

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

Nihan Karatas^{*1}, Soshi Yoshikawa^{*1}, P. Ravindra S. De Silva^{*1} and Michio Okada^{*1}

Abstract – This paper proposes a unique communication approach for an in-car driving agent system by utilizing the usage of a multi-party conversation mechanism. 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.

Keywords : Multiparty conversation, context aware interaction, cognitive workload, driving agent, NAMIDA

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^[1].

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^[2],^[3],^[4]. 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 claim that these technologies are not any less dangerous than the devices that require a driver's eyes or hands.



Fig.1 Multi party conversation between NAMIDA robots.

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^[8]. 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^[9]. From the drivers' perspective, we believe that it is crucial to interact with an IVI system in such a social, natural and famil-

^{*1}: Computer Science and Engineering, Toyohashi University of Technology

iar 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^[5]. Carnaby has been envisioned as a driving assistant to make driving safe and fun^[6]. Moreover, Quin has been developed to handle location services and driver fatigue detection^[7]. 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^[10]. 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 3). 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 conversa-

tion, 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. 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).

2.1 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)^[20]. Context-aware systems use Information Communication Technologies (ICT) to provide a greater awareness of relevant information about the physi-

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

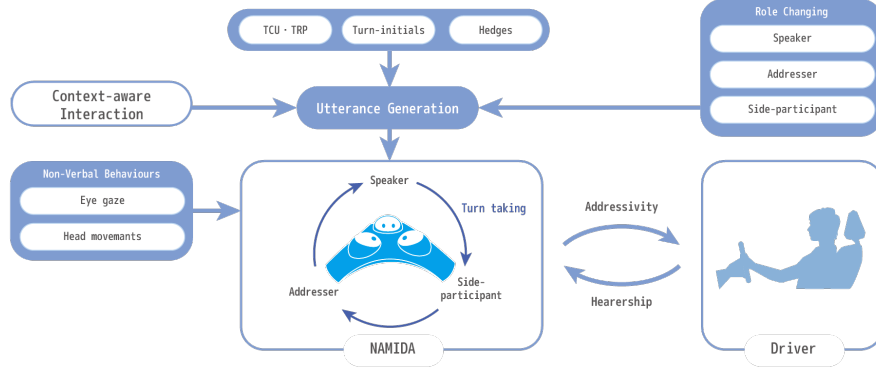


Fig. 2 Depiction of the overall processing architecture for NAMIDA. Utterance generation utilizes a turn-taking process, using turn-initial elements and hedges and taking into account hearership and addressivity status for a multi-party conversation.

cal 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 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^[12]. 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 (hearsership), which creates a conversational burden and mental workload for the user.

Yoshiike et al. studied a system called MAWARI, presents in three social robots who can conduct multi-party conversations as an interactive social medium^[14]. 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^[13]. 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^[27]. Furthermore, Ishizaki et al. discussed that multi-party conversation not only presents new topics/details, but also lessens the stress of conversation initiation^[21].

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 de-

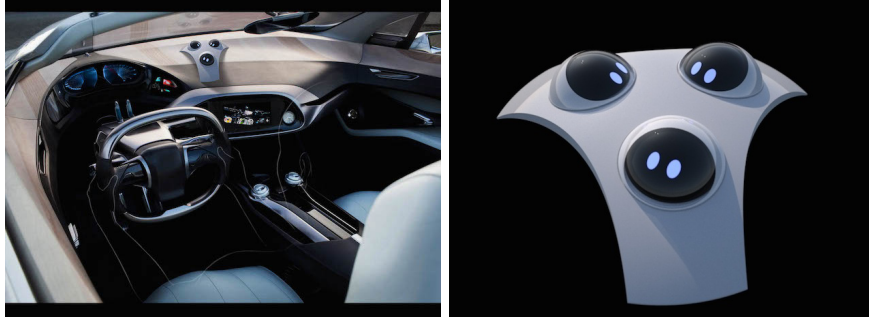


Fig. 3 Base unit of NAMIDA, designed to be secured on the dashboard of the car within the peripheral vision of the driver (*left*). With movable heads and eyes, NAMIDA projects a life-like appearance (*right*).

tails 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.

2.2.1 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^[30]. 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^[13]. 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.

2.3 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^[17]. 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^[19]. Therefore, in the area of HRI, it has become an essential task to develop eye gaze cues to enhance human-robot interaction^[18]. 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.

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 3). 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 pe-

ripheral 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 ([23], [26]). 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.

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 [15], [17]. Psychological studies suggest that the appearance and behavior of machines have the potential to influence human perception and behavior toward machines [31], [32], [33], [34]. According to Leech, the behaviors that allow humans to engage in social interactions in a relatively harmonious atmosphere can be defined as politeness [35]. 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-threatening acts" [36]. 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 [28]. 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 [16]. In our study, the utterances of NAMIDA 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 utter-

ances, 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 [37]. Researches showed that merging the utterance mechanism with bodily movements (head rotating and eye gazing) leads to a persuasive impression (as an example [24]). 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.

3.2 Conversation Structure

During everyday informal conversations, the role of speaker, addresser and other participants intuitively alternates on turn-taking bases. [25] 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 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 con-

Table 1 NAMIDA utterances coupled with nonverbal behaviors (turn-initial elements, hedges and nonverbal behaviors indicated).

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"

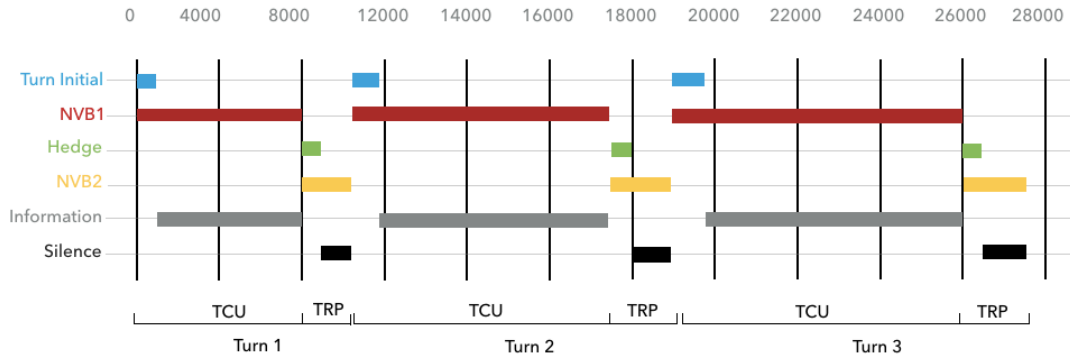


Fig. 4 Figure depicts the utterance generation and non-verbal interaction design of the multi-party conversational agents in a time-scale base. The utterance of the speaking agent starts with a turn initial and ends with a hedge within a one turn.

versation 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. 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 addressee (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.

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. 1). However, in the OOCN case, there is no side-participant, so that during a TCU, the speak-

ing agent directs its eye-gaze only the addressee. The turn-taking occurs 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.

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. 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] (...)

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.

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,

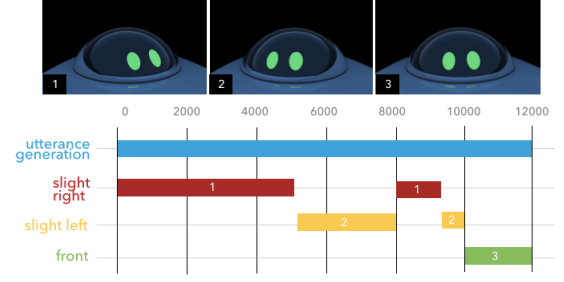


Fig. 5 Figure depicts one turn of the conversation in the Condition 1.

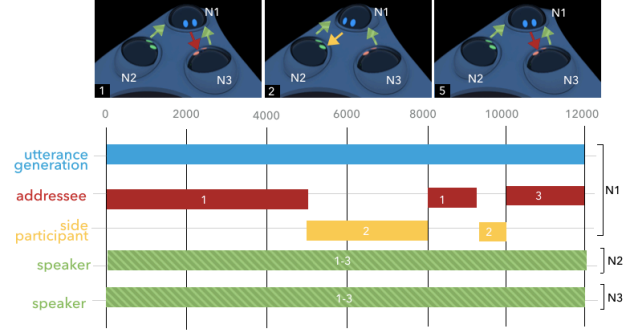


Fig. 6 Figure depicts one turn of the conversation in Condition 2.

meanwhile these two agents direct their eye gaze towards the speaking agent (Fig. 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.

4.3 Experiment Setup

We set a mock driving environment as it is shown in Fig.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-

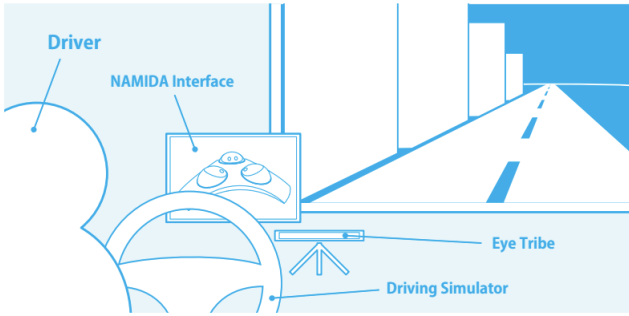


Fig.7 The setup of the experiment. The driver goes along the road while listening to the conversation of NAMIDA.

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 souse 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.

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^[29]. 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.

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.

5.1.1 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,

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

Table 2 Factors of DALI, based on the context and the associated questionnaire with their description.

Dimension	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

which was proposed by^[29] as a revised version of the NASA-TLX⁽⁸¹⁾ 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). 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^[29].

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. 8). According to these results, the OOCN required more attention than the MPCN with regard to remembering 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. 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.

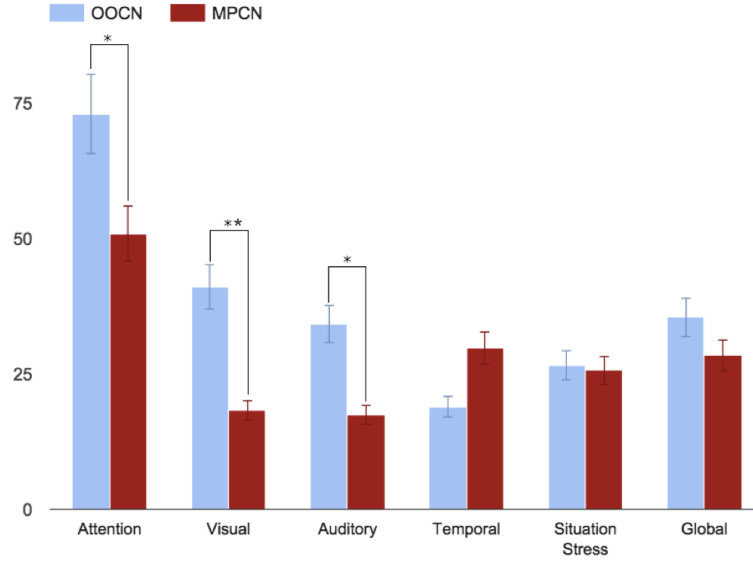


Fig. 8 The figure depicts the results of the DALI factors under the OOCN and MPCN cases. (*: $p < .05$, **: $p < .01$)

Auditory Demand: There was a significant difference in auditory demand ($t(13)=1.83$, $P=0.0449 < 0.05$, significant) (Fig. 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. 8).

Situation Stress: This factor also does not show a significant difference ($t(13)$, $P=0.45992 > 0.05$, non-significant) (Fig. 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. 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. 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.

5.2 Attention Behavior of Driver

The eye gaze movements of a person provide significant cues about their attentional behaviors ([45], [44], [46]). 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 hypothesized that a multi-party

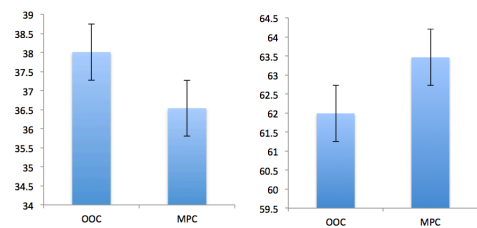


Fig. 9 Results show the percentage of the collected eye gaze data on the NAMIDA system (*left*) and on the simulated road (*right*).

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.

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. 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.

5.2.1 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. 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. 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

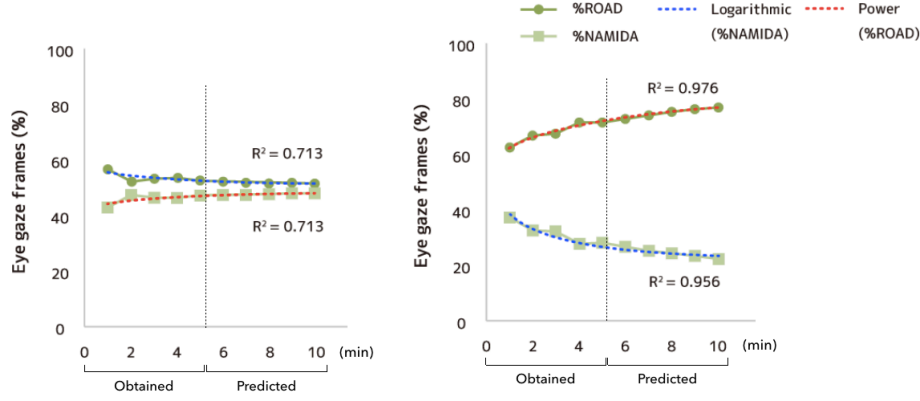


Fig. 10 Figure shows the trendline of the eye gaze behaviours in OOCN (*left*) and MPCN (*right*) conditions. The part with first five minutes shows the eye gaze data collected during the experiment (*Obtained*). The second part shows the predicted eye gaze behaviours for another five minutes (*Predicted*)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

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. 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. 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 important 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.

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 ([37]). Human likeness often corresponds to anthropomorphism, the attribution of human form, human characteristics

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

Table 3 Questionnaire and t-test results of subjective impression on MPCN and OOCN.

Code	Question	Pair-wise t-test $t(d.f.) = t\text{-value}$ $P < 0.05$ (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

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.

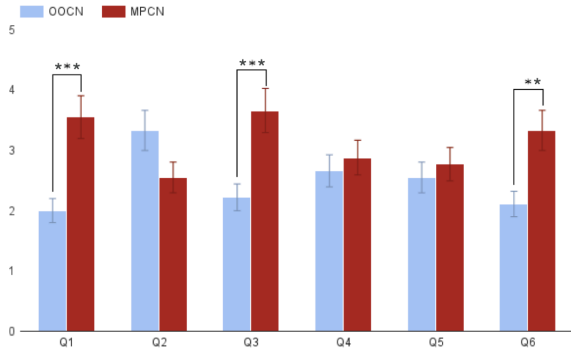


Fig. 11 Figure shows the mean comparison of subjective impression questionnaire on MPCN and OOCN. (*: $p < .05$, **: $p < .01$, ***: $p < .001$).

Research reveals that lifelike creatures can deeply involve users emotionally and that this involvement can be used to influence users^[37]. Heider and Simmel have devoted research to the perceived animacy and intentions of geometric shapes on computer screens^[38]. 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^[41]. 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^[40], 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^[43]. 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.

5.3.1 Results of Subjective Assessment

Table 3 and Fig. 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 3 and Fig. 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.

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 3.1 and 3.2, the proposed design of the multi-party conversation system employs well-established turn-taking and role-changing techniques.

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 sys-

tem, which requires more mental resources, resulting in more attentional behaviors that indicate distraction (e.g., staring at the NAMIDA longer). Fig. 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 [18], 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 ([10]), 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. 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^[13] 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. 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.

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. 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. 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^[47], and also creates a stronger social bond with the driver while transmitting necessary information to them^[10] 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.

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. 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. 11). These results show that the turn-taking mechanism, the different characteristics of the agents, and the lively, sustainable conversa-

tion 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.

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.

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, over-hearing 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.

Acknowledgements

This research has been supported by Grant-in-Aid for scientific research of KIBAN-B (26280102) from the Japan Society for the Promotion of Science (JSPS).

Reference

- [1] Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, 370(1), 54-59.
- [2] Navdy Inc., Feels Like Driving in Future, <https://www.navdy.com/>, 2013
- [3] DiMarzio, Jerome F, Android, *Tata McGraw-Hill Education*, 2008
- [4] Terken, J., Visser, H. J., & Tokmakoff, A. (2011, November). 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* (pp. 21-24). ACM.

- [5] Nissan Motor Company Ltd., PIVO: An in-car robotic assistant, *In Automotto*, 2005
- [6] Pioneer and iXsResearch Corp, Carnaby:anin-carrobotic navigation assistant, *In TechCrunch*, 2010
- [7] BYD Company Ltd., Qin:A social robot for a new plug-in hybrid vehicle, *In China Car Times*, 2012
- [8] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.
- [9] Barton, R. A., & Dunbar, R. I. (1997). Evolution of the social brain. *Machiavellian intelligence II: Extensions and evaluations*, 2, 240.
- [10] Williams, K. J., Peters, J. C., & Breazeal, C. L. (2013, June). 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.
- [11] Deffenbacher, J. L., Oetting, E. R., & Lynch, R. S. (1994). Development of a driving anger scale. *Psychological reports*, 74(1), 83-91.
- [12] Bakhtin, M. M. (2010). Speech genres and other late essays. *University of Texas Press*.
- [13] Suzuki, N., Takeuchi, Y., Ishii, K., & Okada, M. (2000). Talking eye: Autonomous creatures for augmented chatting. *Robotics and autonomous systems*, 31(3), 171-184.
- [14] Yoshiike, Y., De Silva, P. R. S., & Okada, M. (2011, November). MAWARI: a social interface to reduce the workload of the conversation. *In International Conference on Social Robotics* (pp. 11-20). Springer Berlin Heidelberg.
- [15] Andrist, S., Spannan, E., & Mutlu, B. (2013, March). 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* (pp. 341-348). IEEE Press.
- [16] Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. *Studies in interactional sociolinguistics*, 13, 134-184.
- [17] Torrey, C., Fussell, S., & Kiesler, S. (2013, March). How a robot should give advice. *In Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction* (pp. 275-282). IEEE Press.
- [18] Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009, March). 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* (pp. 61-68). ACM.
- [19] Charles, G. (1981). Conversational organization: Interaction between speakers and hearers. *Academic Press*.
- [20] Rakotonirainy, A. (2005). Design of context-aware systems for vehicle using complex systems paradigms.
- [21] Ishizaki, M., & Kato, T. (1998, August). Exploring the characteristics of multi-party dialogues. *In Proceedings of the 17th international conference on Computational linguistics- Volume 1* (pp. 583-589). Association for Computational Linguistics.
- [22] Hayashi, K., Sakamoto, D., Kanda, T., Shiomi, M., Koizumi, S., Ishiguro, H., ... & Hagita, N. (2007, March). Humanoid robots as a passive-social medium-a field experiment at a train station. *In Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on* (pp. 137-144). IEEE.
- [23] Matsumoto, N., Fujii, H., Goan, M., & Okada, M. (2005, August). Minimal design strategy for embodied communication agents. *In Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on* (pp. 335-340). IEEE.
- [24] Ohshima, N., Ohyama, Y., Odahara, Y., De Silva, P. R. S., & Okada, M. (2015). Talking-Ally: The influence of robot utterance generation mechanism on hearer behaviors. *International Journal of Social Robotics*, 7(1), 51-62.
- [25] Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *language*, 696-735.
- [26] Bartneck, C., Kuli, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1), 71-81.
- [27] Todo, Y., Nishimura, R., Yamamoto, K., & Nakagawa, S. (2013, September). Development and evaluation of spoken dialog systems with one or two agents through two domains. *In International Conference on Text, Speech and Dialogue* (pp. 185-192). Springer, Berlin, Heidelberg.
- [28] Itani, R. (1995). Semantics and pragmatics of hedges in English and Japanese (*Doctoral dissertation, University of London*).
- [29] Pauzi, A., Manzan, J., & Dapzol, N. (2007, July). 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.
- [30] Goffman, E. (1979). Footing. *Semiotica*, 25(1-2), 1-30.
- [31] Moon, Y., & Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication research*, 23(6), 651-674.
- [32] Burgoon, J. K., Bonito, J. A., Ramirez, A., Dunbar, N. E., Kam, K., & Fischer, J. (2002). Testing the interactivity principle: Effects of mediation, propinquity, and verbal and nonverbal modalities in interpersonal interaction. *Journal of communication*, 52(3), 657-677.
- [33] Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1), 151-181.
- [34] Moon, A., Parker, C. A., Croft, E. A., & Van der Loos, H. M. (2013). Design and impact of hesitation gestures during human-robot resource conflicts. *Journal of Human-Robot Interaction*, 2(3), 18-40.
- [35] Leech, G. (2007). Politeness: is there an East-West divide?. *Journal of Politeness Research. Language, Behaviour, Culture*, 3(2), 167-206.
- [36] Brown, P., & Levinson, S. C. (1987). Politeness: Some universals in language usage (Vol. 4). *Cambridge university press*.

- [37] Ham, J., & Midden, C. J. (2014). 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.
- [38] Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*, 57(2), 243-259.
- [39] Bartneck, C., Croft, E., & Kulic, D. (2008, March). Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. In *Metrics for HRI workshop, technical report* (Vol. 471, pp. 37-44).
- [40] Bartneck, C., Croft, E., & Kulic, D. (2008, March). Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. In *Metrics for HRI workshop, technical report* (Vol. 471, pp. 37-44).
- [41] Clark, N. K., & Rutter, D. R. (1985). Social categorization, visual cues, and social judgements. *European Journal of Social Psychology*, 15(1), 105-119.
- [42] Karatas, N., Yoshikawa, S., De Silva, P. R. S., & Okada, M. (2015, August). Namida: Multiparty conversation based driving agents in futuristic vehicle. In *International Conference on Human-Computer Interaction* (pp. 198-207). Springer, Cham.
- [43] Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4), 285-293.
- [44] Fletcher, L., & Zelinsky, A. (2009). Driver inattention detection based on eye gaze—Road event correlation. *The international journal of robotics research*, 28(6), 774-801.
- [45] Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4), 694.
- [46] Ishikawa, T. (2004). Passive driver gaze tracking with active appearance models.
- [47] Tanaka, T., Fujikake, K., Yonekawa, T., Yamagishi, M., Inagami, M., Kinoshita, F., ... & Kanamori, H. (2017, October). 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.

(received Dec. 18, 2017, revised Apr. 15, 2018)

Biography

Nihan Karatas



is a PhD student in Department of Computer Science and Engineering at Toyohashi University of Technology (TUT) specializing in HRI. Her research interest includes multiparty interaction, conversation models, mutual adaptation, cognitive science and minimalist robots.

Soshi Yoshikawa



is a Master student in Department of Computer Science and Engineering at Toyohashi University of Technology (TUT) specializing in Human Interface Design.

P. Ravindra S. De Silva



is an assistant professor in the Department of Computer Science and Engineering at the Toyohashi University of Technology (TUT). His research interest includes social robotics, human robot interactions and user modeling based on statistical approaches.

Michio Okada



is a professor of Interactions and Communication Design (ICD) Lab at Toyohashi University of Technology (TUT), specializing in HRI. He was a researcher at NTT Basic Research Laboratories. Then, he has been with ATR International, where he was a senior researcher and department head in Department of Ecological Communications at ATR Network Informatics Laboratories. He is a visiting associate professor in Graduate School of Informatics at Kyoto University. His research interests include social robotics, human robot interaction and cognitive science, especially ecological socio-cultural approach to human cognition.