Novel In-Vehicle Gesture Interactions: Design and Evaluation of Auditory Displays and Menu Generation Interfaces

Moustafa Tabbarah
tabbarah@vt.edu
Mind Music Machine Lab, Virginia Tech
Blacksburg, Virginia, USA

Yusheng Cao
cyschris@vt.edu
Mind Music Machine Lab, Virginia Tech
Blacksburg, Virginia, USA

Ahmad Abu Shamat
ahmadabushamat@vt.edu
Mind Music Machine Lab, Virginia Tech
Blacksburg, Virginia, USA

Ziming Fang
joannafg@vt.edu
Mind Music Machine Lab, Virginia Tech
Blacksburg, Virginia, USA

Lingyu Li
lingyu99@vt.edu
Mind Music Machine Lab, Virginia Tech
Blacksburg, Virginia, USA

Myounghoon Jeon
myounghoonjeon@vt.edu
Mind Music Machine Lab, Virginia Tech
Blacksburg, Virginia, USA

ABSTRACT

Using in-vehicle infotainment systems degrades driving performance and increases crash risk. To address this, we developed air gesture interfaces using various auditory displays. Thirty-two participants drove a simulator with air-gesture menu navigation tasks. A 4x2 mixed-model design was used to explore the effects of auditory displays as a within-subjects variable (earcons, auditory icons, spearcons, and no-sound) and menu-generation interfaces as a between-subjects variable (fixed and adaptive) on driving performance, secondary task performance, eye glance, and user experience. The adaptive condition centered the menu around the user's hand position at the moment of activation, whereas the fixed condition located the menu always at the same position. Results demonstrated that spearcons provided the least visual distraction, least workload, best system usability and was favored by participants; and that fixed menu generation outperformed adaptive menu generation in driving safety and secondary task performance. Findings will inform design guidelines for in-vehicle air-gesture interaction systems.

CCS CONCEPTS

• **Human-centered computing → Gestural input**; **User interface design**.

KEYWORDS

HCI, Driving distraction, In-air gesture controls, Auditory display, Adaptive user interface, Spearcon

ACM Reference Format:
Moustafa Tabbarah, Yusheng Cao, Ahmad Abu Shamat, Ziming Fang, Lingyu Li, and Myounghoon Jeon. 2023. Novel In-Vehicle Gesture Interactions: Design and Evaluation of Auditory Displays and Menu Generation Interfaces. In *15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '23), September 18–22, 2023, Ingolstadt, Germany*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3580585.3607164

AutomotiveUI '23, September 18–22, 2023, Ingolstadt, Germany

ACM ISBN 979-8-4007-0105-4/23/09.
https://doi.org/10.1145/3580585.3607164

1 INTRODUCTION

Driver distraction is a major cause of car crashes and the use of in-vehicle infotainment systems (IVIS) exacerbates crash risk with visual, manual, and cognitive distractions [1]. To reduce this risk, researchers have explored air gesture interactions as an alternative to touchscreens [36, 40, 41]. Gesture-based interfaces offer advantages, such as minimal visual attention and lower driver workload, by decreasing off-road glances [21, 48]. Multiple Resources Theory (MRT) states that dual-task performance is degraded when the same resource is needed for both tasks [53]. MRT suggests that transmitting secondary task information using non-visual modalities can minimize the competition for visual attention while driving and thus, can lower crash rates [51]. In this line, auditory displays have shown to decrease visual distraction [22, 34, 45]. Previous studies have shown that a multimodal display combining both gesture-based interaction and auditory feedback can be optimal in reducing driver distractions [43, 48]. However, careful consideration should be given when choosing auditory feedback in cars to avoid mental demand and long glances away from the road. To best utilize their potentials, auditory displays must be made sufficiently simple and present accurate feedback to reduce the driver's cognitive demand [10]. Accordingly, we built an air-gesture menu navigation system (see Figure 1) and tested with auditory displays, including earcons, auditory icons, and spearcons. On the other hand, the driving task requires manual control, which still overlaps with gesture controls. To explore the better option of physical location of the menu, we developed two menu generation interfaces: in an adaptive menu generation interface (AMGI), the menu is adjusted based on the driver's hand position, whereas in a fixed menu generation interface (FMGI), the menu is fixed in a particular space above the hand tracker. Our study aims to enhance driving safety, usability, and workload in gesture navigation by investigating the effectiveness of different auditory displays and menu navigation prototypes.

2 RELATED WORK

2.1 Auditory supported air gestures in vehicles

Research on in-vehicle gesture applications has been prolific during the past decade. Efforts evaluated the usability of gesture systems compared to traditional touchscreen interactions, e.g., [21, 29, 36, 48, 54] and the feedback modalities, e.g., auditory [36, 43–45, 48, 49], tactile [43, 44], or peripheral visual [43, 44]. May et al. [37] evaluated a simple one-dimensional menu system comparing air gesture systems with and without speech sound, and a touch system. They found that driving performance using air gesture interfaces was significantly worse than the baseline with no secondary task, but similar to the touch interface. Even though all visual glance behaviors conformed with guidelines developed by National Highway Traffic Safety Administration (NHTSA), air gesture prototypes resulted in a significantly higher number of off-road glances and increased workload. Sterkenburg et al. [48–50] conducted multiple evaluations to examine the effects of speech displays on in-vehicle air gesture controls for a two-dimensional grid menu. They showed that the auditory and visual feedback lowered the frequency of off-road glances and reduced driver workload. However, the addition of auditory displays did not have any significant impact on lane departures or secondary task performance. All three evaluations [48–50] showed a significant improvement of visual distraction without degrading driving performance, presenting a common inference that prototypes with auditory displays are a clear improvement on prototypes without. In subsequent experiments, Shakeri et al. [44] showed that auditory feedback resulted in better secondary task performance than tactile feedback but worse than visual feedback, and significantly improved time spent looking away from the road. Shakeri et al. [45] then provided conforming results as the bimodal auditory-ultrasound condition provided less time looking away from the road than visual and visual-ultrasound conditions. In both studies, driving performance was similar across all conditions. Although mixed results exist, the potential of auditory displays to improve driving safety when multitasking with an IVIS air-gesture interface is undeniable. Nonetheless, Sterkenburg et al. [48] argue that trust in auditory displays alone is still lacking and suggest providing bimodal auditory and visual feedback, with auditory information mirroring visual information. While menu navigation using air gestures in vehicles has been either one-dimensional [36] or two-dimensional [48], no three-dimensional menus have been evaluated yet. Thus, there still exists an ambiguity concerning how different auditory displays should be designed and selected to improve driving safety and secondary task performance. The present study targets the existing gap by conducting exploratory research by adding the third dimension to the menu to ultimately provide informed design guidelines.

2.2 Auditory displays

Much research has been conducted on the use of in-vehicle auditory displays, either to support the driving task (e.g., warning signals [19]) or secondary tasks (e.g., navigation of infotainment systems[22, 47]). Researchers classify auditory displays as two labels: non-speech displays such as earcons and auditory icons, and speech sounds. Earcons are non-verbal synthetic sounds that are usually expressed as abstract musical tones or sound patterns. Earcons can be used with structured combinations such as menus [5]. Earcons' recognition rate is known as 40% when the number of earcons exceeds seven [49], but [6] proposed that earcons are a more effective means of communication compared to unstructured bursts of sounds. Design guidelines suggested the use of timbre with multiple harmonics over a simple tone to help user perception and avoid masking [6, 7]. Auditory icons [17] are non-verbal brief sounds associated with objects, functions, or actions; they use elements of the analogic sound of the referent [28]. Spearcons ("Speech-based Earcons") [52] are brief auditory cues that are created through a time-compression of speech to generate a faster speech without altering pitch. Spearcons provide a direct mapping to the item they represent. Although spearcons are based on speech, they are classified as non-speech cues. Prior research has been conducted to evaluate learnability and recognition of different auditory displays [18, 42, 50]. Spearcons and TTS were found similarly the easiest to learn and earcons the most difficult when compared to auditory icons [13]. Spearcons were found to have significantly higher recall than music and earcons [50]. Sabic et al. [42] examined the recognition of auditory icons, spearcons at two compression speeds (40% and 60% of original length) and TTS as car warning signals. They showed that auditory icons had lower recognition accuracy than TTS, while spearcons' accuracy was not significantly different from either. Sabic et al. [41] assessed the effectiveness of spearcons, TTS and auditory icons under various background noise conditions while driving in terms of recognition accuracy, reaction time and inverse-efficiency scores. Overall, auditory icons were the least efficient, and spearcons only outperformed TTS in quiet environments without added noise sources such as music or talk-radio. In terms of menu navigation tasks, using spearcons enhanced selection speed and accuracy in two-dimensional menus [51] and in one-dimensional list menus while performing a primary driving task [23]. Additionally, participants preferred auditory cue enhancements such as spearcons, showing improved perceived workload and not diminishing primary task performance [23]. Larsson and Niemand [33] evaluated the effect of earcons, spearcons, and a baseline condition (no auditory display) on the navigation of a list-menu button-controlled display, driving performance and visual distraction. They showed that spearcons reduced total duration and frequency of off-road glances by almost 50% compared to the earcon and the baseline, and improved subjective driver performance ratings. The present experiment is the first to directly compare all auditory conditions mentioned above for a comprehensive evaluation: earcon, auditory icon, spearcon, and no-sound condition.

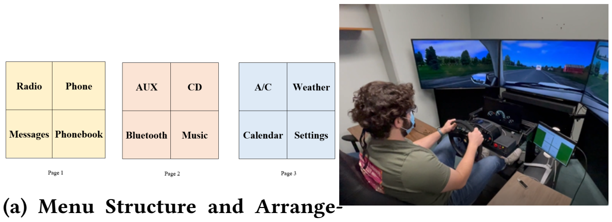
2.3 Novel and Adaptive User Interfaces

As the complexity of user interfaces increases the number of features, a need to manage cluttered interfaces has emerged. There are two general approaches to personalize interfaces and achieve targeted delivery: adaptive and adaptable interfaces. An adaptive system automatically adjusts to support the user, whereas an adaptable system gives the user control over customization mechanisms [36]. Adaptive user interfaces (AUIs) generally achieve adaptation through spatial techniques, graphical techniques, or a combination of both. A common example of spatial adaptation is the reorganization of visual menu items based on frequency of use to reduce navigation time [38]. A post-WIMP (window, icon, menu, pointing device) example that utilizes the adaptive system is marking menus. Marking menus display their menu items around an activation point, and a menu selection is made by performing a directional movement towards a target menu item [40]. Applications of marking menus include air-gesture menu selection for the visually impaired [29] and touchscreen in-vehicle infotainment systems (IVIS) in a driving context [14]. Dim et al. [29] showed the potential of eyes-free air gesture interactions with non-fixed menus whose center point is defined by the user's hand position at the beginning of each selection task. In a simulated driving environment, Ecker et al. [14] evaluated the effect of the menu prototype on dual-task performance and showed that using their adaptive prototype increased secondary-task performance without reducing driving performance. Kurtenbach and Buxton [12] evaluated marking menus with pen-based surface input and emphasized that making a mark selection requires a physical movement identical to selecting the same menu item from a traditional hierarchical menu. They argued that users rehearsed the physical movement that performed a mark every time a selection was made, and thus, the marking was learned through repetition of the same movement and muscle memory. In VR Lou et al. [34] evaluated an adaptive menu positioning technique, where the adaptive user interface was built in the vicinity of the user's preferred hand. When compared to the static user interface layout, the adaptive user interface resulted in significantly greater interaction efficiency, lower perceived workload, less physical exertion, and greater acceptability by participants in VR environments. Nonetheless, participants in VR environments had no restrictions on visual distraction and were able to freely utilize the visual feedback provided. Although adaptive user interfaces are system-controlled and aim to satisfy the needs of the user, they do not always outperform their non-adaptive alternative. Findlater and McGrenere [15] compared the efficiency of three menu conditions: static, adaptive and adaptable. The adaptive menu resulted in significantly slower selection times compared to the static condition, and significantly slower than the adaptable menu in certain conditions. Cockburn et al. [9] predicted and confirmed that although an adaptive menu gradually enlarges frequent items, it offers little advantage over non-adaptive menus. Kim et al. [31] decomposed the manipulation of three-dimensional menus into two subtasks: "positioning" and "making a command" and evaluated spatial-adaptivity of different menu selection systems for virtual environments. They discovered that fixing a menu's location in space reduces selection task performance. Spatial adaptation has been employed for a menu's placement or positioning. Aforementioned adaptive interfaces [9, 14, 25] provide implicit human-computer interaction where adaptation is based on the analysis of contextual information gathered by the system from frequent use [8, 38]. Also, explicit interactions are based on the direct manipulation of an interface by a user, such as pressing a button or a gesture command [8, 38]. According to Dachselt and Hübner's taxonomy of three-dimensional menus [11], we are manipulating the "Placement" criterion of 3D menus. Our Adaptive Menu Generation Interface's (AMGI) placement in space is based on the user's hand position at activation, while the Fixed Menu Generation Interface (FMGI) has a fixed placement in the vehicle environment. FMGI hence relies on finding the menu in the physical environment–"positioning", while AMGI heavily depends on executing directional movements around the point of activation–"making a command." Adaptive user interfaces utilizing spatial adaptation provided mixed results regarding interaction efficiency compared to fixed alternatives. However, users utilizing either adaptive or fixed interfaces for comparison did not have restriction on or competition for visual attention. Most adaptive menu navigation user interfaces are based on implicit interactions using contextual information. Literature in adaptive air-gesture interfaces is either limited to virtual/augmented reality environments or scarce, with the exception of Babic et al. [2] who presented eyes-free air gesture interaction for target acquisition. However, their evaluation did not extend to driving environments. In the present study we addressed this significant research gap by comparing an adaptive user interface with a fixed user interface to minimize visual distraction and improve driving and menu navigation performance. The concept of adaptivity is vast/ambiguous in the context of dynamic environments. Adaptivity is an umbrella term that comprises all implicit (context-based) and explicit (input-based) interactions. An example of implicit adaptivity is Khan and Khusro's [32] work in which they proposed a context-adaptive user interface which adapts the user interface based on dynamic sensing metrics (i.e., traffic, speed, noise, etc.), and found that their proposed solution significantly minimized driver distraction. As far as our knowledge goes, research in explicit adaptivity in dynamic environments is scarce and limited to XR environments with one exception of Ecker et al. [14]'s work.

3 EXPERIMENT

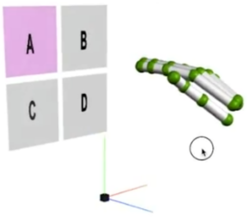
3.1 Menu and interaction design

We developed a 2x2 grid menu selection system (Figure 1) with four square targets: each measuring 10x10 centimeters in the air gesture space, inspired by Sterkenburg et al.'s [48] design that was most efficient compared to a 4x4 grid menu. Be that as it may, real IVIS menus include a larger selection. Basic in-vehicle Android Auto and Apple Carplay displays contain between 8 and 16 menu items, with a median of 12. To expand the number of menu items, we created two additional pages to allow the user to access 3 (pages) * 4 (options) = 12 menu choices; providing a better representation of high-level main menu structures in real in-vehicle displays. To preserve the level of fidelity established, each of the 12 menu items represented an IVIS option present in commercial vehicles: radio, phone, messages, phonebook, AUX, CD, Bluetooth, music, air-conditioning, weather, calendar, and settings. Our gesture menu selection system consists of four gestures mapped to IVIS actions: System Activation (grabbing and releasing fist), Search and Navigation (hovering), Page Switching (swiping), and Selection (grabbing fist). To avoid inadvertent gestures, an activation gesture was introduced. The design employed a swiping motion for page navigation and the swipe right gesture corresponded to a "next page" command. We designed this familiar unidirectional swipe motion to keep the menu system simple and easy to use. The air gesture menu interface consisted of two components: a LEAP Motion controller using an infrared sensor to track and recognize the hand features of the driver, and a visual display comprised of four target boxes arranged in a 2x2 grid (see Figure 1). After performing the activation gesture, the background turned green indicating the system was activated. As the user positions their hand above the LEAP Motion controller and navigates the menu structure, a red highlight of one of the target boxes indicates user's hand position within the 3D menu, accompanied by playing the corresponding auditory display file of that particular menu item. For all conditions, two sound cues were implemented as feedback to inform a successful gesture execution. A "swoosh" sound was provided after a swiping gesture, and a generic digital confirmation sound was provided after a selection was made.



(a) Menu Structure and Arrangement of Menu Items

(b) Experimental Setup



(c) Developer View

Figure 1: Experiment Design

3.2 Auditory Displays

Earcons: We designed earcons to tackle a common usability challenge of both air gesture interfaces and auditory menus: tracking. In both FMGI and AMGI prototypes, a user's hand position is indicative of the relative location within the menu structure of a single page. Although our menu structure contains 3 dimensions (vertical, horizontal, and page), we adhered to design guidelines and manipulated only two sound attributes: timbre and length that varied distinctively. Each **earcon** consisted of four of four consecutive tones and is identified by two sound variables: instrument (page) and tone emphasis (location of the menu item) e.g., first page – piano, first item – Beep Bi Bi Bi; second page – bell, last item – Bi Bi Bi Beep. **Auditory icons**: We mapped each of the twelve menu items to its representative analogic sound acquired from freesound.org, an open-source website, e.g., typing sound for Messages. **Spearcons**: Spearcons were created by using an online text-to-speech (TTS) engine with an American male voice, followed by applying the SOLA algorithm [52] to generate spearcon WAV files.

3.3 Menu Generation Interfaces

Spatial awareness is a significant challenge for air-gesture interfaces, as users often struggle to know where to interact with the interface and perform gestures [4, 16]. To address this issue, we developed the adaptive menu generation interface prototype, which centered the menu around the user's hand position at the moment of activation (Figure 2). This spatial adaptation was achieved by rearranging the position of the entire menu as a single entity. However, the adaptive menu may cause confusion to users because every time the menu location can change, and users need to reconfigurate their behavioral space. In contrast, the fixed menu generation interface located the menu above the LEAP motion at a fixed 10 cm elevation clearance from the sensor to the bottom of the menu, which may develop users' muscle memory. Figure 2 provides a visual representation of the operation of both menus.

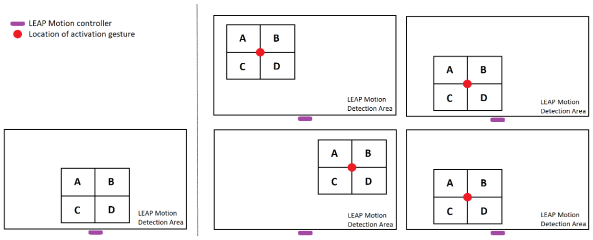


Figure 2: Left: fixed menu generation interface (FMGI), Right: adaptive menu generation interface (AMGI)

3.4 Research Questions

The objective of this experiment is to identify how different attributes of auditory displays and menu generation interfaces affect driving distraction and safety while interacting with in-vehicle information systems (IVISs) using mid-air gesture controls. Driving distraction – a major contributor to crash risk, encompasses visual, manual and cognitive distractions [1]. Eye glance behavior was hence used as a measure of visual distraction, and workload as a measure of both manual and cognitive distraction. Moreover, driving performance is used as an indicator of driving safety, alongside secondary task performance and user experience metrics that reflect usability which is crucial and essential to the operation of a secondary system in a safety-critical environment. Accordingly, three primary research questions (denoted by prefix 'RQ') were formulated. RQ1: How do earcon, auditory icon, and spearcon feedback affect air-gesture IVIS interaction in terms of driving performance, secondary task performance, eye glance behavior, perceived workload, and system usability? RQ2: How do a spatially-adaptive and a fixed air-gesture IVIS compare in their influence on driving performance, secondary task performance, eye glance behavior, perceived workload, and system usability? RQ3: How do different auditory displays and spatial-adaptivity interact together and influence driving performance, secondary task performance, eye glance behavior, perceived workload, and system usability?

3.5 Methods

Participants: A power analysis was conducted to determine the sample size needed to achieve 80% power with a medium effect size; a minimum sample size of 24 was necessary. A total of 33 participants were recruited from the university. One participant was excluded from the study and compensated after showing symptoms of motion sickness. The 32 participants (19 males and 13 females; age: $M = 21.97$, $SD = 2.29$) whose data were included in the analysis came from 12 different countries. All participants had a driver's license for more than a year to avoid novice effects. Each participant's session lasted a maximum of 1 hour and 30 minutes. They received $10/hr compensation. **Experimental Design and Independent Variables**: In this study, Auditory Display was a within-subjects variable and Menu Generating Interface was a between-subjects variable. Each participant experienced four auditory feedback conditions, including earcons, auditory icons, spearcons, and no auditory feedback condition as a baseline. Participants used a gesture system prototype with either fixed menu generation (N=16) or adaptive menu generation (N=16). Adaptive User Interfaces have often been criticized for diminished predictability resulting from changing system behaviors [26, 46]. Höök [24] and Graefe et al. [20] emphasized that predictability is directly correlated with a match between a user's mental model and the system's behavior. We hence found that it would be unfit to introduce two different system behaviors (adaptive and fixed) to the same participant in a single session, and thus, opted to have Menu Generation Interface as a between-subjects variable. **Apparatus**: The study used a NADS MiniSim driving simulator for six-minute car-following scenarios on a suburban route with low to moderate traffic. Participants were instructed to maintain a safe distance while following the lead vehicle. Tobii Pro Glasses 2 eye-tracking device (sampling rate of 50 Hz) was used to capture participants' glance behavior during the study. **Procedure**: After the consent procedure approved by the university's Institutional Review Board (IRB), participants were briefed and completed a short training scenario to test for simulation sickness [31]. They then watched a tutorial on the menu-gesture system and practiced as needed. Before each of the four driving scenarios with different auditory feedback, participants were introduced to the auditory display and given time to become familiar with it. Participants completed 12 trials of a secondary menu navigation task with a pre-recorded voice prompt (e.g., "Find calendar") during each scenario. After each scenario, participants completed a workload assessment and the System Usability Scale questionnaire, then indicated their preferred auditory display condition. The order of auditory conditions was counterbalanced using the balanced Latin square design to minimize order effects. Finally, they were asked about their preferred auditory condition.

4 RESULT

For all data, we conducted a 4 (auditory display conditions) x 2 (menu generation types) mixed-model (or repeated measures) analysis of variance (ANOVA). To ensure the reliability of the obtained results, we checked for the parametric assumptions of the mixed-model ANOVA: normality of residuals, homogeneity of variances, and sphericity for the within-subject variable. Table 1 summarizes significant results.

Table 1: Summary of ANOVA results

	Menu Generation Interface		Auditory Display Condition	
Measurement	**F-value**	**p-value**	**F-value**	**p-value**
Driving Performance				
Lane Deviation	8.16	0.0077*	0.113	0.952
Steering Wheel Angle	9.34	0.0047*	0.625	0.601
Vehicle Speed	8.63	0.0063*	0.843	0.472
Secondary Task Performance				
Selection Accuracy	11.8	0.0018*	0.678	0.568
Eye Glance Behavior				
Short Glance Frequency**	0.0004	0.984	8.57	<0.0001*
Medium Glance Frequency**	2.41	0.131	2.02	0.121
Long Glance Frequency**	1.82	0.187	4.83	0.0041*
Dwell Time**	1.04	0.316	11.0	<0.0001*
Perceived Workload				
Mental Demand	1.13	0.296	3.84	0.0125*
Performance	0.0002	0.989	3.55	0.0177*
Effort	0.950	0.362	4.46	0.0058*
Frustration	0.0003	0.987	2.74	0.0481*
Total Workload	0.585	0.451	3.71	0.0147*

* indicates significant results (p <0.05), ** indicates significant interaction effect

4.1 Driving Performance

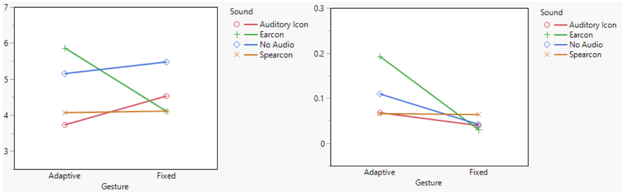
Lane Deviation: ANOVA showed significantly lower variability of lane deviation for FMGI ($M = 0.975$, $SD = 0.167$) than AMGI ($M = 1.160$, $SD = 0.223$). There was no main effect of auditory display or interaction effect. **Steering Wheel Angle**: ANOVA showed significantly lower variability of steering wheel angle for FMGI ($M = 2.78$, $SD = 0.45$) than AMGI (M=3.32, $SD = 0.63$). There was no main effect of auditory display or interaction effect. **Vehicle Speed**: ANOVA showed significantly lower variability of vehicle speed for FMGI ($M = 4.07$, $SD = 0.60$) than AMGI ($M = 4.58$, $SD = 0.51$). There was no main effect of auditory display, or interaction effect. **Following Distance**: There was no main effect or interaction effect in following distance.

4.2 Secondary task performance

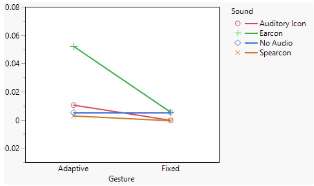
Selection Accuracy: An exponential transformation was applied on selection accuracy data to satisfy parametric assumptions. Drivers using the FMGI ($M = 83.85\%$, $SD = 14.00\%$) achieved a significantly higher selection accuracy than drivers using the AMGI ($M = 72.36\%$, $SD = 15.87\%$). There was no main effect of auditory display or interaction effect. There was no main effect or interaction effect in Selection Time.

4.3 Eye glance behavior

Short glance frequency: A square root transformation was applied to short glance frequency data to satisfy parametric assumptions. ANOVA showed significantly different frequency of short glances for auditory display. There was no main effect of menu generation interface. But there was an interaction effect between auditory display and menu generation interface, $F(3, 1377) = 5.67$, $p = 0.0007$, $\eta_p^2 = 0.012$. Paired samples t-tests with a Bonferroni corrected alpha value of 0.0083 performed between auditory display conditions revealed that spearcons ($M = 4.01$, $SD = 4.18$) resulted in significantly lower frequency of short glances than the no auditory display condition ($M = 5.31$, $SD = 4.90$), $t\ (67.98) = 4.02$, $p = 0.0002$. Auditory icons ($M = 4.02$, $SD = 4.05$) also had significantly lower frequency of short glances than the no auditory display condition, $t\ (67.98) = 4.45$, $p < 0.0001$ (Figure 3). **Medium glance frequency**:

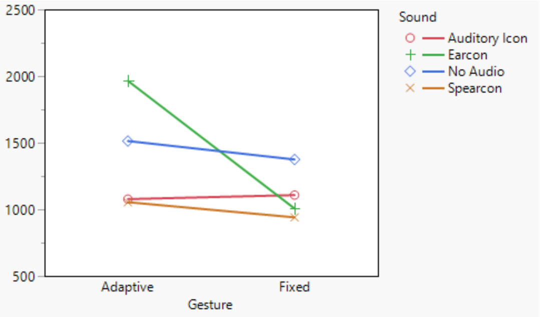


(a) Short Glace Frequency (b) Medium Glace Frequency

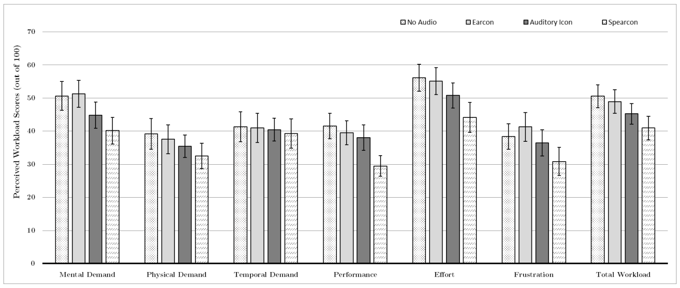


(c) Long Glace Frequency

Figure 3: Interaction effect on short (a), medium (b), and long (c) glance frequency(glace/selection)



(a) Interaction effect on dwell time (milliseconds)



(b) NASA-TLX[23] subscales and total workload scores

Figure 4: Reuslts

ANOVA showed that there was no significant main effect in the frequency of medium glances for auditory display or menu generation interface. However, there was an interaction effect between auditory display and menu generation interface, $F(3, 1377) = 5.1456$, $p = 0.0015$, $\eta_p^2 = 0.010$. **Long glance frequency**: ANOVA showed significantly different frequency of long glances for auditory display. There was no main effect of menu generation interface. But there was an interaction effect between auditory display and menu generation interface, $F(3, 1377) = 3.4404$, $p = 0.0163$, $\eta_p^2 = 0.007$. **Dwell time**: A square root transformation has been applied to dwell time data to satisfy parametric assumptions. ANOVA showed significantly different dwell times for auditory display. There was no main effect of menu generation interface. But there was an interaction effect between auditory display and menu generation interface, $F(3, 1377) = 7.81$, $p < 0.0001$, $\eta_p^2 = 0.022$. Paired samples t-tests with a Bonferroni correction revealed that spearcons (M = 1.010s, SD = 1.249s) resulted in significantly shorter dwell time than both earcons (M = 1.529s, SD = 2.160), $t\,(66.98) = 3.618$, $p = 0.0006$, and the no auditory display condition (M = 1.444s, SD = 1.519s), $t\,(66.98) = 34.808$, $p < 0.0001$. Auditory icons (M = 1.065s, SD = 1.199s) also had significantly shorter dwell times than earcons, $t\,(66.98) = 3.136$, $p = 0.0025$, and the no auditory display condition, $t\,(66.98) = 4.36$, $p < 0.0001$ (Figure 4(a)).

4.4 Workload

ANOVA showed significantly different **Mental Demand** scores for auditory display. There was no main effect of menu generation interface or interaction effect. Paired samples t-tests with a Bonferroni correction revealed that spearcons ($M = 40.17$, $SD = 21.87$) resulted in significantly lower mental demand than the no auditory display condition ($M = 50.65$, $SD = 24.04$), $t\,(29) = 3.072$, $p = 0.0046$ (Figure 4(b)). ANOVA showed significantly different **Performance** scores for auditory display. There was no main effect of menu generation interface or interaction effect. Paired samples t-tests with a Bonferroni correction revealed that spearcons ($M = 29.5$, $SD = 17.09$) resulted in significantly higher performance scores than earcons ($M = 39.52$, $SD = 20.06$), $t\,(29) = 2.84$, $p = 0.0082$, and the no auditory display condition ($M = 41.61$, $SD = 21.42$), $t\,(29) = 3.56$, $p = 0.0013$. ANOVA showed significantly different **Effort** scores for auditory display. There was no main effect of menu generation interface or interaction effect. Paired samples t-tests with a Bonferroni correction revealed that spearcons ($M = 44.17$, $SD = 24.81$) resulted in significantly lower effort scores than the no auditory display condition ($M = 56.13$, $SD = 22.68$), $t\,(29) = 3.309$, p=0.0025. ANOVA showed significantly different **Frustration** scores for auditory display. There was no main effect of menu generation interface or interaction effect. Paired samples t-tests with a Bonferroni correction revealed that spearcons ($M = 30.39$, $SD = 23.23$) resulted in significantly lower frustration than earcons ($M = 41.29$, $SD = 24.36$), $t\,(29) = 3.01$, $p = 0.0054$. ANOVA showed significantly different Total Workload for auditory display. There was no main effect of menu generation interface or interaction effect. Paired samples t-tests with a Bonferroni correction revealed that spearcons ($M = 40.95$, $SD = 19.39$) resulted in significantly lower overall workload than the no auditory display condition ($M = 50.56$, $SD = 19.39$), $t\,(29) = 3.38$, p=0.0021. There were no main effects or interaction effects in Physical Demand and Temporal Demand.

4.5 Usability and Preference

Spearcons showed the highest SUS score in both menu generation conditions: Spearcons (81), Auditory icons (75), No Audio (67.8), and Earcons (62.8) in AMGI; Spearcons (76.3), Auditory icons (70), No Audio (69.8), and Earcons (59.8) in FMGI. At the end of the study, participants were asked to rank the auditory displays based on their preference. Spearcons were favored most (21), followed by Auditory Icons (7), Earcons (3), and No Audio (1).

5 DISCUSSION

To explore the effects of auditory displays and menu-generation interfaces in in-vehicle gesture interactions, we conducted a mixed design experiment using the driving simulator. The menu-generation interfaces significantly influenced driving performance and secondary task performance (Fixed outperformed Adaptive menus) (RQ2), whereas auditory displays significantly influenced eye glance behavior, perceived workload, system usability, and preference (Spearcons outperformed other auditory conditions) (RQ1). There were also significant interaction effects in glance frequency and dwell time (Earcons showed significantly higher glance frequencies and dwell time in the adaptive menu system) (RQ3). **Driving Performance**: Results showed that the menu generation interface affected driving performance (longitudinal and lateral vehicle control), but auditory displays had no influence on driving performance. Drivers using AMGI had significantly higher standard deviation of lane deviation, steering wheel angle, and vehicle speed than drivers using FMGI, and numerically greater standard deviation of following distance (by 16.3%.). These results agree with most prior research on auditory-supported air gesture controls in vehicles. Past studies [25, 43, 47, 48] found that the addition of auditory displays did not influence driving performance. The present study corroborated it and auditory displays also did not differ between each other. In terms of spatial-adaptivity, results do not conform with prior research conducted by Ecker et al. [14] who evaluated adaptive menu navigation on a touchscreen IVIS and found no effect on driving performance. The difference in findings can be explained by the inherent differences of the interaction modality. Although Ecker et al. utilized adaptive menu navigation in their study, the location of the touchscreen was still fixed in space while our AMGI was not. Moreover, their touchscreen interface had a haptic feedback at interaction while our AMGI did not. **Secondary Task Performance**: The menu generation interface led to the significant effect. FMGI outperformed AMGI in accuracy by more than 11%. Auditory displays did not affect selection accuracy. In terms of selection time, neither the menu generation interface nor the auditory display condition affected the time it took participants to select. Results are consistent with prior research [48, 49], which found that the speech-based feedback did not provide any advantage on visual-only displays in terms of selection time and accuracy. They also conform with [33], which evaluated auditory displays in a button-controlled list-menu navigation while driving and found no effect of sound on selection time. The present study utilized air gesture controls for multi-paged grid-menu navigation. We can cautiously infer that secondary task performance does not differ between different auditory displays regardless of input control method or menu type. Drivers using FMGI had an 11% higher accuracy in selection tasks than drivers using AMGI. However, spatial-adaptivity did not influence selection time. These results do not conform to the ones found in literature [2, 36]. Both studies, however, evaluated gesture selection as a stand-alone task without multitasking, while the current study evaluated gesture selection as a secondary task concurrent to the primary driving task. Their participants were thus able to allocate all their cognitive resources to the gesture selection task. Moreover, Lou et al.'s [36] participants were able to freely utilize the visual feedback provided as they had neither restrictions in nor competition for visual demand. Another study that found adaptive menu interfaces more efficient than fixed ones was conducted by Ecker et al. [14]. They evaluated an in-vehicle touchscreen interface with either an adaptive pie-menu or a fixed point-touch menu and found that the adaptive alternative resulted in numerically more accurate main menu navigation and significantly faster selection times than the fixed menu. However, the two menu alternatives differed in the modality of the controls and the size of menu items. The adaptive menu selection was done via directional swipe gestures towards bigger size targets, whereas the fixed menu selection task consisted of point-touch on a traditional WIMP-GUI menu. In either menu alternatives, interface interactions occurred in the same visible location – a 7" touchscreen. Their adaptive aspect was the "navigation". Contradictorily, gesture controls and menu navigation were uniform in the current study, while the adaptive aspect was the "location" of interaction. Therefore, participants in our study had to rely on proprioceptive resources and spatial cognition more than Ecker et al.'s participants. We can cautiously infer that the adaptive menu required our participants to recalibrate their behavioral space, whereas the fixed menu allowed them to get familiar with the location as the trials went and the participants might develop their muscle memory. **Eye Glance Behavior**: Results showed that the auditory displays significantly influenced eye glance behavior (short and long glance frequencies, and total dwell time per selection task), but menu generation interface had no effects on eye glance behavior. Spearcons and auditory icons resulted in fewer short glances and shorter total dwell times than earcons and no auditory condition. There was a consistent pattern of interaction effect between the auditory display condition and menu generation interface, with earcons performing drastically worse (higher frequency of glances and longer dwell time) in the AMGI condition compared to the FMGI condition. Those results suggest that the combination of AMGI and earcon feedback has a detrimental effect on eye glance behavior. This pattern of the results might come from different characteristics of auditory displays. Spearcons and auditory icons are categorized as the item-level auditory displays (i.e., each sound can represent a specific meaning), whereas earcons are categorized as the structure-level auditory displays (i.e., a group of auditory displays provide where information in the entire menu structure) [28]. Thus, we can cautiously infer that when the menu position keeps updated adaptively, the structure-level auditory displays may not help users locate their hand appropriately. Results concerning spearcons conform with previous research for speech-based auditory displays. Sterkenburg et al. [48–50] found that the addition of speech menus reduced the number of off-road glances. Results of our study also conform with Larsson and Niemand's [35] results from evaluating in-vehicle list-menu navigation using a button. They found that spearcons resulted in a shorter dwell time and a lower frequency of off-road glances than both earcons and the baseline without auditory displays, similar to the results found in the present study. They also found no difference in eye glance behavior between the earcons and the no auditory displays conditions, as was revealed by the results of our study. On the other hand, Shakeri et al. [44, 45] found that earcons improved visual distraction compared to equivalent prototypes without auditory feedback, unlike the results from our study. However, there are major differences between Shakeri et al.'s prototypes and ours. Shakeri et al. evaluated gesture commands with direct mappings while we evaluated menu navigation. Shakeri et al.'s auditory displays were played after the task was done, while our auditory displays were used for navigation before a selection gesture was made. **Workload**: Results showed that auditory displays influenced mental demand, performance, effort, frustration and total workload. Spearcons led to lower mental demand, effort, and total workload than the no auditory display condition. Spearcons led to lower frustration than earcons. Also, spearcons led to higher performance than both the no auditory and earcon conditions. In contrast, the menu generation interface did not influence perceived workload at all. Results conform with the limited workload results from prior research. Literature shows that spearcons decreased perceived workload while navigating a button-controlled list menu in a driving context [27]. Sterkenburg et al. showed that speech-based feedback resulted in lower overall workload [48], and mental demand [49, 50] when compared to visual-only alternatives. The present study showed that spearcons generally outperformed the no auditory display condition in perceived workload measures. On the other hand, Shakeri et al. [45] found that using earcon feedback presented lower physical demand than prototypes with visual feedback. Similarly, the no auditory display condition in the present study resulted in the numerically highest physical demand. Regarding the adaptivity of the user interface, the results do not conform with prior research [36]. They found that their adaptive interface resulted in lower overall workload, mental demand, physical demand, performance, effort and frustration. Results from the current study do not reflect any influence from the menu generation interface. Differences between the two were highlighted under the Secondary Task Performance (stand-alone task without multitasking). A very interesting finding is that there was neither difference in visual glance behavior nor perceived workload between the two menu generation types. Participants using AMGI did neither look away from the road on to the visual display more than participants using FMGI nor did they report a higher level of workload, but results revealed worse driving performance and secondary task performance for AMGI. One possible explanation is that participants were not aware of that limitation on dual-task performance and did not try to mitigate for it by compensating with visual resources. A lack of or decrease in awareness about performance describes a dangerous task situation – unfavored in the context of the safety-critical driving environment. **System Usability Scale (SUS)**[30]: Mean SUS scores[30] indicate that the subjective usability of the air gesture prototype depends on the auditory display. Systems having a SUS score superior to 72.75 are described as "good", and systems scoring above 70 are considered "acceptable [3]. Based on this, the spearcon prototypes had "good" usability, while the earcon and no audio prototypes had "ok" prototypes. The auditory icon prototype paired with AMGI had a "good" usability, but barely missed on the "good" range and resulted in "ok" when paired with FMGI. Nonetheless, spearcon and auditory icon prototypes resulted in "acceptable" usability based on the interpretation of SUS scores [3]. The no audio prototypes were labeled having "high marginal", and earcon prototypes presented "low marginal" usability. It is worth noting that the menu generation interface did not affect perceived usability of the air gesture prototype. **MRT, Modality, and Code**: Throughout this experiment, MRT was utilized to rationalize experimental manipulations and discuss the findings. MRT offers a dichotomous representation of perceptual modalities: visual and auditory. As mentioned above, spearcons and auditory icons were classified as item-level auditory displays, but earcons were classified as a structure-level auditory display. Thus, earcons possessed a spatial component. Based on our findings, we can cautiously infer that earcons interfere with visual-spatial resources. We hence postulate that when it has a spatial component, perceived auditory information goes beyond the dichotomous modality model, which may mean that it interferes with the spatial/verbal code. It challenges the assumption that the separation of information by modality is enough to eliminate conflicts in information processing. Moreover, MRT implies that auditory modality equals the verbal processing code [54]. But our findings show that auditory perception is more complex than how it is described by the MRT (both earcons and auditory icons are not verbal). Therefore, the auditory modality can be further refined into subgroups or subcomponents on a more granular level in MRT.

6 DESIGN RECOMMENDATIONS

This section includes general guidelines that have been derived from the lessons learned from the experiment. **Provide simple auditory displays that do not require additional information processing**. One of the main goals of this research is to reduce driver distraction. This study has shown that a major limitation of auditory displays in menu navigation is the amount of cognitive workload required to process the information. To address this issue, designers should strive to create simple auditory displays that do not require excessive information processing. This will help ensure that drivers are able to use the displays effectively without compromising driving safety. In the context of the prototypes designed for this research, speech-based item-based auditory displays are recommended for a safer driving experience. **Avoid using auditory displays to relay visuospatially-interpreted information**. Much of this research is aimed at understanding how different aspects of auditory displays affect driving safety in general. As shown from the results regarding earcon, auditory displays that relay information that is either visual or spatial imposed high cognitive load on drivers – increasing driving distraction.

7 LIMITATIONS, FUTURE WORK, AND CONCLUSION

This experiment examined how auditory displays and menu generation interfaces impact driving distraction and safety when using mid-air gesture controls for in-vehicle information systems. To continue improving menu prototypes, future research should carefully consider technological innovations. We note the following limitations. This study was conducted using a medium fidelity driving simulator which may be considered a 'static' driving environment. Results may differ in a dynamic context such as in a high-fidelity simulator or a real-driving environment where vehicle conditions (i.e., vibrations) may affect the results. Most research about in-vehicle gesture interactions, including the present study, evaluates air gesture prototypes in Level 0 automation. Nonetheless, most new vehicles nowadays are considered to have Level 2 automation as central driving functions are partially automated (e.g., lane keeping). Accordingly, novel air-gesture IVISs will be implemented in the future into vehicles with at least partial automation, which should encourage researchers to evaluate applications of air-gesture IVISs in automated vehicles. This experiment evaluated high-level menu navigation, equivalent to a main menu in the IVIS. Future research should evaluate deeper menu structures as the increasing volume of menu items and gesture commands would exponentially add the complexity of the air-gesture IVISs which can affect driver distraction differently. The limitations of existing technologies can impede innovation, as seen in our experiment where drivers' lack of knowledge of their hand position and the lack of tactile feedback with air gesture interfaces resulted in increased cognitive load and distraction. However, recent advances in haptic technology have shown promising results in reducing visual distraction and mental workload while operating air-gesture interfaces [39]. For example, the Ultraleap STRATOS™ Explore, which combines Ultrahaptics and Leap motion technologies, offers a potential solution for multimodal air-gesture IVISs and could significantly improve driving safety. In this experiment, the spearcon-enhanced menu resulted in the lowest visual distraction and workload, while the fixed menu interface had the best dual-task performance. Our findings suggest that spearcons (speech-based and item-level auditory display) can effectively reduce visual distraction and workload while maintaining high dual-task performance in driving environments, with potential implications for in-vehicle infotainment system design and auditory display use. Findings offer theoretical insights into auditory cue processing in multitasking scenarios.

REFERENCES

[1] National Highway Traffic Safety Administration. 2020. *Overview of the National Highway Traffic Safety Administration's Driver Distraction Program.*

[2] Teo Babic, Harald Reiterer, and Michael Haller. 2017. GestureDrawer: one-handed interaction technique for spatial user-defined imaginary interfaces. In *Proceedings of the 5th Symposium on Spatial User Interaction.* 128–137.

[3] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[4] Victoria Bellotti, Maribeth Back, W Keith Edwards, Rebecca E Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 415–422.

[5] Meera M Blattner, Denise A Sumikawa, and Robert M Greenberg. 1989. Earcons and icons: Their structure and common design principles. *Human–Computer Interaction* 4, 1 (1989), 11–44.

[6] Stephen A Brewster, Peter C Wright, and Alastair DN Edwards. 1994. A detailed investigation into the effectiveness of earcons. In *SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-*, Vol. 18. Addison-Wesley Publishing CO, 471–471.

[7] Stephen A Brewster, Peter C Wright, and Alistair DN Edwards. 1995. Experimentally derived guidelines for the creation of earcons. In *Adjunct Proceedings of HCI*, Vol. 95. 155–159.

[8] Simon Butscher. 2014. Explicit & Implicit Interaction Design for Multi-Focus Visualizations. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces.* 455–460.

[9] Andy Cockburn, Carl Gutwin, and Saul Greenberg. 2007. A predictive model of menu performance. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 627–636.

[10] Joel M Cooper and Donald L Strayer. 2015. Mental workload of voice interactions with 6 real-world driver interfaces. In *Driving Assessment Conference*, Vol. 8. University of Iowa.

[11] Raimund Dachselt and Anett Hübner. 2007. Three-dimensional menus: A survey and taxonomy. *Computers & Graphics* 31, 1 (2007), 53–65.

[12] Nem Khan Dim, Kibum Kim, and Xiangshi Ren. 2017. An Exploratory Study of Marking Menu Selection by Visually Impaired Participants. In *2017 IEEE International Conference on Smart Computing (SMARTCOMP).* IEEE, 1–7.

[13] Tilman Dingler, Jeffrey Lindsay, Bruce N Walker, et al. 2008. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the 14th International Conference on Auditory Display, Paris, France.* 1–6.

[14] Ronald Ecker, Verena Broy, Andreas Butz, and Alexander De Luca. 2009. pieTouch: a direct touch gesture interface for interacting with in-vehicle information systems. In *Proceedings of the 11th international Conference on Human-Computer interaction with Mobile Devices and Services.* 1–10.

[15] Leah Findlater and Joanna McGrenere. 2004. A comparison of static, adaptive, and adaptable menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 89–96.

[16] Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2016. Do that, there: an interaction technique for addressing in-air gesture systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 2319–2331.

[17] Thomas M Gable and Bruce N Walker. 2013. *Georgia tech simulator sickness screening protocol.* Technical Report. Georgia Institute of Technology.

[18] W GAWER. 1986. Auditory icons: Using sound in computer interfaces. *Human-Computer Interactions* 2 (1986), 167–177.

[19] Gregory Goth. 2011. Brave NUI world. *Commun. ACM* 54, 12 (2011), 14–16.

[20] Julia Graefe, Doreen Engelhardt, and Klaus Bengler. 2021. What does well-designed adaptivity mean for drivers? A research approach to develop recommendations for adaptive in-vehicle user interfaces that are understandable, transparent and controllable. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 43–46.

[21] Robert Graham, SJ Hirst, and C Carter. 1995. Auditory icons for collision-avoidance warnings. In *Intelligent Transportation: Serving the User Through Deployment. Proceedings of the 1995 Annual Meeting of ITS America. ITS America.*

[22] Lisa Graichen, Matthias Graichen, and Josef F Krems. 2019. Evaluation of gesture-based in-vehicle interaction: user experience and the potential to reduce driver distraction. *Human factors* 61, 5 (2019), 774–792.

[23] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology.* Vol. 52. Elsevier, 139–183.

[24] Kristina Höök. 1998. Tutorial 2: Designing and evaluating intelligent user interfaces. In *Proceedings of the 3rd international conference on Intelligent user interfaces.* 5–6.

[25] STI MSON JA, MB MACKUEN, and RS ERIKSON. 1995. Dynamic Representation []. *The American Polit-ical Science Review* (1995).

[26] Anthony Jameson. 2011. Systems that adapt to their users. *Decision Making* 2 (2011), 23.

[27] Myounghoon Jeon, Benjamin K Davison, Michael A Nees, Jeff Wilson, and Bruce N Walker. 2009. Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. In *Proceedings of the 1st international conference on automotive user interfaces and interactive vehicular applications.* 91–98.

[28] Myounghoon Jeon, Thomas M Gable, Benjamin K Davison, Michael A Nees, Jeff Wilson, and Bruce N Walker. 2015. Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload. *International Journal of Human-Computer Interaction* 31, 1 (2015), 1–16.

[29] Myounghoon Jeon, Bruce N Walker, and Abhishek Srivastava. 2012. "Spindex"(Speech Index) Enhances Menus on Touch Screen Devices with Tapping, Wheeling, and Flicking. *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, 2 (2012), 1–27.

[30] Brooke John. 1996. SUS: a" quick and dirty" usability scale. *Usability evaluation in industry* (1996), 189–194.

[31] Tzur M Karelitz and David V Budescu. 2004. You say" probable" and I say" likely": improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology: Applied* 10, 1 (2004), 25.

[32] Inayat Khan and Shah Khusro. 2020. Towards the design of context-aware adaptive user interfaces to minimize drivers' distractions. *Mobile Information Systems* 2020 (2020).

[33] Namgyu Kim, Gerard Jounghyun Kim, Chan-Mo Park, Inseok Lee, and Sung H Lim. 2000. Multimodal Menu Presentation and Selection in Immersive Virtual Environments.. In *vr.* Citeseer, 281.

[34] Gordon Kurtenbach and William Buxton. 1993. The limits of expert performance using hierarchic marking menus. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems.* 482–487.

[35] Pontus Larsson and Mathias Niemand. 2015. Using sound to reduce visual distraction from in-vehicle human–machine interfaces. *Traffic injury prevention* 16, sup1 (2015), S25–S30.

[36] Xiaolong Lou, Xiangdong A Li, Preben Hansen, and Peng Du. 2021. Hand-adaptive user interface: improved gestural interaction in virtual reality. *Virtual Reality* 25 (2021), 367–382.

[37] Xiaolong Lou, Xiangdong A Li, Preben Hansen, and Peng Du. 2021. Hand-adaptive user interface: improved gestural interaction in virtual reality. *Virtual*

Reality 25 (2021), 367–382.

[38] Andrew Luxton-Reilly, Emma McMillan, Elizabeth Stevenson, Ewan Tempero, and Paul Denny. 2018. Ladebug: an online tool to help novice programmers improve their debugging skills. In *Proceedings of the 23rd annual acm conference on innovation and technology in computer science education*. 159–164.

[39] Patricia Ivette Cornelio Martinez, Silvana De Pirro, Chi Thanh Vi, and Sriram Subramanian. 2017. Agency in Mid-air Interfaces.. In *CHI*. 2426–2439.

[40] Jeffrey Mitchell and Ben Shneiderman. 1989. Dynamic versus static menus: an exploratory comparison. *ACM SigCHI Bulletin* 20, 4 (1989), 33–37.

[41] Gang Ren and Eamonn O'Neill. 2012. 3D marking menu selection with freehand gestures. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 61–68.

[42] A Riener. 2012. Hand and finger gestures in vehicular applications. *Computer* 45, 4 (2012), 42–47.

[43] Andrew Sears and Ben Shneiderman. 1994. Split menus: effectively using selection frequency to organize menus. *ACM Transactions on Computer-Human Interaction (TOCHI)* 1, 1 (1994), 27–51.

[44] Gözel Shakeri, John H Williamson, and Stephen Brewster. 2017. Novel multimodal feedback techniques for in-car mid-air gesture interaction. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. 84–93.

[45] Gözel Shakeri, John H Williamson, and Stephen Brewster. 2018. May the force be with you: Ultrasound haptic feedback for mid-air gesture interaction in cars. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. 1–10.

[46] Ben Shneiderman. 1997. Direct manipulation for comprehensible, predictable and controllable user interfaces. In *Proceedings of the 2nd international conference on Intelligent user interfaces*. 33–39.

[47] Martha Irene Smith. 2018. *Informing Design of In-Vehicle Augmented Reality Head-Up Displays and Methods for Assessment.* Ph. D. Dissertation. Virginia Tech.

[48] Jason Sterkenburg, Steven Landry, and Myounghoon Jeon. 2017. Eyes-free In-vehicle gesture controls: Auditory-only displays reduced visual distraction and workload. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. 195–200.

[49] Jason Sterkenburg, Steven Landry, and Myounghoon Jeon. 2019. Design and evaluation of auditory-supported air gesture controls in vehicles. *Journal on Multimodal User Interfaces* 13 (2019), 55–70.

[50] Jason Sterkenburg, Steven Landry, Myounghoon Jeon, and Joshua Johnson. 2016. Towards an in-vehicle sonically-enhanced gesture control interface: a pilot study. (2016).

[51] Kay Tislar, Zackery Duford, Brittany Nelson, Madeline Peabody, and Myounghoon Jeon. 2018. Examining the learnability of auditory displays: Music, earcons, spearcons, and lyricons. (2018).

[52] T Victor. 2014. Strategic Highway Research Program Safety Focus Area. *Transportation Research Board, & National Academies of Sciences* (2014).

[53] Bruce N Walker, Jeffrey Lindsay, Amanda Nance, Yoko Nakano, Dianne K Palladino, Tilman Dingler, and Myounghoon Jeon. 2013. Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors* 55, 1 (2013), 157–182.

[54] Christopher D Wickens, Juliana Goh, John Helleberg, William J Horrey, and Donald A Talleur. 2003. Attentional models of multitask pilot performance using advanced display technology. *Human factors* 45, 3 (2003), 360–380.