Multi-party Conversation of Driving Agents: The Effects of Overhearing Information on Lifelikeness and Distraction

Nihan Karatas, Shintaro Tamura, Momoko Fushiki, Michio Okada
Interaction and Communication Design Lab
Toyohashi University of Technology
Toyohashi, Aichi, Japan
{karatas,tamura,fushiki17}@icd.tut.ac.tut.jp,okada@tut.jp

ABSTRACT

Recently, the applications of conversational robots have been gaining popularity due to their potential in providing information while engaging the user in a conversation. However, when a user's attention is already focused on a task, engaging them in conversation may be difficult or even risky. The Human-Robot Interaction (HRI) field should consider interaction methods where a conversational robot can keep the user informed but without the obligation of engagement in the conversation. In this study, we discuss such an approach within a driving scenario by utilizing a multi-party social robot platform that comprises three minimalistic conversational robots, which possesses the feature of being able to decrease the number of directed utterances toward a driver through a turn-taking process among the robots. The results of this study revealed that overhearing information from the multi-party conversation of driving agents is perceived as possessing more life-like characteristics compared to a conventional, one-to-one communication-based approach that directly addresses the driver. Moreover, the proposed approach reduced the distraction level and increased the enjoyment of the drivers.

KEYWORDS

Multi-party conversation, driving agent, life-likeness, distraction

ACM Reference Format:

Nihan Karatas, Shintaro Tamura, Momoko Fushiki, Michio Okada. 2018. Multi-party Conversation of Driving Agents: The Effects of Overhearing Information on Lifelikeness and Distraction. In *6th International Conference on Human-Agent Interaction (HAI '18), December 15–18, 2018, Southampton, United Kingdom.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3284432.3284466

1 INTRODUCTION

Conversational social robots have received broad attention within various domains such as education [1], health care [2], entertainment [3], 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 [4–7]. It has been demonstrated that a robotic driving agent is more noticeable, familiar, and acceptable [8], and also creates a stronger social bond with the driver while transmitting necessary information to them [5] compared to voice-only and display-based driving agents. 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 [9] discussed that this kind of unilateral interaction creates a "master-slave" relationship and can be seen in industrial or professional service robots that lack social interaction abilities. Norman [10] 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 commonly experienced 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 complete and meaningful utterances. However, it would be difficult for one robot to avoid unexpected responses from the interlocutor. Another problem is that during the one-to-one communication occurring 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 [11] 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 mental workload; and, therefore, potentially create 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

HAI '18, December 15–18, 2018, Southampton, United Kingdom

ACM ISBN 978-1-4503-5953-5/18/12...$15.00.
https://doi.org/10.1145/3284432.3284466

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 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 natural, coherent, enjoyable and lively compared to a single robot conversation [12–16]. 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 [17], and broadcasting information as a passive-social medium in public places [18, 19]. 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, thereby 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 [20]. 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 life-likeness 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 [21], 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 multi-party conversational agents rather than one-to-one conversational agent. In the current study, we employed our robotic driving agent platform, NAMIDA, 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. With this research, we evaluated this interaction model in terms of the subjective and objective qualifications of the system.

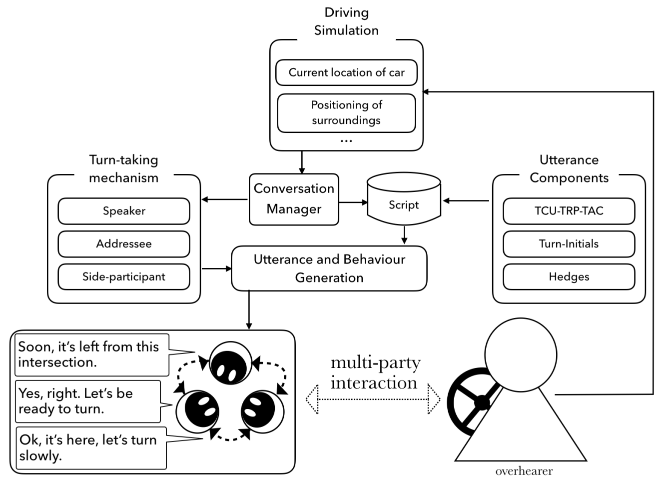

Figure 1: The internal structure of NAMIDA and its communication protocol.

2 DESIGN OF NAMIDA PLATFORM

2.1 Appearance

The NAMIDA platform was built as designed in the previous study [21]. 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 servomotors inside each head attached to the main board. The robots have a conversation mechanism that allows them to generate verbal and non-verbal behaviors within each turn-taking.

2.2 Conversation Mechanism

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 [22]. One of the important points of the multi-party conversation of robots in being persuasive is smooth turn-taking. In this manner, the driver can believe 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. In such 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 [23]. 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.

We adopted the multi-party conversation approach that consists of nonverbal behaviors, turn-initials and hedges by utilizing the turn-taking components [24] as we had in our previous study [25]. Ford and Thompson suggest that humans employ turn-initials for changing direction, error handling and enhancement to maintain the liveliness of a conversation [26]. In our study, the utterances of the robots emerged from informal, polite Japanese language, utilizing the turn-initials and hedges (shown in Table 1), 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 addressee based on the conversation phase seen in casual human conversations [25]. In the NAMIDA system, when the speaker robot generates an utterance, it directs its eye gaze toward the hearers (the other two robots, which are addressee and side-participant, or just one robot, which is the addressee). Accordingly, the hearers (addressee and side-participant) incline their heads toward the speaker.

Fig. 1 represents the structure of NAMIDA system and the communication protocol among the driving simulation, conversation manager, script, turn-taking mechanism, and the utterance and behavior generation of the robots. The script, that was prepared with the utterance components, is triggered by the conversation manager that is activated when the current location of the car gets closer to the designated situations on the simulated road. The conversation manager controls the turn-taking by assigning conversational roles (speaker, addressee and side-participant) to the robots. The roles change after each conversational turn.

3 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 it. 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.

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.

3.1 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 changes and gave information about the driver'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 agent directs its eye-gaze only to the addressee. The turn-taking occurred only in the MPCN condition between the three agents by allocating the script among them.

Figure 2: 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: 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. 2). 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 slightly toward 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). During 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 agent 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 agent.

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

- (H1) A multiple robot environment will exhibit higher life-like qualifications (e.g., anthropomorphism).
- (H2) In a multiple robot environment, the drivers' attentional focus will not shift toward the robot as frequently compared to one with a single robot.
- (H3) A multiple robot environment will be perceived as more positively (more enjoyable, less distractive, etc.) compared to a single robot situation.

3.2 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. Each condition took approximately 10 minutes. All participants had a driving license. The experiment was set up as a counterbalanced, within-subject study. Upon arrival, the participants were introduced to 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 at a specified train station in the simulated town. Their task was to listen to the instructions and information from NAMIDA, and drive accordingly. The experimenter carefully avoided mentioning about the expected responses of the participants for both conditions in order to prevent unconscious changes of their behaviors. The participants were then given a demographic questionnaire. Before starting the experiment, the participants were given five minutes to practice on driving simulation. After each session, there was an approximately 15 minute break, during which, the participants were asked to complete the subjective assessment questionnaires.

3.3 Experiment Setup

We set up our experiment in a driving simulator environment, using the UC-win/Road Drive Simulator. 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 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 conditions on that day) and finally the goal/destination, which was in front of a 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 50 km/h in the city environment, and the total route was approximately 10 km long. The approximate time for arriving at the destination was 10 - 13 minutes.

3.4 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 [27]. To analyze their subjective impressions, they were given a questionnaire that consists of 10 regarding the interaction's perceived annoyance, enjoyment, distraction and conversational burden. For the objective evaluation part, we analyzed the attentional focus of the participants by tracking their eye gaze movements, using Tobii Pro Glasses.

4 RESULTS

4.1 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 (Table 1). The results revealed a significant difference in anthropomorphism $(t(21)=-3.914,\ ^{***}p=0.000)$, 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 $(t(21)=1.311, p=0.203)$.

However, when we carefully looked at the results of each session, we suspected that there had been a carry-over effect of the subjects' responses given on the questionnaire. That is, for example, the subjects who were exposed to the OOCN in the latter session rated the NAMIDA system lower than the subjects who were exposed to the same session earlier. Conversely, the subjects who were exposed to the MPCN condition in the latter session rated the NAMIDA system higher than the subjects who were exposed to the same session earlier. A carry-over effect seemed to have occurred for some of the dimensions in the questionnaire. We conducted a mixed two-way ANOVA with the exposure order (first and second; between groups) and the experiment condition (OOCN and MPCN; within groups) as independent variables and questionnaire ratings as the dependent variable. According to the analysis, the questionnaire ratings showed a significant result on the interaction of the session order and the conditions for anthropomorphism ($F(1, 20) = 8.763,$

Table 1: The means and standard deviations of OOCN and MPCN conditions for Godspeed dimensions.

	OOCN		MPCN	
	mean	std. dev.	mean	std. dev.
Anthropomorphism	2.527	0.643	3.2	0.755
Animacy	3.06	0.847	3.654	0.659
Likability	4.109	0.767	4.463	0.361
P. Intelligence	3.418	0.859	3.39	0.851
P. Safety	3.393	0.596	3.787	0.773

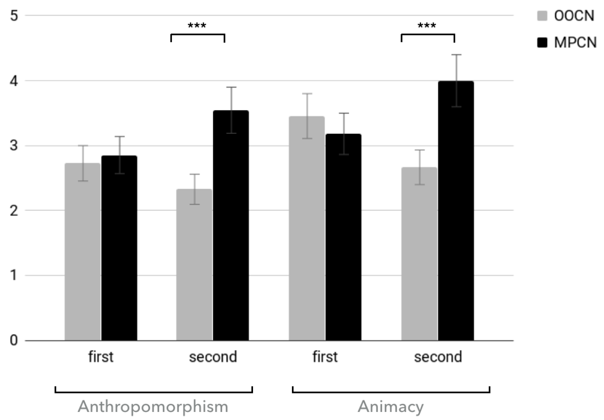

Figure 4: Graph indicates the main effects of order and condition for anthropomorphism and animacy dimensions. (*p<0.05, **p<0.01, ***p<0.001)

Table 2: The main effect of order and condition for Anthropomorphism and Animacy dimensions.

	Anthropomorphism			Animacy		
	OOCN	MPCN	Ave.	OOCN	MPCN	Ave.
First	2.727	2.854	**2.79**	3.454	3.181	**3.317**
Second	2.327	3.545	**2.936**	2.666	3.999	**3.332**
Ave.	**2.527**	**3.2**		**3.06**	**3.59**	

$^{**}p = 0.007, \eta^2=0.304$). Simple main effects analysis showed that the exposure order had a significant effect on MPCN condition ($F(1, 20) = 2.625, {}^{*}p=0.028, \eta^2=0.218$); also, the experiment condition had significant effect on the *second* exposition order ($F(1, 20) = 54.478, {}^{***}p=0.000, \eta^2=0.844$). These results indicated that the participants who exposed to the OOCN as first and the MPCN as second condition rated the anthropomorphism significantly higher than the the participants who exposed to the OOCN as second and the MPCN as first (Fig. 4 and Table 2).

Similarly, a mixed two-way ANOVA showed a significant result on the interaction of the session order and the conditions for animacy ($F(1, 20) = 17.294, {}^{***}p=0.000, \eta^2=0.463$). Simple main effects

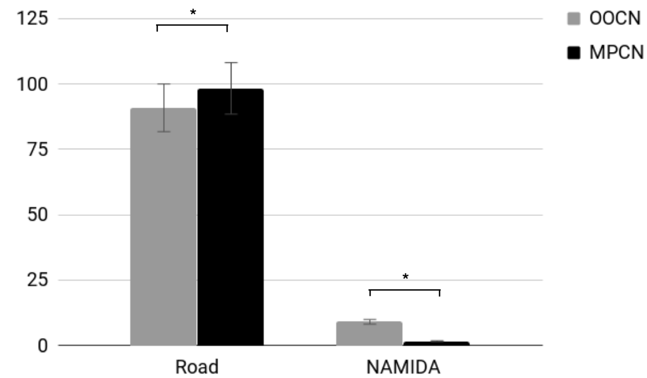

Figure 5: Graph indicates the mean values (%) and results of the rates of eye gaze positions. (*p<0.05, **p<0.01, ***p<0.001).

Table 3: The Wilcoxon signed-rank test results of the eye gaze data rates observed on the NAMIDA and the road.

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

analysis showed that the exposure order had a significant effect on OOCN ($F(1, 20) = 5.847, {}^{*}p=0.025, \eta^2=0.226$) and MPCN ($F(1, 20) = 11.685, {}^{**}p=0.002, \eta^2=0.368$); conditions. In addition, the experiment condition had significant effect on the *second* exposition order ($F(1, 20) = 36.268, {}^{***}p=0.000, \eta^2=0.783$). These results indicated that the participants who exposed to the OOCN condition as first rated the animacy higher compared to the MPCN condition when it was exposed as first. Also, the MPCN condition was rated higher when it was exposed as second (Fig. 4 and Table 2). These results on the anthropomorphism and animacy dimensions were consistent with the possibility of a carry-over effect.

4.2 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 during each session. The gathered data could not 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). We observed that in the MPCN condition, the participants' eye gaze was on the road significantly more than in the OOCN condition ($Z=-3.392, {}^{***}p=0.000$) (Fig. 5). Furthermore, in the OOCN condition, the participants looked at the NAMIDA platform significantly more than in the MPCN condition ($Z=-3.392, {}^{***}p=0.000$) (Fig. 5).

According to the results, the MPCN system could help a driver to focus on the road more than the OOCN system. The reason for this may point to our hypothesis that a one-to-one communication system directs the utterances toward the driver, and this immediate

Table 4: The content of the subjective impression questionnaire.

Code	Group	Question
SQ1	Enjoyability	I think the driving was enjoyable.
		I think the driving was fascinating.
SQ2	Annoyance	I think NAMIDA's conversation was noisy.
		I was annoyed with NAMIDA's conversation
SQ3	Conversational burden	While NAMIDA was talking, I felt like I had to give an answer.
		While NAMIDA was talking, I felt like I had to look at it.
SQ4	Distraction	I was distracted from driving.

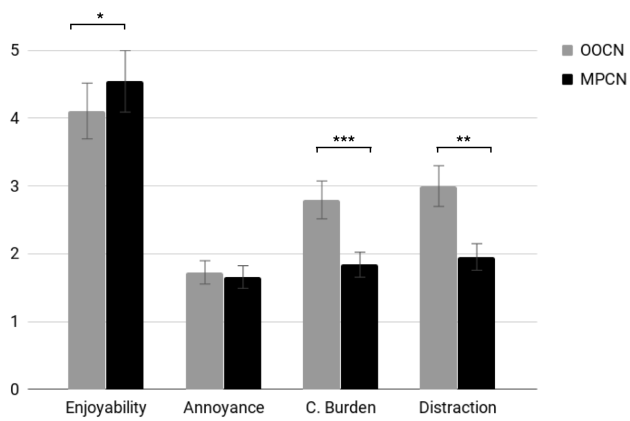

Figure 6: Graph indicates the mean value and results of paired t-test of subjective impression questionnaire. (*p<0.05, **p<0.01, ***p<0.001).

Table 5: The results of Wilcoxon signed-rank test of subjective impression questionnaire. SQ1: Enjoyability, SQ2: Annoyance SQ3: Conversational Burden, SQ4: Distraction

			OOCN		MPCN	
	N	Z	mean	std. dev.	mean	std. dev.
SQ1	15	-2.271	4.106	0.744	4.454	0.604
SQ2	16	-0.31	1.727	0.869	1.659	0.917
SQ3	19	-3.219	2.795	1.191	1.84	0.917
SQ4	16	-2.74	2.59	1.402	1.954	1.174

communication causes drivers to lose their attention on the road. However, a multi-party conversation-based system allocates the conversation burden among the other robots in the conversation. Thus, the driver's loss of attention could be reduced.

4.3 Evaluation of the Subjective Impression

This section focuses on exploring the subjective evaluations of the questionnaire, which was designed to compare and evaluate the effectiveness of the communication method in both conditions. 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). The questionnaire was designed using a 5-point, Likert scale (1=strongly disagree to 5=strongly agree). First, we conducted an internal reliability analysis for the first three items of the questionnaire. For the OOCN condition, the results revealed α*=0.63*, α*=0.935* and α*=0.837* for annoyance, enjoyability and conversational burden, respectively. For the MPCN condition, the results revealed α*=0.822*, α*=0.898* and α*=0.817* for annoyance, enjoyability and conversational burden, respectively. The gathered data could not pass the Shapiro-Wilk normality test in these three categories for both conditions at p=0.05 level, therefore we used the Wilcoxon signed-rank test to compare the conditions for each factor (Table 5). We found no significant difference for annoyance (*Z=-0.3103, p=0.378*) (Fig 6). On the other hand, we found significant differences for enjoyability (*Z=-2.2718, *p=0.011*), conversational burden (*Z=-3.219, ***p=0.000*) and distraction (*Z=-2.74, **p=0.003*) (Fig.6).

These results revealed that even though, in the MPCN condition, there were slightly more lines in the conversation script than in the OOCN condition, which may 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 the persuasive turn-taking nature of the system. In addition, we conducted a two-way ANOVA, including the order as an independent variable for each comparison to investigate any carry-over effects. We found that the p-value for the interaction between the following dimensions and the experiment order was not significant, we also did not observe any main effect for these dimensions.

5 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 looked at drivers' perceptions about the robots' life-likeness and attention manipulation through the subjective evaluation of participant drivers. 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 life-like, enjoyable and less distractive environment compared to a one-to-one conversation model where the robot's utterances were solely directed toward the driver.

We found through the Godspeed questionnaire that the anthropomorphism, animacy and likability dimensions were significantly higher in the multi-party conversation condition. However, we observed that the interaction effect of the exposure order and the condition was significant on anthropomorphism and animacy which leads us to interpret these results as they possess a carry-over effect on these dimensions. Specifically, when the participants were exposed to the MPCN after the OOCN condition, they found the MPCN more anthropomorphic and more animated compared to the times when they were exposed to MPCN before OOCN. Also, the participants rated anthropomorphism and animacy higher when they were exposed to the OOCN first compared to times when they were exposed to the OOCN after.

It can be said that the participants rated MPCN by comparing their previous experience with OOCN where the lively turn-taking based multi-party conversation exhibited better anthropomorphic and animatic features when they compare it with one-to-one conversation. 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 which incorporates directly addressing the participants, which they had already experienced. Likewise, compared to the previous interaction experience, the participants rated MPCN better as the robots exhibited the verbal and non-verbal behaviors that induced the perception of them as having more animacy, which refers to being that of life-like creatures. Piaget [28] emphasized that the major factors of lifelikeness consists of movements and intentional behaviors, and that perceiving something as life-like allows humans to distinguish humans from machines. Therefore, the consistent utterances and turn-taking behaviors contributed by the robots to the MPCN led to a perception of the robots as more life-like. As for the likability dimension, we did not observe neither an interaction effect nor a main effect in the order or in the condition. 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.

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 demonstrated research [17], 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 indirect interactions within the peripheral vision of the drivers have persuasive effects on the users. The results also indicated that there was no significant difference with the annoyance aspect between the two conditions. In the MPCN condition, slightly more utterances were given, thus, there was slightly more auditory output to comprehend for the participants leading to a chance that the robots might be perceived as more annoying. However, even though the participants had to lend an ear to the robots more often in the MPCN condition, the coherent and enjoyable multi-party conversation was perceived as being less annoying, almost the same as with the OOCN condition.

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 in 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 addressee, the driver was obligated to be in direct interaction with the system which emerged as visual and auditory efforts 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. 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 an individual is not aware of while they are doing them. The results of perceived distraction support our hypothesis on the objective analysis of reflexive attention shifts that associates with the diverted eye gaze of participants.

5.1 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.

6 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 through a context-aware interaction during driving. We believe that this conversational approach is life-like and enjoyable as it requires less attention 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 had an influence on anthropomorphism and animacy ratings that the participants evaluated the second condition by considering the first condition they were exposed to. In this respect, a carry-over effect seemed likely to have occurred. 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, the MPCN system exhibited an enjoyable performance and stress-free driving environment. For a more enjoyable and 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, within a between-subject study in order to avoid a possible carry-over effect. 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.

ACKNOWLEDGMENT

This research has been supported by Grant-in-Aid for scientific research of KIBAN-B (18H03322) from the Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] D. Conti, S. Di Nuovo, S. Buono, and A. 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*, vol. 9, no. 1, pp. 51–62, 2017.

[2] R. Q. Stafford, B. A. MacDonald, X. Li, and E. Broadbent, "Older peopleâĂŹs prior robot attitudes influence evaluations of a conversational robot," *International journal of social robotics*, vol. 6, no. 2, pp. 281–297, 2014.

[3] N. T. Fitter, H. Knight, N. Martelaro, and D. Sirkin, "What actors can teach robots," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 574–580, ACM, 2017.

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

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

[6] Y.-H. Nho, J.-H. Seo, J.-Y. Yang, and D.-S. 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*, pp. 480–485, IEEE, 2015.

[7] Y. Okajima, H. Masuta, M. Okumura, T. Motoyoshi, K. Koyanagi, T. Oshima, and E. Takayama, "Utterance generation based on driving evaluation for driving assistance robot," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 20, no. 5, pp. 836–844, 2016.

[8] T. Tanaka, K. Fujikake, T. Yonekawa, M. Yamagishi, M. Inagami, F. Kinoshita, H. Aoki, and H. Kanamori, "Driver agent for encouraging safe driving behavior for the elderly," in *Proceedings of the 5th International Conference on Human Agent Interaction*, pp. 71–79, ACM, 2017.

[9] S. Thrun, "Toward a framework for human-robot interaction," *Human-Computer Interaction*, vol. 19, no. 1, pp. 9–24, 2004.

[10] D. Norman, *The design of everyday things: Revised and expanded edition.* Basic Books (AZ), 2013.

[11] M. 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.

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

[13] T. Arimoto, Y. Yoshikawa, and H. Ishiguro, "Cooperative use of multiple robots for enhancing sense of conversation without voice recognition," *SIG-SLUD*, vol. 5, no. 2, pp. 76–77, 2015.

[14] T. Iio, Y. Yoshikawa, and H. Ishiguro, "Pre-scheduled turn-taking between robots to make conversation coherent," in *Proceedings of the Fourth International Conference on Human Agent Interaction*, pp. 19–25, ACM, 2016.

[15] T. Iio, Y. Yoshikawa, and H. Ishiguro, "Starting a conversation by multi-robot cooperative behavior," in *International Conference on Social Robotics*, pp. 739–748, Springer, 2017.

[16] T. Arimoto, Y. Yoshikawa, and H. Ishiguro, "Multiple-robot conversational patterns for concealing incoherent responses," *International Journal of Social Robotics*, pp. 1–11, 2018.

[17] K. Hayashi, T. Kanda, T. Miyashita, H. Ishiguro, and N. Hagita, "Robot manzai: Robot conversation as a passive–social medium," *International Journal of Humanoid Robotics*, vol. 5, no. 01, pp. 67–86, 2008.

[18] D. Sakamoto, K. Hayashi, T. Kanda, M. Shiomi, S. Koizumi, H. Ishiguro, T. Ogasawara, and N. Hagita, "Humanoid robots as a broadcasting communication medium in open public spaces," *International Journal of Social Robotics*, vol. 1, no. 2, pp. 157–169, 2009.

[19] Y. Pan, H. Okada, T. Uchiyama, and K. Suzuki, "On the reaction to robotâĂŹs speech in a hotel public space," *International Journal of Social Robotics*, vol. 7, no. 5, pp. 911–920, 2015.

[20] Y. Yoshiike, P. R. S. De Silva, and M. Okada, "Mawari: a social interface to reduce the workload of the conversation," in *Social Robotics*, pp. 11–20, Springer, 2011.

[21] N. Karatas, S. Yoshikawa, P. R. S. De Silva, and M. Okada, "Namida: Multiparty conversation based driving agents in futuristic vehicle," in *Human-Computer Interaction: Users and Contexts*, pp. 198–207, Springer, 2015.

[22] N. Suzuki, Y. Takeuchi, K. Ishii, and M. Okada, "Talking eye: Autonomous creatures for augmented chatting," *Robotics and autonomous systems*, vol. 31, no. 3, pp. 171–184, 2000.

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

[24] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *language*, pp. 696–735, 1974.

[25] N. Karatas, S. Yoshikawa, and M. Okada, "Namida: Sociable driving agents with multiparty conversation," in *Proceedings of the Fourth International Conference on Human Agent Interaction*, pp. 35–42, ACM, 2016.

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

[27] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[28] D. Parisi and M. Schlesinger, "Artificial life and piaget," *Cognitive Development*, vol. 17, no. 3-4, pp. 1301–1321, 2002.