Are Two Heads Better than One? Exploring Two-Party Conversations for Car Navigation Voice Guidance

Briane Paul V. Samson
Future University Hakodate
Hakodate, Japan
De La Salle University
Manila, Philippines
brianepauls@acm.org

Yasuyuki Sumi
Future University Hakodate
Hakodate, Japan
sumi@acm.org

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
http://dx.doi.org/10.1145/3334480.3382818

Abstract

Voice guidance for car navigation typically considers drivers as docile actors. Recent works highlight limitations to this assumption which make drivers rely less on given directions. To explore how drivers can make better navigation decisions, we conducted a pilot Wizard-of-Oz study that gives turn suggestions in conversations between two voice 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 their route choices, perceived workload and utterances while driving. We 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. On the other hand, two-party conversations allowed drivers to better reflect on their choices after trips. We conclude by discussing preliminary design implications for car navigation voice guidance specifically and recommender systems in general.

Author Keywords

Voice Guidance; Voice Agents; Two-Party Conversation; Navigation Applications; Recommender Systems; Driving

CCS Concepts

•**Human-centered computing** $\rightarrow$ **Human computer interaction (HCI);** *Empirical studies in HCI;*

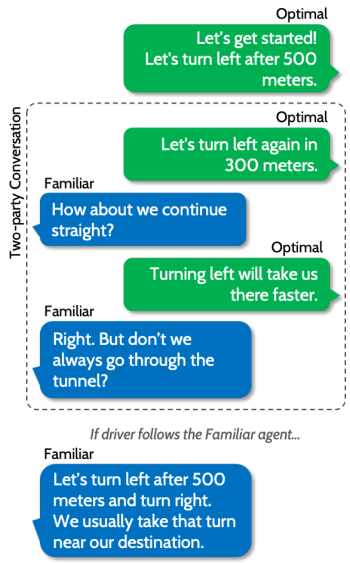

Figure 1: A sample sequence of turn suggestions given in the OF (*Optimal-Familiar*) condition. It has a two-party conversation between the Optimal and Familiar voice agents. In this sequence, turn suggestions are first given by the 1st voice agent in the pair. They also start the conversation with the 2nd voice agent. After choosing a suggestion between the two, the trip continues with turn suggestions from the chosen voice agent, in this the Familiar.

Introduction

In recent years, in-car navigation systems and their mobile application counterparts have gained popularity among drivers [1, 21], especially those driving in cities and other urban areas with increasingly complex road networks. As their core routing service became more advanced with machine learning and sensing capabilities, they are now integral in many commutes to monitor regular routes, to discover new ones and sometimes to avoid traffic congestion. By default, drivers are recommended the fastest route to their destinations, with alternative routes either shown up front (i.e. Google Maps) or hidden for you to discover (i.e. Waze). While many people agree and say that they do want fast or short routes when asked at any given day, asking them again in actual driving contexts shows otherwise [14]. This is further supported by empirical evidence from GPS tracks and recorded actual trips that show drivers' repeated non-preference of recommended fastest routes [15, 23, 19, 8, 4, 17]. While there is great support for drivers to make decisions before starting a trip, there are gaps in current systems and applications that fail to consider changing needs, contexts and preferences, which ultimately affect their compliance and trust on the recommendation.

After choosing to follow optimal route recommendations, drivers tend to deviate because of *normal, natural troubles* that they experience with GPS devices [4], road unfamiliarity, and perceived impracticality and driving unsuitability [17]. Brown & Laurier [4] and Samson & Sumi [17] made extensive accounts and implications of these problems, and encourage designers to not think of drivers as docile actors and to focus more on helping them make *instructed actions* when designing voice guidance. Thus, our design goal is to support a driver's ability to interpret and analyze new route guidance and information in order to help them make better navigation decisions. Specifically, we focus on exploring how to provide ample route information and alternative suggestions for some turns during a trip, providing them agency.

Essentially, navigation is a social activity among drivers and navigators [7, 13]. 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]. However, actively engaging the driver might pose a distraction and increase cognitive workload [10]. As a step towards supporting *instructed actions* by drivers, we explore a concept that use two-party conversation between voice agents, but with the driver as an observer and not engaged in the conversations. We conducted a pilot Wizard-of-Oz study in a within-subject design with 30 participants. Participants were asked to drive 9 times under different conditions (3 without and 6 with conversation) and we recorded their navigation choices, workload, and confidence with their choices. In this pilot exploration, we found that two-party conversations can encourage drivers to follow appropriate but with the right combination of voice agents.

Related Work

Experience with Navigation Systems – Specific to navigation applications, recent studies focused more on their effects on driving and navigation performance. Early work by Dingus et. al. [5] found that voice is the safest modality for receiving turn-by-turn guidance. However, drivers in general still show difficulty at times when following their guidance [5, 12]. Brown & Laurier elaborate on the *normal, natural troubles* with GPS devices such as complex routes, redundant and ill-timed turn instructions, and sensing inaccuracies [4]. Observing different decision-making practices, information used, and the type of trip and trip context, Samson & Sumi highlighted why deviations happen [17]. The

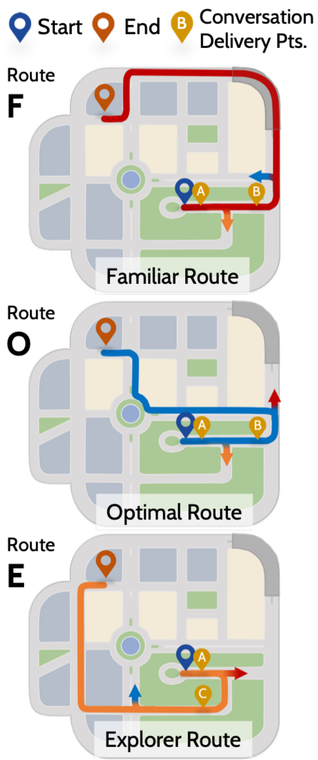

Figure 2: The selected routes from the map. The start and end points are the same for all routes. The orange markers are where the conversations are delivered, only once per trip. The 2 diverging arrows from each route show the alternative turns given in the conversations, colored to represent the type of route they lead to.

last two studies echo the need to support a driver's agency. 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.

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 [18] and Todo et. al. [20] 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. [22], 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. [11] found that engaging drivers in one-to-one conversations with a digital assistant can reduce driver fatigue while Karatas et. al. [10] 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.

Concept: Two-Party Conversations

Route Suggestions – All routes in Figure 2 resemble a home-to-work trip and starts in the residential area of the map. They all had the same destination, which is opposite diagonally from the start point. This pair of points allowed us to identify routes based on Zhu & Levinson and Tang & Cheng's categories of trips that drivers usually take [19, 23]. *Route F* is straightforward and has a prominent landmark (i.e. a tunnel) that participants can easily remember and recognize [3]. *Route O* uses the roundabout to avoid long waits at traffic signals [16, 17]. It makes early turns compared to the Familiar route and is relatively the shortest. Lastly, *Route E* is the longest and uses roads that are farther from the end pt on the other side of the map. This was based on the way modern apps suggest novel routes that are not short distance but algorithmically determined to be faster to avoid busy routes [17].

Voice Agents – For this pilot, we created four voice agents in Table 1. All give out route descriptors for next turns and sometimes the distance before the turn. The Generic voice agent is patterned after popular navigation applications like Waze and Google Maps, using direct and authoritative phrasing (i.e. *Turn Right* and *Go Straight*). It is only used for the familiarization step in the protocol, not in the actual conditions. The rest of the voice agents are designed to sound suggestive, mimicking human collaborative navigators [2]. They always begin their instructions with "Let's," which is the shortest phrase we can add to route descriptors without making them too long. The Familiar, Optimal and Explorer voice agents also include the rationale for their suggestion in their instructions.

Conversation Design – Following Goffman's Participation Framework [9], our driver drives with two collaborative passengers acting as navigators. In Karatas et al. [10], they compared two conditions: an agent conversing with driver and multi-party conversation between 3 agents only. Similar to theirs, our driver is 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. In conversations, each voice agent speaks in two turns. They speak in polite and friendly tones [22] and acknowledges the suggestion of the other agent. We did not want them to sound confrontational despite presenting totally different sugges-

Table 1: The four voice agents and their sample turn-by-turn instructions. Both Generic and Familiar voice agents use Route F. Optimal and Explorer voice agents use Routes O and E respectively. Each line of navigation instruction is synthesized using Google Cloud's Text-to-Speech API. They are played by the researchers during the sessions.

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

tions. In their first turn, they say their suggested direction followed by their rationale in the second turn. They do this alternately as shown in Figure 1.

Delivery as Voice Guidance – In the conversation conditions, drivers hear a conversation once, which can be at the beginning or in the middle of the trip. Before a conversation, they hear only one voice agent giving route information and is the first voice agent in the upcoming conversation. After the conversation is played, they continue hearing route information from the voice agent that they chose. Figure 1 shows the sequence of voice guidance for the whole trip in the OF condition.

Method

Participants – We recruited 30 participants (14 women, M=29 y.o., SD=10.6 y.o.) with at least 1 year of driving experience and has a driver's license. They are comprised of 12 Filipinos and 18 Japanese nationals but we do not compare between nationalities. All do not drive as part of their occupation (non-professional) and only use navigation applications when going to an unknown destination. Only one use it almost everywhere they go. Regarding voice guidance, 18 of them do not regularly use it. For those that do, they frequently use it when they go on trips to new or seldom visited places.

Setup – The physical driving setup in Figure 3 uses one wide screen monitor and a Logitech G29 Driving Force steering wheel and pedals. We used ordinary speakers, positioned on their left. To record what the participants are saying while driving and thinking aloud, we also set up a GoPro Hero 7. 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 used its Town3 map (Figure 2) because of the grid-like layout with many options for alternative routes. It also features distinct land use areas and buildings that participants can easily distinguish (i.e. residential, commercial areas). The Town3 map was used as is. For every participant session, we generate 60 random autonomous vehicles around the map.

Conditions – There are three conditions without conversations namely, *PF* for Familiar only, *PO* for Optimal only, and *PE* for Explorer only. Six conversation conditions use combinations of agents: *FO* (Familiar+Optimal), *FE* (Familiar+Explorer), *OF* (Optimal+Familiar), *OE* (Optimal+Explorer), *EF* (Explorer+Familiar) and *EO* (Explorer+Optimal). The suggestion of the 2nd agent in conversations is the expected choice (*appropriate*).

Protocol – To reduce any ordering effect, conditions were counterbalanced using Latin Square design and randomly assigned. In the room, there is the participant and the researcher. For Japanese participants, there is a student assistant to help translate during orientation only. During actual driving sessions, the researcher and assistant cannot talk nor react to the participant.

We oriented them about the project, obtained their consent to do the study and asked to answer a pre-trial questionnaire. Then, we let them get used to the steering wheel and pedals, and the simulation environment. They were shown a map and we gave them 3 minutes to drive around and get comfortable with the controls. They were asked their preferred language for the voice guidance and all chose their mother tongue (Japanese and Filipino). We then played sample instructions given by the Familiar, Optimal and Explorer voice agents in the chosen language. To familiarize them with a route (Route F) that served as their regular route to the end point, we played the Generic voice agent. They followed this route three times. Before doing

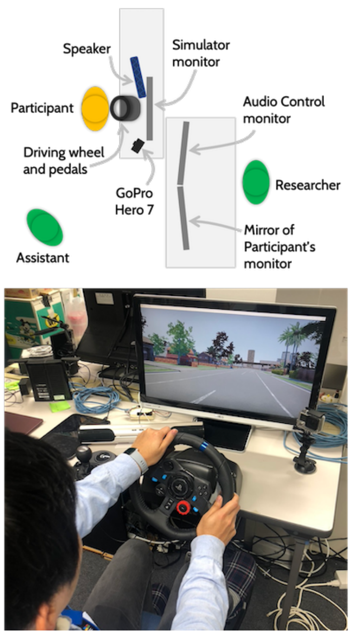

Figure 3: The Wizard-of-Oz setup. [Top] The overhead view of the room with the location of participant, researcher and assistant. The researcher sees a mirror of the participant's monitor. Every time the driver comes near a decision point, the researcher plays the recorded instructions and conversations. [Bottom] A participant using the setup. In front are: 1 monitor for driving simulator, 1 GoPro camera for capturing utterances while thinking aloud, 1 steering wheel set, and 1 speaker.

the 9 conditions, we reminded them that they are free to choose anything or none of the suggestions given. They were 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*. Because participants do not know that we expect specific choices to made in each condition, scenarios aim to contextualize the decision-making of the drivers and encourages a more appropriate choice. 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 NASA-TLX questionnaire. We asked participants to assess based on the following: a) listening to the voice guidance, b) choosing a direction after hearing the agents, and c) checking where to make the turn. Additionally, participants shared the reason behind their navigation choices (free text field) and how confident they were after choosing them (1-7 Likert scale). Each session lasted around 75 to 90 minutes.

Results

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 condition, along with a discussion of their reasons, and then discuss how combinations of these voice guidance affected their choices.

We tallied the participants' choices and found that all suggestions were chosen at least once by the participants in each scenario, with some choosing neither of the given choices. 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 (Figure 4) shows that the type and combination of voice guidance was associated with their navigation choices (χ^2 = 229.87, $p<0.05$).

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.

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, σ = 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, σ = 1.7) and 5.4 for the EF condition (M = 5.5, σ = 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 (μ = 5.5, M = 6, σ = 1.6) and FE (μ = 5.6, M = 6, σ = 1.3) conditions compared to

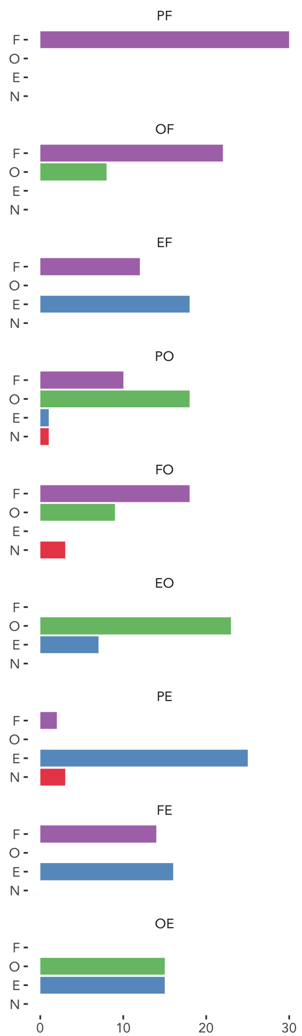

Figure 4: Distribution of navigation choices per condition.

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.

Choosing Alternatives. We look at how confident the participants were when they chose the alternative suggestion over the expected one based on the scenario. 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 less appropriate choice, they must have realized after performing the task that they should have chosen the *Familiar* suggestion instead.

Discussion

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.

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.

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.

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 pilot study suggests the potential of this modality in encouraging drivers to follow certain suggestions with the right combination of voice agents. The participants' low reported confidence suggests after making wrong choices shows potential to encourage them into making better navigation choices in succeeding drives. A longitudinal study on the repeated use of such modality should be explored.

REFERENCES

[1] 2018. Google Maps - GPS Navigation App Ranking and Store Data | App Annie. (2018). https://www.appannie.com/en/apps/ios/app/google-maps/

[2] Vicki Antrobus, Gary Burnett, and Claudia Krehl. 2017. Driver-Passenger Collaboration as a basis for Human-Machine Interface Design for Vehicle Navigation Systems. *Ergonomics* 60, 3 (3 2017), 321–332. DOI: http://dx.doi.org/10.1080/00140139.2016.1172736

[3] Vicki Antrobus, David Large, Gary Burnett, and Chrisminder Hare. 2019. Enhancing Environmental Engagement with Natural Language Interfaces for In-Vehicle Navigation Systems. *Journal of Navigation* 72, 3 (2019), 513–527. DOI: http://dx.doi.org/10.1017/S037346331800108X

[4] Barry Brown and Eric Laurier. 2012. The normal natural troubles of driving with GPS. In *Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems - CHI '12*. ACM, New York, New York, USA, 1621–1630. DOI: http://dx.doi.org/10.1145/2207676.2208285

[5] Thomas A. Dingus, Melissa C. Hulse, Michael A. Mollenhauer, Rebecca N. Fleischman, Daniel V. Mcgehee, and Natarajan Manakkal. 1997. 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* 39, 2 (1997), 177–199. DOI: http://dx.doi.org/10.1518/001872097778543804

[6] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. CARLA: An Open Urban Driving Simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*. 1–16.

[7] Jodi Forlizzi, William C. Barley, and Thomas Seder. 2010. Where should i turn. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, New York, New York, USA, 1261. DOI: http://dx.doi.org/10.1145/1753326.1753516

[8] Takumi Fujino, Atsushi Hashimoto, Hidekazu Kasahara, Mikihiko Mori, Masaaki Iiyama, and Michihiko Minoh. 2018. Detecting Deviations from Intended Routes Using Vehicular GPS Tracks. *ACM Transactions on Spatial Algorithms and Systems* 4, 1 (6 2018), 1–21. DOI: http://dx.doi.org/10.1145/3204455

[9] Erving Goffman. 1979. Footing. *Semiotica* 25, 1-2 (1979), 1–30. DOI: http://dx.doi.org/10.1515/semi.1979.25.1-2.1

[10] Nihan Karatas, Soshi Yoshikawa, P. Ravindra S. De Silva, and Michio Okada. 2018. How Multi-Party Conversation Can Become an Effective Interface While Driving. *The Transactions of Human Interface Society* 20, 3 (2018), 371–388. DOI: http://dx.doi.org/10.11184/his.20.3_371

[11] D. R. Large, G. Burnett, V. Antrobus, and L. Skrypchuk. 2018. Driven to discussion: engaging drivers in conversation with a digital assistant as a countermeasure to passive task-related fatigue. *IET Intelligent Transport Systems* 12, 6 (2018), 420–426. DOI: http://dx.doi.org/10.1049/iet-its.2017.0201

[12] Abdullah Al Mahmud, Omar Mubin, and Suleman Shahid. 2009. User experience with in-car GPS navigation systems: comparing the young and elderly drivers. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 1–90. DOI: http://dx.doi.org/10.1145/1613858.1613962

[13] Kohei Matsumura and Yasuyuki Sumi. 2014. What Are You Talking About While Driving?. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '14*. ACM Press, New York, New York, USA, 1–8. DOI: http://dx.doi.org/10.1145/2667317.2667417

[14] Bastian Pfleging, Alexander Meschtscherjakov, Stefan Schneegass, and Manfred Tscheligi. 2014. Experience Maps: Experience-Enhanced Routes for Car Navigation. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '14*. ACM, New York, New York, USA, 1–6. DOI: http://dx.doi.org/10.1145/2667239.2667275

[15] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The shortest path to happiness. In *Proceedings of the 25th ACM conference on Hypertext and social media - HT '14*. ACM Press, New York, New York, USA, 116–125. DOI: http://dx.doi.org/10.1145/2631775.2631799

[16] Madlen Ringhand and Mark Vollrath. 2017. Investigating urban route choice as a conflict between waiting at traffic lights and additional travel time. *Transportation Research Procedia* 25 (2017), 2432–2444. DOI: http://dx.doi.org/10.1016/j.trpro.2017.05.258

[17] Briane Paul V. Samson and Yasuyuki Sumi. 2019. Exploring Factors That Influence Connected Drivers to (Not) Use or Follow Recommended Optimal Routes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 371, 14 pages. DOI: http://dx.doi.org/10.1145/3290605.3300601

[18] Yasuyuki Sumi and Kenji Mase. 2001. AgentSalon: Facilitating Face-to-face Knowledge Exchange Through Conversations Among Personal Agents. In *Proceedings of the Fifth International Conference on Autonomous Agents (AGENTS '01)*. ACM, New York, NY, USA, 393–400. DOI: http://dx.doi.org/10.1145/375735.376344

[19] W. Tang and L. Cheng. 2016. Analyzing multiday route choice behavior of commuters using GPS data. *Advances in Mechanical Engineering* 8, 2 (2016), 1–11. DOI: http://dx.doi.org/10.1177/1687814016633030

[20] Yuki Todo, Ryota Nishimura, Kazumasa Yamamoto, and Seiichi Nakagawa. 2013. Development and Evaluation of Spoken Dialog Systems with One or Two Agents through Two Domains. In *Text, Speech, and Dialogue*, Ivan Habernal and Václav Matoušek (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 185–192.

[21] Waze. 2016. *Driver Satisfaction Index*. Technical Report. Waze. https://inbox-static.waze.com/driverindex.pdf

[22] Yuta Yoshiike, P. Ravindra S. De Silva, and Michio Okada. 2011. MAWARI: A Social Interface to Reduce the Workload of the Conversation. In *Proceedings of the Third International Conference on Social Robotics (ICSR'11)*. Springer-Verlag, Berlin, Heidelberg, 11–20. DOI: http://dx.doi.org/10.1007/978-3-642-25504-5_2

[23] Shanjiang Zhu and David Levinson. 2015. Do People Use the Shortest Path? An Empirical Test of Wardrop's First Principle. *PLoS ONE* (2015). DOI: http://dx.doi.org/10.1371/journal.pone.0134322