Vol.2019-HCI-185 No.16
Vol.2019-UBI-64 No.16
2019/12/11

Incorporating Conversations in Car Navigation Voice Guidance to Support Instructed Actions

Briane Paul V. Samson[1,2,†1,a)] Yasuyuki Sumi[1,b)]

Abstract: Voice guidance for car navigation typically assume drivers as docile actors. However, recent research highlight limitations that make drivers rely less on given directions. To explore how drivers can make better navigation decisions, we conducted a Wizard of Oz study that delivers turn suggestions in a conversation between two agents. We asked 30 participants to drive in a simulation environment using voice guidance that gives three types of suggestions: familiar, optimal, and new routes. We examined how these affect their choices and found that while most drivers followed directions appropriate for the given scenarios, they were more likely to make inappropriate choices after hearing alternatives in conversations. We also show how these two-party conversations may enable drivers to better reflect on their choices based on reported confidence after their trips. We conclude by discussing the design implications for car navigation voice guidance specifically and recommender systems in general.

1. Introduction

In 2050, we will see almost 70% of the global population move to cities, increasing car ownership and potentially affecting our goals of achieving sustainability. These additional vehicles will slowly congest denser urban environments and complex road networks, worsening traffic conditions and bring forth a number of negative consequences [16]. While our current road networks and transportation systems are still keeping up with the rising demand, modern navigation applications such as Waze and Google Maps are offering a slight reprieve and have become an integral part of a driver's commute with the hopes of circumnavigating daily traffic congestion. With the growing usage of such applications [1], [24], several studies have investigated how drivers follow route guidance and how it affects their navigation and driving behavior [4], [5], [14], [20]. In Brown & Laurier's ethnomethodological study [4], they enumerated five *normal, natural troubles* of driving with GPS devices. And in order for a driver to have better *instructed actions*, developers should focus more on supporting their interpretation and analysis of new route guidance and information. Aside from providing more empirical evidence supporting drivers' repeated non-preference of recommended fastest routes [4], [9], [18], [22], [27], a related work by Samson & Sumi [20] also found that drivers have differences in practice, sets of information used and decision making depending on the type of trip and trip context.

Going back to first principles, navigation is a social activity among drivers and navigators [8], [15]. And despite our growing reliance on modern navigation systems, we still perform better in terms of navigation and route learning when we are with an active collaborative partner in the task [2], [3], [4]. As a step towards supporting *instructed actions* by drivers, we conducted a Wizard of Oz study that explores the use of a two-party conversation in giving turn-by-turn guidance. In a within-subject design with 30 participants, we asked them to drive 9 times under different conditions (3 pure and 6 conversation). We recorded their navigation choices, workload, and confidence with their choices, and found that two-party conversations can encourage drivers to follow appropriate but with the right combination of voice agents.

2. Related Work

2.1 HCI of Recommender Systems

Recommender systems have become critical components on most modern services these days that they gained the interest of the HCI community to focus on user-centered approaches and user studies in order to build better systems. Looking at user perceptions, Knijnenburg et. al. [12] found that prolonged usage of such systems may indicate positive experiences, but that can still change over time. Aside from that, positive experiences can also be achieved if there is enough diversity in the recommendations [7], [12] and if they have a good match with our behavior history [25]. Specific to navigation applications – which can be considered a multi-criteria recommender system – recent studies have

[1] Future University Hakodate, Hakodate, Hokkaido 041–8655, Japan
[2] Center for Complexity and Emerging Technologies, De La Salle University, Taft, Manila 1004, Philippines
[†1] Presently with Future University Hakodate
[a)] b-samson@sumilab.org
[b)] sumi@acm.org

focused more on investigating how they affect driving and navigation performance. Early work by Dingus et. al. [5] found that voice is the safest modality for receiving turn-by-turn guidance. But even with this feature in most modern navigation applications, drivers in general still show difficulty at times when following their guidance [5], [14]. Recent works by Brown & Laurier [4] and Samson & Sumi [20] made extensive accounts and implications of these problems, emphasizing the need for designers to focus on helping drivers make *instructed actions* instead. Our work builds on their findings by exploring a concept for a new generation of navigation applications that focuses on helping drivers interpret and analyze turn-by-turn guidance through two-party conversations.

2.2 Conversation as a Modality

Recent works on HCI and human-robot interaction have explored using conversational user interfaces and multi-party conversations in various contexts. The early works of Sumi & Mase [21] and Todo et. al. [23] show how advantageous multi-party conversations can be in engaging users and giving new information about a topic. In the work of Yoshiike et. al. [26], they even saw reduced workload and conversational burden from users when they listened to a conversation between three social robots. In the car context, Large et. al. [13] found that engaging drivers in one-to-one conversations with a digital assistant can reduce driver fatigue while Karatas et. al. [11] found that keeping the driver as a bystander in a multi-party conversation between social robots can help them find good places to go while keeping their focus on the road. We build on this body of work by focusing our attention to the time critical task of turn-by-turn guidance and see whether it can maintain a reduced workload for drivers while helping them compare the value of two suggestions.

3. Method

3.1 Participants

We recruited participants with at least 1 year of driving experience and has a driver's license mainly through word-of-mouth from a public university and local communities. Because not many students have a driver's license, we also used snowball sampling wherein our early participants recommended other people they know that fits our recruitment criteria. We successfully recruited 30 participants with an almost equal mix of people who identify as men (N = 16) and women (N = 14). Their ages range from 19 to 64 years old (M = 29, SD = 10.6). They comprise of 12 Filipinos and 18 Japanese nationals. The Filipino participants do not drive in their current place of residence but they drive in their country of origin. All participants do not drive as part of their occupation. Thirteen are students while eleven are foreign workers. When asked about their driving experiences, three have been driving for more than 10 years while the rest are only driving for 1 to 5 years. In terms of application usage, they have experienced using a mix of Google Maps (N = 25), in-car navigation systems (N = 8), Waze (N = 4) and NaviTime (N = 1). Three of them have not used a navigation application before. Two Japan residents have been using these applications for more than 5 years while the rest are using them for less than 5 years. All of them use navigation applications only when going to an unknown destination and only one participant use it almost anywhere they go. When it comes to using voice guidance, 18 of them do not use it. For those that do, they frequently use it when they go on trips to new or seldom visited places.

3.2 Setup

The physical driving setup uses one wide screen monitor and a Logitech G29 Driving Force steering wheel and pedals. We used ordinary speakers for playing the voice guidance and this was placed in front of driver, positioned on their left. To record what the participants are saying while driving and thinking aloud, we also set up a GoPro Hero 7 that faces the driver. We only start recording once the actual driving sessions have started.

We used the open-source CARLA simulator [6] as our virtual driving environment. We selected its Town3 map (**Fig. 1**a) because of its grid-like layout with many options for alternative routes. The map also features distinct land use areas and buildings that participants can easily distinguish (i.e. residential, commercial and industrial areas) for easy orientation in the environment. The virtual driving environment was used as is. For every participant session, we generate 60 random vehicles of different types around the map and they drive autonomously.

3.3 Conditions

Using the routes discussed in the Routes subsection, we designed the study to have three pure conditions and 6 conversation conditions. The pure conditions use only one voice agent namely, *PF* for Familiar voice agent only, *PO* for Optimal voice agent only, and *PE* for Explorer voice agent only. The conversation conditions use combinations of voice agents and are named the following: *FO* (Familiar+Optimal), *FE* (Familiar+Explorer), *OF* (Optimal+Familiar), *OE* (Optimal+Explorer), *EF* (Explorer+Familiar) and *EO* (Explorer+Optimal).

3.4 Protocol

We conducted a within-subject Wizard of Oz study which tasks each participant to drive 9 times under different navigation conditions. To reduce any ordering effect, we prepared 30 randomly ordered sequences of the 9 conditions and randomly assigned the participants to them. In the room, there is the participant and the experimenter. For the Japanese participants, a Japanese student assistant who is knowledgeable about the study and can translate English to Japanese is also present. For the duration of the actual driving sessions, the experimenter and assistants cannot talk nor react to the participant.

Orientation. At the beginning of each session, we briefed them about the project and the purpose of the experiment they are about to perform. Then, we obtained their consent to the procedures of the study and their answers to a pre-trial questionnaire that asks about their demographic information, driving background and experience with using navigation applications and voice guidance. Then, we oriented them about the steering wheel and pedals, and the simulation environment. For Japanese participants, it was emphasized that the environment is configured for driving on the right side of the road, which was different from what they are used to. We also mentioned the presence of a roundabout which does not exist in Japanese roads. During the whole orientation, we showed them a map. We gave them 3 minutes to drive around and get comfortable with the controls.

Familiarization with Voice Agents. After they became comfortable with the controls and environment, we asked what language they prefer for the voice guidance. All participants chose to use the local language versions, with nobody using the English voice guidance. We told them that they will hear 3 types of voices during the driving sessions and then played them the synthesized voices. Each voice was assigned a name and a number just for this step. To check how well they can differentiate the voices, we played them again but this time, they had to tell which voice was speaking (i.e. first voice, Tanaka-san, Olive). This step was done in order for them to easily detect when a conversation is being played already.

Remembering a Regular Route. Once they are familiar with the voices, the next step required them to familiarize with a route that served as their regular route to the destination. We showed them a map with the route drawn in red and they made two trips in the simulation following it. We played voice guidance so they can focus on the road and practice hearing the guidance. After this, they were asked to drive again to the destination but without voice guidance. In this step, we wanted to check how familiar they were with the route we asked them to follow. Once they reach the destination, we asked them to rate in a 7-point Likert scale how good they think the route is. For this question, we wanted to know later if their score affects how often they follow this route.

Trial of NASA-TLX. Since this was the first time that the participants have done this kind of study, we gave them a trial. We asked them to rate their workload using the NASA Task Load Index (TLX) questionnaire after following the voice guidance in the route familiarization step.

Driving Sessions. Each participant drove 9 times. Before anything, we reminded them that they are not obliged to follow the directions given by the voice guidance. After hearing the suggestions and conversations, it is up to them to decide which direction to go based on the given scenario and their personal preference. At the beginning of each drive, they were told to forget their previous drives and assume that they are hearing the voice guidance for the first time. They were also asked to think aloud. While we were starting their environments, we told them to internalize one of the following scenarios: *Regular Day*, *In a hurry*, and *Lots of time*.

Each scenario corresponds to a set of conditions. The Regular Day scenario is given in the PF, OF and EF conditions while the In a hurry scenario is given in the PO, FO and EO conditions. Lastly, the Lots of time scenario is given in the PE, FE and OE conditions. After each drive, they answered a post-task questionnaire that includes questions discussed in the following subsection. They can choose to have a break anytime during each session. After all 9 drives, they accomplished the Source-of-Workload Comparison sheet to complete the workload assessment. Each session lasted around 75 to 90 minutes.

3.5 Post-Task Questionnaire

First, participants assessed their amount of workload using the standard NASA TLX questionnaire. We did not use a modified version like in Karatas et al. [11] because this study does not intend to measure driving workload per se. To focus their assessment, we asked participants to assess based on the following aspects of the navigation task: a) listening to the voice guidance, b) choosing a direction after hearing the agents, and c) checking where to make the turn. The questionnaire was translated to Japanese following the work of Miyake [17]. Additionally, participants shared the reason behind their navigation choices (free text field) and how confident they were after choosing them (1-7 Likert scale).

4. Concept: Two-Party Conversations

For our Wizard of Oz study, we identified different routes that will be suggested, designed the voice agents and the two-party conversations, and planned when we will deliver them during the trip.

4.1 Route Suggestions

All trips resembled a home-to-work trip and starts in the residential area that is indicated by the red arrow in Figure Fig. 1a. They all went to the same destination indicated by the yellow arrow, which is opposite diagonally from the origin. The origin and destination pair was strategically chosen because they allowed us and the participants to select various routes to follow.

The following routes were identified based on the categories of trips that drivers usually take [22], [27]:

- **Route F** (Fig. 1b) - This route is straightforward and has a prominent landmark (i.e. a tunnel) that participants can easily remember and recognize [3].
- **Route O** (Fig. 1c) - This route uses the roundabout to avoid long waits at traffic signals [19], [20]. It makes early turns compared to the Familiar route and is relatively the shortest among the three routes.
- **Route E** (Fig. 1d) - This route is the longest among the three. It uses roads that are farther from the destination and goes through the other side of the map.

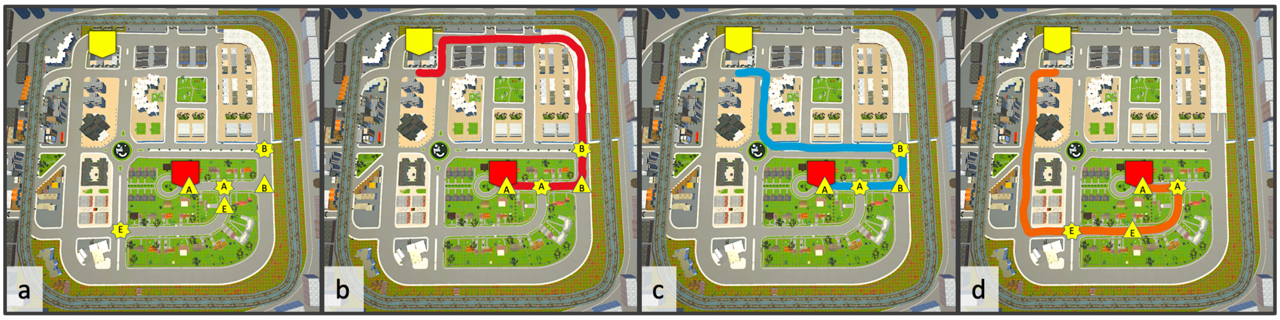

Fig. 1 The Town3 map with the 3 selected routes. The origin and destination are indicated by the red and yellow arrows, respectively. The triangles represent the locations where the conversations will be played. The stars represent the locations where suggested turns will take place. Symbols with similar letters represent the pairing of locations.

4.2 Voice Agents

We created four voice agents that deliver turn-by-turn instructions to the participants, two for Route F and one each for Route O and E.

Table 1 The four voice agents, their assigned routes and their sample give turn-by-turn instructions.

Voice Agent	Route	Sample Instruction
Generic	Route F	In 500 meters, turn left.
Familiar	Route F	Let's turn left after 500 meters. We take that direction on most days.
Optimal	Route O	We can turn left again in 300 meters. It will take us faster.
Explorer	Route E	Let's turn right. I think we haven't gone in this direction before.

Table 1 shows the four voice agents used in this study, along with their assigned routes. All voice agents give out route descriptors for next turns and sometimes an absolute distance towards the next turn. The Generic voice agent give instructions patterned after the instructions commonly delivered by current navigation applications like Waze and Google Maps. Its phrasing is direct and authoritative (i.e. *Turn Right* and *Go Straight*). On the other hand, the Familiar, Optimal and Explorer voice agents are designed to sound more suggestive and promotes a partnership between the voice agent and the driver, mimicking the way a human collaborative navigator would give out instructions [2]. We also phrased them as such because we are aiming for drivers to have agency, make instructed actions, and for them to not panic as much when they miss turns [4], [20]. To achieve this effect, we designed them to always start their instructions with "Let's," which is the shortest phrase we can add to the route descriptors without making them too long.

Aside from the typical route descriptors, the instructions given by the Familiar, Optimal and Explorer voice agents also include the rationale for their suggestion. The Familiar voice agent says a phrase or sentence that reminds how regular the driver takes a road (i.e. We take that direction on most days). The Optimal voice agent adds a phrase or sentence to emphasize fastness or having less waits on traffic signals (i.e. It will take us faster). Lastly, the Explorer voice agent adds a phrase or sentence that highlights the novelty of the suggestion (i.e. *I think we haven't gone in this direction before*).

4.3 Conversation Design

The main goal of this study is to explore how turn-by-turn instructions delivered in two-party conversations affect the navigation choices of drivers. Following the Participation Framework [10], we assume the scenario of a driver driving with two collaborative passengers acting as navigators. Similar to Karatas et al. [11], the driver participates as a bystander or a passive addressee to remove the conversational burden and to not distract the driver from driving. The active interlocutors are two voice agents which give different types of suggestions.

We designed the conversations to have each voice agent speak in two turns, for a total of four turns. Each voice agent speaks in polite and friendly tones [26] and acknowledges the suggestion of the other agent. The intention was to not make the conversation sound confrontational even though the voice agents may be presenting totally different suggestions. The voice agents split the typical route information they provide in two turns. They say their suggested direction in their first turn followed by their rationale in the second turn, and they do this alternately. Here is a sample conversation between the Familiar and Explorer voice agents in the FE condition:

Table 2 The conversation flow between the Familiar and Explorer voice agents when activated in the FE condition.

Turn	Voice	Instruction
T1	Familiar	"Let's go straight and then turn left."
T1	Explorer	"How about turning right before that?"
T2	Familiar	"That's possible. But we take a left on most days."
T2	Explorer	"That's true. But we haven't gone in this direction before."

In **Table 2**, the first voice agent (Familiar) suggests a direction followed by a counter-suggestion from the second voice agent (Explorer). In most cases, the counter-suggestion is also phrased as a question (i.e. Explorer: "How about turning right before that?"). In their second turn (T2), each voice agent shares the rationale behind their suggestion. They usually start with an affirmation or another question (i.e. Optimal: "Are you sure? Turning left will take us faster"), followed by the rationale. All route information shared in conversations are the same as when they are giving suggestions by themselves (i.e. pure conditions).

4.4 Delivery as Voice Guidance

In the conversation conditions, participants heard a conversation only once, which was either at the beginning or in the middle of the trip. Before a conversation, they heard only one voice agent giving route information. This is the first voice agent in the upcoming conversation. After the conversation is played, they continued hearing route information from the voice agent that they chose. **Table 3** shows the sequence of voice guidance for the whole trip in the OF condition. The voice guidance is started by the Optimal voice agent followed by the conversation. Assuming that the participant chose the Familiar suggestion, the voice guidance continued with the Familiar voice agent. Once they reach the destination, they heard the message "We've arrived at our destination." If they deviate from the designed routes, there are also generic route information prepared for each voice agent (i.e. "Let's turn left," "Let's go straight.").

Table 3 The sequence of route information given by the voice agents in the OF condition.

Turn	Voice	Instruction
	Optimal	"Let's get started!"
	Optimal	"Let's turn left after 500 meters."
T1	Optimal	"Let's turn left again in 300 meters."
T1	Familiar	"How about we continue straight?"
T2	Optimal	"Turning left will take us there faster."
T2	Familiar	"Right. But don't we always go through the tunnel?"
	Familiar	"Let's turn left after 500 meters and then turn right. We usually take that turn near our destination."
	Familiar	"We've arrived at our destination."

5. Results

5.1 Impact on Choices

In this pilot study, one of our main goals is to explore the impact and limitations of adding conversations in making navigation choices. We analyzed how associated their choices were for each given scenario and condition, along with a discussion of their reasons, and then discuss how combinations of these voice guidance affected their choices.

First, we wanted to see how aligned the participants' choices were based on the scenario that was given to them at the beginning of each condition. Looking at the contingency table of choices versus scenario in **Fig. 2**, a chi-square test shows that choices made by participants are dependent on the current context of their driving (χ^2=123.35, $p<0.05$).

Examining this association further, a chi-square test of the breakdown of choices made by participants under each condition (**Fig. 3**) shows that the type and combination of voice guidance was associated with their navigation choices (χ^2 = 229.87, $p<0.05$). Many participants navigating under the PF, FO and OF conditions were likely to choose the

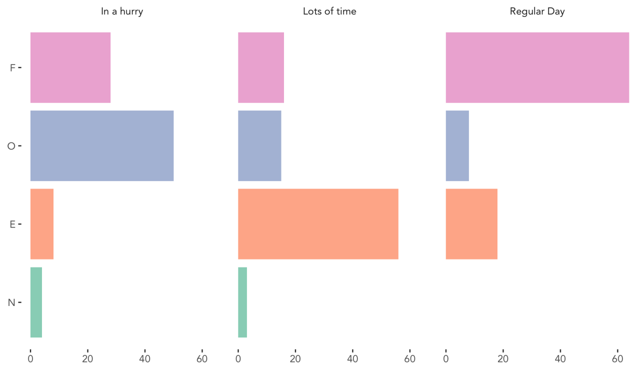

Fig. 2 Distribution of navigation choices per scenario. F refers to those who chose the Familiar suggestion, O for Optimal suggestion, E for Explorer suggestion, and N for those who chose neither of the given suggestions.

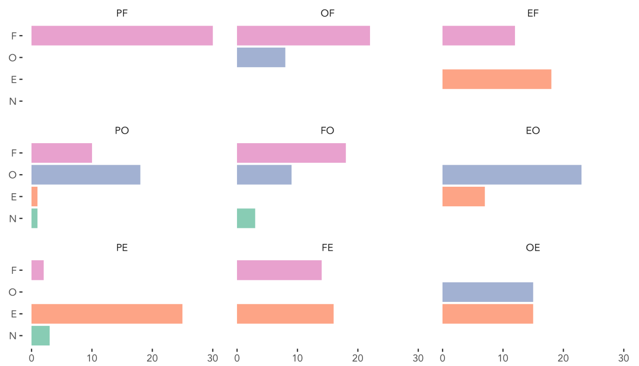

Fig. 3 Distribution of navigation choices per condition. The first row shows the conditions under the *Regular Day* scenario, followed by the conditions in the *In a hurry* and *Lots of time* scenarios.

Familiar suggestion, while those under the PO and EO conditions were likely to choose the Optimal suggestion. In the PE, EF and FE conditions, participants were attracted to choosing the Explorer suggestion, while both Optimal and Explorer suggestions were positively associated with the OE condition.

5.1.1 Regular Day Conditions

Given the prompt in the Regular Day scenario, the Familiar suggestion comprise almost 3/4 of the choices (N = 64) suggesting a strong association. And although there were those who chose the Optimal scenario, it was only chosen 8 times across the three conditions (PF, OF, EF).

In the PF condition, all 30 participants chose the Familiar suggestion. When suggestions were given in two-party conversations (OF and EF), their positive association with the Familiar suggestion was only maintained in the OF condition. Even in a two-party conversation with the Optimal voice agent, participants still chose the Familiar suggestion because they thought it was correct (P12), easier to follow (P7, P22-23) and familiar (P13, P16, P18, P26). In the EF condition, more participants chose the Explorer suggestion (N = 18) than the Familiar suggestion. They did not see it as a bad choice (P30) while some actually chose it because they wanted to explore a new route (P4, P6-7, P20, P24-25, P27) and they had ample time (P18, P26).

5.1.2 In a hurry Conditions

In the In a hurry scenario, participants were most attracted to choosing the Optimal suggestion with more than half of the choices made (N = 50). However, almost a third (N = 28) of the choices were still Familiar suggestions, indicating participants' tendency to default to regular routes even though they were in a hurry [20], [22], [27].

In the PO condition, more than half of the participants chose the Optimal suggestion (N = 18) primarily because they believed that the suggestion will take them faster towards the destination. After listening to a two-party conversation in the FO condition, the number of participants who chose the Optimal suggestion drops to less than a third in the pure condition (N = 9). Participants had the same reasons why they continued to follow the Optimal suggestion (i.e. belief that it is faster, with less waits). Having the opposite effect, participants in the EO condition followed the Optimal suggestion the most number of times in the In a hurry scenario (N = 23). This time, most of the participants agreed with the rationale of the Optimal voice agent and decided that it was more appropriate.

5.1.3 Lots of time Conditions

In the Lots of time scenario, they mostly chose the Explorer suggestion (N = 56) when they were told that they had much time to spare. This preference was consistently observed in the PE condition wherein 25 of them chose the Explorer suggestion. Two participants (P6, P8) followed it because that was the only suggestion given while six others were just open to the suggestion given the scenario they are in (P10, P12, P18, P24, P26, P29).

When the Explorer suggestion was given in the FE condition, there were less participants who chose it because they were also reminded with the Familiar suggestion. In the OE condition, participants were evenly split between the Optimal and Explorer suggestions. Everyone who chose the Explorer suggestion are driven by the non-urgency of the scenario, making them more open to explore new routes. However for those who chose the Optimal suggestion, while they also considered the non-urgency of the situation, they prioritized comfort (P12, P20, P23) and familiarity (P14, P21, P23, P28) in choosing.

5.2 Impact on their Confidence with Choices

Overall, confidence in their choices was generally lower when suggestions were given in conversations. When the *Familiar* suggestion was given on its own (PF condition), average confidence was 5.9 (M = 6.5, $\sigma = 1.41$) – the highest among conditions – with half of the participants giving a score of 7. Compared with other conditions given in the *Regular Day* scenario, their average confidence then drops to 5.6 for the OF condition (M = 6, $\sigma = 1.7$) and 5.4 for the EF condition (M = 5.5, $\sigma = 1.5$).

When participants heard suggestions that are different from what they are familiar with, they self-reported relatively lower confidence with their choices. The only increases happened when the familiar route suggestion was also given in conversations in the FO ($\mu = 5.5$, M = 6, $\sigma = 1.6$) and FE ($\mu = 5.6$, M = 6, $\sigma = 1.3$) conditions compared to when it was only the *Optimal* and *Explorer* suggestions mentioned. These suggests that the addition of novel suggestions, *Optimal* and *Explorer*, in conversations for all scenarios negatively affects how they perceive their choices.

5.2.1 High Confidence on Familiar and Optimal Suggestions

Based on a chi-square test, the self-reported confidence of drivers are choice-dependent ($\chi^2 = 23.90$, $p < 0.05$). In trips where the *Familiar* suggestion was followed (N=108), participants had an average confidence rating of 5.58 and this choice has a positive association with the confidence rating of 7, primarily because it is what they are already familiar with.

For all *Regular Day* scenario trips, participants who chose the *Familiar* suggestion were more confident (N=64, $\mu = 5.84$, M = 6). In the OF condition, while many trips chose the *Familiar* suggestion (N = 22) over the *Optimal* suggestion (N = 8), participants were equally as confident in choosing the *Optimal* suggestion ($\mu = 5.6$, M = 6.5, $\sigma = 1.85$) compared to choosing *Familiar* ($\mu = 5.5$, M = 6, $\sigma = 1.68$). Due to the low stakes nature of the scenario, even though the participants chose an unnecessarily faster suggestion, they did not mind it as much unlike when they chose the totally novel *Explorer* suggestion ($\mu = 4.89$, M = 5) overall. The low stakes nature of the *Lots of time* scenario also made participants more confident in following *Familiar* even though it was followed less often (N=16, $\mu = 5.63$, M = 6) and intentionally less appropriate compared to the *Explorer* suggestion (N=56, $\mu = 5.23$, M = 5). However in the *In a hurry* scenario, confidence with the *Familiar* suggestion was remarkably lower ($\mu = 4.96$, M = 5) even though it was chosen by participants in 47% of the trips in the PO and FO conditions. Despite being chosen the least among the three suggestions, the *Optimal* suggestion (N=73) was still positively associated with an average confidence of 5.75 after choosing the *Optimal* suggestion.

In the *In a hurry* scenario, participants who chose the appropriate *Optimal* suggestion reported a 5.8 average confidence (N=50, M = 6) which was consistently the highest in the PO ($\mu = 5.94$, M = 6, $\sigma = 1.3$), FO ($\mu = 6.33$, M = 1, $\sigma = 1$) and EO ($\mu = 5.48$, M = 6, $\sigma = 1.27$) conditions. Half of the participants were also confident with choosing the *Optimal* suggestion in the OE condition ($\mu = 6.33$, M = 1, $\sigma = 1$) despite it being the less appropriate choice.

5.2.2 Low Confidence for Explorer Suggestion

Choosing the *Explorer* suggestion is strongly positively associated with the confidence rating of 5 and strongly negatively associated with confidence rating 7. The novel nature of the suggestion made drivers less confident in their choices. This is consistent with previous works. Participants also felt unsure whether they will receive continuous guidance if they deviate from the familiar route.

5.2.3 Choosing Alternatives

We also looked at how confident the participants were when they chose the alternative suggestion over the appropriate ones. In the EF condition, participants started self-reporting low confidence scores of 1 to 4 (N = 4) after choosing the *Explorer* suggestion (μ = 4.89, M = 5, σ = 1.57) compared to those that chose the *Familiar* suggestion, who mostly reported scores between 5 to 7. In the *Regular Day* scenario, we expect them to prefer the *Familiar* suggestion over the *Explorer* one. It shows that even though they made a wrong choice, they must have realized after performing the task that they should have chosen the *Familiar* suggestion instead. The same lower level of confidence was also reported after participants chose the *Familiar* suggestion in the FE (μ = 5.43, M = 6, σ = 1.65) and FO (μ = 5.11, M = 5, σ = 1.75) conditions. They were not expected to prefer the *Familiar* suggestion, but 14 and 18 participants did in the FE and FO conditions respectively. And although some of them self-reported scores of 6 to 7 — because they are familiar with it — we also observed more participants reporting lower scores from 2 to 4 (N_{FE}= 4, N_{FO}= 7).

5.2.4 Good and Bad Pairs

Pairing novel suggestions in conversations made participants less confident with their choices even when they made appropriate ones. Participants in the EO condition who chose the *Optimal* suggestion reported confidence scores (μ = 5.48, M = 6, σ = 1.27) with seven of them giving scores between 3 and 4. This is lower compared to when they chose the same suggestion in the PO condition (μ = 5.94, M = 6, σ = 1.30) with only two participants reporting scores between 2 and 4. This was also the case in the OE condition where the average confidence score of 4.87 (M = 5, σ = 1.41) after choosing the *Explorer* suggestion was lower compared to the 5.12 average confidence score in the PE condition (M = 5, σ = 1.67). Four more participants gave scores between 1 and 4 in the OE condition. This is consistent with previous works that highlighted people's tendency to not prefer suggestions when they are too novel, putting them under a lot of uncertainty [7].

On the other hand, the delivery of the *Familiar* suggestion as an alternative in the FO and FE conditions made participants feel more confident in choosing the *Optimal* and *Explorer* suggestions. Even though less participants chose the *Optimal* suggestion in FO (N=9) compared to PO (N=18), and the *Explorer* suggestion in FE (N=16) compared to PE (N=25), they felt relatively more confident with average scores of 6.33 (M = 7, σ = 1) and 5.75 (M = 5.5, σ = 1) respectively. Including the *Familiar* suggestion gave participants a recognizable point of comparison. This was in contrast to their experience in the all-novel conditions (EO, OE) wherein they had to process two new suggestions and also recall their regular choices.

5.3 Impact on Workload

Because our concept gives more information than the typical voice guidance, we also wanted to see how much the two-party conversations impact the workload of the participants. The total NASA TLX scores show that the PF condition (M = 26.84, σ = 17.31) resulted to less workload compared to the PO (M = 47.5, σ = 20.8) and PE (M = 37.5, σ = 19.86) conditions. A Student's Paired lower-tailed t-tests between PF and PO, and PF and PE, indicates significant decrease in the PF condition, p¡0.001 and p¡0.05 respectively. Comparing between PO and PE, a Student's Paired upper-tailed t-test resulted in p¡0.01 indicating a significant increase in workload for the PO condition.

6. Towards Better Voice Guidance

Our pilot study provides initial insights into how voice guidance delivered as two-party conversations can impact the way drivers make navigation decisions.

6.1 Supporting Instructed Actions

Just by looking at the distribution of navigation choices made by participants, we can see clear patterns of choices being made in the pure conditions than in the conversations. When alternative suggestions get mentioned, their choices changed as well. While this can be considered as a negative result, we see it supporting our initial goal of encouraging drivers to have instructed actions [4]. Although we designed our scenarios to give more reasons for the participants to choose and follow certain suggestions (i.e. We expect the Optimal suggestion to be chosen more in the In a hurry scenario), we certainly do not consider choosing the alternative suggestions as a wrong decision. Our intent is to design and explore a new modality that will empower them with a handful of choices, rather than constrain them into following something that was already decided for them.

6.2 Better Reflection

The two-party conversations were designed to deliver an alternative suggestion followed by the suggestion appropriate for the scenario. Despite participants making less appropriate choices in some scenarios, the low self-reported confidence on their choices shows the potential of such conversations to support and encourage proper reflection for drivers. The delivery of two suggestions gave drivers a concrete and recent point of comparison which might be difficult if they try to recall choices in previous trips. Their late realization might positively impact their future choices when they encounter similar suggestions under the same circumstances.

7. Limitations

In this study, our within-subject design required participants to make 9 trips in one 90-minute session. Although we gave them some breaks in between drives and asked them to forget their previous drives before starting a new one, there might still be learning effects. Second, Our physical setup only used one monitor which may have made it difficult for the participants to verify the suggested turns, especially when they take the outer lanes. Considering the best options for Route O (*optimal*), we were limited by the existing roads in the simulation environment. A minor lengthening of that road segment where most participants ignored the Optimal suggestion may change the preference for it. Lastly, we acknowledge that the scenarios were few and could have been worded vaguely, leaving it to interpretation.

8. Conclusion and Future Work

Motivated by supporting drivers to make *instructed actions*, we introduced a nascent concept of a navigation application that integrates a two-party conversation in its voice guidance. Our within-subject Wizard of Oz study suggests the potential of this modality in encouraging drivers to follow certain suggestions with the right combination of voice agents. Although the conversations contributed to a higher workload unlike in previous studies that used the same modality, the participants' reported confidence suggests the potential to encourage them into making better navigation choices in succeeding drives. For future work, we would like to implement a prototype of this concept and explore in a longitudinal study whether the repeated use of such modality can actually change their navigation choices.

References

[1] : Google Maps - GPS Navigation App Ranking and Store Data — App Annie (2018).

[2] Antrobus, V., Burnett, G. and Krehl, C.: Driver-Passenger Collaboration as a basis for Human-Machine Interface Design for Vehicle Navigation Systems, *Ergonomics*, Vol. 60, No. 3, pp. 321–332 (online), DOI: 10.1080/00140139.2016.1172736 (2017).

[3] Antrobus, V., Large, D., Burnett, G. and Hare, C.: Enhancing Environmental Engagement with Natural Language Interfaces for In-Vehicle Navigation Systems, *Journal of Navigation*, Vol. 72, No. 3, p. 513–527 (online), DOI: 10.1017/S037346331800108X (2019).

[4] Brown, B. and Laurier, E.: The normal natural troubles of driving with GPS, *Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems - CHI '12*, New York, New York, USA, ACM, pp. 1621–1630 (online), DOI: 10.1145/2207676.2208285 (2012).

[5] Dingus, T. A., Hulse, M. C., Mollenhauer, M. A., Fleischman, R. N., Mcgehee, D. V. and Manakkal, N.: Effects of Age, System Experience, and Navigation Technique on Driving with an Advanced Traveler Information System, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 39, No. 2, pp. 177–199 (online), DOI: 10.1518/001872097778543804 (1997).

[6] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A. and Koltun, V.: CARLA: An Open Urban Driving Simulator, *Proceedings of the 1st Annual Conference on Robot Learning*, pp. 1–16 (2017).

[7] Ekstrand, M. D., Harper, F. M., Willemsen, M. C. and Konstan, J. A.: User Perception of Differences in Recommender Algorithms, *Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14*, New York, New York, USA, ACM Press, pp. 161–168 (online), DOI: 10.1145/2645710.2645737 (2014).

[8] Forlizzi, J., Barley, W. C. and Seder, T.: Where should i turn, *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, New York, New York, USA, ACM Press, p. 1261 (online), DOI: 10.1145/1753326.1753516 (2010).

[9] Fujino, T., Hashimoto, A., Kasahara, H., Mori, M., Iiyama, M. and Minoh, M.: Detecting Deviations from Intended Routes Using Vehicular GPS Tracks, *ACM Transactions on Spatial Algorithms and Systems*, Vol. 4, No. 1, pp. 1–21 (online), DOI: 10.1145/3204455 (2018).

[10] Goffman, E.: Footing, *Semiotica*, Vol. 25, No. 1-2, pp. 1–30 (online), DOI: 10.1515/semi.1979.25.1-2.1 (1979).

[11] Karatas, N., Yoshikawa, S., Silva, P. R. S. D. and Okada, M.: How Multi-Party Conversation Can Become an Effective Interface While Driving, *The Transactions of Human Interface Society*, Vol. 20, No. 3, pp. 371–388 (online), DOI: 10.11184/his.20.3_371 (2018).

[12] Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H. and Newell, C.: Explaining the user experience of recommender systems, *User Modeling and User-Adapted Interaction*, Vol. 22, No. 4, pp. 441–504 (online), DOI: 10.1007/s11257-011-9118-4 (2012).

[13] Large, D. R., Burnett, G., Antrobus, V. and Skrypchuk, L.: Driven to discussion: engaging drivers in conversation with a digital assistant as a countermeasure to passive task-related fatigue, *IET Intelligent Transport Systems*, Vol. 12, No. 6, pp. 420–426 (online), DOI: 10.1049/iet-its.2017.0201 (2018).

[14] Mahmud, A. A., Mubin, O. and Shahid, S.: User experience with in-car GPS navigation systems: comparing the young and elderly drivers, *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ACM, pp. 1–90 (online), DOI: 10.1145/1613858.1613962 (2009).

[15] Matsumura, K. and Sumi, Y.: What Are You Talking About While Driving?, *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '14*, New York, New York, USA, ACM Press, pp. 1–8 (online), DOI: 10.1145/2667317.2667417 (2014).

[16] Mehndiratta, S. and Quiros, T. P.: Traffic jams, pollution, road crashes: Can technology end the woes of urban transport? (2017).

[17] MIYAKE, S.: Special Issues No. 3 : Measurement Technique for Ergonomics, Section 3 : Psychological Measurements and Analyses (6), *The Japanese Journal of Ergonomics*, Vol. 51, No. 6, pp. 391–398 (online), DOI: 10.5100/jje.51.391 (2015).

[18] Quercia, D., Schifanella, R. and Aiello, L. M.: The shortest path to happiness, *Proceedings of the 25th ACM conference on Hypertext and social media - HT '14*, New York, New York, USA, ACM Press, pp. 116–125 (online), DOI: 10.1145/2631775.2631799 (2014).

[19] Ringhand, M. and Vollrath, M.: Investigating urban route choice as a conflict between waiting at traffic lights and additional travel time, *Transportation Research Procedia*, Vol. 25, pp. 2432–2444 (online), DOI: 10.1016/j.trpro.2017.05.258 (2017).

[20] Samson, B. P. V. and Sumi, Y.: Exploring Factors That Influence Connected Drivers to (Not) Use or Follow Recommended Optimal Routes, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA, ACM, pp. 371:1–371:14 (online), DOI: 10.1145/3290605.3300601 (2019).

[21] Sumi, Y. and Mase, K.: AgentSalon: Facilitating Face-to-face Knowledge Exchange Through Conversations Among Personal Agents, *Proceedings of the Fifth International Conference on Autonomous Agents*, AGENTS '01, New York, NY, USA, ACM, pp. 393–400 (online), DOI: 10.1145/375735.376344 (2001).

[22] Tang, W. and Cheng, L.: Analyzing multiday route choice behavior of commuters using GPS data, *Advances in Mechanical Engineering*, Vol. 8, No. 2, pp. 1–11 (online), DOI: 10.1177/1687814016633030 (2016).

[23] Todo, Y., Nishimura, R., Yamamoto, K. and Nakagawa, S.: Development and Evaluation of Spoken Dialog Systems with One or Two Agents through Two Domains, *Text, Speech, and Dialogue* (Habernal, I. and Matoušek, V., eds.), Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 185–192 (2013).

[24] Waze: Driver Satisfaction Index, Technical report, Waze (2016).

[25] Wu, M.: Hybrid user perception model: comparing users' perceptions toward collaborative, content-based, and hybrid recommender systems, PhD Thesis, Iowa State University (2015).

[26] Yoshiike, Y., De Silva, P. R. S. and Okada, M.: MAWARI: A Social Interface to Reduce the Workload of the Conversation, *Proceedings of the Third International Conference on Social Robotics*, ICSR'11, Berlin, Heidelberg, Springer-Verlag, pp. 11–20 (online), DOI: 10.1007/978-3-642-25504-5_2 (2011).

[27] Zhu, S. and Levinson, D.: Do People Use the Shortest Path? An Empirical Test of Wardrop's First Principle, *PLoS ONE*, (online), DOI: 10.1371/journal.pone.0134322 (2015).