The evolution of a theory of mind.* na, 1999.

[8] Simon Baron-Cohen, Ruth Campbell, Annette Karmiloff-Smith, Julia Grant, and Jane Walker. Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, 13(4):379–398, 1995.

[9] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[10] Dunbar R. M. Barton, R. A. 9 evolution of the social brain. *Machiavellian intelligence II: Extensions and evaluations*, 2:240, 1997.

[11] Beatrice Beebe, Steven Knoblauch, Judith Rustin, Dorienne Sorter, Theodore J Jacobs, and Regina Pally. *Forms of intersubjectivity in infant research and adult treatment.* Other Press, 2005.

[12] Bernard Berelson. Communications and public opinion. 1950.

[13] Jane Bierstedt, Aaron Gooze, Chris Gray, Josh Peterman, Leon Raykin, and Jerry Walters. Effects of next-generation vehicles on travel demand and highway capacity. *FP Think Working Group*, pages 10–11, 2014.

[14] James P Bliss and Sarah A Acton. Alarm mistrust in automobiles: how collision alarm reliability affects driving. *Applied ergonomics*, 34(6):499–509, 2003.

[15] C. L. Breazeal. *Designing sociable robots.* MIT press, 2004.

[16] Cynthia Lynn Breazeal. *Sociable machines: Expressive social exchange between humans and robots.* PhD thesis, Massachusetts Institute of Technology, 2000.

[17] Ingar Brinck. The objects of attention. *European Society of Philosophy and Psychology*, 2003.

[18] Rechele Brooks and Andrew N Meltzoff. The development of gaze following and its relation to language. *Developmental science*, 8(6):535–543, 2005.

[19] Penelope Brown and Stephen C Levinson. *Politeness: Some universals in language usage*, volume 4. Cambridge University Press, 1987.

[20] Judee K Burgoon, Joseph A Bonito, Artemio Ramirez, Norah E Dunbar, Karadeen Kam, and Jenna Fischer. Testing the interactivity principle: Effects of mediation, propinquity, and verbal and nonverbal modalities in interpersonal interaction. *Journal of communication*, 52(3):657–676, 2002.

[21] George Butterworth and Nicholas Jarrett. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British journal of developmental psychology*, 9(1):55–72, 1991.

[22] Robert B Cialdini and Melanie R Trost. Social influence: Social norms, conformity and compliance. 1998.

[23] Noel K Clark and Derek R Rutter. Social categorization, visual cues, and social judgements. *European Journal of Social Psychology*, 15(1):105–119, 1985.

[24] Daniela Conti, Santo Di Nuovo, Serafino Buono, and Alessandro Di Nuovo. Robots in education and care of children with developmental disabilities: a study on acceptance by experienced and future professionals. *International Journal of Social Robotics*, 9(1):51–62, 2017.

[25] Dipankar Das, Md Golam Rashed, Yoshinori Kobayashi, and Yoshinori Kuno. Supporting human–robot interaction based on the level of visual focus of attention. *IEEE Transactions on Human-Machine Systems*, 45(6):664–675, 2015.

[26] Kerstin Dautenhahn and Aude Billard. Bringing up robots or—the psychology of socially intelligent robots: From theory to implementation. In *Proceedings of the third annual conference on Autonomous Agents*, pages 366–367. ACM, 1999.

[27] Kerstin Dautenhahn and Aude Billard. Studying robot social cognition within a developmental psychology framework. In *Advanced Mobile Robots, 1999.(Eurobot'99) 1999 Third European Workshop on*, pages 187–194. IEEE, 1999.

[28] Gedeon O Deák, Ian Fasel, and Javier Movellan. The emergence of shared attention: Using robots to test developmental theories. In *Proceedings 1st International Workshop on Epigenetic Robotics: Lund University Cognitive Studies*, volume 85, 2001.

[29] Daniel Clement Dennett. *The intentional stance.* MIT press, 1989.

[30] J. F. DiMarzio. *Android.* Tata McGraw-Hill Education, 2008.

[31] Thomas A Dingus, Steven K Jahns, Avraham D Horowitz, and Ronald Knipling. Human factors design issues for crash avoidance systems. *Human factors in intelligent transportation systems*, pages 55–93, 1998.

[32] Anup Doshi and Mohan Manubhai Trivedi. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):453–462, 2009.

[33] David M Driscoll, Scotty D Craig, Barry Gholson, Matthew Ventura, Xiangen Hu, and Arthur C Graesser. Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, 29(4):431–450, 2003.

[34] Mica R Endsley. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3):462–492, 1999.

[35] Naomi T Fitter, Heather Knight, Nikolas Martelaro, and David Sirkin. What actors can teach robots. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 574–580. ACM, 2017.

[36] Judy J Fleiter, Alexia Lennon, and Barry Watson. How do other people influence your driving speed? exploring the 'who'and the 'how'of social influences on speeding from a qualitative perspective. *Transportation research part F: traffic psychology and behaviour*, 13(1):49–62, 2010.

[37] Frank O Flemisch, Catherine A Adams, Sheila R Conway, Ken H Goodrich, Michael T Palmer, and Paul C Schutte. The h-metaphor as a guideline for vehicle automation and interaction. 2003.

[38] Luke Fletcher and Alexander Zelinsky. Driver inattention detection based on eye gaze - road event correlation. *The international journal of robotics research*, 28(6):774–801, 2009.

[39] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002.

[40] Thompson S. A. Ford, C. E. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. *Studies in interactional sociolinguistics*, 13:134–184, 1996.

[41] Jodi Forlizzi and Carl DiSalvo. Service robots in the domestic environment: a study of the roomba vacuum in the home. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 258–265. ACM, 2006.

[42] Alexandra Frischen, Andrew P Bayliss, and Steven P Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4):694, 2007.

[43] E. Goffman. Footing. *Semiotica*, 25(1-2):1–30, 1979.

[44] Christian Gold, Daniel Damböck, Lutz Lorenz, and Klaus Bengler. "take over!" how long does it take to get the driver back into the loop? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 1938–1942. SAGE Publications Sage CA: Los Angeles, CA, 2013.

[45] Charles Goodwin. *Conversational organization: Interaction between speakers and hearers*. Academic Press, 1981.

[46] Nicole Gridling, Alexander Meschtscherjakov, and Manfred Tscheligi. I need help!: exploring collaboration in the car. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, pages 87–90. ACM, 2012.

[47] Jaap Ham and Cees JH Midden. A persuasive robot to stimulate energy conservation: the influence of positive and negative social feedback and task similarity on energy-consumption behavior. *International Journal of Social Robotics*, 6(2):163–171, 2014.

[48] Staveland L. E. Hart, S. G. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[49] Kotaro Hayashi, Takayuki Kanda, Takahiro Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. Robot manzai: Robot conversation as a passive–social medium. *International Journal of Humanoid Robotics*, 5(01):67–86, 2008.

[50] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. The influence of social presence on acceptance of a companion robot by older people. 2008.

[51] Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, pages 243–259, 1944.

[52] Pamela J Hinds, Teresa L Roberts, and Hank Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1):151–181, 2004.

[53] Philipp Hock, Johannes Kraus, Marcel Walch, Nina Lang, and Martin Baumann. Elaborating feedback strategies for maintaining automation in highly automated driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 105–112. ACM, 2016.

[54] Chien-Ming Huang. *Joint attention in human-robot interaction.* PhD thesis, Georgia Institute of Technology, 2010.

[55] Chien-Ming Huang and Andrea L Thomaz. Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *RO-MAN, 2011 IEEE*, pages 65–71. IEEE, 2011.

[56] Takamasa Iio, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Pre-scheduled turn-taking between robots to make conversation coherent. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, pages 19–25. ACM, 2016.

[57] Takamasa Iio, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Starting a conversation by multi-robot cooperative behavior. In *International Conference on Social Robotics*, pages 739–748. Springer, 2017.

[58] Takahiro Ishikawa. Passive driver gaze tracking with active appearance models. 2004.

[59] Kato T. Ishizaki, M. Exploring the characteristics of multi-party dialogues. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 583–589. Association for Computational Linguistics, 1998.

[60] Shoji Itakura. An exploratory study of gaze-monitoring in nonhuman primates 1. *Japanese Psychological Research*, 38(3):174–180, 1996.

[61] Reiko Itani. *Semantics and pragmatics of hedges in English and Japanese*. PhD thesis, University of London, 1995.

[62] Frédéric Kaplan and Verena Hafner. The challenges of joint attention. 2004.

[63] Frederic Kaplan and Verena V Hafner. The challenges of joint attention. *Interaction Studies*, 7(2):135–169, 2006.

[64] Nihan Karatas, Nozomi Kina, Daiki Tanaka, Naoki Ohshima, P Ravindra S De Silva, and Michio Okada. Designing sociable culot as a playground character. In *International Conference on Universal Access in Human-Computer Interaction*, pages 571–580. Springer, 2014.

[65] Nihan Karatas, Soshi Yoshikawa, P Ravindra S De Silva, and Michio Okada. Namida: Multiparty conversation based driving agents in futuristic vehicle. In *Human-Computer Interaction: Users and Contexts*, pages 198–207. Springer, 2015.

[66] Nihan Karatas, Soshi Yoshikawa, and Michio Okada. Namida: Sociable driving agents with multiparty conversation. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, pages 35–42. ACM, 2016.

[67] Nihan Karatas, Soshi Yoshikawa, Shintaro Tamura, Sho Otaki, Ryuji Funayama, and Michio Okada. Sociable driving agents to maintain driver's attention in autonomous driving. In *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*, pages 143–149. IEEE, 2017.

[68] Yasuhiro Katagiri, Toru Takahashi, and Yugo Takeuchi. Social persuasion in human-agent interaction. In *Second IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, pages 64–69. Citeseer, 2001.

[69] Sara Kiesler, Aaron Powers, Susan R Fussell, and Cristen Torrey. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition*, 26(2):169–181, 2008.

[70] S. G. Klauer, F. Guo, and Ouimet M. C.-Lee S.E. Dingus T. A. Simons-Morton, B. G. Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, 370(1):54–59, 2014.

[71] Hideki Kozima, Yuriko Yasuda, and Cocoro Nakagawa. Social interaction facilitated by a minimally-designed robot: Findings from longitudinal therapeutic practices for autistic children. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 599–604. IEEE, 2007.

[72] John D Lee, Joshua D Hoffman, and Elizabeth Hayes. Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 65–72. ACM, 2004.

[73] Victor Lee and P das Gupta. Children's cognitive and language development. 1996.

[74] Geoffrey Leech. Politeness: is there an east-west divide. *Journal of Foreign Languages*, 6(3), 2005.

[75] Christina Lichtenthäler, Tamara Lorenzy, and Alexandra Kirsch. Influence of legibility on perceived safety in a virtual human-robot path crossing task. In *RO-MAN, 2012 IEEE*, pages 676–681. IEEE, 2012.

[76] Andrea Lockerd and Cynthia Breazeal. Tutelage and socially guided robot learning. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 4, pages 3475–3480. IEEE, 2004.

[77] BYD Company Ltd. Qin:asocialrobotforanewplug-inhybrid vehicle, 2012.

[78] Fujii H. Goan M. Okada-M. Matsumoto, N. Minimal design strategy for embodied communication agents. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 335–340. IEEE, 2005.

[79] Natasha Merat, A Hamish Jamson, Frank CH Lai, Michael Daly, and Oliver MJ Carsten. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, 27:274–282, 2014.

[80] AJung Moon, Chris AC Parker, Elizabeth A Croft, and HF Machiel Van der Loos. Design and impact of hesitation gestures during human-robot resource conflicts. *Journal of Human-Robot Interaction*, 2(3):18–40, 2013.

[81] Youngme Moon and Clifford Nass. How "real" are computer personalities? psychological responses to personality types in human-computer interaction. *Communication research*, 23(6):651–674, 1996.

[82] Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro 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*, pages 69–76. ACM, 2009.

[83] Shiwa T. Kanda T. Ishiguro-H. Hagita N. Mutlu, B. 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*, pages 61–68. ACM, 2009.

[84] Frederik Naujoks, Christoph Mai, and Alexandra Neukum. The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Advances in Human Aspects of Transportation*, 7(Part I):431, 2014.

[85] Feels Like Driving in Future Navdy Inc. https://www.navdy.com/, 2013.

[86] Young-Hoon Nho, Ju-Hwan Seo, Jeong-Yean Yang, and Dong-Soo Kwon. Driving situation-based real-time interaction with intelligent driving assistance agent. In *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on*, pages 480–485. IEEE, 2015.

[87] In Automotto Nissan Motor Company Ltd., PIVO: An in-car robotic assistant, 2005.

[88] DA Norman. 18. the design of everyday things, 1990.

[89] Don Norman. *The design of everyday things: Revised and expanded edition.* Constellation, 2013.

[90] JC Nunnally. Psychometric theory (2nd edit.) mcgraw-hill. *Hillsdale, NJ*, 1978.

[91] Naoki Ohshima, Yusuke Ohyama, Yuki Odahara, P Ravindra S De Silva, and Michio Okada. Talking-ally: The influence of robot utterance generation mechanism on hearer behaviors. *International Journal of Social Robotics*, 7(1):51–62, 2015.

[92] Michio Okada. Muu: Artificial creatures as an embodied interface. In *Proceedings of the 27th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000)*, 2000.

[93] Yoshikazu Okajima, Hiroyuki Masuta, Masatoshi Okumura, Tatsuo Motoyoshi, Ken'ichi Koyanagi, Toru Oshima, and Eiichi Takayama. Utterance generation based on driving evaluation for driving assistance robot. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 20(5):836–844, 2016.

[94] Yadong Pan, Haruka Okada, Toshiaki Uchiyama, and Kenji Suzuki. On the reaction to robot's speech in a hotel public space. *International Journal of Social Robotics*, 7(5):911–920, 2015.

[95] Raja Parasuraman and Victor Riley. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2):230–253, 1997.

[96] Domenico Parisi and Matthew Schlesinger. Artificial life and piaget. *Cognitive Development*, 17(3-4):1301–1321, 2002.

[97] Roy D Patterson. *Guidelines for auditory warning systems on civil aircraft.* Civil Aviation Authority, 1982.

[98] Annie Pauzié, J Manzan, and Nicolas Dapzol. Driver's behavior and workload assessment for new in-vehicle technologies design. In *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Stevenson, Washington*, 2007.

[99] Pioneer and iXsResearch Corp. Carnaby:anin-carrobotic navigation assistant, 2010.

[100] Sharon L Poczter and Luka M Jankovic. The google car: driving toward a better future? *Journal of Business Case Studies (Online)*, 10(1):7, 2014.

[101] Ioannis Politis, Stephen Brewster, and Frank Pollick. Language-based multi-modal displays for the handover of control in autonomous cars. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 3–10. ACM, 2015.

[102] A. Rakotonirainy. Design of context-aware systems for vehicle using complex systems paradigms. 2005.

[103] Byron Reeves and Clifford Nass. *How people treat computers, television, and new media like real people and places.* CSLI Publications and Cambridge university press, 1996.

[104] Byron Reeves and Clifford Ivar Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press, 1996.

[105] Christina M Rudin-Brown and Heather A Parker. Behavioural adaptation to adaptive cruise control (acc): implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(2):59–76, 2004.

[106] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking for conversation. *language*, pages 696–735, 1974.

[107] Hayashi K. Kanda T. Shiomi-M. Koizumi S. Ishiguro H. Ogasawara T. Hagita N. Sakamoto, D. Humanoid robots as a broadcasting communication medium in open public spaces. *International Journal of Social Robotics*, 1(2):157–169, 2009.

[108] Eva-Lotta Sallnäs, Kirsten Rassmus-Gröhn, and Calle Sjöström. Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(4):461–476, 2000.

[109] Allison Sauppé and Bilge Mutlu. The social impact of a robot co-worker in industrial settings. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3613–3622. ACM, 2015.

[110] Steven J Sherman, Karin Ahlm, Leonard Berman, and Steven Lynn. Contrast effects and their relationship to subsequent behavior. *Journal of Experimental Social Psychology*, 14(4):340–350, 1978.

[111] Gabriel Skantze, Anna Hjalmarsson, and Catharine Oertel. Turn-taking, feedback and joint attention in situated human–robot interaction. *Speech Communication*, 65:50–66, 2014.

[112] Dan Sperber and Deirdre Wilson. Précis of relevance: Communication and cognition. *Behavioral and brain sciences*, 10(4):697–710, 1987.

[113] Rebecca Q Stafford, Bruce A MacDonald, Xingyan Li, and Elizabeth Broadbent. Older people's prior robot attitudes influence evaluations of a conversational robot. *International journal of social robotics*, 6(2):281–297, 2014.

[114] Noriko Suzuki, Yugo Takeuchi, Kazuo Ishii, and Michio Okada. Effects of echoic mimicry using hummed sounds on human–computer interaction. *Speech Communication*, 40(4):559–573, 2003.

[115] Satoshi V Suzuki and Seiji Yamada. Persuasion through overheard communication by life-like agents. In *Intelligent Agent Technology, 2004.(IAT 2004). Proceedings. IEEE/WIC/ACM International Conference on*, pages 225–231. IEEE, 2004.

[116] Takeuchi Y. Ishii K. Okada-M. Suzuki, N. Talking eye: Autonomous creatures for augmented chatting. *Robotics and autonomous systems*, 31(3):171–184, 2000.

[117] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving robot readability with animation principles. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, pages 69–76. IEEE, 2011.

[118] Takahiro Tanaka, Kazuhiro Fujikake, Takashi Yonekawa, Misako Yamagishi, Makoto Inagami, Fumiya Kinoshita, Hirofumi Aoki, and Hitoshi Kanamori. Driver agent for encouraging safe driving behavior for the elderly. In *Proceedings of the 5th International Conference on Human Agent Interaction*, pages 71–79. ACM, 2017.

[119] Jacques Terken, Henk-Jan Visser, and Andrew Tokmakoff. Effects of speech-based vs handheld e-mailing and texting on driving performance and experience. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 21–24. ACM, 2011.

[120] Sebastian Thrun. Toward a framework for human-robot interaction. *Human–Computer Interaction*, 19(1-2):9–24, 2004.

[121] Linda Tickle-Degnen and Robert Rosenthal. The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4):285–293, 1990.

[122] Nishimura R. Yamamoto K. Nakagawa-S. Todo, Y. Development and evaluation of spoken dialog systems with one or two agents through two domains. In *Text, Speech, and Dialogue*, pages 185–192. Springer, 2013.

[123] Michael Tomasello and Malinda Carpenter. Shared intentionality. *Developmental science*, 10(1):121–125, 2007.

[124] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(05):675–691, 2005.

[125] Michael Tomasello, C Moore, and PJ Dunham. Joint attention as social cognition. *Joint attention: Its origins and role in development*, pages 103–130, 1995.

[126] Cristen Torrey, Susan Fussell, and Sara Kiesler. How a robot should give advice. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 275–282. IEEE Press, 2013.

[127] Toyota. http://toyota.jp/kirobo_mini/, 2017.

[128] Tom Vanderbilt. Let the robot drive: The autonomous car of the future is here. *Wired Magazine, Conde NAST, www. wired. com*, pages 1–34, 2012.

[129] Adam Waytz, Joy Heafner, and Nicholas Epley. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52:113–117, 2014.

[130] Christopher D Wickens. Multiple resources and mental workload. *Human factors*, 50(3):449–455, 2008.

[131] Peters J. C. Breazeal-C. L Williams, K. J. Towards leveraging the driver's mobile device for an intelligent, sociable in-car robotic assistant. In *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pages 369–376. IEEE, 2013.

[132] P. De Ravindra S. Okada-M. Yoshiike, Y. Mawari: a social interface to reduce the workload of the conversation. In *Social Robotics*, pages 11–20. Springer, 2011.

[133] Robert B Zajonc. Social facilitation. *Science*, 149(3681):269–274, 1965.

[134] Jordan Zlatev. The epigenesis of meaning in human beings, and possibly in robots. *Minds and Machines*, 11(2):155–195, 2001.