May the Force Be with You: Ultrasound Haptic Feedback for Mid-Air Gesture Interaction in Cars

Gözel Shakeri, John H. Williamson, Stephen Brewster
Glasgow Interactive Systems Group
University of Glasgow, Scotland
g.shakeri.1@research.gla.ac.uk, {johnh.williamson,stephen.brewster}@glasgow.ac.uk

ABSTRACT
The use of ultrasound haptic feedback for mid-air gestures in cars has been proposed to provide a sense of control over the user's intended actions and to add touch to a touchless interaction. However, the impact of ultrasound feedback to the gesturing hand regarding lane deviation, eyes-off-the-road time (EORT) and perceived mental demand has not yet been measured. This paper investigates the impact of uni- and multimodal presentation of ultrasound feedback on the primary driving task and the secondary gesturing task in a simulated driving environment. The multimodal combinations of ultrasound included visual, auditory, and peripheral lights. We found that ultrasound feedback presented uni-modally and bi-modally resulted in significantly less EORT compared to visual feedback. Our results suggest that multimodal ultrasound feedback for mid-air interaction decreases EORT whilst not compromising driving performance nor mental demand and thus can increase safety while driving.

CCS Concepts
•**Human-centered computing** → **Haptic devices; Auditory feedback; Gestural input;**

Author Keywords
ultrasound haptics; multimodal feedback; mid-air gestures;

INTRODUCTION
Car manufacturers such as BMW, VW, Cadillac, and Hyundai see potential in mid-air gesture interaction in driving situations and are investing in these gesture systems [4]. The advantage of mid-air gesture interaction interfaces is a reduction in mental demand of infotainment systems on the driver [32]. Since these motions do not require accurate hand-eye coordination they decrease mental and visual demands on the driver [30] compared to traditional touch-based interaction [24, 48]. By merely executing a coarse gesture in air, the driver can increase the in-car temperature, select the next song, or reject a phone call [1]. However, there is only partial understanding of the effects of mid-air gestures on driving performance, visual attention, and mental workload due to the novelty of the interaction technique. For instance, the driver's unfamiliarity with a gesture system has a negative impact on lane keeping ability and mental workload [19, 38]. The decoupling of the user's hand from the interface [7] and the lack of sense of control over the touch less interface [8] are additional aspects which lessen user satisfaction, safety and usability of gestures [33].

Feedback from the system can address these problems [33] by informing the driver about how the system interpreted their actions [15]. Feedback informs the user whether the system pays attention to them, classifies the executed gesture correctly, and provides the user with knowledge about system state. This information is necessary to avoid increased mental efforts.

Previous works have shown that there is growing interest in ultrasound feedback for mid-air gesture interaction in cars [40, 8, 17]. Rümelin *et al.* [40] investigated ultrasound feedback for mid-air pointing. They showed that ultrasound haptics can display floating widgets and be utilised for button tapping in air. Car manufacturers like BMW (HoloActive Touch [1]) and Bosch (neoSense [3]) are also using ultrasound feedback in their next generation cars. However, the impact of this feedback technique on driving performance and perceived mental demand has not yet been investigated.

There is a necessity to understand and mitigate the effects of ultrasound feedback for mid-air gestures such that neither driving performance nor safety is negatively impacted by increased workload or distraction of the driver [20, 35]. Therefore, the contribution of this work is multimodal ultrasound feedback and its effect on driving performance, visual attention, and perceived mental demand. The benefit to be gained from this contribution is reduced eyes-off-the-road time compared to visual feedback only.

RELATED WORK
It is necessary to provide effective feedback to mid-air gestures to help users overcome uncertainty [12]. It reduces mental workload [39, 32] and provides a sense of control of the outcomes of one's actions [8]. It decreases eyes-off-the-road time [41] and improves user satisfaction [33]. For these reasons, a growing body of research is investigating feedback for mid-air interaction in driving situations.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *AutomotiveUI '18, September 23–25, 2018, Toronto, ON, Canada* https://doi.org/10.1145/3239060.3239081

Shakeri *et al.* [41] have investigated different unimodal feedback types such as visual, auditory, peripheral visual, and cutaneous push haptic feedback for mid-air gesture interaction in cars. They found that non-visual feedback reduced eyes-off-the-road time significantly and did not influence driving performance negatively. In their study, however, peripheral visual feedback and cutaneous push feedback significantly increased mental demand compared to visual and auditory. May *et al.* [32] found that mental demand did not increase nor did driving performance decrease compared to direct touch interaction when auditory and visual feedback were presented bimodally. Sterkenburg *et al.* [43, 42] also explored bi-modal audio-visual feedback for gesture control and found that it led to significantly less eyes-off-the-road time and lane deviation than uni-modal visual feedback. However, May *et al.* [32] and Sterkenburg *et al.* [43, 48] presented visual feedback in their bi-modal conditions continuing to present information to a channel which is fully occupied by the primary driving task.

A rich body of research [45, 22, 37] shows that distribution of information to sensory channels not used in the primary one, such as hearing and touch, does not decrease driving performance, such as decrease in reaction time when mental workload is increasing [37]. Further, Lee *et al.* [26] showed that well designed multimodal feedback for mid-air gestures has the potential to reduce mental workload without compromising the primary driving task. Since the haptic, auditory and peripheral visual channel are not primary target for driving information they are available for infotainment system output.

Ultrasound Feedback

Technologies such as ultrasound haptics [5] lets users experience tactile sensations in mid-air as they gesture. Ultrasound haptics uses acoustic radiation pressure to create areas of friction on the skin and thus tactile sensations. It offers the new opportunity of "adding touch to a touchless interaction" [13] and therefore increases the sense of control over the user's intended mid-air gesture actions [8]. Additionally, it conveys tactile feedback which can be felt by the unadorned hand in mid-air mitigating the problem of on-body-displays. For these reasons, recent research proposed the use of ultrasound haptics in cars [40, 17].

Rümelin *et al.* [40] conducted a study on pointing gestures and tapping ultrasound buttons in mid-air. Their aim was to determine the optimal modulation frequency and duration of the signal. Therefore, the participants' task was ultrasound stimuli recognition. Sixteen percent of their participants could not discriminate between the presented stimuli and were excluded from data analysis. This is not the only work which reported on stimuli discrimination issues. Long *et al.* [31] too found that ultrasound patterns are challenging to distinguish from each other. They conducted a study where participants were presented with six 3D ultrasound objects which had to be classified. The presented objects were: sphere, pyramid, horizontal prism, vertical prism, and cube. Even though there was no additional task to preoccupy the participants' mental resources, correct identification of objects ranged from 60% (sphere) to $\leqslant$ 95% (pyramid). The authors argued that lack of visual feedback lead to misidentification of the presented objects. This suggests that effective ultrasound feedback needs additional multimodal information such as auditory or peripheral vision lights.

Peripheral Vision Feedback

The peripheral visual channel can be engaged by using interactive lights. It has been shown that peripheral light feedback in event driven and data-rich environments (i.e. aeroplane cockpits) does not interfere with the performance of the primary visual task [34]. Peripheral visual cues are further highly effective in conveying information [23, 44]. These findings have led to growing interest in ambient light feedback in driving situations. *AmbiCar* [28] was the first interactive lights display used in the car to inform the driver about driving related events. Peripheral lights are now being used to inform the driver about lane change decisions [29], current travel speed [46], and intentions of the automated car [27]. Results from these works show that ambient light demands significantly less visual attention than a traditional centre console screen.

Peripheral lights have also been used for mid-air gesture interaction in driving situations [41]. Freeman *et al.* [15] showed that peripheral lights in combination with tactile feedback can successfully surpass the shortcomings of each feedback type and provide an additional modality for mid-air gesture feedback for mobile phones. If combined appropriately with ultrasound feedback, these techniques are promising for in-car gesture applications.

Auditory Feedback

Auditory feedback in driving environments has been shown to reduce looking away time [10] and if presented to mid-air gestures it reduces eyes-off-the-road time without negatively impacting the driving performance [41, 43] nor mental demand [41, 32]. A multimodal combination of ultrasound and audio feedback can surpass the shortcomings of ultrasound haptics for mid-air feedback [8].

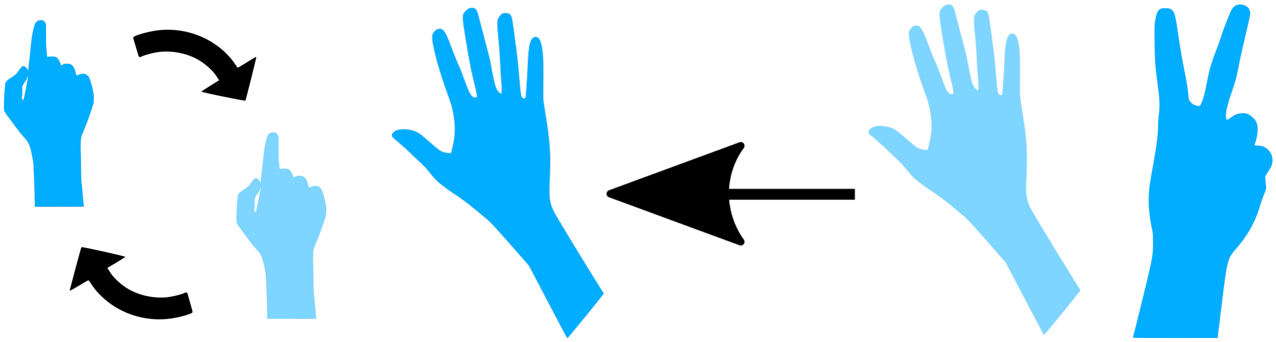
Summary

There is growing interest in multimodal feedback for mid-air gestures, especially the usage of ultrasound haptics which provides tactile sensations in mid-air to the unadorned, gesturing hand. This research aims at answering the question of how to design mid-air gesture feedback which does not increase perceived mental nor visual demands on the driver compared to traditional touch based interaction nor compromises driving performance. We have chosen to focus on ultrasound haptics since automotive manufacturers already show case their latest car designs incorporating this technology. Multimodal presentation of ultrasound can enrich this technology such that driving performance will not be negatively impacted nor the driver distracted.

GESTURE AND FEEDBACK DESIGN

Gesture Design

The set of gestures used for this study were based on mid-air gesture design guidelines [47, 16] and already available ones for in-car interaction (BMW, VW). VW introduced the swipe left/right motion in their gesture enabled user interface [2].



(a) Circle Clockwise (b) Swipe Left (c) Victory

Figure 1. Gesture Set.

BMW use a circular motion to increase/decrease a setting. The gesture is performed by circling with an extended index finger either clockwise or anti-clockwise. BMW also introduced the victory gesture to turn the centre console screen on/off; this gesture is executed by extending the index and middle finger parallel to the tabletop.

These gestures were performed in following ways: Swipe Left 2, 3, 4 times (SL2, SL3, SL4), Swipe Right 2, 3, 4 times (SR2, SR3, SR4), Victory (V), Circle Clockwise 2, 3, 4 times (CW2, CW3, CW4), and Circle Anti Clockwise 2, 3, 4 times (CAW2, CAW3, CAW4). We differentiate the gestures depending on motion and direction, not number of execution; this results in five gesture types overall: SL, SR, CW, CAW, and V. As suggested by previous research [9, 39, 6] to keep the gesture set smaller than eight, we limited it to the above named five.

Duration of a single interaction is made up of three parts: time before gesture execution, gesture execution and the feedback. Before gesture execution, it is recommended [16, 14] to provide information whenever the hand enters the interaction box. In our study, feedback for this information lasts 150 ms. The duration of a single gesture lasts for 750 ms gesture execution and 500 ms gesture feedback. For example, a single swipe motion is required to last for 750 ms. If a participant is instructed to swipe left 4 times (SL4), the entire interaction lasts 150 + 3600 ms (150 ms on interaction box entrance plus $4x$750ms gesture execution and $4x(150+50)$ms feedback) (Table 1). For the swipe and circular motions we used the built in Leap Motion classifiers. The victory gesture was recognised by extending the index and middle finger for at least 750 ms.

Feedback Design

During our pilot studies and as found by previous research [32, 41], participants were prone to make accidental gestures by entrance of the interaction box and caused unwanted system response. To account for falsely classified gestures which can increase mental demand [16], we implemented our system such that it only provided feedback to the expected gesture. If a circular motion was expected, only CA and CAW gestures caused system response. We chose this solution because we were not evaluating the quality of the gesture recogniser but the gesture feedback.

We presented feedback on entering the interaction box due to recommendation by Gable *et al.* [16] and Freeman *et al.* [14]. This feedback assured the user that the system is paying attention and is ready for input.

The feedback was presented *functionally*. It was presented *after* gesture execution in a *discrete* manner instead of presenting it during the execution and continuously. Continuous feedback is important for usability [16, 15]. However, it might overload the driver's mental demand and increase distraction.

We chose Visual feedback as baseline for our study since it has already been used in the literature and in industry for mid-air gesture interaction [32, 43, 41].

Ultrasound Feedback: was presented via an Ultrahaptics array (Figure 2) to the right hand of the user. Functional feedback was presented for 500 ms to the palm of the driver. Each executed gesture was confirmed with a specific feedback pattern. The clockwise motion was confirmed with the presentation of a circular clockwise motion, the anti-clockwise gesture was confirmed with an anti-clockwise circular motion. Victory gesture feedback was provided by a 500 ms long ultrasound pulses to the tip of index and middle finger. Swipe motion was confirmed with a feedback pattern which mimicked the swiping motion of the hand, i.e. if swiped left, the presented feedback was a wall of air moving from right to left across the palm of the driver. Whenever the hand entered the interaction box, a short pulse was presented to the middle of the palm.

Peripheral Light Feedback: was presented on an LED strip from the A-pillar on the left side of the driver to the beginning of centre console.The strip was placed behind the steering wheel where the car instrument cluster would be (as proposed by Löcken *et al.* [28]).

Feedback for the swiping motions SL and SR was a yellow lights animation mimicking the direction of the gesturing hand. Duration of the animation was 500 ms. Successful CW or CAW motion was indicated by blue lights either incrementing to the right (CW) or decrementing to the left (CAW). As long as the hand was inside the interaction box, the blue lights remained alight. V on gesture feedback was presented with an animation of blue lights moving from the ends of the LED strip to the centre. The duration was 500 ms. V off feedback

Figure 2. Experiment set-up.

was an outward animation of red lights (from centre to the ends of the LED strip). We chose yellow and blue colours because they are most distinguishable in the peripheral vision [11]. Additionally, we chose the colour red to avoid issues for users who were colour blind. On entrance of the hand in the interaction box, the strip would pulse briefly (350 ms) in a dim white light.

Visual Feedback: was presented on the centre console to the right of the participant (Figure 2). The GUI design was adapted by Jaguar Landrover's centre console in terms of size of screen display, size of menu items, size of letters, etc (Figure 2). The GUI was a single horizontal scroll bar (from 0 to 10) with the cursor set at 5. A circle gesture caused the cursor to move up or down on the scale depending on whether it was a clockwise or anti-clockwise motion. A swipe left caused the entire scale to shift from 0 - 10 to 1 to 11, with the cursor remaining at the centre position. A swipe right moved the scale in the other direction. The victory gesture turned the screen on / off. Feedback for entrance of the hand in the interaction box was a brightening of the screen from "standby" to "active" mode. The screen turned darker again after the hand exited the interaction box.

Auditory Feedback: was presented via headphones through both speech and non-speech feedback and lasted 500 ms (Table 1). Feedback for the clockwise motion, for example, was the incremental increase of a tone by an octave, and decrease by an octave for the anti clockwise motion. Audio feedback for victory and swipe was presented through presentation of two separate notes. Speech feedback followed non-speech feedback representing an internal count from 0 to 10. Whenever

Gesture	Non-speech	Duration	Speech
V on	g#4 $\rightarrow$ c5	300 ms	on
V off	c5 $\rightarrow$ g#4	300 ms	off
SL	c4 $\rightarrow$ c4	250 ms	$\uparrow \{0-10\}$
SR	c5 $\rightarrow$ c5	250 ms	$\downarrow \{0-10\}$
CW	c4 $\rightarrow$ b4	250 ms	$\downarrow \{0-10\}$
CAW	b4 $\rightarrow$ c4	250 ms	$\uparrow \{0-10\}$

Table 1. Auditory feedback for gestures. The arrow in Non-speech describes the transition from one note to the next. Duration describes the length of each non-speech unit. Speech stands for a read-out gesture feedback; e.g. feedback for SL would be an increment of a number between $\{0-10\}$, and a decrement for SR. In total the feedback lasted 500ms. (CAW: circle anti-clockwise; CW: circle clockwise; SL: swipe left; SR: swipe rigth; V: victory.)

the driver's hand entered the interaction box, they heard the G4 tone (392 Hz) for 150 ms. There was no interaction-box-exit sound. The non-speech tones were generated in Audacity[1] and guided by Shakeri *et al.*'s design [41] (Table 1). The speech feedback was read out loud by a male US American voice (www.cereproc.com/ Voice: Nathan. Accessed 2016-01-31).

STUDY

We designed a within-subject, simulated driving study to evaluate the effectiveness of multimodal ultrasound feedback for three different mid-air gestures. The aim was to gain insight into the amount of distraction from the primary task.

Participants

17 participants (9 females) ranging from 19 to 40 years of age (μ 28.6 σ 6.8) were recruited via our university's student online forum. Seven of the participants had a UK driving license and ten a license from elsewhere. One participant was left handed, and one was ambidextrous; both are from the UK. A total of ten participants indicated that they had no prior experience with mid-air gesture interfaces.

Conditions

We presented following five combinations of feedback modalities: ultrasound (U), ultrasound-visual (UV), ultrasound-auditory (UA), ultrasound-peripheral (UP), and visual (V).

Hypotheses

H1: Multimodal ultrasound feedback decreases eyes-off-the-road time compared to unimodal visual and unimodal ultrasound feedback.

H2: Multimodal ultrasound feedback improves lane keeping ability compared to unimodal feedback.

H3: Feedback types do not influence gesturing task performance.

H4: Mental demand is significantly less in the multimodal feedback conditions.

[1] Audacity Version 2.1.2 http://www.audacityteam.org/ Accessed summer 2016

Experimental Variables

The Independent Variable was mid-air gesture type. There were five levels: Ultrasound, Ultrasound-Visual, Ultrasound-Auditory, Ultrasound-Ambient, and Auditory-Visual. The Dependent Variables were: lane deviation (metres), visual attention to primary task (number of glances off the road, average duration per glance, average time between glances), number of correct gestures (% correct), task duration (ms), perceived workload (NASA TLX), and our own questionnaire (demographics, handedness, preferences of feedback).

Apparatus

The usability lab (Figure 2) was equipped with 1) a computer running the OpenDS simulation, 2) a HD overhead projector, 3) a racing car chair, 4) a Logitech steering wheel with a 15 inch steering wheel, 5) headphones, 6) Neopixel LED strip for peripheral visual feedback, 7) an 8 inch monitor for visual feedback, and 8) a Leap Motion hand tracker, 9) an Ultrasound array, and 10) capacitive sensor on the steering wheel under the driver's right hand. The car set-up was adapted by Jaguar Landrover's interior design in terms of size of centre console, distance from steering wheel to dashboard, height of car instrument cluster, etc. The ultrasound array was placed where the gear stick is located with the Leap motion device at its top. The gesture interaction area was determined by the area in which ultrasound feedback can be perceived optimally (i.e. 10 cm above device, 14*x*14 cm).

The webcam recorded the participants' eye gaze while performing the driving and input tasks. Gaze and head pose data were extracted using OpenFace[2], an open source tool for eye-gaze and head pose estimation. An SVM classifier with a linear kernel was trained on 11,845 images obtained during a pilot study (6 participants). Input data for the classifier were 3D vectors for each eye and head pose rotation. The SVM model classified 94.56% eyes-off-the-road time correctly (10-fold cross validation).

We used OpenDS Version 3[3] to simulate a driving scenario. Participants performed the Lane Change Task (LCT). LCT (ISO standard 26022:2010) is designed to help evaluate any type of in-vehicle technology [36].

Procedure

On arrival, participants were provided with an introduction to the experiment. This included executions of each mid-air gesture in every condition (5 gestures per condition *x* 5 feedback modalities = 25 gesture executions). The experiment consisted of five blocks, one block for each feedback condition. During each block, participants executed 15 mid-air gestures (6 *x* SL/SR, 3 *x* V, 6 *x* CW/CAW). Each block lasted approximately 6 minutes. To counter balance for any learning effect, the conditions were ordered via a Balanced Latin Square.

Participants started the driving task at the outmost left lane of a five lane motorway and had to steer the car into the middle lane (Figure 2). After approximately 20 seconds of stabilised driving in the middle lane, the experiment and recordings of the data started. After these initial 20 seconds, the driver was instructed to gesture (via a.) pop up message box at the bottom of the screen and b.) speech instructions through headphones which the participants were wearing at all times). The message box was displayed for 3 seconds and the accompanying auditory instructions lasted up to 2 seconds. The auditory instructions were "swipe left/right 2-4", "(anti) clockwise 2-4", "victory". The speech instructions were read aloud by a male US American voice (www.cereproc.com/ Voice: Nathan. Accessed 2016-01-31). This reduced the chances of participants missing an instruction.

Simultaneously, the driver was prompted to change lane which was indicated by an arrow on a bridge panel above the motorway. The participant was free to choose which task to prioritise, lane change or gesture execution. However, participants were asked to perform their gestures and lane changes as quickly as possible while maintaining stabilised driving.

The gestures were performed with the right hand (as if driving on the right). The gestures were executed above the area where the gear stick is located (Figure 2).

Participants had a total of 30 seconds to complete a gesture and change lanes before the next trial started. We chose a set time interval for each trial because we could not dynamically manipulate the position of the panels above the motorway. This interval of 30 seconds provided an opportunity to gesture and return the car to the middle of the target lane and regain stabilised driving. After each feedback condition block, participants were asked to fill in a NASA TLX workload questionnaire. Participants were reimbursed with £6 for an hour of their time.

RESULTS

Gaze Behaviour

For all conditions, mean eyes-off-the-roads time (848 ms) across conditions and participants was within the NHTSA guidelines (< 2000 ms). Collected glance data was non-normal. Therefore, we used the Kruskal-Wallis test to analyse the variance in the data. Kruskal-Wallis test showed that gaze behaviour was dependent on condition ($p \leqslant 0.01, \chi^2(4) = 14.32$). Visual and Ultrasound-Visual conditions had highest EORT (Figure 3).

Gaze on Lane Deviation

We used the Root Mean Square Error to obtain the differences between the collected driving data points. Therefore, the data is non-normal. Kruskal-Wallis analysis shows no significant difference in impact of EORT on lane deviation ($p = 0.48, \chi^2(475) = 475.48$) nor the number of glances off the road on lane deviation ($p = 0.46, \chi^2(15) = 14.83$). Further, average duration of glances off the road had no significant effect on lane deviation ($p = 0.47, \chi^2(488) = 489.04$), nor the average duration between two glances on lane deviation ($p = 0.73, \chi^2(395) = 376.73$). Finally, there is no significant influence of gaze duration and number of glances on lane deviation ($p = 0.33, \chi^2(2) = 2.19$).

[2]OpenFace, https://github.com/TadasBaltrusaitis/OpenFace Accessed 2017-04-17

[3]OpenDS Version 3, https://www.opends.eu/ Accessed 2017-04-25

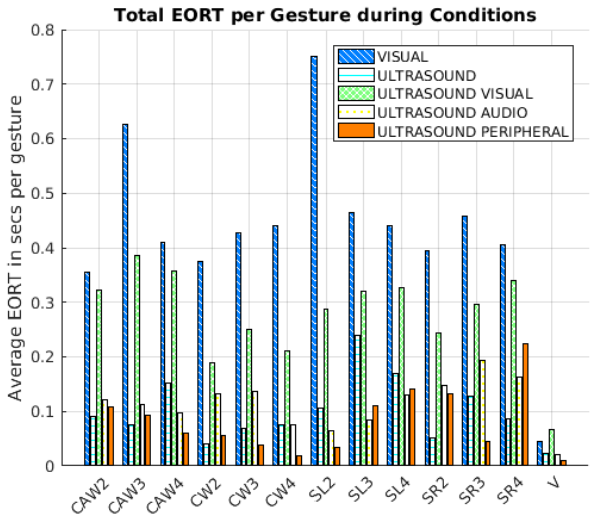


Figure 3. Eyes-off-the-road time per gesture across participants across the five conditions. (CAW: circle anti-clockwise; CW: circle clockwise; SL: swipe left; SR: swipe rigth; V: victory.)

Gaze on Secondary Task Performance

A Kruskal-Wallis analysis revealed a significant impact of number of glances on gesture performance ($p \leqslant 0.01, \chi^2(12) = 37.50$). This is due to a correlation between type of gesture and number of glances ($p \leqslant 0.01, \chi^2(8) = 4.00$) with Victory gesture resulting in least glances off the road and Swipe Left 2 and Circle Clockwise 3 in most (Figure 3).

However, there is no significant difference in gaze duration on secondary task performance ($p = 0.15, \chi^2(54) = 64.56$), nor the average glance duration on secondary task performance ($p = 0.15, \chi^2(54) = 64.56$), nor average time between glances on gesture performance ($p \geqslant 0.05, \chi^2(39) = 54.33$).

Lane Deviation

Lane deviation was measured from the presentation of the lane change until the next presentation of lane change. Therefore, lane deviation in our study is high since we take transition from current lane into next lane into our analysis.

Kruskwal-Wallis analysis showed no significant influence of condition on lane deviation ($p : 0.17, \chi^2(4) = 6.31$) (Figure 5.2). There is no significant difference in lane deviation between the visual (Visual / Ultrasound-Visual) and non-visual (Ultrasound, Ultrasound-Auditory, Ultrasound-Peripheral) conditions ($p : 0.12, \chi^2(1) = 2.33$). Further analysis of lane deviation showed there is no statistically significant difference in lane deviation across gestures ($p \geqslant 0.05, \chi^2(12) = 20.57$).

Secondary Task Performance

A gesture was classified as correct if the executed gesture was performed as instructed. Overall, 43.74% of instructed gestures were executed correctly with the V gesture the best at 99.62% (Figure 5). A multiple comparison of mean ranks showed that all swipe gestures performed significantly worse than Victory and all Circle gesture variations, especially Swipe Left gestures (Table 2). Further, gesture performance was

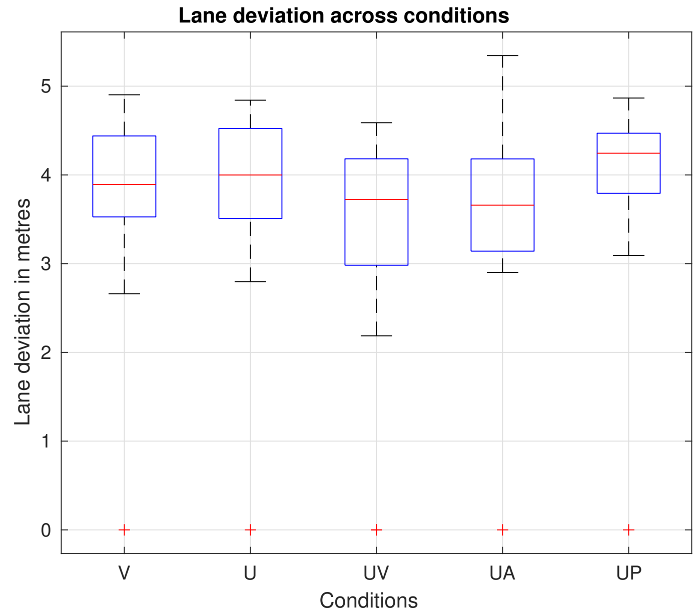


Figure 4. Lane deviation across the conditions. (V: visual; U: ultrasound; UV: ultrasound-visual; UA: ultrasound-audio; UP: ultrasound-peripheral.)

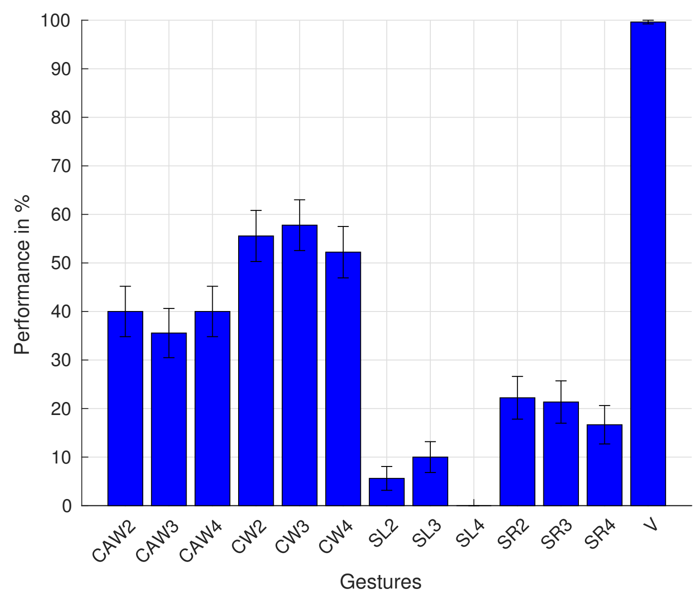


Figure 5. Gesturing performance across all participants across all conditions. (CAW: circle anti-clockwise; CW: circle clockwise; SL: swipe left; SR: swipe rigth; V: victory.)

significantly dependant on feedback type ($p \leqslant 0.01, \chi^2(4) = 73.07$) and Dunn's post-hoc test revealed that Ultrasound feedback resulted in 21.11% correct gesture performance, significantly worse than the other conditions (Table 2).

V	U	UV	UA	UP
51.68%	21.11%	46.29%	52.22%	47.40%
10.06 s	8.22 s	10.16 s	10.58 s	9.98 s

Table 2. Secondary task performance (%) and duration (seconds) depending on condition. (V: visual; U: ultrasound; UV: ultrasound-visual; UA: ultrasound-audio; UP: ultrasound-peripheral.)

Average hands off the wheel duration was 8.87 seconds and thus below the 15s rule [18]. Task duration was significantly dependent on feedback type ($p = 0.01, \chi^2(4) = 12.73$) (Table 2) with the lowest being 8.22 seconds during the Ultrasound condition. Individual gesture durations in seconds were CW2 9.51, CW3 10.69, CW4 12.01, CAW2 8.76, CAW3 10.99, CAW4 11.81, SL2 8.49, SL3 10.28, SL4 10.97, SR2 8.81, SR3 11.02, SR4 11.70, and V 2.34.

Subjective Workload

Analysis of the NASA TLX questionnaire revealed a significant difference in physical demand ($\chi^2(4) = 15.00, p \leqslant 0.01$), with the visual conditions having the highest levels (Figure 6). There were no significant differences in the remaining measures: mental demand ($\chi^2(4) = 9.09, p = 0.06$), temporal demand ($\chi^2(4) = 7.23, p = 0.12$), performance ($\chi^2(4) = 8.22, p = 0.08$), effort ($\chi^2(4) = 10.69, p = 0.03$), and frustration ($\chi^2(4) = 5.25, p = 0.26$).

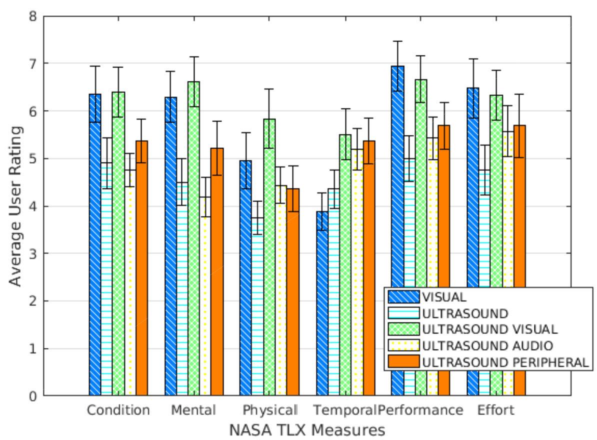


Figure 6. Results of the NASA TLX questionnaire. (V: visual; U: ultrasound; UV: ultrasound-visual; UA: ultrasound-audio; UP: ultrasound-peripheral.)

Preferences

Each participant ranked the feedback types from most to least preferred (Figure 7). Analysis of our questionnaire showed that 47.06% of participants preferred Ultrasound-Audio feedback, followed by Ultrasound-Peripheral, Ultrasound-Visual, and Ultrasound feedback. Visual feedback was ranked as least preferred feedback type by 41.17% of the participants.

DISCUSSION

In this paper, we investigated the effects of multimodal feedback for mid-air gesture interaction on the primary driving task and the secondary gesturing task — with focus on ultrasound haptics. Our results suggest that providing multimodal ultrasound feedback is promising since it reduces eyes-off-the-road time significantly compared to visual feedback. In this section, we will discuss our findings in regards to our hypotheses followed by a discussion of gesturing performance and user preferences.

Hypothesis 1: is rejected since our results show that multimodal ultrasound feedback resulted in less time looking away

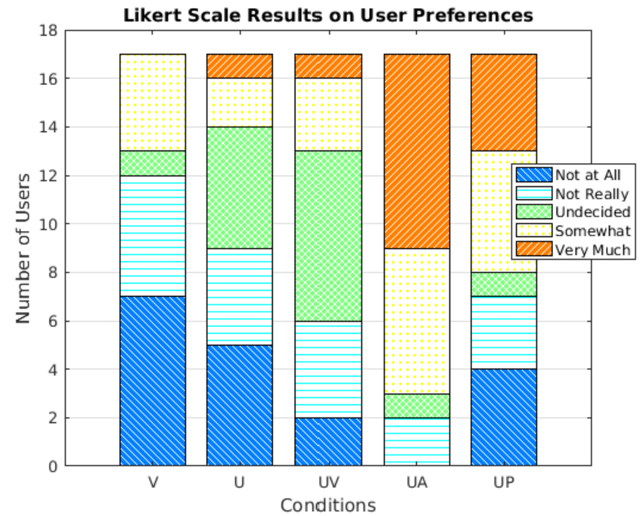


Figure 7. Results of User Preferences of Conditions.

from the road compared to Visual and Ultrasound-Visual feedback (Figure 3); however, not significantly less than unimodal Ultrasound feedback. If Ultrasound is combined with Visual feedback to bimodal Ultrasound-Visual then looking away time is significantly reduced. This suggests that multimodal Ultrasound feedback is more appropriate for in-car usage compared to unimodal Visual feedback.

Hypothesis 2: is rejected since feedback type had no significant effect on lane deviation. Shakeri *et al.* [41] assumed that a more challenging primary task, such as lane change, could influence the driving performance during gesture execution. However, we found this not to be true in our study. This might be due to:

1. the secondary task duration being too short (average glance duration was less than 2000ms). Research has shown that if drivers' glances off the road are shorter than 2 seconds, it has no significant effect on lane deviation [25, 41];

2. over time participants found the optimal steering wheel position for least lane deviation.

Hypothesis H3: is rejected since we found significant differences between secondary task performance depending on feedback condition. Gestures were performed best during the multimodal conditions and during the uni-modal Visual condition (Table 2). We believe the high performance rate during the Visual condition is due to participants being familiar with visual feedback in general. The good performance during the multimodal blocks is due to the presentation of redundant information which improves perception of information in mentally demanding situations [22, 37]. Gesture performance during Ultrasound-Audio was highest, which is in accordance with the literature suggesting that auditory feedback is a suitable alternative for visual feedback when presented to coarse mid-air gestures [41]. Unimodal ultrasound feedback resulted in worst gesture performance of only 21.11% (discussion to why under Hypothesis 4).

Hypothesis 4: is rejected since we did not find a significant difference in mental demand between the unimodal and multimodal conditions. Despite the fact that uni-modal ultrasound feedback was one of the feedback conditions which caused least mental and temporal demand, least frustration, and generally least effort; quantitative analysis of gesture execution shows that performance was worst (more than 30% worse than Ultrasound-Audio and Visual feedback). We think that participants did not distinguish any feedback patterns provided. For example, following scenario is likely to have happened: the participant was instructed to swipe left three times but after the first swipe motion the hand "exited" the interaction box area; on second entrance of the hand into the interaction box the system provided a short pulse to the palm; the participant however expected a swipe left motion across the palm but accepted the "entrance" feedback as swipe left feedback. Participant P1 commented this process with "I was never really sure if it was feedback or trying to tell me it is ready". This suggests that the mental effort in distinguishing the haptic messages on the palm were so great — contrary to the NASA TLX results — that participants eventually did not bother to tell them apart. P16 said "hart to identify if move was correct but except that relax drive with less pressur[e]" which supports our hypothesis. Participants accepted any ultrasound feedback at a certain point without caring whether it was the correct one for the executed gesture. This means, ultrasound stimuli passed the point from being mentally demanding but distinguishable to mentally so demanding that discrimination of stimuli was impossible as secondary task. This also correlates with the subjective feedback in which P12 commented on the ultrasound feedback "couldn't always feel the feedback and struggled to tell the difference between the different types of feedback for each motion". That explains why there is no difference between Visual and Ultrasound-Visual feedback regarding mental demand. The additional ultrasound modality did not contribute enough to lower mental demand.

However, when ultrasound was presented multimodally, e.g. Ultrasound-Visual, the gesturing performance remained stable compared to Visual feedback and EORT decreased significantly. These findings suggest that ultrasound is useful in a multimodal application for mid-air gesture interaction. If used uni-modally it is more useful in a binary scenario.

Secondary Task Analysis: Further analysis revealed a significant difference in secondary task performance across gestures. The V gesture yielded the highest performance accuracy with 99.62%. This might be a result of the V gesture consisting of a single discrete and static motion. Other gestures consisted of two or more motions (e.g. SL2, CW2). Swipe motions performed worst, especially SR. CW motions performed better than CAW gestures. SR and CW gestures being motions where the arm is moving away from the torso of the driver and this "away" movement might have caused greater arm and shoulder fatigue [21]. The difference between the circular motion and the swipe motion was the nature of their continuity. A CW2 motion is one continuously performed gesture. With SL2, the user has to return the hand to the start point and swipe again. This interruption of rhythm — the new alignment of the hand inside the interaction box — might have caused the different performance rates between the gestures. Further, "resetting" the swiping motion — returning the hand to the starting point to swipe again — might have caused misclassification of the intent. The user wanted to swipe right again, thus brought the hand back to the left and this "resetting" was interpreted by the Leap Motion device as a left swipe. This can be confirmed by participant feedback "system takes gestures to[o] quick if you enter the space. feeling of ultrasound hard to identify the difference". The nature of the swipe motion is not suitable for a limited space such as the car cockpit where the driver cannot "reset" the hand outside of the interaction box but has to do it within causing unwanted reverse actions. With the Victory gesture on the other hand — it being static and discrete, users expected an initial feedback on entrance of the Leap Motion sensing area and then a second feedback for gesture confirmation. This resulted in high performance. Swipe and circle gestures required repeated execution for at least two motions to successfully complete the instructed gesture. Once gesture intent and gesture recognition capabilities of the sensing devices will have significantly improved the results will be less frustration and better performance.

Finally, the overall low success rate (43.74%) of correct gesture execution led us to believe that our participants might not have noticed or cared enough to match the gesture execution to the instructed commands. We think this is due to the lack of specific consequences of gesturing — i.e. users did not make any actual selections (e.g. select target three) but followed instructions (e.g. swipe left three times). This might have resulted in them believing that any recognised gesture suffices (e.g. 2 left swipes and 1 right swipe). In the beginning of the experiment we instructed our participants to think of the given task as a selection task but see now that this was not enough. We will account for this in future experiments.

User Preferences: Ultrasound was ranked second least preferred. Ultrasound-Audio was ranked most preferred followed by Ultrasound-Peripheral.

Limitations: We acknowledge that our participants where relatively young which might have had a significant impact on adaptation for gesturing in air and ultrasound feedback as a new interaction technique.

CONCLUSION AND OUTLOOK

This paper contributes multimodal ultrasound feedback techniques for mid-air gesture interaction in driving situations. If used unimodally ultrasound haptics are a useful feedback method for binary mid-air gesture information; these can be a) confirmation that the hand entered the gesture sensing area or b) the system is paying attention to the user and ready to receive input. If used multimodally it significantly reduces eyes-off-the-road time compared to Visual feedback without compromising driving performance nor mental demand and can therefore reduce crash risks.

ACKNOWLEDGEMENTS

This research is funded by the European Union Horizon 2020 research and innovation programme HAPPINESS project (645145). Many thanks to Andrii Matviienko for his relentless feedback.

REFERENCES

1. 2017. BMW's 7-Series 'gesture controls' work pretty well. (2017). https://eu.usatoday.com/story/money/cars/2016/05/16/bmws-7-series-gesture-controls-work-pretty-well/32613369/
2. 2017. Explore Golf R : Volkswagen UK. (2017). http://www.volkswagen.co.uk/new/golf-vii-pa/explore/r
3. 2017. Look out for Ultrahaptics haptic feedback in new cars this year | TechCrunch. (2017).
4. 2018. Car functions now controlled by waving a hand. (2018). https://eu.usatoday.com/story/money/cars/2013/01/10/cartech-gestures-ces/1820453/
5. 2018. Ultrahaptics - Discover a new type of haptics. (2018). https://www.ultrahaptics.com/
6. Micah Alpern and Katie Minardo. 2003. Developing a car gesture interface for use as a secondary task. *CHI EA* (2003), 932. DOI: http://dx.doi.org/10.1145/766077.766078
7. Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: Multi-Point Mid-Air Haptic Feedback for Touch Surfaces. *UIST* (2013), 505–514. DOI: http://dx.doi.org/10.1145/2501988.2502018
8. Patricia Ivette Cornelio Martinez, Silvana De Pirro, Chi Thanh Vi, and Sriram Subramanian. 2017. Agency in Mid-air Interfaces. In *CHI*. DOI: http://dx.doi.org/10.1145/3025453.3025457
9. Nelson Cowan. 2001. The magical number 4 in short term memory. A reconsideration of storage capacity. In *Behavioral and Brain Sciences*, Vol. 24. 87–186. DOI: http://dx.doi.org/10.1017/S0140525X01003922
10. Birsen Donmez, Linda Ng Boyle, and John D Lee. 2010. Differences in Off-Road Glances: Effects on Young Drivers' Performance. *Journal of Transportation and Engineering* 136, 5 (2010), 403–409. DOI: http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000068
11. Capt Eileen Ancman. 1991. Peripherally Located CRTs: Color Perception Limitations. In *NAECON*. DOI: http://dx.doi.org/10.1109/NAECON.1991.165871
12. Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2014a. Illuminating Gesture Interfaces with Interactive Light Feedback Abstract. *NordiCHI* (2014). DOI: http://dx.doi.org/10.1007/978-3-319-22723-8{_}42
13. Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2014b. Tactile Feedback for Above-Device Gesture Interfaces: Adding Touch to Touchless Interactions. *ICMI* (2014), to appear. DOI: http://dx.doi.org/10.1145/2663204.2663280
14. Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2015. Interactive light feedback: Illuminating Above-Device gesture interfaces. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9299. 478–481. DOI: http://dx.doi.org/10.1007/978-3-319-22723-8{_}42
15. Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2016. Do That , There : An Interaction Technique for Addressing In-Air Gesture Systems. In *CHI*. ACM, 2319–2331. DOI: http://dx.doi.org/10.1145/2858036.2858308
16. Thomas M Gable, R May May, and Bruce N Walker. 2014. Applying Popular Usability Heuristics to Gesture Interaction in the Vehicle. In *AutoUI*. 1–7. DOI: http://dx.doi.org/10.1145/2667239.2667298
17. Orestis Georgiou. 2017. Haptic In-Vehicle Gesture Controls. In *AutoUI*. Oldenburg, Germany. DOI: http://dx.doi.org/10.1145/3131726.3132045
18. Paul Green. 1999. The 15-second rule for driver information systems. *Proc of the ITS America* Cd (1999), 1–9. http://www.umich.edu/~driving/publications/ITSA-Green1999.pdf
19. Paul Green. 2000. Crashes Induced by Driver Information Systems and What Can Be Done to Reduce Them. *Society of Automotive Engineers Conference Proceedings* (2000), 27–36.
20. Paul Green. 2004. Driver distraction, telematics design, and workload managers : safety issues and solutions. *Proc. Int. Congr. Transp. Electron* (2004), 165 –180. DOI: http://dx.doi.org/10.1167/3.9.157
21. Juan David Hincapié-ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance : A Metric to Quantify Arm Fatigue of Mid - Air Interactions. (2014), 1063–1072.
22. Cristy Ho, Hong Z. Tan, and Charles Spence. 2005. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 6 (2005), 397–412. DOI: http://dx.doi.org/10.1016/j.trf.2005.05.002
23. John Jonides. 1981. Voluntary Versus Automatic Control Over the Mind'S Eye'S Movement. *Attention and performance IX* 9 (1981), 187–203.
24. Raine Kajastila and Tapio Lokki. 2013. Eyes-free interaction with free-hand gestures and auditory menus. *International Journal of Human Computer Studies* 71, 5 (2013), 627–640. DOI: http://dx.doi.org/10.1016/j.ijhcs.2012.11.003
25. S.G. Klauer, T. A. Dingus, V. L. Neale, J.D. Sudweeks, and D.J. Ramsey. 2006. The Impact of Driver Inattention On Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. *National Highway Traffic Safety Administration* (2006), 226. http://hdl.handle.net/10919/55090
26. Ju-Hwan Lee and Charles Spence. 2008. Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. In *British HCI*.

27. Andreas Löcken, Wilko Heuten, and Susanne Boll. 2016. AutoAmbiCar: Using Ambient Light to Inform Drivers About Intentions of Their Automated Cars Motivation. In *AutoUI*. DOI:http://dx.doi.org/10.1145/3004323.3004329

28. Andreas Löcken, Heiko Müller, Wilko Heuten, and Susanne Boll. 2013. AmbiCar: Towards an in-vehicle ambient light display. In *AutoUI*.

29. Andreas Locken, Heiko Muller, Wilko Heuten, and Susanne Boll. 2015. An experiment on ambient light patterns to support lane change decisions. In *IEEE Intelligent Vehicles Symposium*. DOI: http://dx.doi.org/10.1109/IVS.2015.7225735

30. Sebastian Loehmann, Martin Knobel, Melanie Lamara, and Andreas Butz. 2013. Culturally independent gestures for in-car interactions. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8119 LNCS, PART 3 (2013), 538–545. DOI: http://dx.doi.org/10.1007/978-3-642-40477-1{_}34

31. Benjamin Long, Sue Ann Seah, Tom Carter, and Sriram Subramanian. 2014. Rendering volumetric haptic shapes in mid-air using ultrasound. *ACM Transactions on Graphics* (2014). DOI: http://dx.doi.org/10.1145/2661229.2661257

32. Keenan R May, Thomas M Gable, and Bruce N Walker. 2014. A Multimodal Air Gesture Interface for In Vehicle Menu Navigation. In *AutoUI*. 1–6. DOI: http://dx.doi.org/10.1145/2667239.2667280

33. S Morrison-Smith and J Ruiz. 2014. Using Audio Cues to Support Motion Gesture Interaction on Mobile Devices. *CHI'14 Extended Abstracts on Human Factors . . .* (2014), 1621–1626. DOI: http://dx.doi.org/10.1145/2559206.2581236

34. Mark I. Nikolic and Nadine B. Sarter. 2001. Peripheral visual feedback: a powerful means of supporting effective attention allocation in event-driven, data-rich environments. *Human factors* 43, 1 (2001), 30–38. DOI: http://dx.doi.org/10.1518/001872001775992525

35. Yiyun Peng, Linda Ng Boyle, and Shauna L. Hallmark. 2013. Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. In *Accident Analysis and Prevention*. DOI: http://dx.doi.org/10.1016/j.aap.2012.06.013

36. Matthew J. Pitts, Lee Skrypchuk, Tom Wellings, Alex Attridge, and Mark a. Williams. 2012. Evaluating user response to in-car haptic feedback touchscreens using the lane change test. *Advances in Human-Computer Interaction* 2012 (2012). DOI: http://dx.doi.org/10.1155/2012/598739

37. Ioannis Politis. The Effects of Modality, Urgency and Message Content on Responses to Multimodal Driver Displays. In *AutomotiveUI 2014 Adjunct Proceedings*. Seattle, WA, USA, 1–5. DOI: http://dx.doi.org/10.13140/2.1.4592.3842

38. Bryan Reimer, Bruce Mehler, J. Dobres, and J.F. Coughlin. 2013. *The Effects of a Production Level "Voice - Command" Interface on Driver Behavior : Summary Findings on Reported Workload , Physiology , Visual Attention , and Driving Performance*. Technical Report 17A. Massachusetts Institute of Technology, Cambridge, MA.

39. Andreas Riener, Alois Ferscha, Florian Bachmair, Patrick Hagmüller, Alexander Lemme, Dominik Muttenthaler, David Pühringer, Harald Rogner, Adrian Tappe, and Florian Weger. 2013. Standardization of the in-car gesture interaction space. *AutoUI* (2013), 14–21. DOI: http://dx.doi.org/10.1145/2516540.2516544

40. Sonja Rümelin, Thomas Gabler, and Jesper Bellenbaum. 2017. Clicks are in the Air: How to Support the Interaction with Floating Objects through Ultrasonic Feedback. In *AutoUI*. DOI: http://dx.doi.org/10.1145/3122986.3123010

41. Gözel Shakeri, John H Williamson, and Stephen Brewster. 2017. Novel Multimodal Feedback Techniques for In-Car Mid-Air Gesture Interaction. In *AutoUI*. DOI: http://dx.doi.org/10.1145/3122986.3123011

42. Jason Sterkenburg, Joshua Johnson, Steven Landry, and Myounghoon Jeon. 2016a. Development Tool for Rapid Evaluation of Eyes-free In-vehicle Gesture Controls. In *AutoUI*. DOI:http://dx.doi.org/10.1145/3004323.3004357

43. Jason Sterkenburg, Steven Landry, Myounghoon Jeon, and Joshua Johnson. 2016b. Towards an in-vehicle sonically-enhanced gesture control interface: a pilot study. In *ICAD*. 0–3. DOI: http://dx.doi.org/10.21785/icad2016.015

44. Jan Theeuwes. 1991. Exogenous and endogenous control of attention: the effect of visual onsets and offsets. *Perception & Psychophysics* 49, 1 (1991), 83–90. DOI: http://dx.doi.org/10.3758/BF03211619

45. Jan B F Van Erp and Hendrik a H C Van Veen. 2004. Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behaviour* 7, 4-5 (2004), 247–256. DOI: http://dx.doi.org/10.1016/j.trf.2004.09.003

46. Hanneke Hooft van Huysduynen, Jacques Terken, Alexander Meschtscherjakov, Berry Eggen, and Manfred Tscheligi. 2017. Ambient Light and its Influence on Driving Experience. In *AutoUI*. DOI: http://dx.doi.org/10.1145/3122986.3122992

47. Yu Zhang and Linda Angell. 2014. Pointing Towards Future Automotive HMIs: The Potential for Gesture Interaction. *AutoUI* 22, 3 (2014), 22–29. DOI: http://dx.doi.org/10.1016/j.apergo.2013.10.013

48. Ilka Zöller, Roman Bechmann, and Bettina Abendroth. 2017. Possible applications for gestures while driving. *Automotive and Engine Technology* (2017). DOI: http://dx.doi.org/10.1007/s41104-017-0023-7