Janssen, C.P., Iqbal, S.T., Ju, Y.C. (in press) Sharing a Driver's Context with a Caller via Continuous Audio Cues to Increase Awareness about Driver State. Journal of Experimental Psychology: Applied. In press in 2014.

Sharing a Driver's Context with a Caller via Continuous Audio Cues to Increase Awareness about Driver State

Christian P. Janssen

Microsoft Research and The Smith-Kettlewell Eye Research Institute

Shamsi T. Iqbal and Yun-Cheng Ju

Microsoft Research

Author note

Christian P. Janssen, Microsoft Research and The Smith-Kettlewell Eye Research Institute; Shamsi T. Iqbal, Microsoft Research; Yun-Cheng Ju, Microsoft Research.

Christian P. Janssen is now at The Smith-Kettlewell Eye Research Institute.

We thank Erin Caroll, Mary Czerwinsky, Paul Dietz, Eric Horvitz, Mike Seltzer, Desney Tan, and Andy Wilson for useful discussions regarding the experimental design of this study. We also wish to thank Duncan Brumby, John Krumm, and Jason McCarley for feedback on earlier versions of this manuscript.

Correspondence concerning this article should be addressed to Christian P. Janssen, The Smith-Kettlewell Eye Research Institute, San Francisco, CA 94115. E-mail: cjanssen@ski.org.

Abstract

In an experiment using a driving simulator we investigated if sharing information of a driver's context with a remote caller via continuous audio cues can make callers more aware of the driving situation. Increased awareness could potentially help in making the conversation less distracting. Prior research has shown that while sharing context using video can create such beneficial effects, it also has some practical disadvantages. It is an open question whether other modalities might also provide sufficient context for a caller. In particular, the effects of sharing audio, a cheaper, more salient, and perhaps more practical alternative than video, are not well understood. We investigated sharing context using direct cues in the form of realistic driving sounds (e.g., car honks, sirens) and indirect cues in the form of elevated heartbeats. Sound sharing affected the caller's perception of the driver's busyness. However, this had at most a modest effect on conversation and driving performance. An implication of these results is that although sharing sounds can increase a caller's awareness of changes in the driver's busyness, they need more training or information on how to leverage such context information to reduce disruption to driving. Limitations and implications are discussed.

Keywords: Driver distraction, Attention, Conversation, Adaptation, Context awareness.

Sharing a Driver's Context with a Caller via Continuous Audio Cues to Increase Awareness about Driver State

Multitasking in the car, in particular talking on the phone while driving, is a persistent practice. These practices prevail despite tougher legislation (Insurance Institute for Highway Safety, 2012), media attention and public initiatives (e.g., "Think before you speak: Distracted driving is the new drunk driving," 2011), and over twenty years of research demonstrating the dangers (e.g., Alm & Nilsson, 1995; Brookhuis, De Vries, & De Waard, 1991; Caird, Willness, Steel, & Scialfa, 2008; Horrey & Wickens, 2006; Kearney, Rizzo, & Severson, 2007; AJ McKnight & McKnight, 1993; Strayer & Johnston, 2001). The compromise of hands-free calling is unfortunately also not a panacea as the effort of thinking to generate speech required in regular conversations can still distract (Becic et al., 2010; Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001).

It seems unlikely that banning phone conversations while driving will be successful in practice, given expectations of modern mobile communications and given that occasionally there might be benefits of having a conversation while driving. For example, talking can keep a person more vigilant when performing tasks that are low on demand (Atchley & Chan, 2011), such as driving on a deserted highway. It is therefore important to get a theoretical understanding of how conversations impact driving safety, and applying that understanding to realize how technology can be used to reduce the amount of distraction a conversation may cause during driving.

One promising avenue has been based on the insight that sharing information of the driver's context with a remote caller can make remote conversations more cognizant of the driving situation and subsequently, less distracting (Charlton, 2009; Crundall, Bains, Chapman, & Underwood, 2005; Drews, Pasupathi, & Strayer, 2008; Gugerty, Rakauskas, & Brooks, 2004; Maciej, Nitsch, & Vollrath, 2011; Schneider & Kiesler, 2005). This work has mostly focused on sharing context through video. As we will explain in detail later on, while this approach of sharing context through video appears promising there are some important practical limitations. Hence it is important for research to explore other alternatives.

A possible alternative to video is sharing cues through continuous audio signals. This is cheaper and requires less network bandwidth. However, a theoretical understanding is missing of whether audio signals can indeed successfully convey context (without the visual context offered by video). This paper reports an empirical study that used a medium-fidelity driving simulator to investigate: (a) how sharing of sounds affects the remote caller's perception of the driver's busyness, (b) if this subsequently affects the conversation pattern, and (c) if this leads to relatively better driving performance. Our findings showed that sharing traffic sounds and simulated heartbeat sounds results in more accurate perception of how busy the driver is, as compared to sharing no sounds or nature sounds during a conversation. In experiment 1 we found that the caller's talk segments were shorter and that driving performance on one metric became better when context sounds were shared with the caller. However, this effect did not replicate in experiment 2. This suggests that while callers can get a reliable changed perception of the driver's busyness when sounds are shared, effective usage of that knowledge is yet to be understood.

To motivate this research further, we will first describe current insights on the distracting nature of conversations while driving and on how sharing context might alleviate some of these problems.

How conversations distract driving

Conversations during driving can be generally distracting but vary based on the content of the conversation (Iqbal et al., 2010). The primary reason for distraction is that talking requires cognitive resources. In particular, the more "thinking" is required for talking, the stronger the interference on other tasks such as visual attention tasks (Kunar et al., 2008; Strayer & Johnston, 2001) and driving (Iqbal et al., 2010). The distraction is likely due to a central processing bottleneck (Kunar et al., 2008), not the act of generating sound, as even mentally (and not verbally) reciting a list can interfere with driving performance (Salvucci & Beltowska, 2008).

Due to these fundamental resource limitations (Norman & Bobrow, 1975), eliminating the distracting effects of a conversation on driving performance might be difficult. However, the amount and severity of distraction can change with the nature of the conversation. That is, if the conversation is less dense (e.g., with more pauses and less complex sentences), processing and responding to a spoken message will demand fewer cognitive resources. These resources (e.g., the central processing bottleneck, Kunar et al., 2008; Salvucci & Beltowska, 2008) can then be directed to driving.

Conversations during driving are not all equally distracting. In particular, talking to a passenger can be less distracting than talking to someone over the phone (Lee & Abdel-Aty, 2008; McEvoy, Stevenson, & Woodward, 2007). This can be attributed in part to the lack of shared context between driver and caller. Conversational partners who are aware of a driver's context can use this information to adapt their conversation to this context such that it becomes less demanding. For example, they can reduce the conversation's density (Charlton, 2009; Crundall et al., 2005; Gugerty et al., 2004; Maciej et al., 2011; Schneider & Kiesler, 2005) and complexity (i.e., the number of syllables per word, Drews et al., 2008). This will again reduce the demands on the driver for cognitive resources to process and respond to a conversation. Moreover, callers or passengers who are aware of the context might even make this topic of conversation, and in that way warn the driver for upcoming dangers (Charlton, 2009; Drews et al., 2008).

Sharing Context to Reduce Distraction

The hypothesis of beneficial effects of sharing context between drivers and callers has triggered many investigations into various aspects of context sharing, both when using driving simulators (e.g., Charlton, 2009; Drews et al., 2008; Gugerty et al., 2004; Maciej et al., 2011; Schneider & Kiesler, 2005) and on the road driving (Crundall et al., 2005). The typical finding is that the density of the conversation was reduced and that driving performance improved when the driving context was shared with a remote caller compared to a situation without context sharing.

Studies of context sharing between drivers and remote callers can be further classified based on three aspects: (a) the timing of sharing, (b) who the context is shared with, and (c) the modality and type of the context. In terms of when context is shared, sharing context with the caller before a call takes place (de Guzman, Sharmin, & Bailey, 2007; Grandhi, Schuler, & Jones, 2011; Lindqvist & Hong, 2011) can be beneficial to the caller in deciding whether to make a call or not, for example, based on the knowledge that someone is driving. Systems that provide context during a call (Charlton, 2009; Maciej et al., 2011; Schneider & Kiesler, 2005) complement this approach as they can communicate *changes* in busyness after the conversation has started. Within such systems, information can be shared continuously (Charlton, 2009; Maciej et al., 2011; Schneider & Kiesler, 2005) or occasionally (e.g., to warn for difficult driving sections, Charlton, 2009).

In terms of whom information is shared with, most studies focus on sharing driving context with the remote caller to enrich their context of the driving situation (e.g., Charlton, 2009; Crundall et al., 2005; Gugerty et al., 2004; Maciej et al., 2011; Schneider & Kiesler, 2005). The implicit assumption here is that sharing more information with the driver (e.g., about the context of the remote caller) would create additional cognitive load that could distract from driving, whereas the caller might have the necessary resources to think about the conversation and use those to make the conversation less demanding for the driver.

Some studies have focused on providing the driver with additional driving context by alerting them of upcoming dangerous road segments (e.g., Charlton, 2009; Iqbal, Horvitz, Ju, & Matthews, 2011). One study has looked at the effects of having a two-way video-conversation, by showing a video of the caller to the driver and of the driver to the caller (Kun & Medenica, 2012). This lead to reduced visual attention for the road.

In terms of the modality in which context is presented, there are again several options. As a comparison, passengers have various cues about the driver's context at their disposal, such as visual cues from the road (e.g., traffic density, weather conditions), visual cues from the driver about their state (e.g., are they sweating, do they look stressed), auditory cues from surrounding traffic (e.g., car honks, sirens), and vestibular cues about, for example, the speed of the car. Each modality will put different demands on the remote caller. Multiple resource theories of cognition (Salvucci & Taatgen, 2008; 2011; Wickens, 2002; 2008) would predict that cues that overlap least with those used for the conversation task are beneficial, as the appropriate resource can be used at full to process that input signal. In the case of a passenger (or remote caller) who tries to hold a conversation while also staying aware of a driver's context, presenting context in a different modality than audio seems attractive, given the auditory nature of the conversation.

It is therefore perhaps of no surprise that previous research has mostly looked at sharing visual cues. Visual cues were either a shared view of the road (Charlton, 2009; Crundall et al., 2005; Drews et al., 2008; Gugerty et al., 2004; Maciej et al., 2011; Schneider & Kiesler, 2005), or of the driver's face (Maciej et al., 2011). One study has also looked at the role of vestibular cues, by blindfolding a passenger (Crundall et al., 2005).

Unfortunately, sharing information using these two modalities does create some practical limitations that might negatively influence their success when applied to phones and in-car systems. First, vestibular cues are challenging to share remotely, as they require the caller to experience the motion of the car. Video does not have this requirement, and can be shared remotely. However, video sharing has some practical disadvantages of its own: video screens are not available on all phones, video might not be used on all calls or be ignored (e.g., when the caller is on the move themselves), and video sharing requires a higher data bandwidth to broadcast than just audio. Also, a recent web survey found that people suspect that continuous video information sharing would invade privacy too much (Pfleging, Schneegass, & Schmidt, 2013).

In contrast to video, audio is available on all phones, requires less bandwidth than video, and might be harder to ignore. As less detailed information is shared compared to video, it is less invasive to privacy. Given these potential practical advantages, a theoretical understanding is needed about whether and how sharing auditory cues can help in generating a shared context between a driver and a remote caller and whether this influences the conversation and driving performance. The generated theory can then be used to inform the design of safer in-car technologies.

Sharing context through audio

Surprisingly few studies have looked into how audio signals can be used to convey driving context to a remote caller. Research has shown that providing an explicit warning of upcoming dangerous segments by sharing a discrete audio signal (i.e., beeps) with either the driver (Iqbal et al., 2011), or both driver and caller (Charlton, 2009) is beneficial for driving performance. However, while such just-in-time warnings are useful, without context it is also difficult for the caller to understand the potential dangers (Iqbal et al., 2011), and subsequently *help* the driver by cooperating in reducing load from the driver during an ongoing conversation.

In this paper, we investigate how effective audio cues are in reflecting context to a remote caller. We will look at two types of context cues: direct and indirect. The direct cues are sounds that come directly from a typical driving context, such as car honks, screeching tires, and sirens. Given the direct mapping to the driving context, these cues might make it relatively easy for the caller to understand that the other person is driving. Moreover, the straightforward semantics of the sounds might require less cognitive stages to interpret the sounds and thereby put relatively little additional cognitive workload on the caller (Wickens, 2002; 2008).

The indirect cues we provided were artificial heartbeat sounds, to suggest to the caller the stress level of the driver due to the challenges of driving. Within the human-computer interaction literature, sharing heartbeats has gained recent interest as a way to inform and connect people (Slovák, Janssen, & Fitzpatrick, 2012). Methods for sharing this data are being developed. For example, one system uses measured heartbeats to adjust light conditions to the heartbeat (Lotan & Croft, 2007), another system has used artificial heartbeat sounds to influence social interaction with 3D avatars (J. H. Janssen, Bailenson, IJsselsteijn, & Westerink, 2010).

A challenge in sharing heartbeat sounds is that interpreting the semantics of the heartbeat signal already requires some understanding of the context (Slovák et al., 2012). For example, is a raise in heartbeat due to stress, exercise, or excitement? As such, we suspected that heartbeat signals might be less effective (compared to the direct cues) in communicating driver busyness to the remote caller. We controlled the frequency of the heartbeat (see materials), to create “normal” sounds and “exaggerated” sounds (i.e., heartbeats with a relatively high frequency). We hypothesized that higher frequency heartbeats may make it more salient to the caller that the driver is busy. Saliency is believed to underlie the use of exaggerated signals in animal (Wiley, 2006) and human communication (Stewart & Kreuz, 2003).

Experiment 1: Sharing direct and indirect context cues

The goal of our research was to investigate the effect of sharing realistic continuous audio cues of the driver’s context with a remote caller. The specific research questions were:

RQ1. *Does sharing direct audio cues from the driving environment (e.g., car honks, sirens) and indirect audio cues about the driver's stress level (e.g., elevated heartbeats) increase the caller's subjective assessment of how busy the driver is?*

RQ2. *If sounds are shared, do callers adapt their part of the conversation to make the conversation less demanding for the driver? Do they reduce the conversation's density (Charlton, 2009; Crundall et al., 2005; Gugerty et al., 2004; Maciej et al., 2011; Schneider & Kiesler, 2005), as measured by the length of talk segments?*

RQ3. *In situations where sounds are shared do drivers drive better compared to when no sounds are shared? Previous work found that sharing visual information with a caller improved driving performance. Does audio sharing have similar benefits?*

To answer these questions, we conducted an empirical study using a medium fidelity driving simulator. The caller (situated remotely) talked to the driver over a headset. Driving sounds and heartbeat sounds could be shared with the caller (according to the experimental conditions) to provide context cues of the driver's situation.

Method

Participants

Twenty-four pairs took part, with one participant in each pair acting as the *driver* and the other as the *caller*. The drivers (19 male, 5 female, M_{age} = 36 year, SD_{age} = 9 year) were recruited from a list of people who had expressed interest in the driving simulator and who had a valid driver's license. All drivers were tested in a separate session for motion sickness in the simulator before taking part in the study.

We asked the drivers to bring an acquaintance to be the caller (10 male, 14 female, M_{age} = 35 year, SD_{age} = 11 year). In one instance we arranged a caller. All drivers and callers were native English speakers. Of the 23 pairs who knew each other, 7 were in a relationship, 2 were family, 3 were friends, 4 were colleagues, and 7 were colleagues and friends. Each participant received a gratuity or $50 gift card. To encourage appropriate attention, an additional gratuity was granted to the best performing pair of participants (see also the section on Incentives and priorities).

Tasks & Materials

In the dual-task set-up a driver performed a car following task while also holding a conversation with a remote caller. Figure 1 shows the set-up. A video clip that demonstrates this set-up (without sound sharing) is included in the supplementary materials .

Driving task. The driving task was to follow a lead car and stay at a fixed distance from that car. Car following tasks have been used effectively in prior research (e.g., Alm & Nilsson, 1995; Brookhuis et al., 1991; 1994) and have two important properties for our research questions. First, the task allowed a continuous measure of driving performance (e.g., the distance from the lead car). Alternative settings, such as unexpected events and making turns, are discrete and may not be affected by dual tasking at the moment of occurrence. Second, the task required continuous attention for and reaction to the lead car's changing speed. Therefore, it was more likely to show effects of interference from calling.

We used the STISIM™ driving simulator (Systems Technology Inc., Hawthorne, CA) (Figure 1, top-left). This simulator brand has been used in various research studies and labs (e.g., Atchley & Chan, 2011; Iqbal et al., 2010; Mehler, Reimer, Coughlin, & Dusek, 2009). Our set-up can be considered medium fidelity. The realistic aspects are that it provides realistic interaction with the simulated car through a steering wheel and projects the simulation on large immersive screens. The less realistic aspects are that the driving task is constrained (car following) and without any real risk to the driver. Similar set-ups, using other software and hardware, have been frequently used in driving studies, including most of the studies that looked at context sharing (Charlton, 2009; Drews et al., 2008; Iqbal et al., 2011; Maciej et al., 2011; Schneider & Kiesler, 2005). Meta-analyses suggest that findings with simulators in general translate well to on-the-road settings (Caird et al., 2008; Horrey & Wickens, 2006).

In our simulator, participants sat on an actual car seat and controlled the car using a steering wheel, brakes and gas pedals[1]. Speed could be read on the accelerometer, which was part of a realistic dashboard. The view through the two side windows and the windscreen was projected on three 47-inch screens.

For the simulation (for close-up see Figure 1, bottom-left), the road was a six-lane highway, with three lanes on each side. The background was formed by a mountain landscape. The highway had trees and grass on each side of the highway with posted speed signs every 1000 feet.

The driver drove on the left-most lane of their (right) side of the road. The objective was to follow a lead car (in front of them in the same lane) and maintain a fixed distance of 100 ft from that car. As estimating distances in a simulator can be challenging (e.g., Kearney et al., 2007), a second car drove at the same speed as the driver, at a fixed distance of 100 ft in the adjacent lane (red car in Figure 1). The driver was instructed to keep the back of this car aligned with the back of the lead car.

At 700 ft into the simulation, the lead car appeared 100 ft in front of the driver. This car drove at a constant speed of 40 MpH for 15 seconds, after which it started accelerating and decelerating for 105 seconds. The pattern of the speed changes followed a sinusoid shape with a mean value of 40 MpH. The period (i.e., time for one complete acceleration and deceleration cycle) and amplitude (i.e., influencing maximum and minimum speed) differed between road types.

We used two road types. The *easy* driving condition had a straight road, on which the lead car accelerated relatively infrequently (period of 30 sec) and less steeply (amplitude of 8 ft/sec, resulting in speeds between 34 and 45 MpH). The *difficult* driving condition had alternate left and right curves[2] every 1000 ft and a more challenging following task: the lead car accelerated more frequently (period of 20 sec) and more steeply (amplitude of 10 ft/sec, resulting in speeds between 32.7 and 46.3 MpH).

In line with our definition of difficulty, driving performance on the difficult road was worse compared to the easy road when analyzing single- and dual-task data on three driving metrics (see results).

On all roads there was traffic in the incoming (opposite) lanes. The driver was instructed that this traffic would remain in their lanes, that these cars would make noises, and that these noises were independent of how well the driver was driving.

Conversation task. The objective of the conversation task was to collaboratively generate fictional stories. This task was inspired by research that used engaging game-like tasks (e.g., Crundall et al., 2005; Kun, Shyrokov, & Heeman, 2012) and story-telling tasks (Becic et al., 2010) to study conversations while driving. The game-like aspect in our context was mostly that the caller and driver took turns in the conversation and that the task was engaging compared to other tasks such as question answering.

As in other studies that investigated context sharing, our task allowed free form speech, just like normal conversations (Charlton, 2009; Crundall et al., 2005; Drews et al., 2008; Maciej et al., 2011). However, as we initiated the stories that people talked about, the conversation was more controlled than random conversation, while maintaining ecological validity. More specifically, the task had five important properties that map to normal conversations: (a) it required turn taking with alternations between talking and listening, (b) response length was flexible and could be strategically adapted to the driving demands, (c) the task was engaging and could distract the driver at times, (d) the task had a clear objective without being restrictive in the required answers, and (e) no visual-manual interaction was needed, so distraction came only from the conversation. In addition, a conversation revolving around story-telling would be less likely to provide direct or subtle hints about the driving condition. Since we were interested in understanding how the audio cues helped the caller perceive the driving context, we did not want explicit dialogue on the driving condition to confound the findings.

Our story task is like the one used by Becic and colleagues (2010) in that participants told stories in freeform speech during short intervals. However, in the study by Becic and colleagues, participants retold prerecorded stories to their conversational partners without turn taking and without creating their own content (apart from sentence generation).

Participants spoke with each other through headsets. Their conversation was broadcasted and recorded using a conference call program. On each trial, the caller would read out a sentence that we provided to start the story (occasionally adding one or two sentences of their own). From then on, the driver and caller took turns in adding to the story until the end of the trial (signaled by the driver). Participants were asked not to interrupt each other, but to take turns without providing explicit cues of the end of a turn. In effect, this added another interesting property of normal conversations to the task, namely delay management. An additional benefit was that it made it easier to analyze the conversation (and duration of talk segments) in detail.

Figure 1 (top-right) shows the remote caller and a screenshot of the program that cued sentences and words (bottom-right). Figure 2 shows an example of the start of a conversation. To help the flow of the story, callers had an interface that provided them with seven words that could be used in the story (e.g., in the example of Figure 2 these were: *doctor, pants, point, tooth, book, clock,* and *zoo*; all sentences and their associated words are in the supplementary materials)[3]. The interface could also be used to generate additional random words and names.

Participants were instructed that they would not be judged on the content of their stories but should focus on keeping the story going. Instructions, examples, and hints were given during training on how to achieve this and summarized on a separate sheet for the remote caller[4].

Participants were recommended to both contribute to the story and try not to say more than five sentences at a time. Although such constraints are not present in real conversations, we wanted to minimize situations where people might completely omit a conversation or put the burden on only one of the conversants. In effect, this would make it challenging to compare dual-task performance between participants, as the amount of dual-tasking might differ significantly between pairs. Our parameters were such that observed performance was within the same range as regular unconstrained conversations. For example, an analysis of various English corpuses revealed average turn lengths between 0.4 and 10.7 seconds (Yuan, Liberman, & Cieri, 2006). Interpolation of the data in another driving study (Maciej et al., 2011) suggests silences of maximum 6 seconds - well within our 15 seconds range.

Participants were requested to not discuss the driving task. This allowed us to investigate the direct effect of sounds on a caller's perception of a driver's busyness. Sharing explicit information would confound this manipulation and previous work has already shown that discussing the driving context might improve performance (Charlton, 2009; Drews et al., 2008).

Incentives and priorities. Each participant of the pair received their own priority to focus on during the task, as priorities can strongly affect the trade-offs that people make between the time and effort that is dedicated to tasks in multitasking settings (Brumby, Salvucci, & Howes, 2009; C. P. Janssen & Brumby, 2010; C. P. Janssen, Brumby, & Garnett, 2012). The driver had to prioritize good performance on the car following task. The caller's priority was to help the driver by ensuring that the conversation never fell silent for more than 15 seconds (i.e., by taking over if the driver appeared nonresponsive). These priorities reflect a situation in which a caller wants to receive information from a driver, while being aware that talking can distract them. To encourage adherence to the instructions, an extra gratuity was given to the pair that had the best driving performance and no silences of over 15 sec..

Sound sharing. Two types of sounds were shared with the caller using custom software: Direct cues in the form of driving sounds and indirect cues in the form of heartbeat sounds. The experimenter controlled who heard which sounds.

Direct cues: Driving sounds. As direct cues of the driver's context we shared realistic driving sounds. In practice, the difficulty of the car following task was challenging to capture in sound. Therefore, we used proxy sounds to indicate the difficulty levels. The sounds were mapped to the cars on the other side of the road (e.g., car honks, sirens), which did not impact the driver's task in any way.

The sounds lasted at most 2 seconds. Driving sounds were played at two different intervals, based on the difficulty of the car following task. A new sound was started every 9 seconds for the easy driving condition. In the difficult driving condition this was every 3 seconds. The driver always heard the traffic sounds associated with their road type. However, whether the caller heard the sounds was determined by the experimental manipulations. As the caller was not aware of the exact task of the driver, they could interpret the sounds as being part of the driving task (i.e., suggesting busyness).

Indirect cues: Heartbeat sounds. As indirect cues we shared heartbeat-like sounds, to suggest to the caller the stress level of the driver due to the challenges of driving. The driver did not hear the heartbeat sound, however. We did not share the driver's actual heartbeat with the caller, but instead used various heartbeat frequencies corresponding to the difficulty of driving to explore the effects of heartbeat rate on the caller's perception of the driver's busyness. For each road there were two "normal" sounds and two "exaggerated" sounds (i.e., heartbeats with a relatively high frequency). For the easy road the normal beat was 60 beats-per-minute (bpm) and the exaggerated beat was 90 bpm. For the difficult road the beats were 105 and 135 bpm. Heartbeat sounds were played continuously throughout a trial.

Physiological measurements. We collected heart-rate data using the CMS50E Fingertip Pulse Oximeter (Souteastern Medical Supply, Columbias, SC) and Electro Dermal Activity using the Affectiva Q-curve sensor (Affectiva Inc., Waltham, MA). While previous research has shown that such physiological measurements can provide an indicator of workload level in dual-task driving settings (e.g., Healey & Picard, 2005; Mehler et al., 2009; Mehler, Reimer, & Coughlin, 2012a), our preliminary exploration found no significant differences between single-task and dual-task settings or between conditions within dual-task. For brevity, we omitted the analysis of this data. The lack of effects might be for several reasons. For example, the differences in mental workload might not be noticeable in physiological measures for a setting that is in between lab setting and on-the-road study. Another reason might be that other emotions (e.g., laughing) overshadowed the effect that workload had on physiological response. We leave detailed analysis of this data for future work.

Design

The study was designed as a 2 (Driving sounds shared: shared, not shared) X 3 (Heartbeat sounds shared: normal heartbeat, exaggerated heartbeats, not shared) X 2 (Road type: easy, difficult) within-subjects study.

Measures

For RQ1 we asked participants to rate their perception of the driver's busyness on a scale of 1 to 5 on every trial. The callers were aware that this question referred to the driver's ability to balance the driving task and the conversation task.

For RQ 2 these were turn duration and talk segment duration (see Figure 2 and RQ 2 for definition). A turn was defined as the time between the first onset of an utterance by a speaker (after the previous speaker had finished) and the offset of the last utterance of that speaker (before the next speaker started). Audacity™ (audacity.sourceforge.net) was used to automatically detect silence segments and talk segments within the turns (similar to Maciej et al., 2011). Silences were detected when sound levels were less than 25 dB for at least 0.2 seconds, automatically implying other sections of a turn as talk segments. Talk segments of 100 msec or less were also disregarded. In the majority of cases, these were events that could not be considered part of the conversations (e.g., respirations, sighs).

For the identified talk segments, the experimenters labeled which person (driver or caller) was talking. Unknown cases were labeled as such. When both speakers talked, the main contributor was identified. Only conversation turns that happened when the driver was dual-tasking were analyzed. Due to recording errors, data of at least one trial was missing for the audio of two pairs. We omitted their audio data from the analysis.

For RQ 3 we assessed four measures of success in following the lead car. Three measures were taken from Brookhuis et al. (1994): coherence score, delay, and modulus. Delay reflects how quickly the driver reacted to changes of speed by the lead car (in seconds). Modulus (also referred to as "gain" in the literature) describes whether the driver responded to these changes by braking and accelerating at the same pace as the car (score of 1), or more intense (score higher than 1). These two metrics are best not to be viewed in isolation, as degraded performance on one metric can result in degraded performance on the other metric. For example, a delay in responding to the lead car (i.e., a higher delay score) requires the driver to accelerate to catch up with the lead car (i.e., resulting in a higher modulus score). The third measure, coherence score (Allen & Jex, 1972; Brookhuis et al., 1994) can be considered as capturing both aspects in a single score. It is calculated as the squared correlation between the lead vehicle's pattern of speed changes and the pattern produced by the driver. The maximum score was 1, which would be achieved only when the driver drove perfectly in sync with the lead car at all times. Lower scores reflected worse performance (a delay in reaction time, less accurate speed, or both), a reason for which could be more distraction.

The fourth measure for driving performance was the median absolute distance from the lead car. This measure expresses how well the driver adhered to their objective: to stay at a fixed distance of 100 feet with the lead car. We investigated the median *absolute* deviation from the ideal, as drivers could drive either too fast or too slow while distracted, which could average out to a mean non-absolute deviation of zero. In our measure, higher values reflect worse performance: more discrepancy from the ideal distance. This measure is specific to our setting, where a fixed distance was set as ideal performance. In other driving settings, keeping a larger distance can be safer, as it gives more headway time to respond to sudden events. Such strategic adaptation has been observed before (e.g., Cnossen, Meijman, & Rothengatter, 2004; Iqbal et al., 2010).

Procedure

Upon arrival the driver was put on a heart rate monitor and a device to measure skin conductance, with the caller present. For some, but not all, callers this added to the illusion that shared heartbeat sounds were real, as verbally reported afterwards. The driver then practiced four driving trials (two per driving condition) without the caller present. Participants could use the distance between the "projected" car and the lead car in front of them to estimate distance. The experimenter also provided the driver with verbal feedback on their performance.

Participants then practiced two to three conversations (approximately 1-2 minutes each), followed by two dual-task practice trials remotely. Participants again received feedback on their performance from the experimenters, for example by suggesting ways to keep a conversation going when it got stuck. After these practice trials no feedback was given on performance.

The practice was followed by two single-task driving trials (one per driving condition, order counter-balanced and randomly assigned), followed by two single-task conversation trials (approximately 2.5 minutes each).

This was followed by three experimental dual-task blocks, with four trials each. The blocks varied in the type of heartbeat sound that was shared: none,

normal, or exaggerated. The order of these conditions was randomly assigned and counter-balanced. Within each dual-task block were four trials to accommodate all combinations of the two other factors (driving sounds and road type). On each block, participants experienced the conditions in different orders. Orders were counterbalanced so that across participants, each order occurred equally often for each block position (1st, 2nd, 3rd block) and for each heartbeat condition level.
In between blocks were breaks of 2-5 minutes. After each trial, participants rated their experience on a five-point scale. After the experiment, they filled out a general questionnaire (see: supplementary materials). The experiment lasted about 2 hours.

Results

We present results by research question. For RQ 2 and 3 we compared performance between single- and dual-task using a One-way ANOVA (single- or dual-task) for conversation metrics and a 2 x 2 ANOVA (road type x single- or dual-task) for driving metrics.

When analyzing metrics for the dual-task conditions only (RQ 1, 2, and 3), we used a 2 x 2 x 3 ANOVA (road type x driving sounds sharing x heartbeat sharing), unless noted otherwise. A significance level of p = .05 is used. For each metric we report means and 95% within-subjects confidence intervals, following Morey (2008).

RQ1: Caller's perception of driver's busyness

We first wanted to establish that sharing sounds indeed changed the caller's perception of busyness, as it is a useful measurement in this experiment to confirm that the sound signals indeed had the intended effect of changing the caller's perception of the driver's busyness.

Sharing sounds was effective for callers in assessing the driver's busyness. The caller's perception of driver busyness was significantly influenced by driving sounds sharing, heartbeat sounds sharing, and road-type. The bar plot on the left of Figure 3 illustrates that the caller thought the driver was more busy when driving sounds were shared (M = 3.74, 95% CI = [3.61, 3.86]) compared to when no driving sounds were shared (M = 3.06, 95% CI = [2.94, 3.18]), $F(1, 23) = 48.75$, $p < .001$, $\eta_p^2 = .68$.

There was also a main effect of heartbeat sharing, $F(2, 46) = 3.29$, $p = .046$, $\eta_p^2 = .13$. A holm-corrected post-hoc test found no significant differences between the three heartbeat conditions (all $p > .1$). The trend was that the rating of busyness was lowest when no heartbeat was shared (M = 3.20, 95% CI = [3.00, 3.39]), compared to when a normal heartbeat was shared (M = 3.52, 95% CI = [3.35, 3.69]) and when an exaggerated heartbeat was shared (M = 3.48, 95% CI = [3.28, 3.68]).

There was also a main effect of road type. The driver was rated only slightly more busy when driving on difficult roads (M = 3.53, 95% CI = [3.40, 3.67]) compared to easy roads (M = 3.26, 95% CI = [3.13, 3.40]), $F(1, 23) = 6.46$, $p = .018$, $\eta_p^2 = .22$.

There was an interaction effect between heartbeat sharing and road type, $F(2, 46) = 4.205$, $p = .021$, $\eta_p^2 = .15$. The right side plot in Figure 3 illustrates this interaction. For the easy road (open points), there is hardly any variation between the condition without heartbeat sounds and the two conditions with heartbeat sound sharing. In contrast, for the difficult road, the ratings of busyness were higher when heartbeat sounds were shared. The highest rating was given in the difficult road with exaggerated heartbeat sounds ($M = 3.75$, 95% CI = [3.53, 3.97]), where the mean score was about 0.5 point higher compared to the conditions without heartbeat sound sharing. There were no other significant effects.

Taken together, the trend in the data is that callers rated the drivers as more busy when they were cued about busyness through sounds. Sharing driving sounds increased the rating of the driver's busyness. Sharing heartbeat sounds had a main effect, but through an interaction effect with road-type, had the strongest effect when shared on difficult roads. In the easy road condition, where the normal and exaggerated heartbeats were relatively low, sharing heartbeats did not change the perception of busyness significantly. On the difficult roads, where the normal and exaggerated heartbeats were relatively fast, the perception of busyness increased. When no sounds were shared, the busyness of the driver was rated as 2.59 for the difficult road. In contrast, when both driving sounds and an exaggerated heartbeat were shared, this rating went up by 1.5 points to 4.08. This suggests that when sounds are shared, callers get a more accurate perception of how busy the driver is, compared to when no sounds are shared.

RQ2: Adaptation of the conversation

We next consider adaptation of the conversation, as indicated by the median turn length of the conversation and the length of talk segments within a turn. To get a representative sample of turn length, we analyzed the median turn length per trial, as this is less influenced by occasional very short or very long turns than the mean turn length. The mean values reported below are the means of those medians. As conversational turns go back and forth between the caller and the driver, we included the speaker as factor in the ANOVAs.

Conversational turn length. When comparing performance between single- and dual-task, there was a main effect of speaker, which suggested that the caller's turn was longer than the driver's turn, $F(1,21) = 10.58$, $p = .004$, $\eta_p^2 = .33$. There was also a significant interaction between speaker and single- versus dual-task, $F(1, 21) = 6.85$, $p = .016$, $\eta_p^2 = .25$. In dual-task, the caller's turn length slightly increased ($M = 12.23$, 95% CI = 11.14, 13.32]) compared to single-task ($M = 11.72$, 95% CI = [10.04, 13.41]). In contrast, the driver's turn length slightly reduced in dual-task ($M = 8.44$, 95% CI = [7.02, 9.87]) compared to single-task trials($M = 9.51$, 95% CI = [8.29, 10.73]). Figure 4 visualizes these data. There was no effect of road type, $p > 1$.

For analysis of dual-task performance, we included four factors: Speaker (caller/driver), road type (easy/difficult), driving sound sharing (none/yes), heartbeat sound sharing (none, normal, exaggerated). There was again a main effect of speaker. The caller's turn durations were longer ($M = 12.23$ sec, 95% CI = [10.98 , 13.49]) compared to the driver's turn durations (M = 8.44 sec, 95% CI = [7.19, 9.70]), $F(1, 21) = 19.66$, $p = .0002$, $\eta_p^2 = .48$. There were no other effects, $p > .1$.

Median talk segment length for caller. Within the turns of each of the speakers, we analyzed the median duration of talk segment. We first report an analysis of the caller's talk segments. When comparing single- with dual-task performance, there was no significant difference between talk segment length in single-task (M = 1.77, 95% CI = [1.65, 1.89]) and dual-task (M = 1.62, 95% CI = [1.50, 1.73]), $F(1, 21) = 2.86$, $p = .106$.

When observing performance within dual-task trials, there were differences. For the caller, there was a main effect of heartbeat sound sharing, $F(2, 42) = 5.016$, $p = .011$, , $\eta_p^2 = .19$. Figure 5 (left) shows the caller's talk segment lengths for the three levels of heartbeat sharing. A bonferroni-corrected pairwise t-test did not reveal any differences between conditions ($p > .1$). The trend in the data was that talk segments were longest when no heartbeat sounds were shared (M = 1.74 sec, 95% CI = [1.63, 1.86]). When heartbeat sound were shared: normal heartbeats (M = 1.52 sec, 95% CI =[1.42,1.62) and exaggerated heartbeats (M = 1.59, 95% CI = [1.50, 1.67]) resulted in similar lengths of talk segments.

There was a marginal effect of road type, $F(1, 21) = 3.48$, $p = .076$, $\eta_p^2 = .14$. This effect was affected by a significant interaction effect between road type and driving sounds sharing, $F(1, 21) = 5.86$, $p = .025$, $\eta_p^2 = .22$. Figure 5 (right) illustrates this interaction. Talk segments were shortest on the difficult road when sounds were shared (M = 1.50, 95% CI = [1.40, 1.61]) and longest on the easy road when sounds were shared (M = 1.73, 95% CI = [1.62,1.84]). When no driving sounds were shared, talk segments were roughly equal length in the difficult road (M = 1.61, 95% CI = [1.52, 1.70]) and the easy road (M = 1.62, 95% CI = [1.54, 1.70]). There were no other significant effects, $p > .1$.

Taken together, these results suggest that the caller's talk segments were shorter when sounds were shared. That is, they were shorter when heartbeat sounds or driving sounds were shared, with one exception: in the easy road condition, talk segments were longer in the condition with driving sounds shared (see Figure 5).

Median talk segment length for driver. For the driver, there was no significant difference between the length of the talk segments in single-task (M = 1.72, 95% CI = [1.61, 1.84]) and dual-task trials (M = 1.78, 95% CI = [1.66, 1.90]), $F(1, 21) = 0.394$, $p > .1$.

For the driver, within the dual-task trials there was only a significant interaction effect between heartbeat sound sharing and driving sound sharing $F(2, 42) = 3.982$, $p = .0261$, $\eta_p^2 = .16$. The data did not provide an easy to interpret pattern. The means and 95% Confidence Intervals were as follows when no driving sounds were shared: without heartbeat sharing (M = 1.79, 95% CI = [1.52, 2.07]), with normal heartbeat (M = 2.01, 95% CI = [1.84, 2.18]), with exaggerated heartbeat (M = 1.64, 95% CI = [1.47, 1.80]). When driving sounds were shared the values were as follows: without heartbeat sharing (M = 1.75, 95% CI = [1.62, 1.88]), with normal heartbeat (*M* = 1.68, 95% CI = [1.54, 1.82]), with exaggerated heartbeat (*M* = 1.82, 95% CI = [1.66, 1.98]). There were no other significant effects, $p > .1$

RQ3: Effects on driving performance

The above analyses suggest that the caller's conversation style changed when sounds were shared. Did this impact the driver's performance? Recall that the driver was not told that sounds of their context were shared with the caller – drivers always heard the traffic sounds, but never heard the heartbeat sounds. Any effect on the driver's performance is therefore expected to be a result of task demands and a reaction to the adaptation on the caller's side, not a reaction to the sounds directly.

Driving performance: Coherence score. When comparing performance between single- and dual-task and roadtype, coherence scores were higher in single- (*M* = 0.864, 95% CI = [0.850, 0.878]) compared to dual-task trials (*M* = 0.834, 95% CI = [0.820, 0.848]), $F(1, 23) = 9.573$, $p = .005$, $\eta_p^2 = .29$. There was no significant effect of road type and no interaction effect, $p > .1$.

Within the dual-task trials, drivers had a slightly better coherence score when driving sounds were shared with the caller (*M* = 0.840, 95% CI =[0.834,0.846]) compared to when these sounds were not shared (*M* = 0.829, 95% CI = [0.822, 0.835]), $F(1,23) = 6.07$, $p = .022$, $\eta_p^2 = 0.21$. There was a trend for the coherence score to be slightly better in the difficult driving condition (*M* = 0.845, 95% CI = [0.832, 0.858]) compared to the easy driving condition (*M* = 0.824, 95% CI = [0.811, 0.838]), $F(1, 23) = 3.94$, $p = .059$, $\eta_p^2 = .15$. There were no other effects ($p > .1$).

Driving Performance: Delay in reacting to lead car. When comparing performance between single- and dual-task trials, delays in responding to the lead car were longer in dual-task compared to single-task trials, $F(1, 23) = 7.195$, $p = .013$, $\eta_p^2 = .24$. Delays were also longer on the difficult road compared to the easy road, $F(1, 23) = 34.16$, $p < .001$, $\eta_p^2 = .60$. These effects were influenced by an interaction effect between road type and single-/dual-task, $F(1, 23) = 3.118$, $p = .091$, $\eta_p^2 = .12$. The contrast was particularly large between the dual-task difficult road (M = 1.32 sec, 95% CI = [1.18, 1.45]) and the single-task easy road (M = 0.67 sec, 95% CI = [0.49, 0.85]). The other values were in between these extremes: dual-task easy road (M = 0.78 sec, 95% CI = [0.67, 0.89]) and single-task difficult road (M = 1.01 sec, 95% CI = [0.89, 1.12]).

Within the dual-task trials, delays were longer on the difficult roads compared to the easy roads, $F(1, 23) = 51.08$, $p < .001$, $\eta_p^2 = .69$ (Mean values are reported above). There were no other significant effects, $p > .1$.

Driving Performance: Modulus. Modulus expresses how extreme the driver adapts to speed changes of the car in front of them. Values higher than 1 mean they break and accelerate more abruptly. A value of 1 means they accelerate and decelerate with the same amplitude as the lead car.

When comparing performance between single- and dual-task trials, the modulus was found to be higher on difficult roads (M = 1.24, 95% CI = [1.21, 1.27]) compared to the easy roads (M = 1.17, 95% CI = [1.14, 1.20]), $F(1, 23) = 9.901$, p =

.005, $\eta_p^2 = .30$. There was no effect of single- and dual-task and no interaction effects, $p > .1$.

Within the dual-task trials, there was again only an effect of road type, $F(1, 23) = 36.19$, $p < .001$, $\eta_p^2 = .61$. Modulus was larger on difficult roads ($M = 1.24$, 95% CI = [1.23, 1.26]) compared to the easy roads ($M = 1.17$, 95% CI = [1.16, 1.19]).

Driving Performance: Fixed distance with the lead car. The objective of our task was to keep the car at a fixed distance of 100 ft. We measured how much participants deviated from this ideal measure using absolute distance.

When comparing single-task with dual-task performance, the distance from the ideal was larger in dual-task ($M = 10.44$ ft, 95% CI = [9.76,11.12]) compared to single-task trials ($M = 8.86$ ft, 95% CI = [8.17,9.54]), $F(1, 23) = 11.58$, $p = .002$, $\eta_p^2 = .33$. The distance was also larger on difficult roads ($M = 11.50$ ft, 95% CI = 11.12, 11.88]) compared to easy roads ($M = 7.80$ ft, 95% CI = [7.42,8.17]), $F(1, 23) = 154.6$, $p < .001$, $\eta_p^2 =.87$. There was no significant interaction effect, $p > .1$.

Within the dual-task trials, it was found that drivers deviated more when driving on the difficult road ($M = 12.46$ ft, 95% CI = [11.91, 13.02]) compared to the easy road ($M = 8.42$ ft, 95% CI = [7.86, 8.97]), $F(1,23) = 84.61$, $p < .001$, $\eta_p^2 = .79$. There were no other effects, $p > .10$.

Taken together, the driving metrics suggest that the driving task was more challenging on dual-task trials compared to single-task trials and that the difficult road was indeed more difficult than the easy road. For the coherence metric, we found that there were slightly better scores when driving sounds were shared compared to when no driving sounds were shared. Coherence score captures both the effect of delays in responding to speed changes by the lead car and the modulus with which one reacts to these changes (i.e., how hard one steps on the gas). However, the individual measures of delay and modulus did not show significant differences.

Experiment 2: Sharing irrelevant sounds

In experiment 1 we found that when sounds were shared with the caller, the callers rated the driver as being more busy. Data also suggested that the caller's talk segments were shorter when sounds were shared. An open question is why the talk segments were shorter. Did callers reduce talk segment length because of their assessment of the driver's busyness (and the content of the sounds), or was this an artificial effect because the shared sounds interfered with the conversation? If the change was a response to the content of the shared sounds (i.e., to an assessment of busyness), we should only find such adaptation when content is shared that suggests busyness (e.g., driving sounds), but we should not find such adaptation when other, irrelevant sounds (e.g., nature sounds) are shared at the same pace. In contrast, if the change was due to interference of the sounds, we should find such adaptation regardless of what type of sound is being shared.

We ran a follow-up experiment with 15 participant pairs to test this. The experiment is reported in full in the supplementary materials. The drivers only drove on the difficult road and callers experienced one of three sound sharing conditions: driving sounds (similar to the sounds in the difficult condition in experiment 1), nature sounds (e.g., bird chirps, wolves howling, thunderstorm), or no sounds. The nature sounds were played at the same pace as the driving sounds (i.e., a new sound started every 3 seconds). Nature and driving sounds were played at the same intensity.

Results

In the analyses we will focus on measurements in which we found significant effects in Experiment 1. We will only look at dual-task trials. A comprehensive set of results is reported in the supplementary materials.

RQ1: Caller's perception of driver's busyness

Figure 6 plots the effect of sound sharing on the rating of busyness. As in experiment 1, there was a significant effect of sound type on the rating of busyness, $F(2, 28) = 20.21$, $p < .001$, $\eta_p^2 = .59$. A holm-corrected post-hoc test found that callers rated the driver to be more busy in the condition with driving sound sharing (M = 3.47, 95% CI = [3.18, 3.75]) than in the condition with nature sound sharing (M = 2.55, 95% CI = [2.31, 2.79]), $p = .043$. There was also a significant difference between the driving sound condition and the condition without sounds (M = 2.50, 95% CI = 2.25, 2.75]), $p = .043$. There was no significant difference between the nature sound condition and the condition without sound, $p = .896$. That is, the caller increased the busyness rating on average by 1 point when driving sounds were shared compared to the other two conditions. Sharing nature sounds did not influence the caller's perception of the driver's busyness.

We also asked participants to rate how interfering the nature and driving sounds were on the conversation for each trial. A one-way ANOVA found that driving sounds (M = 3.68, 95% CI = [3.46, 3.90]) were rated as more interfering by half a point on a five point scale than the nature sounds (M = 3.10, 95% CI = [2.88, 3.32]), $F(1, 14) = 16.03$, $p = .0013$, $\eta_p^2 = 0.53$.

These results suggest that there was indeed modest interference from sharing the driving sounds with the caller. However, the sounds were useful: they altered the caller's perception of the driver's busyness - replicating our finding in experiment 1.

RQ2: Adaptation of the conversation

Conversational turn length. Within dual-task trials, the caller's conversational turn (M = 12.54 sec, 95% CI = [10.74, 14.33]) was again longer than the driver's turn (M = 8.79 sec, 95% CI = [7.11, 10.71]), $F(1, 14) = 5.851$, $p = .0298$, $\eta_p^2 = 0.29$. There were no other significant effects, $p > .1$.

Median talk segment length for caller. Within the dual-task trials, no significant differences were found between the three sound sharing conditions, $F(2, 28) = 0.07$, $p > .1$. The mean values and 95% Confidence Intervals were as follows: no sounds (M = 1.90 sec, 95% CI = [1.68, 2.12]), nature sounds (M = 1.87, 95% CI = [1.68, 2.06]), and traffic sounds (M = 1.85, 95% CI = [1.70, 2.01]).

Median talk segment length for driver. Within the dual-task trials, no significant differences were found between the three sound sharing conditions, $F(2, 28) = 0.028$, $p > .1$. The means and 95% Confidence Intervals were as follows: no sounds (M = 1.75 sec, 95% CI = [1.63, 1.86]), nature sounds (M = 1.75 sec, 95% CI = [1.56, 1.95]), and traffic sounds (M = 1.73 sec, 95% CI = [1.59, 1.87]).

Taken together, in this experiment we found no evidence to support the hypothesis that the caller adapted their conversation style based on the sharing of sounds. The data also did not show that the conversation adaptation that was observed in Experiment 1 was due to sound interference. If that was the case, we should have observed a significant difference here between the condition without sounds and the two conditions with sounds (nature and traffic).

RQ3: Effects on driving performance

Driving performance: Coherence score. In experiment 1, only the coherence metric had a significant effect of sound sharing. In experiment 2, we found no significant effect of sound sharing, $p > .1$. The mean values and 95% Confidence Intervals were as follows: no sounds (M = 0.847, 95% CI = [0.841, 0.853]), nature sounds (M = 0.844, 95% CI = [0.836, 0.851]), and traffic sounds (M = 0.853, 95% CI = [0.847, 0.859]).

On the other driving metrics, we also found no significant effect of sound sharing, similar to the findings in experiment 1. Taken together, this suggests that the influence of sound sharing on driving performance is not present, or moderate at best.

Discussion

Sound sharing increases awareness of the driver's state

We found that sharing audio cues from a driver's context with a remote caller strongly influenced a remote caller's perception of the driver's busyness (RQ1). The remote callers rated the driver as being more busy when traffic sounds were shared and when high frequency heartbeats were shared (i.e., as in the difficult road condition). For example, for the difficult road condition in experiment 1, the mean rating of busyness increased by 1.5 points on a five-point scale when both traffic sounds and an exaggerated heartbeat sound were shared, compared to the situation without sound sharing.

In experiment 1 we found a significant effect of sound sharing on the length of the caller's talk segments (RQ2), in which the trend in the data was that callers had shorter talk segments when sounds were shared. However, this effect did not replicate in experiment 2. This suggests that this effect is a modest effect at best.

In both experiments we found for most driving metrics (RQ3) a significant difference between single-task and dual-task performance, with performance being worse in dual-task trials. When comparing performance within the dual-task trials, we found a significant effect of sound sharing on one driving metric (coherence score) in experiment 1. The trend in the data was that driving performance on this metric was better when driving sounds were shared with the remote caller. However, this effect did not replicate in experiment 2. In both experiments, no other metrics of driving performance showed an effect of sound sharing. This suggests that in this driving task there might not be a beneficial effect (or at best a modest effect) of sound sharing on driving performance. However, further testing is needed.

Alternative ways of conversation adaptation

In both experiments we found that sharing sounds changed the caller's perception of the driver's busyness. The results suggest that the effect on the caller's talk segment length was at most modest. Did participants perhaps apply other conversational strategies to support the driver? We asked the callers in a post-study questionnaire whether they used any particular strategies to maintain the conversation and to make it easier for the driver to take part in the conversation while driving safely. A variety of answers was given. Across the two studies, feedback that was given by more than one participant was as follows: 12 callers tried to talk about a familiar topic or context, 9 callers tried to end their conversational turn with something that was easy to respond to (e.g., an open or easy question), 8 callers talked longer when they thought the driver was busy (i.e., increasing turn length), and 5 callers tried to keep the story simple. Only 5 callers (4 in experiment 1, 1 in experiment 2) indicated that they did not use any strategy.

It is encouraging that callers did their best to respond to the driver's busyness. However, these attempts did not seem to help driving performance drastically.

An important applied implication of this work is therefore that although callers can be made more aware of the driver's busyness through sound sharing, they need to be trained or be informed on how to act on the sounds. Further testing is needed of what alternative strategies (e.g., as suggested by the callers) might be most effective in reducing the distracting effects of the conversation. In a setting outside of the lab, perhaps the most effective way might be to decide to postpone the conversation to a later time.

Sharing sounds: A complementary approach

Sharing sounds can complement other approaches that are aimed at reducing the deleterious effects of engaging in a cell phone conversation while driving. For instance, systems that share context before the call (de Guzman et al., 2007; Grandhi et al., 2011; Lindqvist & Hong, 2011), systems that provide video information during the call (Charlton, 2009; Maciej et al., 2011; Schneider & Kiesler, 2005), and system that provide explicit auditory warnings for upcoming dangerous roads (Charlton, 2009; Iqbal et al., 2011) could be augmented by also including auditory context sharing. As explained in the introduction, audio sharing has multiple practical advantages. For example, it is harder to ignore by callers compared to video.

A practical implication is that our findings open up the question whether current trends for filtering out background noise in cars and (in-car) phones are wise. In particular, in-car noise reduction is receiving increased focus in the automobile industry (Bennett, 2014). In-car noise reduction benefits the driver, as it gives them a quieter ride. Noise reduction in phone conversations is also essential for the primary purpose of communication: to convey a clear, audible speech signal. However, leaving in, or adding, some of the noise from the car to an in-car phone call might help the caller understand that they're distracting someone from driving. This might encourage callers to postpone the call to a later, safer, time.

Theoretical challenges

A theoretical challenge when sharing context through audio is that resource conflicts might occur (Wickens, 2002; 2008). Callers have to split their cognitive resources between two tasks that use audio: taking part in the conversation and listening and reacting to context sounds. In our set-up we expected that indirect cues (i.e., heartbeats) might create more resource conflicts than direct cues (i.e., driving sounds), as indirect cues might require more cognitive stages to interpret the sounds (Wickens, 2002; 2008). Our results did not provide evidence for this though. Both cue types raised the caller's awareness that the driver was busy (see Figure 3).

Limitations and alternative considerations

In the Method section of experiment 1 we motivated our choices for the study design and how this balanced realism with experimental control. We will discuss some further considerations and limitations below. In general, the generality of our findings can be tested in other conversation settings other driving tasks (e.g., responding to sudden events), including driving on a real road.

In our conversation task participants were not allowed to discuss the driving task during the experiment. This allowed us to test how sound sharing alone can make conversations less distracting. In more realistic settings, the explicit sharing of context is likely to facilitate discussion of the driving context, which can improve driver safety (Charlton, 2009; Drews et al., 2008). We wanted to avoid this confound.

We analyzed the conversation using a combination of automatic and manual labor. The automatic detection of talk segments (cf. Maciej et al., 2011) provided a consistent, replicable criterion for detecting turns without experimenter bias. However, it might occasionally have led to discrepancies in detecting the exact start or end of a talk segment (e.g., if initial vocalization was too soft to detect). The effect of such occasional discrepancies on the statistical results was most likely small, as we used the algorithm for all conversations and because we used a within-subjects design. For each segment we manually labeled the speaker. This ensured that incorrect inferences by the algorithm about the speaker could be corrected.

Our conversation metrics have also been assessed in other work (Charlton, 2009; Maciej et al., 2011). However, alternative analyses are possible. For example, the number of turns (Drews et al., 2008), interruptions (Maciej et al., 2011), pauses (Iqbal et al., 2010; Maciej et al., 2011), and references to traffic (Charlton, 2009; Drews et al., 2008). Production rate has also been measured (Drews et al., 2008; Schneider & Kiesler, 2005). Some have analyzed the complexity of speech (Drews et al., 2008). For specific conversation tasks, previous work has also reported measures such as the accuracy in telling or remembering a story (Becic et al., 2010; Iqbal et al., 2010).

We could not employ content specific metrics, as story topics varied too much for a consistent evaluation criterion (e.g., a recall test). We explored production rate metrics (words per second) using transcribed data from the second half of experiment 2. These analyses did not reveal any significant effects on production rate and were not explored further.

Alternative analyses might be valuable in other studies and expanded beyond the set that is already explored in the literature. For example, future work could develop measurements for the strategies that our participants mentioned in the post-study questionnaire.

Callers rated the driver's busyness on each trial. This provided a subjective assessment of workload. Multi-item questionnaires such as the NASA-TLX (Hart & Staveland, 1988) could give a more detailed measurement. However, this had two drawbacks for our set-up. First, this assessment would have required more time than was available per trial. Second, the caller did not have sufficient information about the driver to rate performance on all required aspects of the NASA-TLX (e.g., to rate frustration) - the NASA-TLX is more typically used for self-rating.

The construct busyness was intuitive to participants, but is less well defined than for example workload. Although the construct workload itself is being debated (for a review in the context of driver distraction see Mehler, Reimer, & Zec, 2012b), our work offered objective measures of workload in terms of conversation performance, driving performance, and physiological responses. In our analyses we sticked with the term "busyness", as used in the questionnaire.

Future work & implementation considerations

Our results imply that although callers can be made more aware of the driver's busyness through sound sharing, they need to be trained or be informed on how to act on the sounds. Future work should explore what ways of conversation adaptation (for example, as suggested in the post-study questionnaire) are most effective in reducing the driver's workload. This can then inform callers in general on how to cope with a driver that faces a high workload.

We captured the driver’s context in proxy sounds, as our car following task was not directly associated with sound – a limitation in common with other driving tasks such as lane changing, freeway entry, making turns, and approaching traffic lights. Future work could further investigate how the type and frequency of proxy sounds influences performance.

How frequently proxy sounds are played might be based on an automated decision about the driver’s workload. The decision can be based on offline sources such as road accident statistics. Online sources can also be used, for example using car sensors (e.g., driving speed, number of surrounding cars) or cognitive and physiological measures, such as pupil dilation (Iqbal, Adamczyk, Zheng, & Bailey, 2005; Wierda, van Rijn, Taatgen, & Martens, 2012), or heart rate (Healey & Picard, 2005; Mehler et al., 2012a).

An understanding about a driver’s context and state might also be derived from the driver’s voice (e.g., by considering pitch). Automatic detection systems might be needed for these purposes, as conversations without context sharing are consistently found to be more distracting, suggesting that callers did not detect potential cues in the driver's voice (Charlton, 2009; Crundall et al., 2005; Drews et al., 2008; Gugerty et al., 2004; Maciej et al., 2011; Schneider & Kiesler, 2005).

An alternative to adding proxy sounds to a conversation stream is to share *actual* context sounds of a driver (i.e., the "real" traffic sounds). Future research should investigate how this can be done, as it is going against the new industry practice of noise reduction (Bennett, 2014), and how these sounds can be made more salient to a caller. This allows callers to discuss the context with the driver (Charlton, 2009; Drews et al., 2008), as the context is real.

Other modalities than audio or video can be used for context sharing (e.g., tactile feedback). In addition, it is an empirical question how easily callers can integrate information from multiple modalities (e.g., audio and video) to generate a rich understanding of a driver's context.

Finally, none of the studies that investigated context sharing between drivers and callers have looked at how useful this technique is over longer periods of time and whether beneficial effects persist. Moreover, it is an open question how useful context sharing is when signals (i.e., frequency, intensity of sound) *change* during a phone call - for example, to convey that a driver's context changed from "relaxed" to "busy". These follow-up questions can be investigated now that we have established that sharing context sounds can in principle alter a caller's perception of the driver's busyness.

References

Allen, R. W., & Jex, H. R. (1972). A simple fourier analysis technique for measuring the dynamic response of manual control systems. *IEEE Transactions on Systems, Man and Cybernetics*, *2*(5), 638–643.

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Analysis & Prevention*, *27*(5), 707–715. doi:10.1016/0001-4575(95)00026-V

Atchley, P., & Chan, M. (2011). Potential Benefits and Costs of Concurrent Task Engagement to Maintain Vigilance: A Driving Simulator Investigation. *Human Factors*, *53*(1), 3–12. doi:10.1177/0018720810391215

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin and Review*, *17*(1), 15–21.

Bennett, J. (2014, January 8). The Quest for a Quieter Ride. *The Wall Street Journal*, p. D1 & D3.

Brookhuis, K., De Vries, G., & De Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention*, *23*(4), 309–316. doi:10.1016/0001-4575(91)90008-S

Brookhuis, K., Waard, D. de, & Mulder, B. (1994). Measuring driving performance by car-following in traffic. *Ergonomics*, *37*(3), 427–434.

Brumby, D. P., Salvucci, D. D., & Howes, A. (2009). Focus on Driving: How Cognitive Constraints Shape the Adaptation of Strategy when Dialing while Driving (pp. 1629–1638). Presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, *40*(4), 1282–1293. doi:10.1016/j.aap.2008.01.009

Charlton, S. G. (2009). Driving while conversing: cell phones that distract and passengers who react. *Accident Analysis & Prevention*, *41*(1), 160–173. doi:10.1016/j.aap.2008.10.006

Cnossen, F., Meijman, T., & Rothengatter, T. (2004). Adaptive strategy changes as a function of task demands: a study of car drivers. *Ergonomics*, *47*(12), 218–236. doi:info:doi/10.1080/00140130310001629757

Crundall, D., Bains, M., Chapman, P., & Underwood, G. (2005). Regulating conversation during driving: a problem for mobile telephones? *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(3), 197–211.

de Guzman, E. S., Sharmin, M., & Bailey, B. P. (2007). Should I call now? Understanding what context is considered when deciding whether to initiate remote communication via mobile devices. In C. G. Healey and E. Lank (Eds.), *Proceedings of Graphics Interface* (pp. 143–150). New York: ACM Press.

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, *14*(4), 392–400.

Duggan, G. B., Johnson, H., & Sørli, P. (2013). Interleaving tasks to improve performance: Users maximise the marginal rate of return. *International Journal* *of Human-Computer Studies, 71(5)*, 533-550.

Grandhi, S. A., Schuler, R., & Jones, Q. G. (2011). Telling calls: facilitating mobile phone conversation grounding and management. In D. Tan, G. Fitzpatrick, C. Gutwin, B. Begole & W. Kellogg (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2153–2162). New York: ACM Press.

Gugerty, L., Rakauskas, M., & Brooks, J. (2004). Effects of remote and in-person verbal interactions on verbalization rates and attention to dynamic spatial scenes. *Accident Analysis & Prevention*, *36*(6), 1029–1043. doi:10.1016/j.aap.2003.12.002

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, *1*(3), 139–183.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2), 156–166.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors*, *48*(1), 196–205.

Insurance Institute for Highway Safety. (2012). Cellphone and texting laws. Retrieved September 7, 2012, from http://www.iihs.org/laws/cellphonelaws.aspx

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an index of opportunity: Understanding Changes in Mental Workload during Task Execution. In W. Kellog, S. Zhai, G. van der Veer, and C. Gale (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 311–320), New York: ACM Press.

Iqbal, S. T., Horvitz, E., Ju, Y., & Matthews, E. (2011). Hang on a Sec! Effects of Proactive Mediation of Phone Conversations while Driving. In D. Tan, G. Fitzpatrick, C. Gutwin, B. Begole & W. Kellogg (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 463–472). New York: ACM Press.

Iqbal, S. T., Ju, Y.-C., & Horvitz, E. (2010). Cars, Calls, and Cognition: Investigating Driving and Divided Attention (pp. 1281–1290). In S. E. Hudson, G. Fitzpatrick, W. K. Edwards, T. Rodden & E. Mynatt (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1281-1290). New York, NY: ACM Press.

Janssen, C. P., & Brumby, D. P. (2010). Strategic Adaptation to Performance Objectives in a Dual-Task Setting. *Cognitive Science*, *34*(8), 1548–1560. doi:10.1111/j.1551-6709.2010.01124.x

Janssen, C. P., Brumby, D. P., & Garnett, R. (2012). Natural Break Points: The Influence of Priorities, and Cognitive and Motor Cues on Dual-Task Interleaving. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 5–29. doi:10.1177/1555343411432339

Janssen, J. H., Bailenson, J. N., IJsselsteijn, W. A., & Westerink, J. H. D. M. (2010).

Intimate heartbeats: Opportunities for affective communication technology. *IEEE Transactions on Affective Computing*, *1*(2), 72–80.

Kearney, J. K., Rizzo, M., & Severson, J. (2007). Virtual reality and neuro-ergonomics. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics* (pp. 253–274). Oxford: Oxford University Press.

Kun, A. L., & Medenica, Z. (2012). Video Call, or Not, that is the Question In J. A. Konstan, E.H. Chi, and K. Höök. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - Extended Abstracts* (pp. 1631–1636). New York: ACM Press.

Kun, A. L., Shyrokov, A., & Heeman, P. A. (2012). Interactions between human-human multi-threaded dialogues and driving. *Personal and Ubiquitous Computing*, 1–10. doi:10.1007/s00779-012-0518-1

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin and Review*, *15*(6), 1135–1140. doi:10.3758/PBR.15.6.1135

Lee, C., & Abdel-Aty, M. (2008). Presence of passengers: Does it increase or reduce driver's crash potential? *Accident Analysis & Prevention*, *40*(5), 1703–1712.

Lindqvist, J., & Hong, J. (2011). Undistracted driving: A mobile phone that doesn't distract. In A. Lamarca & L. Cox (Eds.), *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications* (pp. 70-75). New York: ACM Press.

Lotan, G., & Croft, C. (2007). imPulse. In B. Begole, S. Payne, E .Churchill, R. St. Armant, D. Gilmore & M.B. Rosson (Eds.), *Proceedings of the SIGCHI Conference* *on Human Factors in Computing Systems - Extended Abstracts* (pp. 1983-1988). New York, NY: ACM Press.

Maciej, J., Nitsch, M., & Vollrath, M. (2011). Conversing while driving: The importance of visual information for conversation modulation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(6), 512–524. doi:10.1016/j.trf.2011.05.001

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The contribution of passengers versus mobile phone use to motor vehicle crashes resulting in hospital attendance by the driver. *Accident Analysis & Prevention*, *39*(6), 1170–1176.

McKnight, AJ, & McKnight, A. (1993). The effect of cellular phone use upon driver attention. *Accident Analysis & Prevention*, *25*(3), 259–265.

Mehler, B., Reimer, B., & Coughlin, J. F. (2012a). Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand From a Working Memory Task: An On-Road Study Across Three Age Groups. *Human Factors*, *54*(3), 396–412. doi:10.1177/0018720812442086

Mehler, B., Reimer, B., & Zec, M. (2012b). Defining workload in the context of driver state detection and HMI evaluation. In: A.L. Kun, L. Boyle, B. Reimer, A. Riener, J. Healey, W. Zhang, B. Pfleging and M. Kurz (Eds.). *Proceedings of AutomotiveUI - International Conference on Automotive User Interfaces and Intervehicular Applications* (pp. 187–191). New York, NY: ACM Press.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*, 61–64.

Norman, D. A., & Bobrow, D. G. (1975). On Data-limited and Resource-limited Processes. *Cognitive Psychology*, *7*, 44–64.

Pfleging, B., Schneegass, S., & Schmidt, A. (2013). Exploring User Expectations for Context and Road Video Sharing While Calling and Driving. In J. Terken (Ed.), *Proceedings of AutomotiveUI - International Conference on Automotive User Interfaces and Intervehicular Applications* (pp. 132–139). New York, NY: ACM Press.

Salvucci, D. D., & Beltowska, J. (2008). Effects of memory rehearsal on driver performance: experiment and theoretical account. *Human Factors*, *50*(5), 834–844.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101–130. doi:10.1037/0033-295X.115.1.101

Salvucci, D. D., & Taatgen, N. A. (2011). *The Multitasking Mind*. New York, NY: Oxford University Press.

Schneider, M., & Kiesler, S. (2005). Calling while driving: effects of providing remote traffic context. In W. Kellog, S. Zhai, G. van der Veer, and C. Gale (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 561–569), New York: ACM Press.

Slovák, P., Janssen, J. H., & Fitzpatrick, G. (2012). Understanding heart rate sharing: towards unpacking physiosocial space In J. A. Konstan, E.H. Chi, and K. Höök. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp.859–868). New York: ACM Press. doi:10.1145/2207676.2208526

Stewart, C. O., & Kreuz, R. J. (2003). On the communicative function of exaggeration: How to be a million times clearer. *Communication Research Reports*, *20*(4), 331–340.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: dual-Task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*(6), 462–466.

Think before you speak: Distracted driving is the new drunk driving. (2011, April 13). Think before you speak: Distracted driving is the new drunk driving. *The Economist*.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors*, *50*(3), 449–455. doi:10.1518/001872008X288394

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456–8460.

Wiley, R. H. (2006). Signal detection and animal communication. *Advances in the Study of Behavior*, *36*, 217–247.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Interspeech* (pp 541-544).

Footnotes

[1] Vehicle dynamics were as follows. The simulator uses STISIM's "simple dynamics" setting. For the steering wheel controls we did not use an oversteer coefficient. The yaw rate scale factor determines the angle the simulated vehicle takes per second given the angle of the steering wheel and was set to 0.00015 Rad/Sec/Deg.

For acceleration, an acceleration limit was set at 0.35 g's. This is the maximum acceleration that is exerted when the gas pedal is fully pressed. According to the STISIM manual this value is slightly higher than the value range expected of a family size Sedan (0.1 to 0.25 g's) and more comparable to the lower range of a sports car (0.2 to 0.4 g's). The simulator distinguished 195 input levels for the accelerator.

For deceleration, a deceleration limit was set at -0.65 g's. This is the maximum deceleration that is exerted when the brake is pressed fully. According to the STISIM manual, this value is comparable to the deceleration of a normal family sedan (0.6 to 0.8 g's). The simulator distinguished 130 input levels for the brake.

[2] Curves became visible 200 ft before the curve appeared. They had an entry and exit spiral of 100 ft to create a smooth transition from straight to the desired curve of 0.008 per foot and back to the straight road. The curved section (0.008 per foot) was 600 ft.

[3] To create sentences and theme words, we gathered inspiration from a website that lists words for family games. We used the “easy” sets for Catchphrase,

Charades, and Pictionary on http://www.thegamegal.com/printables/. From this set we removed duplicates and difficult words (e.g., “jumping jack”), and simplified words (e.g., “swimming pool” became “pool”). We randomized the remaining list of 129 words and split it into 18 sets of 7 words. This was repeated 8 times, resulting in 144 unique sets, which were printed on cards.

We asked 24 people to judge 18 randomly drawn cards. Raters were identified using an opportunity sample at a social event. Each rater was instructed to read the words on a card, and to think whether there was a coherent theme amongst the words – “if they had to tell a story using the majority of these words, what would that story be about?” Raters were also asked to rate the coherence of the words on a five-point likert-scale, with 1 indicating no coherence, and 5 indicating strong coherence.

For the experiment we used cards where (1) the median coherence rating was at least 2, and less than 4, and (2) at least one theme was identified. Based on the themes, the experimenter then developed a starting sentence. Each sentence contained the name of one person. Half the sentences had a boy’s name, half a girl’s name (taken from the top-50 girls and boys names listed on http://baby-names.familyeducation.com/popular-names/). All sentences were approximately the same length. There were 78 sentences in total. Order of sentences was randomly determined for each participant. None of the participants used all sentences. Sentences and words are listed in the supplementary materials.

[4] Examples of hints for keeping the conversation going were to take turns in speaking (e.g., making use of each other's creativity), moving conversation away from an initial topic if this was too challenging (e.g., "At that point, John started to dream about …"), and talking about the main character in the third person (i.e., so as to not limit imagination to personal experiences).

Figure 1. The set-up of the experiment. Top-left: driver driving in simulator with headset on. Top-right: remote caller operating interface with sentences and words. Bottom-left: detail view of the follow task. The driver tries to align the car they are following (police car) with a “projected” red car in front of them, and one lane to the right. Bottom-right: screenshot of the interface of the remote caller.

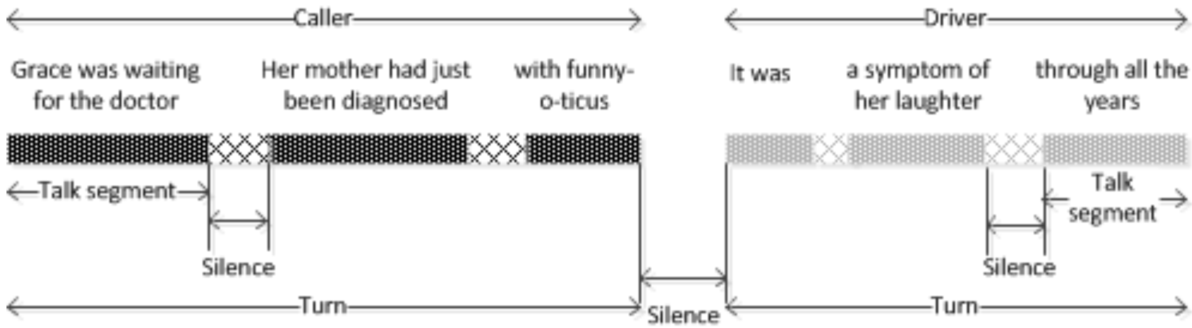

Figure 2. Illustration of an example conversation with turns, talk segments and silence segments marked.

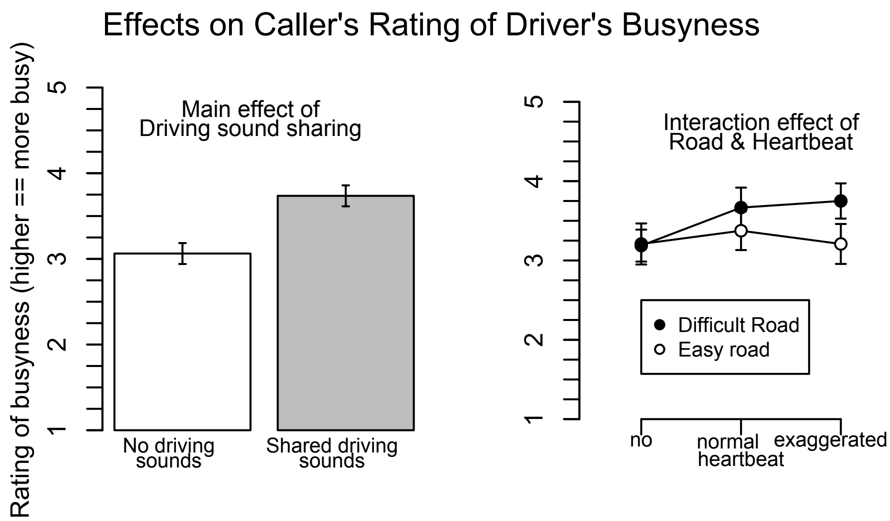

Figure 3. Effects on the caller's rating of the driver's busyness, with higher values indicating a perception of more busyness. Left: Bar plot illustrating the main effect of driving sound sharing: drivers were perceived as more busy when driving sounds were shared (grey bar) compared to when no sounds were shared (white bar). Right: scatterplot of rating as a function of heartbeat sharing (horizontal axis) and road type, with white points denoting the easy road and black points denoting the difficult road. Drivers were perceived as more busy on the difficult road when heartbeats were shared. Error bars indicate 95% within-subject confidence intervals.

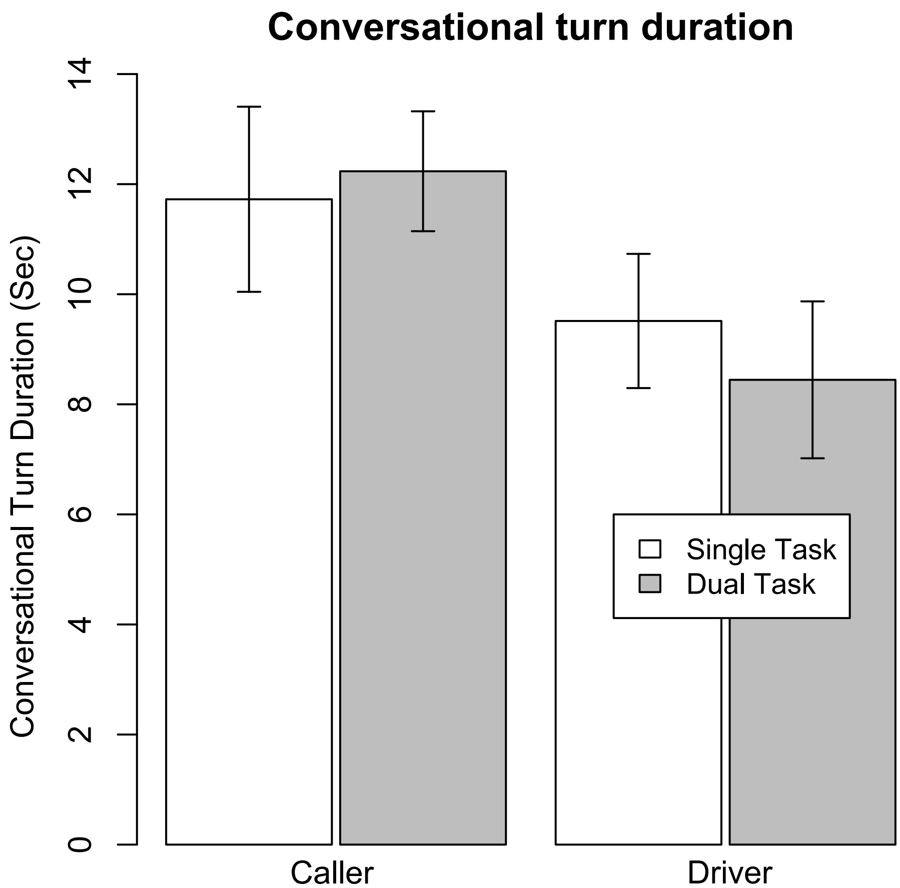

Figure 4. Bar plot of conversational turn length in single-task (white bars) and dual-task (grey bars). The caller's conversational turns (left two bars) are longer in dual-task trials, compared to single-task trials. In contrast, the driver's conversational turns (right two bars) are shorter in dual-task. Error bars indicate 95% within-subjects confidence intervals.

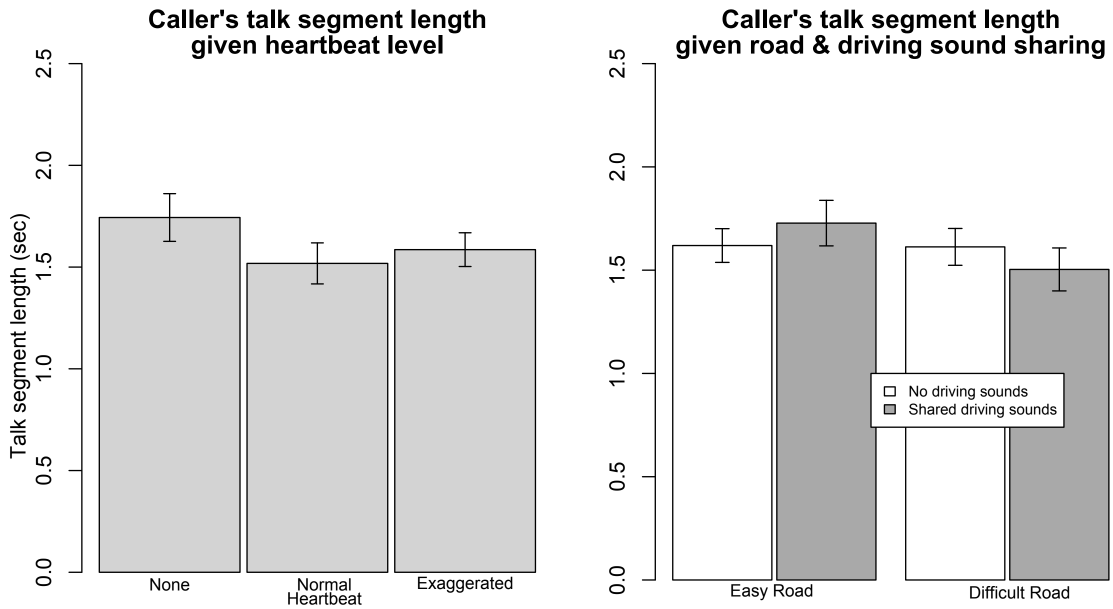

Figure 5. Bar plots of the length of the caller's talk segments. The left plot shows segment length for various heartbeat sharing levels. The trend in the data was that talk segments were shorter when heartbeat sounds where shared. The right plot shows talk segment length based on whether driving sounds were shared (grey bars) or not (white bars) for the easy road (left two bars) and the difficult road (right two bars). When sounds were shared talk segments were shorter on the difficult road and longer on the easy road. Error bars indicate 95% within-subjects confidence intervals.

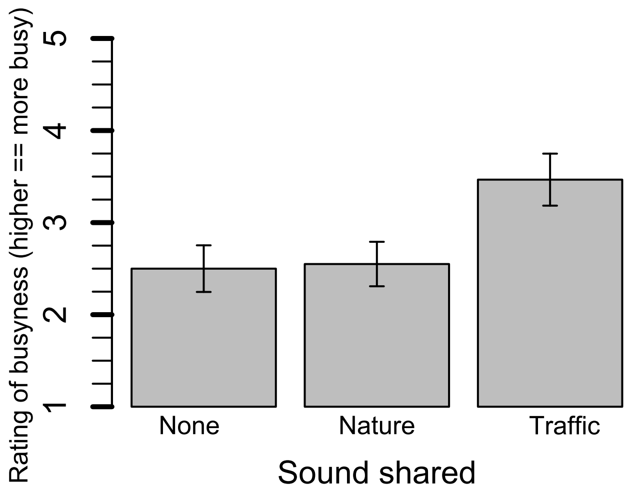

Figure 6. Bar plot of the perceived busyness of the driver by the caller, given the sounds that were shared: no sounds, nature sounds, or traffic sounds (data from Experiment 2). The caller rated the driver as being more busy when traffic sounds were shared. Error bars indicate 95% within-subjects confidence intervals.