

Route Guidance Modality for Elder Driver Navigation

SeungJun Kim¹, Jin-Hyuk Hong¹, Kevin A. Li², Jodi Forlizzi¹, and Anind K. Dey¹

¹ Human-Computer Interaction Institute, School of Computer Science,
Carnegie Mellon University, Pittsburgh, PA 15213, USA

² AT & T Research, Labs, Florham Park, NJ 07932, USA
{sjunikim, forlizzi, anind}@cs.cmu.edu,
kevinli@research.att.com, hjinh7@gmail.com

Abstract. Differences in perceptual and cognitive abilities between the young and elderly have implications for in-car tasks. As a primary example, although in-car navigation systems enhance situational awareness, this comes at the cost of increasing visual distraction and cognitive load. To address these shortcomings, this paper explores the efficacy of multi-modal cues for providing route guidance information. We present the results of a study evaluating the impact of multi-modal feedback on driving performance and cognitive load. We found that the full combination of visual, auditory, and haptic feedback was generally most useful to reduce way-finding errors. However, our study highlighted a number of differences between elder and younger drivers for their *safer* navigation. Adding more modalities strained the already high workload of elder drivers. In contrast, adding haptic feedback to traditional audio and visual feedback led to more attentive driving by younger drivers. Therefore, for elder drivers, navigation systems need to be personalized to enhance the benefit of auditory feedback without increasing the number of sensory feedbacks. For younger drivers, it is necessary to incorporate new non-visual feedback to minimize distractions caused by visual feedback. We demonstrate these results through task performance-based measures, subjective workload measures and through objective workload measures that use psychophysiological responses of participants to predict a driver's cognitive load in near real-time.

Keywords: Elderly driver, Car navigation, Cognitive load, Divided attention, Haptics, Psycho-physiological measurement.

1 Introduction

As our society ages, the number of drivers over the age of 65 is rapidly growing. However, the cognitive effects of aging can force them to relinquish control of their cars. Unfortunately, quality of life is acutely linked with the ability to maintain independence in mobility [1, 29]. Thus, reduced mobility combined with the need for mobility independence has substantial negative impact on an individual as well as their family, who often takes on the burden of care [24].

Decay in vision, hearing, and general mobility collectively reduce the performance of elderly drivers (*e.g.*, [10]). Age-related decreases in spatial cognition ability leads to challenges for elderly individuals in accurately forming a mental representation of

a spatial environment and efficiently navigating such environments. For example, it has been found that older adults have difficulty in understanding ‘you-are-here’ maps, which are used to plan simpler driving routes, even though they may increase driving time [30]. In an on-road driving assessment [15], older adults forgot to check blind spots and made errors when asked to report road marking and traffic signs as they drove. In addition, it has been found that older adults are affected more when taking their eyes off of the road, and thus do not use secondary displays in cars as commonly as younger drivers do [18, 20]. These secondary displays, such as in-car navigation systems, typically enhance drivers’ situation awareness, at the cost of increased visual attention and cognitive load. These shortcomings are harder for elderly drivers to overcome; technologies such as GPS systems are often considered too difficult for older drivers to use effectively as a driving aid [20].

We are interested in improving the driving performance of elderly drivers. The relationship between workload and performance is complex. Performance can be affected by workload being too high or too low [27], resulting in a saturation of cognitive capability, the loss of situational awareness or a reduced sense of alertness. Multiple resource theory proposes that the cognitive burden from information overload can be reduced by utilizing multiple modalities to present information [33]. This allows users to process information in parallel rather than sequentially [4].

This paper examines how the usage of multi-modal route guidance cues can lead to safer driving by studying the impact of different combinations of modalities on driving performance for elder and younger drivers.

1.1 Study Overview

The focus of this paper is an investigation of the most effective combinations of feedback modalities for younger and elder drivers’ navigation. We compare driver performance for four different combinations of sensory feedback: 1) *visual plus auditory*, 2) *visual plus haptic*, 3) *auditory plus haptic*, and 4) *visual plus auditory plus haptic*.

Based on previous studies [18], we hypothesized that elder drivers will exhibit lower driving performance than younger drivers, independent of modality combination. Informed by multiple resource theory, we also hypothesized that different combinations of multi-modal feedback would reduce the workload required to process navigation information, and that this would be observed through differences in driving performance and cognitive load. To evaluate these combinations, we implemented a driving simulation test-bed with auditory and visual feedback. We also instrumented a steering wheel with vibrotactile feedback. Thirty-three participants (17 elderly) performed a series of tasks using our simulation with each of the feedback combinations described above. We measured driving performance as well as subjective and objective measures of task workload to compare the effectiveness of the different feedback combinations. We define *effectiveness* as providing route guidance without decreasing task performance or increasing cognitive load: a more *effective* guidance system should result in *safer* driving.

We make two contributions. First, we evaluated the effectiveness and safety benefits of different combinations of multi-modal navigation cues for elder and younger drivers. Second, our mixed-methods assessment approach based on task performance, divided attention and induced workload allowed us to evaluate models for predicting

cognitive load based on psycho-physiological responses of the two age groups in real time.

We begin our paper with a discussion of impoverished attention in driving environments and the benefits of using multiple attention resources. We then describe the test-bed we developed for our study to evaluate four multi-modal route guidance combinations. We describe the results of our study, where we found that the *visual plus auditory plus haptic* condition generally led to improved way-finding performance with higher driver satisfaction. Surprisingly, we found that for elder drivers, *auditory plus haptic* was safer than using all three modalities. These results highlighted the need to be sensitive to the already heavier workload for older drivers. We also found that for elders, auditory feedback was indispensable, while for our younger drivers, the addition of haptic feedback to the typical combination (*i.e.*, *visual plus auditory*) was effective for safer driving. We conclude with a discussion of our results and implications for designing multi-modal route guidance systems.

2 Background Work

In this section, we discuss related work on performance and attention in driving, the effect of using multiple modalities in interfaces, and on the objective measurement of cognitive load.

Impoverished-Attention in a Driving Environment. The split attention effect while driving stems from the cognitive load from secondary tasks. Impact of these tasks has been assessed through measurement of eyes-off-the-road occurrence, driver reaction time, and accuracy within a dual-task paradigm (*e.g.*, [7], [13], [21]). Artificial tasks such as short term memory tests are often employed as secondary in-vehicle tasks during a primary simulated driving task, to assess the impact of the secondary tasks on driving performance, using, for example, eye-gaze data. Similar approaches have been used to evaluate the effects of heterogeneous secondary tasks (*e.g.*, mobile phone use *vs.* eating in car) [7, 13, 28], and different dashboard and navigation displays [18, 19].

GPS route guidance systems are a canonical secondary information source in the car, and can both positively and negatively impact workload. They enhance drivers' situational awareness while causing drivers to take their eyes off the road. As observed in prior work [18], the cost and benefits of such systems depend on a driver's cognitive capabilities.

Multiple Attention Resources. Human attention is finite, however we can enhance one's mental processing capability leveraging multiple cognitive channels. From the Multiple Resource Theory [33], we know that information services that use multiple channels can facilitate perception of information more effectively than single-modal services. In learning, it has been shown that learners can accommodate more new information when it is presented using auditory and visual channels [22]. Most conventional GPS systems use both, which should improve drivers' ability to process navigation information over using either channel alone.

Even though such an approach is useful, existing studies of in-car information services have focused on single modalities (*e.g.*, [14], [31]). Even fewer have focused on elderly drivers and their cognitive workload, even though multi-modal approaches for in-vehicle secondary information systems could be especially beneficial in reducing attention demands for this population. For younger adults, the effect of multi-modality and the use of tactile feedback are advantageous for navigating, particularly when used as a back-up or confirmation of information available from other modalities [17]. Thus, we focus our efforts on studying multi-modal feedback during navigation tasks, using assessments based on performance and cognitive load, especially with respect to elderly drivers.

Cognitive Load Assessment. Performance-based methods are frequently used to assess experimental conditions in dual-task settings. In driving situations, these methods examine the ways that a driver's responses deteriorate (*e.g.*, lag in reaction time or increase in errors) when using finite cognitive resources to perform two or more tasks. This objective approach has been demonstrated to have a strong link with cognitive load [3]. However, recent work has shown that driver response may be less sensitive to subtle differences in cognitive load [9]. Different combinations of modality feedback may not impact performance in a measurable way, so we also focus on measures of cognitive load.

Subjective rating-based methods such as the NASA TLX [11] or Likert-scales use participants' own judgment of their efforts. This approach is applied *post-hoc*, is non-intrusive, and is a reliable indicator of users' preferences for particular test conditions [3]. However, even when users struggle to complete a task in a timely fashion, they may self-report the task as having a low workload, if they believe they did not make any errors [25]. This calls for a more objective measure of workload.

Psycho-physiological response analysis can provide such a measure. Some believe that this approach can more sensitively assess cognitive load over a continuous time frame [3], allowing for detection of changes in cognitive load even when no deterioration in task performance is demonstrated. We also note that psycho-physiological measures can be applied across a variety of stimuli modalities, and with different numbers of stimuli [32]. This approach also supports assessment for tasks like navigation that leverage major cognitive processes such as perception, memory and reasoning [2]. An advantage of physiological responses is that they do not require an overt response by the operator, while at the same time, most cognitive tasks do not require overt behavior. While this approach is promising, it has yet to be validated with a driving task such as we are proposing in this paper.

Finally, the frequency and length of eye fixations is usually correlated with changes in workload (*e.g.*, [6]). However, the sum of these measures could linearly increase in cases where task duration is prolonged due to an individual's abilities or the difficulty of a given task. Therefore, the proportion of these measure's totals or overall task durations have been used to provide more meaningful results.

Accordingly, our study uses measures of task performance, and subjective and objective ratings of cognitive load, induced by different navigation feedback mechanisms, to determine which are most appropriate for younger and elder drivers. We additionally validate whether psycho-physiological responses can be used in assessment in such situations.

3 Experimental Design

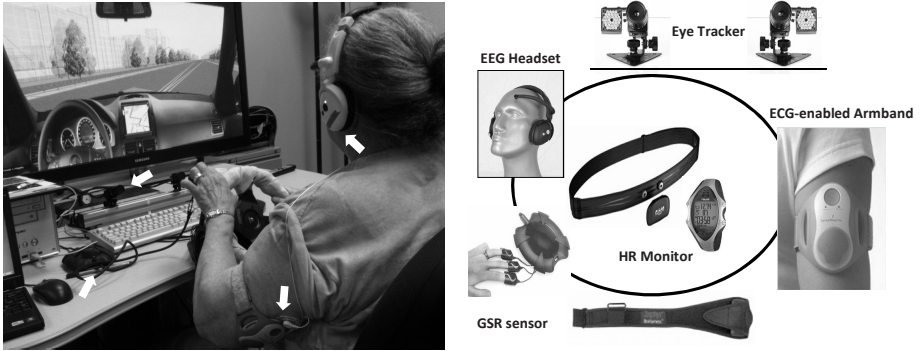


Fig. 1. Our experimental test-bed includes a three-dimensional driving simulator with a wheel joystick set, contactless gaze tracker, and multiple psycho-physiological sensors

3.1 Test-Bed Implementation

We implemented a three-dimensional driving simulator to conduct our experiment. In the simulator, drivers are given specific routes to navigate. Geospatial information from Google Maps for Pittsburgh is graphically rendered in this simulator on a 58-inch widescreen LCD HDTV (screen size: 1.290m \times 0.723m, resolution 1280 \times 1024). Participants can navigate through the simulated cities using a virtual wheel and two foot pedals (see Fig. 1). We automatically record simulator state, driver input from the virtual wheel and foot pedals and psycho-physiological responses for later analysis of driving performance.

Each navigation route is 3.36km long and includes 12 intersections: 4 right turns, 4 left turns and 4 which drivers pass straight through. During each driving condition, participants encounter 12 light signals, 3 stop signs, 5 pedestrians (such as a person pushing a baby carriage) crossing the road from right to left and 5 other pedestrians (such as a man wearing a business suit, holding a suitcase) crossing the road from left to right, whom they are expected to avoid. In the case of missed turns, a U-turn must be made to get back on the route again. Route guidance information is provided using three different types of sensory feedback as follows (see Fig. 2.).

Visual Feedback (V). We implemented a 2D bird's-eye-view map display mode (on-screen size: 12cm \times 12cm, see Fig. 2.), commonly employed in GPS navigation systems and usually installed on the lower-right side of a driver's head. The map display is synchronized with car movement in real-time and the suggested route is highlighted. It supports both global awareness (*i.e.*, a driver's understanding of nearby road networks) and local guidance for use as navigation aids. However, given the nature of this sensory feedback, drivers need to consistently monitor the device (increasing eyes-off-the-road time) in order to determine where they are supposed to turn. Earlier work found that elder drivers had greater difficulty in using systems with

only visual feedback [18]. We hypothesize that drivers using multi-modal navigation systems that incorporate visual feedback will be able to leverage an alternative feedback modality, and only selectively pay attention to the visual feedback, improving overall performance.

Auditory Feedback (A). We implemented three kinds of voice commands: two for route guidance: 'turn left' and 'turn right', and one for driving instructions in case of missed turns: 'wrong way, make a U-turn' (Fig. 2.). In the case of no missed turns during a driving session, twelve voice commands prompt route guidance information, designed to occur at a constant distance of 45m in front of every intersection where drivers should make turn decisions (45m was determined empirically as early enough to initiate turns at intersections considering typical driving speed, while not so early as to forget the provided turn information). When using this sensory feedback, drivers do not have to take their eyes off the road to obtain route guidance information, however they may have to be careful not to miss or forget the provided information since this feedback supports only local guidance (*i.e.*, the information is not persistent) and is triggered at a specific moment for each intersection. We hypothesize that auditory feedback will be the easiest and most comfortable modality for elder drivers.

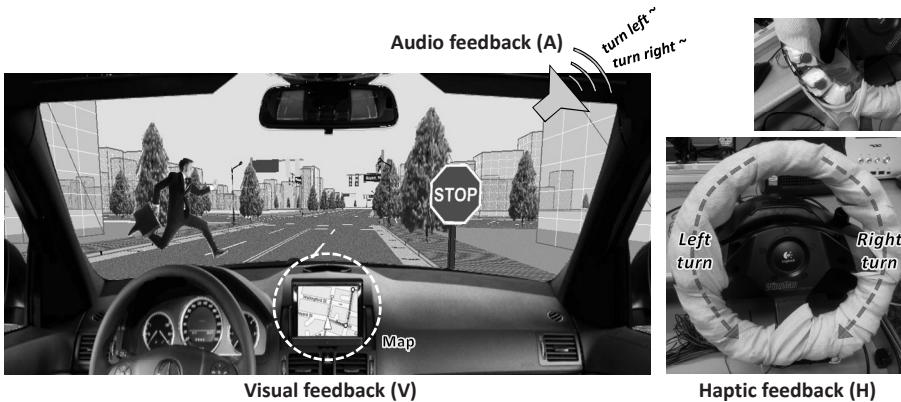


Fig. 2. Traffic events, including pedestrians, lights, and stop signs are incorporated into the driving simulation test-bed. Route guidance information is provided using three different types of sensory feedback: visual feedback ('V'), auditory feedback ('A'), and haptic feedback ('H'). Haptic wheel that triggered turn right cues with clockwise vibrations and turn left cues with counterclockwise vibrations (bottom right). Vibration motors mounted on steering wheel, padded with memory foam (top right).

Haptic Feedback (H). We developed an enhanced prototype of a vibrotactile steering wheel that provides haptic feedback, as shown in Fig. 2 (right). It provides a confusion-free vibrotactile localization with the use of memory foam (*e.g.*, unresolved issue in other settings such as [14]) and a higher resolution by using a larger number of actuators (*e.g.*, more than six epicenters employed in [17]). We installed twenty vibration motors onto the front face of the wheel. In initial pilot studies we found that

vibrations generated from one motor would get transferred all along the core of the wheel, making it difficult to determine which particular motor had been turned on. To localize the vibrations, each vibration motor was padded with memory foam, solving the problem (See Fig. 2.). We used an Arduino Mega 2560 to control 20 motors embedded in the steering wheel to produce vibration patterns.

For right-turn information, the wheel creates a clockwise vibration from the one o'clock to five o'clock positions of the wheel, while for left-turn information it provides a counterclockwise vibration from eleven o'clock to seven o'clock. We activated and deactivated vibration motors appropriately to generate these different directions. By triggering spatially separated vibrotactile motors, we hoped to elicit *sensory saltation* [8] whereby users would perceive a direction of motion. Each vibration sequence was repeated 8 times to ensure the driver would not miss the cue. All participants responded that there was no confusion about the turn direction cue provided by the wheel. Cues were triggered at a constant distance (45m) from the turn.

3.2 Pilot Study

We conducted two stages of a pilot study. In the first stage, we focused on validating each type of sensory feedback, especially the usability of the vibrotactile wheel and auditory voice-commands that we developed for this study. In particular, we validated the vibrotactile localization and representations for effective turn cues, additional auditory cues for the case of missed turns, and the service timings of vibrotactile and auditory cues. Four pilot subjects evaluated a number of different versions of cue representations, and based on their feedback, we iterated on the design of the auditory and haptic cues.

The second stage of the study was used to decide which combinations of feedback modalities to explore in the actual study, as well as to re-validate our cue design. Seven pilot subjects, including the four subjects from the first stage of the study were asked to perform a virtual driving task using all possible combinations of the three modalities: *visual only*, *auditory only*, *haptic only*, *visual plus auditory*, *visual plus haptic*, *auditory plus haptic*, and *visual plus auditory plus haptic*. After all the tests, we collected pilot participants' responses for three aspects (preference, cognitive load, and performance) through Likert-scale questions as well as the NASA-TLX assessment after each test. We found that in almost all cases, the multi-modal conditions outperformed the single-modal conditions, in terms of preference, cognitive load and performance. There was only one exception: *visual only* was tied for the third most preferred condition, with *auditory plus haptic*, ranking ahead of *visual plus haptic*. Based on these results, we chose to focus on the multi-modal conditions in our main study.

3.3 Participants in the Main Experiment

We recruited 33 study participants: 16 younger participants (age $M=25.4$, $SD=5.46$, age range: 19-36, gender: female 31% and male 69%); 17 older participants, over the age of 65, with normal or corrected-to-normal vision and hearing (age $M=73.8$, $SD=7.48$, age range: 65-91, gender: female 65% and male 35%). Participants were

recruited through lifelong learning institutes at two local universities, a local center for behavioral decision research, and through bulletin board ads placed around the university campus and local apartments. Participants were compensated with \$15 (US), with an additional \$5 for parking as necessary.

3.4 Task, Test Conditions, and Procedure

Participants were asked to wear six sensor devices during their driving tasks and were asked to execute all driving tasks in front of two cameras installed at the bottom of the simulation screen in order to capture gaze tracking (see Fig. 1). Participants were shown how to use our test-bed, and how to respond to each of the traffic events, driving rules and regulations. Participants then performed one iteration of practice driving with each feedback modality. Users had no trouble interpreting any of the directional cues. To ensure participants felt the vibrotactile cues, we instructed them to hold the steering wheel with both hands at all times.

Each participant then performed virtual driving tasks using the 4 multi-modal combinations of route guidance feedback: 1) *visual plus auditory feedback* (V+A), 2) *visual plus haptic feedback* (V+H), 3) *auditory plus haptic feedback* (A+H), and 4) *all three forms of feedback* (V+A+H). Order of presentation was counter-balanced using a Latin square method.

After each condition, participants filled out a questionnaire including the NASA-TLX assessment. After all four tests, they filled out a post-questionnaire where we collected demographics, information about driving experience, and their comparative evaluations of the four multi-modal conditions.

3.5 Measurement

Measure 1: Task Performance. For driving and navigation task performance measures, in this paper we used the following metrics: *driving time*, *lateral lane deviation* (distance between the center of the right-hand lane and a car's current position), *the number of missed turns*, *traffic signal violations*, *stop sign violations*, and *the number of incidents that placed pedestrians in danger*, which indicates moments when the car intersected with pedestrians.

Measure 2: Divided Attention. As a measure for divided attention, we measured eye gaze movement and extracted the coordinates where a subject's gaze intersected with the screen. In this study, we focused on metrics related to eyes-off-the-road issues, as follows: *eyes-off-the-road time*, which indicates how long a participant's gaze remained on the secondary display or off the road (e.g., rear-view mirror, stereo, dashboard); *eyes-off-the-road frequency*, which indicates how many times gaze is drawn to the secondary display; *eyes-off-the-road time at a glance*, which indicates the average length of a single glance on the secondary display (i.e., equals eyes-off-the-road time / eyes-off-the-road frequency); *eyes-off-the-road percent*, which indicates the proportion of the eyes-off-the-road duration in comparison to the duration of driving time; *gaze movement distance*, which indicates the distance in which a participant's gaze travels over the screen during a driving task; *gaze movement distance per 1-minute driving*, which indicates the distance over which a participant's gaze travels

over the screen per minute. Given different individual driving times, in this paper we focused our analysis on three metrics: eyes-off-the-road time at a glance, eyes-off-the-road percent, and gaze movement distance per 1-minute driving.

Measure 3: Subjective Ratings. After each driving condition, we subjectively measured workload by having participants use the NASA-TLX workload assessment tool. This is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales: *Mental Demand*, *Physical Demand*, *Temporal Demand*, *Own Performance*, *Effort* and *Frustration*. After completing all driving sessions, participants were asked to rate the most annoying and most useful modality condition, and to rate their agreement with the following statements: 1) Preference - “*I would use the following navigation mode.*”, 2) Cognitive load - “*The following mode was easy for me to use.*”, and 3) Performance - “*I drove well using the following mode.*” They responded using a 6-point Likert scale: *disagree strongly* (score: 1) - *disagree moderately* (2) - *disagree slightly* (3) - *agree slightly* (4) - *agree moderately* (5) - *agree strongly* (6).

Measure 4: Psycho-physiological Responses. We measured participants’ psycho-physiological responses using six sensor devices: a contactless eye tracker (SmartEye 5.6), an ECG-enabled armband (SenseWear Pro3), a wireless EEG headset (NeuroSky mindset kit), a wireless HR monitor belt (Polar RS800CX HR monitor), another wireless physiological monitor belt (BioHarness™ BT) and a GSR (galvanic skin response) finger sensor (LightStone biofeedback sensors). These devices measure and record information about pupil size, blink rate, GSR, heat flux (rate of heat transfer on the skin), heart rate and heart rate variability, inter-beat (RR) interval, ECG (electrocardiography, electrical activity of the heart over time) and EEG (electroencephalography, electrical activity of the brain). The value of these psycho-physiological signals in assessing cognitive load have been demonstrated in [9].

4 Results

For the task performance, eye tracking, and NASA-TLX assessment measures, we performed one-way repeated measures ANOVA analysis (route guidance modality as a within-subject factor and age group as a between-subjects factor) and the Bonferroni *post-hoc* test. For the analysis of Likert-scale rating data, Friedman tests and a Wilcoxon Signed Rank *post-hoc* test were conducted.

4.1 Elder Drivers’ Navigation

As hypothesized, elderly drivers exhibited significantly lower task performance and higher task workload than younger drivers across all conditions (see Fig. 3).

Specifically, elders took 1.7 times longer to complete the driving tasks ($F_{1,27}=23.6$, $p<.001$), had 2.6 times more missed turns ($F_{1,27}=4.81$, $p=.037$), and reported 1.3 times higher task workload using the NASA TLX assessment ($F_{1,27}=4.31$, $p=.047$). There were no significant differences in eye-off-the-road measures between age groups. The only exception to these navigation measures was that elder drivers obeyed stop signs significantly better than younger drivers ($F_{1,27}=7.04$, $p=.013$).

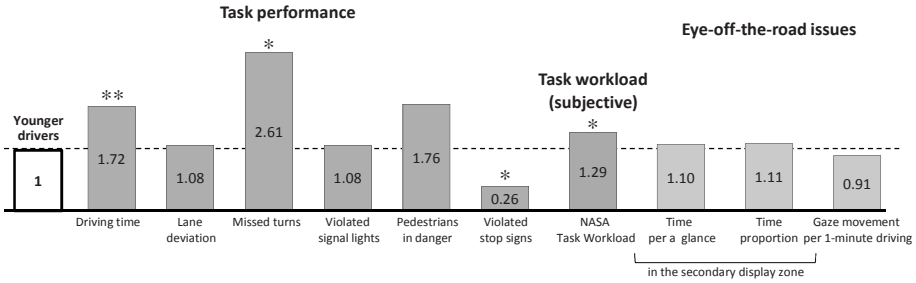


Fig. 3. When driving using route guidance systems, elder drivers exhibited worse task performance, and higher task workload, compared to younger drivers. All the measured values from the younger group were set as 1 (the left-most bar). The numbers in each bar represent the proportions of the measured values from elder group to the values from younger group. For example, elder drivers missed 2.61 times more numbers of turns at intersections than younger drivers. * $p < .05$. ** $p < .001$.

4.2 Effects of Route Guidance Modalities

In our analysis, we treat *visual plus auditory* to be the baseline condition, as it is what is found in most conventional GPS systems. Route guidance modalities significantly impact eye-off-the-road issues across all participants (eye-off-the-road time per glance: $F_{3,81}=23.9, p < .001$, eye-off-the-road time proportion: $F_{3,81}=33.3, p < .001$); however they did not have much impact on navigation performance (i.e., mean differences in the number of missed turns, $F_{3,81}=0.605, p = .614$). In particular, when using *auditory plus haptic* feedback (A+H), participants' eye gazes dwelled off the road almost 0.2 seconds less per glance independent of age group ($p < .001$), and eye-off-the-road time proportion (= total eye-off-the-road time / total driving time) was also significantly reduced (elder drivers: 4% less, $p = .009$, younger drivers: 9% less, $p < .001$ - differences, not ratio, between the proportion measures), than when using the baseline condition. The significant reductions in eye-off-the-road issues by the use of 'A+H feedback' were consistently found within each age group and also in the post-hoc comparisons with other modalities (i.e., A+H vs. either V+A, V+H, or V+A+H).

Interestingly, we found that younger drivers' eye-off-the-road metrics associated with visual feedback were reduced when using haptic feedback. When younger drivers used V+A+H, eye-off-the-road time proportion was significantly lower (by 3.1%, $p = .048$), than when using the baseline condition V+A; no such phenomenon was shown for the elder group (see Fig. 4). In order of impact, this eye-related issue differed significantly as follows: V+A (greatest issue) \geq V+H \geq V+A+H $>$ A+H (least issue). We further note that this result is due to the impact of visual feedback, rather than auditory feedback. When using haptic feedback instead of visual feedback (A+H rather V+A), younger drivers' eye-off-the-road time proportion was also significantly reduced ($p < .001$). In addition, there were no significant eye-off-the-road issues when adding audio feedback (V+A+H vs. V+H), nor when replacing audio feedback with haptic feedback (V+A vs. V+H).

These results indicate that the degree of drivers' engagement with a particular feedback modality differs by age group and by the combinations of feedback used; in

other words, there are significant differences in cognitive engagement with different route guidance modalities for the two age groups. Service providers should take this into account and consider the adoption of a novel modality into conventional GPS systems based on the expected user population.

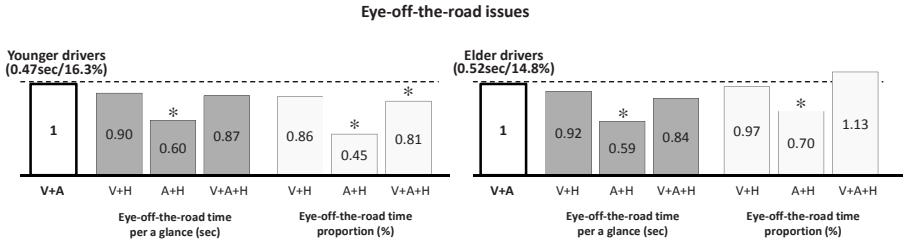


Fig. 4. Route guidance modalities have a significant impact on eye-off-the-road issues within each age group ($*p < 0.05$). The numbers in each bar graph represent the proportions of the measured values when using each navigation mode to the measured values when using the baseline mode, V+A (the left-most bars for each age group). For example, the first seven bars for younger group (left) show that younger drivers' eye-off-the-road time proportion (%) was significantly lower when using A+H or V+A+H than when using the baseline (V+A), 0.45 factor and 0.81 factor, respectively. The next seven bars for elder group (right) show that elder drivers exhibited such significant reduction only when using A+H.

4.3 Modality Differences between Age Groups

From participants' self-reports of preference, performance and load, we found that elder drivers relied on auditory feedback, while younger drivers relied on visual feedback. Elder drivers significantly disliked and had difficulty using the 'V+H'-based route guidance (with no auditory feedback) even though when using this modality their driving and navigation performance was not worse and their eye-off-the-road issues were not greater, compared to those from the other modalities (e.g., V+A+H vs. V+H - Likert preference: $Z = -3.02$, $p = .003$; Likert cognitive load: $Z = -2.66$, $p = .008$; Likert performance: $Z = -2.75$, $p = .006$). In contrast, younger drivers disliked the most, had the most difficulty using, and thought they performed the worst when using the 'A+H'-based route guidance (with no visual feedback), even though the effects were not always significant. In particular, in all three aspects (preference, cognitive load, and performance) younger drivers rated the A+H modality as significantly worse than the baseline modality, V+A (Likert preference: $Z = -3.13$, $p = .002$; Likert cognitive load: $Z = -2.36$, $p = .018$; Likert performance: $Z = -2.38$, $p = .017$).

We saw consistent results in our summative questions about sensory feedback. 71% of elder drivers thought the auditory modality was the most useful and 59% thought the visual modality was the most annoying. In contrast, 63% of younger drivers thought the visual modality was most useful and 50% of them thought the auditory modality was most annoying. Both groups ranked haptic feedback between auditory and visual feedback.

4.4 Psycho-physiological Task Workload

For each age group, we examined the psycho-physiological assessment of drivers' task workloads (high, medium, and low, labeled from the NASA-TLX task workload) induced by the four multi-modal route guidance conditions, and built a model that measures the task workload based on these psycho-physiological measurements. Participants who experienced a measurement problem (*e.g.*, no eye gaze tracking data for a significant amount of driving time) were excluded from the analysis, leaving 12 younger and 11 elder participants. Data from each psycho-physiological sensory channel was segmented every 3 seconds, and the mean (μ) and standard deviation (σ) were calculated for each segment. All told, there were a total of 59 extracted features. The features were normalized for each participant, and discretized into five states.

Table 1. Classification accuracies and high-ranked physiological features for each of age group. Task workload classes were determined based on the NASA-TLX results.

	Younger group	Elder group
Bayesian Networks	82.3%	76.2%
Naïve Bayes	53.3%	51.0%
Neural Networks	64.2%	58.4%
High-ranked features	GSR (μ), heat flux (μ)	heat flux (μ), BR (μ)
	BR.Amp (μ , σ)	blink (σ , μ), HR (μ)
	EEG β_2, γ_1 (μ), pupil size (μ)	GSR (μ)

We used three machine learning techniques 1) Bayesian Networks¹, 2) the Naïve Bayes¹ classifier, and 3) neural networks, to build two models, each of which estimates the task workload based on the psycho-physiological features and the NASA TLX results (ground truth) for one of the age groups. The performance of the models was evaluated using five-fold cross validation, where the data were randomly divided into five parts with four parts used for training and one part for testing. We report the average performance from repeating this process five times.

Table 1 shows the prediction performance of the models for each age group. Among the three techniques, Bayesian networks had the best performance, with accuracies of 82% and 76% for the younger and the elder groups, respectively. With all 3 techniques, younger drivers were modeled better than elder drivers. In terms of features, the psycho-physiological features that were most predictive of load, differed between the two age groups, although both GSR (μ) and heat flux (μ) were highly ranked for both groups. For example, eye-blink had higher information gain for the elder group, while EEG features had higher information gain for the younger group.

When analyzing individual prediction results, we found that the prediction rates are quite high, which indicates the feasibility of applying our psycho-physiological

¹ We implemented the Bayesian Networks and the Naïve Bayes classifiers based on the SMILE reasoning engine for graphical probabilistic models contributed to the community by the Decision Systems Laboratory, University of Pittsburgh (<http://dsl.sis.pitt.edu>). In particular, for the Bayesian Networks, we first learned the structure with the K2 algorithm and then applied the EM algorithm to learn the parameters.

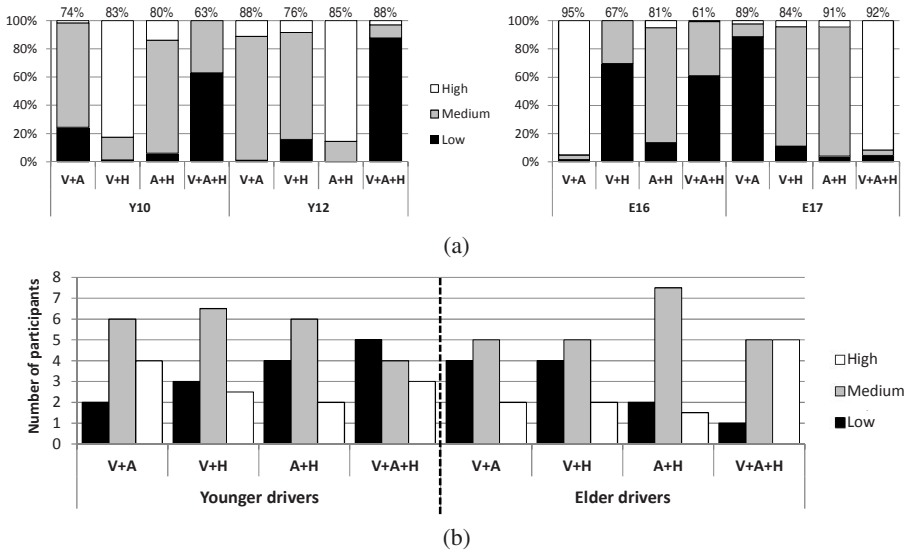


Fig. 5. Psycho-physiological assessment: (a) Two participant examples from each age group (each bar shows the ratio of classified samples during the driving with the corresponding navigation feedback modality. For example, 83% of Y10's samples indicated V+H condition induced *high* task workload.), (b) Across all participants divided by age group.

approach for assessing task workload in driving tasks. Fig. 5(a) shows the results for 4 of our participants, in which the prediction of task workload is correctly assessed except for one case. S16 identified V+A+H as having *medium* workload (through the NASA-TLX assessment), but our Bayesian network model classified the workload as *low* based on his psycho-physiological measurements. (Note that in this instance, almost 40% of the driving sensor samples were classified as representing *medium* workload, meaning that our psycho-physiological assessment is very capable, despite the incorrect assessment). Our proposed psycho-physiological assessment approach also allows us to rank the navigation feedback modality for each participant, *e.g.*, the ranking of Y10 is {V+A+H, V+A, A+H, V+H} while the ranking of E17 is {V+A, V+H, A+H, V+A+H}. In addition, this approach allows for an assessment of workload every 3 seconds while driving (as opposed to the one-time, post hoc NASA TLX), allowing it to be extended and applied to a real-time monitoring system. Fig. 5(b) shows the ranking of the modalities across all participants based on the psycho-physiological approach. There were differences between the two age groups, with many younger drivers expressing lower task workload than elder drivers with V+A+H but higher task workload with V+A. Also, individual differences were found in ranking the four navigation feedback modalities, *e.g.*, 5 younger participants had low workload with V+A+H but 3 other younger participants had high workload with this combination. These results, while useful, suggest future work to personalize the psycho-physiological assessment of task workload.

4.5 Discussion

Several important issues emerged from our study of multi-modal navigation feedback, and we discuss those here, along with a discussion of the limitations of our study.

Elderly benefit from multi-modal feedback, but have more difficulty using navigation systems

Older drivers made a higher number of secondary task errors and experienced a higher task workload, compared to younger drivers. However, our study also revealed that the existence of visual feedback did not cause drivers much trouble in managing their visual attention. This result implies that multi-modal services can support elderly drivers in their interaction with navigation systems, especially in managing their visual attention, without negatively impacting their driving performance. Older drivers were able to use other modalities to navigate safely. However, when using these other forms of feedback, their workload remained higher and their secondary task performance was still lower than that of the younger drivers. Older drivers still experience greater workload than younger drivers, even when using multi-modal services. Although the reluctance of elder drivers to use navigation systems can be reduced through the provision of multi-modal services, they may still face greater difficulties than younger drivers in using route guidance information. Therefore, developers of systems for elder drivers may need to focus more on reducing drivers' cognitive burden than on resolving effects of divided attention.

Different combinations of feedback worked best for elder and younger drivers

It is clear from the results in Table 2 that different combinations of feedback worked best for younger and elder drivers. In terms of subjective and objective workload reduction and safe driving, the combination of all modalities worked best for younger drivers, which was predicted by multiple resource theory. It also had them driving the slowest and the fewest instances of taking their eyes off the road (except for the non-visual multi-modal condition). It is quite interesting that although younger drivers preferred the conventional feedback of *visual plus auditory*, this combination ranked lowest overall. Incorporating haptic feedback appears to be the most important for younger drivers, with the three combinations including haptic feedback were best.

Younger drivers stated that visual feedback was the most useful and auditory feedback was the most annoying. Despite this preference for visual feedback, it did not improve their task performance and did create issues of divided attention.

For elder drivers, the results were more mixed. The non-visual multi-modal combination, *auditory plus haptic*, was the best in terms of task performance and divided attention. It did not rank particularly high on the workload measures, although it did induce high cognitive load in the fewest number of drivers. In stark contrast to the younger drivers, the combination of all modalities ranked lowest over all measures, although both this combination and *visual plus haptic* ranked lower than the other combinations. It is also noteworthy that, similar to the younger drivers, the most preferred combination ranked lowest over all measures. For elder drivers, including auditory feedback was the most useful, with *auditory plus haptic* and *visual plus auditory* performing the best.

Table 2. Comparison of the multi-modal conditions for each of our measures, for both age groups. A (conventional): V+A, B: V+H, C: A+H, D: V+A+H. A sign of inequality indicates which condition was better for safer navigation (except for driving time). For example, $C>A$ in elder drivers' *lane deviation* means that elder drivers made smaller lane deviation with C than A.

		Younger	Elder
Subjective Reporting	Stated preferences (Likert)	A>D>B>C	D>A>C>B
	NASA-TLX	D>B>C>A	B>A>C>D
Task Performance	Missed turns	D>B=C=A	C=D=A>B
	Violated regulations	D>C>B>A	A>C>B>D
	Lane deviation	D>B>A>C	C>A>D>B
	Driving Time (faster to slower)	A>C>B=D	B>C>A>D
Eye Gaze	Off-time at a glance	C>D>B>A	C>D>B>A
	Off-time percent	C>D>B>A	C>B>A>D
	Move per 1-min driving	A>C=B>D	C>D>A>B
Physiological	As focusing on the 'Low's in Fig.5	D>C>B>A	A=B>C>D
	As focusing on the 'High's in Fig.5	C>B>D>A	C>A=B>D

Elder drivers stated that visual feedback was most annoying, and this is consistent with past work [15]. This could indicate why elder drivers rated *visual plus haptic* as being significantly less preferable for them despite there being no deterioration in performance or divided attention. This may not be due to the existence of visual feedback, as *auditory plus haptic* was not rated highly either (Table 2). In terms of multi-modal feedback, this may not imply that elder drivers are reluctant to use visual feedback, but instead, that they prefer to have auditory feedback included. As the overall results for elders were mixed, so are the recommendations for improved navigation systems. It may be that navigation systems need to be customized according to the cognitive and perceptual abilities of a particular elder. Their needs or preferences must be taken into account as well; our results would recommend a different navigation system if one wanted safer navigation *vs.* higher satisfaction.

How to incorporate haptic feedback in navigation systems

Our results indicate that a vibrotactile steering wheel can be used as a navigation aid for reducing eyes-off-the-road issues, but not for reducing task performance errors. To incorporate haptic feedback in navigation systems for elder drivers, it could be used as a replacement for visual feedback since older drivers significantly rely on auditory feedback. For younger drivers, haptic feedback should be added to conventional systems, as the combination of all modalities resulted in more attentive driving.

Workload can be estimated in near real-time

Another issue we explored was psycho-physiological task workload assessment. Our results showed that workload models for each age group were able to predict the task workload of individual drivers with quite high accuracies during driving tasks which include dependence on multi-modal navigation systems. This psycho-physiological approach allows for workload assessment every three seconds while driving, making it a promising approach for a near real-time workload monitoring system. This approach has great potential applicability in the design of intelligent route guidance systems. As seen in the learning domain [27], loads which are either too high or too

low can lead to low task performance. If this phenomenon also exists in the driving domain (*e.g.*, suffering from too noisy traffic), our route guidance system may be able to provide services with relatively lower noise through use of visual or haptic feedback without auditory feedback. If drivers' workload is too low (*e.g.*, if driving on the highway becomes monotonous), in-vehicle information systems may be used to provide more detailed information to help maintain attentive driving (*e.g.*, preventing drivers from becoming drowsy).

Limitations. We acknowledge the potential shortcomings of our study here. First, this study was conducted in a custom-built driving simulator in a lab setting. To replicate this study in a real environment, more experimentation is needed to understand the set of sensory distractors. A challenging issue will be in conducting field research with elderly subjects (*e.g.*, ethical issues such as losing decision-making capacity during the course of a study [12]). The video game-like task implemented in this study may have resulted in differences in virtual driving abilities between two age groups, since younger drivers may have had more exposure to video games. There may be potential gender differences in the driving task, since males may play more driving games than females. Other differences may be caused by the difference between local and non-local residents in knowing a route, in years of driving or frequency of computer usage.

While we didn't technically control for the experience in using simulated driving, we found little effect of the age factor. We provided sufficient practice time for both age groups, with no use of any test modalities as well as in all of the test modalities. Our simulated driving task using no test modalities showed no variation between age groups in the task performance and the divided attention measures; the only exception was the driving time (*i.e.*, slower driving of the elderly). All elderly participants had more than 30 years of driving experience; this factor did not result in worse virtual-driving. In this study, we included only the age factor in the analysis and balanced our populations only by this factor. We focused our analysis on the evaluation of route guidance modalities within each of age groups rather than the comparison of age groups within each of driving condition. We used other demographic factors collected as a criterion to filter out additional participants.

Lastly, in the current stage of this study we did not use the standard lane-change test [15, 22] to regulate the use of fixed driving speed condition or extra road signs that can stimulate extra cognitive load or attention management. Use of this standard test would allow others to replicate our results more easily.

5 Conclusions

We conducted a study exploring the modalities used to provide route guidance feedback for drivers. Our study revealed that 'multi-modal' route guidance systems can enhance older drivers' visual attention management. However, these drivers still experienced higher task workloads and exhibited lower navigation performance than younger drivers for such in-vehicle dual-tasks.

We learned that the combination of modalities should be designed differently for older drivers than for younger drivers. For younger drivers, *visual plus auditory plus haptic* was generally the best in terms of improved task performance and reduced induced workload. When designing a navigation system for this population, haptic

feedback should be included, as it improved performance across almost all measures. For elder drivers, while the results were more mixed, *auditory plus haptic* was the best. Due to these mixed results, a customized solution may be necessary depending on their needs to be safer or more satisfied. Auditory feedback should be included in any route guidance system for elder drivers.

Our results indicate the importance of understanding differences in cognitive engagement with different route guidance modalities across age groups. To address this issue, we demonstrated that our psycho-physiological response-based assessment is a highly-accurate approach for predicting drivers' cognitive workloads when driving with navigation systems. We leave the personalization of predictive models and the personalization of route navigation systems to future work.

Acknowledgments. This work was generously funded by General Motors, the National Science Foundation and Quality of Life Engineering Research Center (EEEC-540865). We would also like to thank Eija Haapalainen for her efforts with the modeling of cognitive load using psycho-physiological signals.

References

1. Ball, K., Owsley, C.: Increasing mobility and reducing accidents of older drivers. In: Schaie, K.W., Pietrucha, M. (eds.) *Mobility and Transportation in the Elderly*, pp. 213–250. Springer, New York (2000)
2. Beatty, J.: Task evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91, 276–292 (1982)
3. Cegarra, J., Chevalier, A.: The use of tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. *Behavior Research Methods* 40(4), 988–1000 (2008)
4. Chandler, P., Sweller, J.: The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology* 62(2), 233–246 (1992)
5. Colliá, D., Sharp, J., Giesbrecht, L.: The 2001 national household travel survey: A look into the travel patterns of older Americans. *Journal of Safety Research* 34, 461–470 (2003)
6. De Waard, D.: The measurement of drivers' mental workload. PhD Thesis, University of Groningen (1996)
7. Dingus, T.A., Klauer, S.G.: The relative risks of secondary task induced driver distraction. Society of Automotive Engineers, Technical Paper Series 2008-21-0001 (2008)
8. Geldard, F.A.: *Sensory Saltation: Metastability in the Perceptual World*. Lawrence Erlbaum Associates, Hillsdale (1975)
9. Haapalainen, E., Kim, S., Forlizzi, J., Dey, A.: Psycho-physiological Measures for assessing Cognitive Load. In: *Proc. Ubicomp 2010*, pp. 301–310 (2010)
10. Hakamies-Blomqvist, L., Siren, A., Davidse, R.: Older drivers - a review. VTI rapport 497A, Swedish National Road and Transport Research Institute (2004)
11. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, pp. 139–183. Amsterdam, North-Holland (1988)
12. High, D.M., Doole, M.M.: Ethical and legal issues in conducting research involving elderly subjects. *Behav. Sci. Law* 13(3), 319–335 (1995)
13. Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J., Brown, J.: Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention* 38, 185 (2006)

14. Hwang, S., Ryu, J.: The Haptic steering Wheel: Vibro-tactile based navigation for the driving environment. In: IEEE Intl' Conf. PERCOM Workshops, pp. 660–665 (2010)
15. International Organization For Standardization. ISO-26022: Road Vehicles – Ergonomic Aspects of Transport Information and Control Systems -Simulated Lane Change Test to Assess in-Vehicle Secondary Task Demand. International Standard. Geneve (2010)
16. Kay, L., Bundy, A., Clemson, L., Jolly, N.: Validity and reliability of the on-road driving assessment with senior drivers. *Accid. Anal. Prev.* 40, 751–759 (2008)
17. Kern, D., Marshall, P., Hornecker, E., Rogers, Y., Schmidt, A.: Enhancing Navigation Information with Tactile Output Embedded into the Steering Wheel. In: Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B., Tobe, Y. (eds.) *Pervasive 2009*. LNCS, vol. 5538, pp. 42–58. Springer, Heidelberg (2009)
18. Kim, S., Dey, A.K.: Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In: Proc. CHI 2009, pp. 133–142 (2009)
19. Kim, S., Dey, A.K., Lee, J., Forlizzi, J.: Usability of car dashboard displays for elder drivers. In: Proc. CHI 2011, pp. 493–502 (2011)
20. Kostyniuk, L., Streff, F., Eby, D.: The older driver and navigation assistance systems. The University of Michigan Transportation Research Institute, Final report UMTRI-97-47 (1997)
21. Lengenfelder, J., Schultheis, M.T., Ali-Shihabi, T., Mourant, R., DeLuca, J.: Divided Attention and Driving: A Pilot Study Using Virtual Reality Technology. *Journal of Head Trauma Rehabilitation* 17(1), 26–37 (2002)
22. Mattes, S., Hallén, A.: Surrogate Distraction Measurement Techniques: The Lane Change Test. In: Regan, M.A., Lee, J.D., Young, K.L. (eds.) *Driver Distraction: Theory, Effects, and Mitigation*, p. 107. CRC Press (2008)
23. Mayer, R.E.: *Multimedia Learning*, 2nd edn. Cambridge University Press (2009)
24. McCarthy, D.P.: Elder drivers and technology. In: Mann, W.C. (ed.) *Smart Technology for Aging, Disability and Independence*, pp. 247–283. John Wiley & Sons, N.J. (2005)
25. Mital, A., Govindaraju, M.: Is it possible to have a single measure for all work? *International Journal of Industrial Engineering Theory* 6, 190–195 (1999)
26. Mourant, R.R., Tsai, F.-J., Al-Shihabi, T., Jaeger, B.K.: Measuring Divided Attention Capability of Young and Older Drivers. *TRR Journal* 1779, 40–45 (2001)
27. Nachreiner, F.: Standards for ergonomics principles relating to the design of work systems and to mental workload. *Applied Ergonomics* 26(4), 259–263 (1995)
28. Strayer, D.L., Drews, F.A., Crouch, D.J.: A comparison of the cell phone driver and the drunk driver. *Hum. Factors* 48, 381–391 (2006)
29. Suen, S.L., Sen, L.: Mobility Options for Seniors. In: *Transportation in an Aging Society: A Decade of Experience*, pp. 97–113. Transportation Research Board (2004)
30. Eby, D.W., Trombley, D.A., Molnar, L.J., Shope, J.T.: The assessment of older drivers capabilities: A review of the literature. UMTRI-98-24 (1998)
31. Van Erp, J.B.F., Van Veen, H.A.H.C.: Vibro-Tactile Information Presentation in Automobiles. In: *Eurohaptics 2001 Conference*. University of Birmingham, Birmingham (2001)
32. Veltman, J.A., Gaillard, A.W.K.: Physiological indices of workload in a simulated flight task. *Biological Psychology* 42, 323–342 (1998)
33. Wickens, C.D.: Processing resources in attention. In: Parasuraman, R., Davies, R. (eds.) *Varieties of Attention*, pp. 63–101. Academic Press, New York (1984)
34. Yeh, Y., Wickens, C.D.: Dissociation of performance and subjective measures of workload. *Human Factors* 30(1), 111–120 (1988)