Review article

Touchscreen interfaces in context: A systematic review of research into touchscreens across settings, populations, and implementations

Andreas K. Orphanides [a], Chang S. Nam [b, *]

[a] User Experience Department, North Carolina State University, Raleigh, NC 27695 USA
[b] Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695 USA

ARTICLE INFO

ABSTRACT

Although many studies have been conducted on the human factors and ergonomics (HFE) of touchscreens, no comprehensive review has summarized the findings of these studies. Based on a schema (three dimensions of understanding critical for successful display selection) presented by Wickens et al. (2004), we identified three dimensions of analysis for touchscreen implementations: touchscreen technology, setting and environment of implementation, and user population. We conducted a systematic review based on the PRISMA protocol (Moher et al., 2009), searching five article databases for relevant quantitative literature on touchscreens. We found that all three dimensions of analysis have a significant effect on the HFE of touchscreens, and that a selection for or against touchscreens must take into consideration the specific context of system interaction in order to maximize safety, performance, and user satisfaction. Our report concludes with a set of specific recommendations for systems designers considering touchscreens as input/output devices, and suggestions for future study into the HFE of touchscreens.

© 2017 Published by Elsevier Ltd.

Contents

---

* Corresponding author.
  E-mail addresses: akorphan@ncsu.edu (A.K. Orphanides), csnam@ncsu.edu (C.S. Nam).

1. Introduction

1.1. Background and motivation

A touchscreen interface is a combined display/input device; the screen displays a graphical interface, and a user's physical touching of the screen is interpreted as an input or interaction with the interface, at the point of contact. Touchscreen interfaces were proposed over 50 years ago (Johnson, 1965), were first put into practice for use in the "Conseil Européen pour la Recherche Nucléaire", or European Council for Nuclear Research (CERN) particle accelerator in 1976 (Beck and Stumpe, 1973), and were in commercial production by 1983 (Hewlett-Packard, 2012). Since that time touchscreens have been implemented in a variety of devices and using an assortment of technologies, but in many ways, the release of the Apple iPhone in 2007 (Apple, 2007) heralded a new age for touchscreens, both as a milestone in touchscreen-based personal ubiquitous computing, and also as a hallmark of the possibilities that touchscreen interfaces permit in industrial and aesthetic design. As a result, in the last decade touchscreens have become ubiquitous across industrial, commercial, and consumer applications, in devices as diverse as industrial controls, medical equipment, photocopiers, refrigerators, home thermostats, desktop computers, and, of course, smartphones.

Touchscreens have obvious appeal from a systems design perspective, as they combine display and input affordances (thus providing space and design efficiency), while simultaneously offering human factors and ergonomics (HFE) benefits, such as the direct mapping of inputs to targets. However, the decision to use a touchscreen as a system input results in certain HFE compromises as well, including concerns such as a lack of tactile input affordances, ergonomic compromises leading to fatigue and discomfort (Kang and Shin, 2014; Shin and Zhu, 2011; Young et al., 2012), and factors that compromise the quality of direct mapping of inputs and targets such as the "fat finger problem" (e.g., Siek et al., 2005) and parallax concerns due to physical separation of the input and display layers of the device (Leahy and Hix, 1990). This suggests that choosing a touchscreen as an input device may incur a tradeoff from an HFE perspective, which may influence the safety, task

performance, and satisfaction of the users of the system. Further, the precise effect of touchscreen-as-input may depend on the context of the system, in particular the system's setting, the user population interacting with the system, and touchscreen implementation selected.

Despite the current dominance of touchscreens in contemporary systems engineering, and despite extensive human factors research into touchscreen interfaces in a multitude of contexts, there has been to date no comprehensive review of formal research into the human factors and ergonomics of touchscreens. Such a review would be an invaluable asset to a system designer, allowing her or him to make a determination about whether and how to integrate a touchscreen into the system, given the system's environment and tasks, the properties of the system's users, and the possible touchscreen implementations available. For this reason, we present a review of the human factors and ergonomics of touchscreens, examining the HFE concerns of touchscreens and touchscreen systems through the lenses of system setting, user population, and touchscreen implementation. We intend this review to serve two purposes: (1) for the system practitioner described above, to inform the process of evaluating whether to integrate a touchscreen interface into a system, based on the system's settings and tasks, the users of the system, and the touchscreen implementations under consideration; and (2) for the human factors researcher, to identify the contexts (i.e., settings and tasks, populations, and touchscreen implementations) that have been well-studied, those that have had promising initial studies conducted and may hold opportunities for additional research, and (through their absence) contexts that are ripe for new research.

1.2. Review objectives

This review is concerned with evaluative studies of touchscreen interfaces from a HFE perspective. Of specific concern are the effects of touchscreen interfaces on safety, task performance, and user satisfaction in relation to their contexts of use. We consider contexts of use along three dimensions: the implementation of the touchscreen, the task the touchscreen is serving, and the characteristics of the user interacting with the touchscreen. We base these

dimensions on a schema provided by Wickens et al. (2004), to drive the selection and development of displays, which identifies three dimensions of understanding critical for successful display selection: (1) the **physical implementation** of the display, (2) the **task** that the display supports, and (3) the **characteristics of the human** who interact with the system. Wickens et al. (2004) intended this schema for display technologies specifically; however these same dimensions provide a useful taxonomy of concerns for system inputs as well. A touchscreen by its very definition is both display and input, so this schema provides a useful framework for critical investigation into HFE touchscreen research. For the purposes of analyzing touchscreens, we understand the three dimensions as follows:

- **Physical implementation** entails the properties that are intrinsic to the touchscreen, in particular with respect to its hardware and software design; this includes the physical components, the drivers, the information the screen capable of communicating, the inputs it is capable of receiving, etc. Aspects under consideration here include the technology used to sense and interpret touches, the ability to provide visual, auditory, or haptic feedback, the capability of receiving and interpreting gestures (tap, drag, two-finger rotate, pinch zoom, etc.), as well as physical size, display resolution, etc.
- **Task** includes the setting and environment in which the touchscreen is being used, the tasks that are being performed on the touchscreen or in conjunction with its use, environmental factors that may affect human capability in the context of touchscreen use, etc. Examples of this dimension include settings such as driving an automobile, as well as environmental variables such as ambient temperature or uncontrolled external motion.
- **Characteristics of the human**, for the purpose of this review, include factors intrinsic to the user that they have no or limited capability to change at will, which may have an effect on the user's capabilities in HFE terms. These characteristics include age, physical ability, handedness, etc.

A pictorial representation of a human-touchscreen system, including identification of the key system components outlined above, is provided in Fig. 1.

To this end, we offer the following research questions based on these dimensions of analysis:

- RQ1. What aspects of touchscreen implementation — hardware selection, software implementation, interface design, interaction design, etc. — have been evaluated? More specifically, how do they affect safety, performance, and user satisfaction? Do particular implementation choices yield better outcomes?
- RQ2. What environmental and task settings have touchscreen interfaces been evaluated in? More specifically, how does the use of a touchscreen in each context affect the safety, performance, and user satisfaction of the task in question or associated tasks? What conclusions can be drawn about settings and tasks for which touchscreens are well or poorly suited?
- RQ3. What populations have been evaluated as touchscreen users? More specifically, how does the use of a touchscreen by each population affect safety, performance, and user satisfaction? Are touchscreens more appropriate for some users than for others given these criteria?

1.3. Review structure

The remainder of the review is structured as follows. In section

2, we describe our search methodology, covering the information sources searched, the criteria for inclusion (including both general principles and inclusion criteria specific to individual research questions), the search strategies used in database searching and search limits applied, and a summary of search results and result counts based on the PRISMA 2009 structured search methodology (Moher et al., 2009), demonstrating the steps that resulted in the 59 articles included in this review. In section 3 we provide detailed findings for each of our three research questions, including summaries of each article that addresses the research question, narrative summaries of the research, comparison and contrast of article findings, identification of common themes, and gathering of common results. In section 4 we summarize our findings, highlight the most relevant and notable trends, suggest avenues for future research into this topic, and present our conclusion.

2. Search methodology

We conducted this search using the PRISMA 2009 structured search methodology (Moher et al., 2009), a phased approach to systematic literature reviews and meta-analyses which guides the search process through four phases: *identification*, in which candidate articles are identified primarily through exhaustive database searching (plus manual selection, serendipitous discovery, etc.); *screening*, where candidate articles are deduplicated and evaluated based on their abstracts and other metadata; *eligibility*, where articles are evaluated through full-text analysis; and *inclusion*, where the final set of included articles is identified for review. We applied the PRISMA protocol to identify literature relating to touchscreen interfaces and their safety, efficiency, and user experience across different contexts. By design, our initial search criteria and search environments were extremely broad, in order to ensure the diversity of research settings and perspectives included in the review.

2.1. Information sources

We searched the following article databases: (1) Compendex/INSPEC searched together (taking advantage of the databases' built-in federated searching feature) for a broad engineering perspective, including ergonomics and human factors, (2) IEEE Xplore, for a computer engineering perspective, (3) ACM Digital Library, for a computer science and user experience design perspective, (4) Health and Safety Science Abstracts, for additional perspectives on ergonomics and safety, and (5) Web of Science, for a cross-disciplinary perspective.

2.2. Inclusion and exclusion criteria

We included original, quantitative research on the ergonomics and human factors of touchscreens, including safety, task efficiency and success rate, and user satisfaction. We included only articles published within the last 10 years, to ensure that the included research reflects current technologies, user populations, and implementation approaches; the 2007 cutoff also coincides with the launch date of the first iPhone, which we see as an informal marker of the dawn of the current touchscreen era. Research articles were required to be in English and published in scholarly, peer-reviewed journals. Conference papers were excluded, even if republished in a scholarly journal, as these papers often do not receive the same level of scrutiny as original papers. In addition, studies on the following topics were excluded: (1) Novel or prototype hardware, interfaces, or interaction designs, not in general use, (2) Handwriting recognition, (3) Validation of existing human factors or HCI models or proposals of new or updated models (some model validation studies are included if the experiment conducted
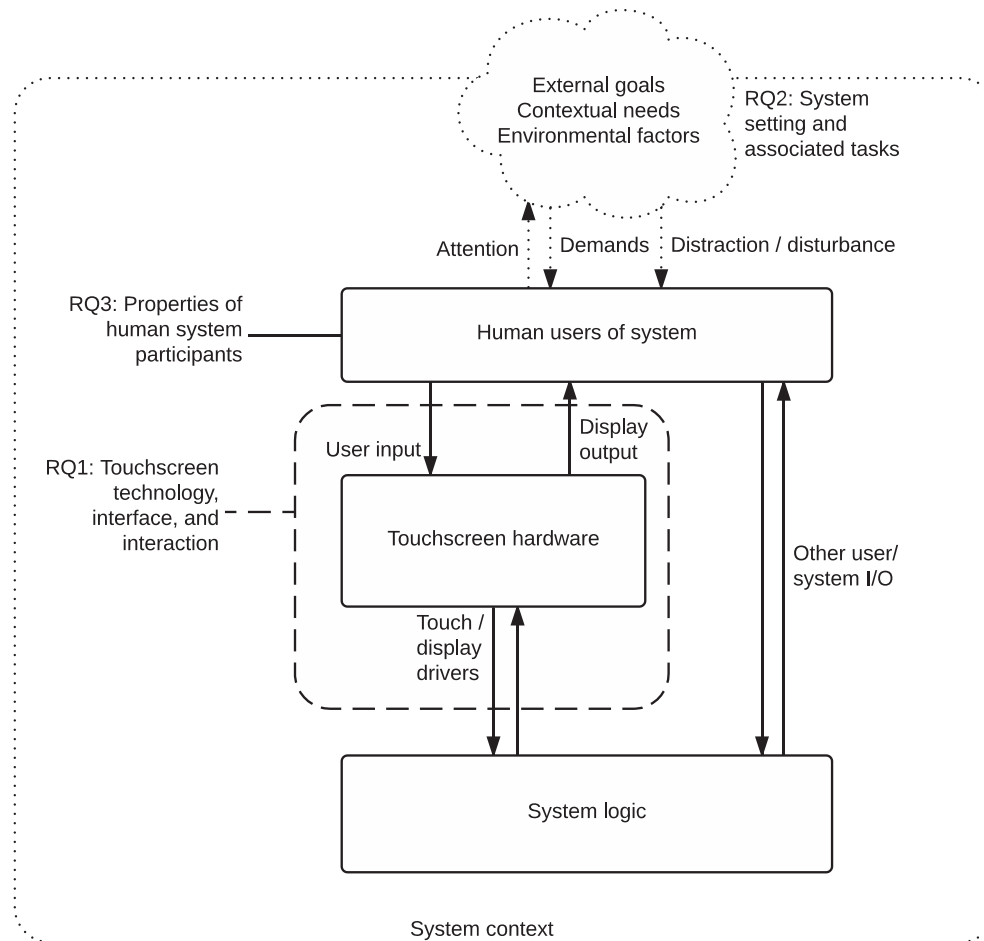
Fig. 1. Flow diagram representing the major components of a system including touchscreen display/input. Components corresponding to research questions in this review are: 1) the touchscreen technology used, including interface design, interaction design, hardware, and drivers, represented by the elements enclosed by the dashed line; 2) the system setting and context, including other tasks contemporaneous to the system tasks, the demands of the setting, environmental factors that may interfere with system interaction, etc.; key features of the system context are represented in dotted lines, though the entirety of the system is ipso facto included in the context; 3) intrinsic properties of the human users of the system.

to validate the model contributes useful findings to one of the research questions independent of model validation), and (4) Purely qualitative studies, ethnographic studies, case studies, etc.

Full-text review of articles included the following analyses of populations, interventions, comparisons, outcomes, and study design:

- Populations: All populations were considered. Articles that named a specific population (e.g., "older adults", "children") as a population of interest in their research question, or compared two or more populations as a study factor, and otherwise met the criteria for inclusion, were considered under RQ3.
- Interventions: No specific dimensions of intervention were screened for, but interventions varying touchscreen implementation were considered under RQ1, and interventions varying contextual task settings were considered under RQ2.
- Comparisons: Any comparison that could reasonably inform the decision of whether to use a touchscreen input vs. an alternative, or one touchscreen implementation vs. another, was included.
- Outcomes: Studies were required to include a quantitative measure of safety, task performance, or user satisfaction. It was acceptable for user satisfaction measures to be based on

subjective questionnaires (e.g., Likert scale evaluations of preference).
- Study design: Given the breadth of our research questions, variations in accepted study design across disciplines, and the relative paucity of large-n studies, all quantitative study types and sample sizes were considered. Any study which adequately described the details of the study design, whose design had reasonable face validity, and which indicated its method of statistical analysis of results was included.

In addition, for inclusion in our review, a study's research question had to explicitly inquire along one of our dimensions of analysis, i.e., implementation, setting, or population. Papers that did not examine touchscreens through one of these lenses were excluded. In particular:

- For inclusion in RQ1, a study's research question had to investigate the implementation of a touchscreen itself: evaluating hardware or driver technologies, user interfaces in a touchscreen-specific context, or interaction design in a touchscreen setting.
- For inclusion in RQ2, a study's research question had to inquire about the fitness or design of touchscreens or touchscreen devices in a specific setting or environment, such as in a clinical

care environment, during road driving, or for teleoperation of robots.

- For inclusion in RQ3, a study's research question had to inquire about the use of touchscreens by a specific population (e.g., age group, mental or physical disorder, etc.), or had to compare touchscreen use across two or more populations. For a population to be of interest, the characteristic defining the population had to be intrinsic to that population, such as age, handedness, or physical capability (i.e., not something that the population members could change with relative ease of their own accord).

Finally, in cases where it was clear that the same research had been published in two or more articles, the article that appeared to provide the richest detail into the research was kept, while the duplicate-reporting articles were removed from consideration.

2.3. Search strategy and limits

We sought a broad perspective on human factors concerns related to the selection, implementation, and use of touchscreens. To that end, our search terms were high-level and cross-disciplinary. The basic search query structure for each database was: (1) Search terms to capture the notion of touchscreens, such as "touchscreen", "touch display", and "touch-sensitive screen", including plural variants, and (2) Search terms related to our key concerns, including terms such as "human factors", "ergonomics", "user experience", "safety", "efficiency", "task completion", etc.

These terms were sought across all available metadata fields. In addition, when possible, we included limits in the form of additional search terms or database filters to limit searches to those meeting the inclusion criteria. We also applied any available filters to limit to articles meeting our search criteria; in particular we applied filters to include only results in English, to include results from peer-reviewed publications only, and/or to exclude non-article results.

All databases were able to support at least the key search terms and the date range limits. Further, the Compendex/INSPEC search performs a federated search of both the Compendex and INSPEC databases, and offers a deduplication feature; results counts for Compendex/INSPEC in Section 2.4 reflect the count after automatic deduplication.

2.4. Search results

As shown in Fig. 2, initial database searches yielded 1237 results, including 444 from Compendex/Inspec, 174 from IEEE Xplore, 121 from ACM Digital Library, 38 from Health and Safety Science Abstracts, and 454 from Web of Science. Manual removal of duplicate records excluded 299 records, yielding 938 unique articles for consideration. Examination of abstracts and other record metadata excluded a further 712 articles, all due to not adequately conforming to any of our research questions, leaving 226 articles for full-text analysis. Full-text analysis of these articles excluded a further 167 records for not meeting one or more of the requirements laid out in section 2.2. Thus, a total of 59 articles addressing one or more of the research questions were identified. Of these, 30 articles addressed RQ1, 18 addressed RQ2, and 26 addressed RQ3 (16 articles addressed two of our research questions while no article addressed all three).

3. Review results

We identified 59 articles that address at least one of our three research questions. These articles reported a total of 68 studies that are included in this analysis. Tables 2–4 summarize the identified studies with respect to each of the three research questions.

3.1. Study characteristics

3.1.1. Number of participants

The largest study included in this review had $n = 841$ (Herrick et al., 2013), and the smallest had $n = 10$ (Lin et al., 2010). The average sample size across all 68 studies was 48.20. Several of the highest-n studies (including Herrick et al., 2013) were cross-sectional studies conducted in a high-volume clinical practice; omitting the four cross-sectional studies, the largest studies had $n = 90$ (Chang et al., 2014; Jochems et al., 2013) and the average study size was 32.77. Participants in studies in RQ2 with setting-specific tasks were typically familiar with the task domain. Studies not in RQ3 all focused on adult populations.

3.1.2. Study design

With the exception of four cross-sectional studies (Hammon et al., 2014; Herrick et al., 2013; Pal et al., 2013), all studies were within-subjects, between-subjects, or mixed factorial design; typical design was full-factorial within-subjects except when population comparisons necessitated a between-subjects or mixed design: across all mixed and between-subjects studies (20 studies), only four non-population study factors were between-subjects (Eliav et al., 2011; McLaughlin et al., 2009; Sonderegger et al., 2016; Young et al., 2014). In addition, one study (Huang and Chen, 2010) made use of the method of adjustment for one of its research questions; otherwise it was a within-subjects study.

3.1.3. Touchscreen hardware

All but 14 of the studies specified the make or model of the touchscreen(s) used in the study. All but three studies either provided the size of the devices studied or included sufficient contextual information to determine the approximate size of the touchscreen evaluated. Twenty-five studies examined phone-size touchscreens (less than 6″), 19 studies examined tablet-size screens (less than 12″), 25 studies examined desktop monitor-size screens (less than 25″), and two studies examined large-format displays (25" or larger). Touch-sensing technology was generally not explicitly stated; seven studies indicated that capacitive screens were under study, six indicated resistive screens, and the four studies from Gao and Sun (2015) compared screens with optical and surface acoustic wave (SAW) touch sensing. For other studies that reported touchscreen make and model, this is presumably enough information to identify the size and technology of the screen in use, but in our summary tables we have provided size and sensing technology only when it was reported in the original study.

3.1.4. Input method comparisons

Eleven studies evaluated two or more touchscreens or touchscreen technologies in comparison. In addition, 27 studies compared touchscreens to a variety of alternate inputs. Six studies compared touchscreens to mice; six to physical keypads or keyboards; five to rotary controllers; four to trackballs; four to stylus input; two to eye tracking, and one each to hand gestures and voice input. In addition, five studies compared touchscreen devices holistically to alternatives.

3.2. Research question 1

Thirty articles addressed RQ1. The aspects of touchscreen implementations that were evaluated fell into four categories:
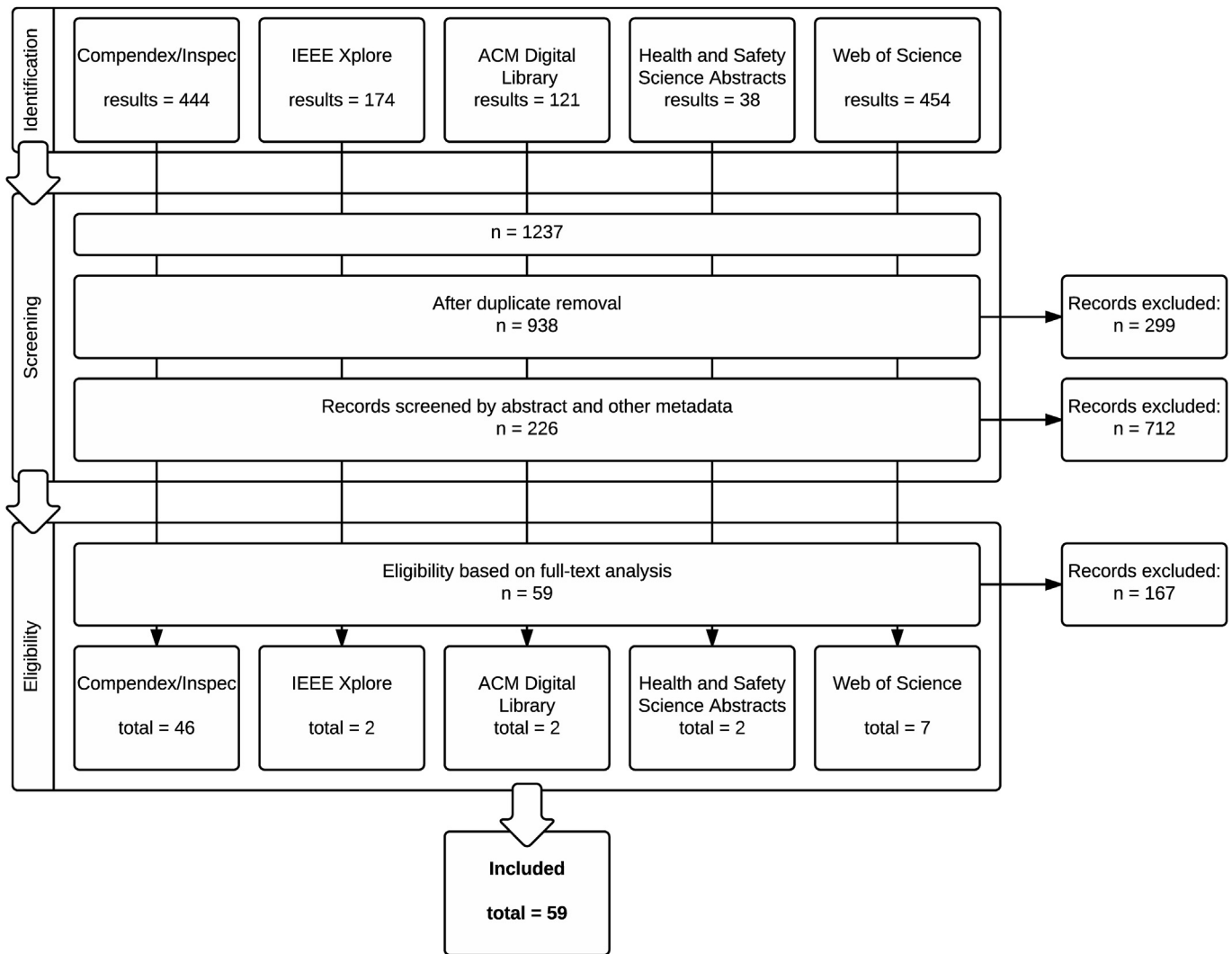
Fig. 2. PRISMA flow diagram for this review.

- *Hardware and drivers*: aspects relating to the technical capabilities of touchscreens that are intrinsic to their manufacture or the low-level driver software. Examples include screen size, touch-sensing technology, palm rejection, etc.
- *Interface*: aspects relating to the design of interfaces, including the dimension and configuration of buttons or other control features of the display. Examples include size and arrangement of touch targets.
- *Interaction − input*: aspects relating to interaction design, specifically variations on input paradigms. Examples include touch gestures and single vs. multitouch interaction.
- *Interaction − output*: aspects relating to interaction design, specifically variations on touch response. Examples include visual, auditory, or tactile feedback in response to user touch.

3.2.1. Hardware and drivers

Five articles evaluated a touchscreen's hardware and drivers. Two (Chang et al., 2014; Chung and Shin, 2015) addressed physical aspects of the touchscreen itself; two (Burnett et al., 2013; Gao and Sun, 2015) compared touch input technologies; and one (Camilleri et al., 2013) examined driver technology.

Physical touchscreen implementations were evaluated with respect to size and touch surface treatment. With regard to size, Chang et al. (2014) used a performance value measure (Fitts, 1992) to compare the usability of different touchscreen sizes for users of different ages (30 children, 30 young adults, and 30 elderly adults), evaluating separately for tap, tap-and-drag, zoom, and rotate gestures. All users, especially children, encountered difficulty with the rotation gesture at screen sizes of 23″ and greater; meanwhile, dragging was problematic for all users at smaller screen sizes (4.3″ and 10″), but especially for elderly participants. Chung and Shin (2015) evaluated whether an elasticized surface on a touchscreen would improve performance, fatigue or user satisfaction in a sequential tapping task on a handheld (4.8″) device; no significant difference was found between a hard and elastic surface for finger input.

Resistive, capacitive, optical, and surface-acoustic wave (SAW) touch input technologies were evaluated in the identified studies. Burnett et al. (2013) compared resistive and capacitive touchscreens in an in-vehicle information system (IVIS) used for completing secondary tasks during simulated driving, evaluating the performance of 38 experienced (2–38 years driving) adults. The capacitive touchscreen yielded better performance of secondary tasks, improving task times by 27% over resistive touchscreens, and better user perceptions of quality and performance, while

Table 1a
Articles identified under RQ1 relating to hardware and driver selection.

	Burnett et al., 2013	Camilleri et al., 2013	Chang et al., 2014	Chung and Shin, 2015
Affiliation of Primary Author	University of Nottingham, Nottingham, UK	University of California, Berkeley, CA, USA	Chang Gung University, Taoyuan, Taiwan	Sungkyunkwan University, Seoul, South Korea
Aim	Evaluate capacitive and resistive input in a touchscreen IVIS	Evaluate ergonomics of palm rejection	Evaluate touchscreen gesture usability across age groups	Evaluate effect of surface elasticity on touchscreen usability
Study design	Within-subjects 2 × 9	Within-subjects 2 × 2	Mixed 3 × 4	Within-subjects 2 × 3
Devices considered	Elonex eTouch (resistive) (10″); Motorola Xoom (capacitive) (10.1″)	3M 2256 PW (capacitive) (22″)	HTC EVO 4G (4.3″); ASUS Eee Pad Transformer TF101 (10.1″); Acer T231H (23″); G-Touch MI Series (42″)	Samsung Galaxy S3 (4.8″); Acrylic stylus; rubber stylus
Activity	Road driving (simulated) – follow lead vehicle while completing secondary tasks	one-touch, two-touch, and three touch dragging tasks	Drag, scale, rotate tasks	Sequential tapping task
Participants	20; driving experience 2—38 years	31; familiar with touchscreen use	30 children, 30 young adults, 30 elderly	44: all touchscreen smartphone owners
Age	38 (24—63)	40.8 ± 15 (18—65)	children: 12.4 (11—14) adults: 23.5 (20—28) elderly: 70.6 (65—84)	23.25 ± 2.08 (19—28)
Independent variables	Touchscreen type (2); task (9)	Palm rejection technology (enabled, disabled); task configurations require palm motion (yes, no)	Touchscreen size (4); age group (3)	Touchscreen surface (bare glass, elastic coating); input method (finger, hard stylus, rubberized stylus)
Dependent variables	Task time; error count and type; glance measures; standard deviation of lane position; subjective ratings of touchscreen sensitivity, quality, preference	Task speed; force; shoulder movement; self-selected display position; subjective discomfort; subjective productivity	Performance value	Task performance; subjective satisfaction, sensation, and fatigue measures
Statistical approach	Paired t-test	Repeated-measures ANOVA	ANOVA, t-test	Two-way repeated-measures ANOVA

	Gao and Sun, 2015			
Affiliation of Primary Author	Tsinghua University, Beijing, China			
	Study 1	Study 2	Study 3	Study 4
Aim	Evaluate touchscreen gesture usability across age groups (clicking)	Evaluate touchscreen gesture usability across age groups (dragging)	Evaluate touchscreen gesture usability across age groups (zooming)	Evaluate touchscreen gesture usability across age groups (rotation)
Study design	Mixed 2x2x4x4x4	Mixed 2x2x4×5	Mixed 2 × 2 × 4	Mixed 2 × 2 × 4
Devices considered	HP TouchSmart 600-1168cn (optical) (23″); Lenovo A700 (surface acoustic wave (SAW)) (23″)	HP TouchSmart 600-1168cn (optical) (23″); Lenovo A700 (SAW) (23″)	HP TouchSmart 600-1168cn (optical) (23″); Lenovo A700 (SAW) (23″)	HP TouchSmart 600-1168cn (optical) (23″); Lenovo A700 (SAW) (23″)
Activity	Target acquisition task (click)	Drag and drop task (drag)	Object resizing task (zoom)	Object rotation task (rotate)
Participants	40 older adults; 40 younger adults	40 older adults; 40 younger adults	40 older adults; 40 younger adults	40 older adults; 40 younger adults
Age	Older: 64.53 ± 7.40 (52—81) Younger: 21.62 ± 1.13 (19—24)	Older: 64.53 ± 7.40 (52—81) Younger: 21.62 ± 1.13 (19—24)	Older: 64.53 ± 7.40 (52—81) Younger: 21.62 ± 1.13 (19—24)	Older: 64.53 ± 7.40 (52—81) Younger: 21.62 ± 1.13 (19—24)
Independent variables	Touch technology (optical, SAW); age group (younger, older); inclination angle (4); button size (4), button spacing (4)	Touch technology (optical, SAW); age group (younger, older); inclination angle (4); drag direction (4); drag distance (5)	Touch technology (optical, SAW); age group (younger, older); inclination angle (4)	Touch technology (optical, SAW); age group (younger, older); inclination angle (4)
Dependent variables	Error rate; task time; subjective measures	Error rate; task time; subjective measures	Error rate; task time; subjective measures	Error rate; task time; subjective measures
Statistical approach	Mixed-model ANOVA	Mixed-model ANOVA	Mixed-model ANOVA	Mixed-model ANOVA

simultaneously resulting in a significant reduction in eyes-off-road time. On the strength of these results, the authors unequivocally recommended capacitive technology over resistive, but noted that capacitive screens are incompatible with most gloves. Gao and Sun (2015) compared optical and SAW touch sensing in four gesture-specific studies (tap, drag, zoom, rotate) among older adults (52—81) and younger adults (19—24). For both groups, clicking, dragging, and rotating tasks took longer on the optical screen than the SAW screen, and the optical screen was more error-prone for clicking and rotating, but optical screens were perceived as faster, less forceful, and less fatiguing. The authors recommended SAW for cases where unintentional activation should be avoided, but

otherwise recommended optical detection on the merits of its strong subjective acceptance; they also reminded the reader that interfaces should be designed to match the strengths and limitations of the underlying touch technology.

The only driver technology evaluated was palm rejection, in Camilleri et al. (2013). This study evaluated the effect of palm rejection technology on participant biomechanics, comfort, and task performance while executing one-, two-, and three-point gestures. Enabling palm rejection yielded greater wrist extension and touch force, less reported discomfort, and (for multitouch gestures) improved task speed. One of the authors' hypotheses was that users would rest their palms on the screen surface when palm

Table 1b
Articles identified under RQ1 relating to interface design.

	Garcia-Lopez et al., 2015	Huang and Chen, 2010	Hwangbo et al., 2013	Im et al., 2015
Affiliation of Primary Author	University of Alcalá, Spain	Southern Taiwan University, Taiwan R.O·C	Yonsei University, Seoul, Korea	Korea University, Seoul, Republic of Korea
			Study 1	
Aim	Evaluate content navigation methods in mobile touchscreens	Evaluate depth cues and contrast in touchscreen icons	(See Table 3a)	Evaluate icon layout in icon image to target ratio in search task
Study design	Within-subjects 2 × 3	Method of adjustment followed by within-subjects (3 levels)		Within-subjects 4 × 3
Devices considered	iPhone 4 (3.5″)	3M LCD/MD17-5MS (17″)		HTC Desire HD (4.3″)
Activity	Information retrieval task	Icon adjustment and selection		Icon search-and-touch task
Participants	23	30: 15 novices, 15 experienced with touchscreens		40: experienced in using mobile touch devices
Age	Precise age demographics not provided	25.5 ± 2.66 (22—32)		27.5 (22—36)
Independent variables	Content distribution method (paging, internal linking, vertical scrolling); target distance (near, far)	Icon bevel depth cue (adjustment variable); icon visual cue region (3)		Icon layout (4); icon ratio (3)
Dependent variables	Error rate; task completion time; subjective satisfaction measures	Icon bevel depth; distribution of touched locations		Input offset; hit rate; task completion time; subjective preference
Statistical approach	Friedman test	Frequency statistics, ANOVA, Duncan test		4 × 3 repeated-measures ANOVA

	Irune and Burnett, 2007		Kim et al., 2014a	Kim et al., 2014b
Affiliation of Primary Author	University of Nottingham, Nottingham, UK		Northern Illinois University, DeKalb	Pohang University of Science and Technology, Pohang, South Korea
	Study 1	Study 2		
Aim	Evaluate control element configuration in touchscreen IVIS	Evaluate control element configuration in touchscreen IVIS	Evaluate ergonomics of key size in virtual keyboards	Evaluate element size in touchscreen IVIS
Study design	Within-subjects 3 × 3 × 3	Within-subjects 4 × 16	Within-subjects (4 levels)	Within-subjects 5 × 3
Devices considered	Unspecified 8.5″ touch display	Unspecified 8.5″ touch display	Acer Iconia (14″)	Digital Windows Communication dwcom17s (17″)
Activity	Road driving (simulated) – right hand drive vehicle: Follow lead vehicle, complete secondary tasks	Road driving (simulated) – right hand drive vehicle: Follow lead vehicle, complete secondary tasks	Typing task	Road driving (simulated) – maintain a target speed while completing secondary task
Participants	18; held drivers' license for at least 2 years	20; held drivers' license for at least 2 years	21; all with typing speed over 40wpm	30; average 4.7 ± 3.7 years driving experience (1—15 years)
Age	(18—36)	(18—45)	24.5 (18—49)	28 ± 3.6 (22—38)
Independent variables	Cluster size (3); control size (3); control spacing (3)	Cluster size (4); cluster arrangement (16)	Virtual keyboard key size (4)	Virtual keyboard key size (4); driving speed (3)
Dependent variables	Difficulty rating; task time; glance frequency; glance measures	Task time; glance measures	Typing forces; muscle activity; wrist posture; typing performance; subjective preference	Task time; task error rate; NASA-TLX; subjective size preference; driving safety measures
Statistical approach	Unspecified correlation analysis	Unspecified correlation analysis	Mixed model with restricted maximum likelihood estimation	ANOVA

	Kujala and Salvucci, 2015		Park and Han, 2010a	Park and Han, 2010b	Sesto et al., 2012
Affiliation of Primary Author	University of Jyväskylä, Jyväskylä, Finland		Pohang University of Science and Technology, Pohang, Republic of Korea	Pohang University of Science and Technology, Pohang, Republic of Korea	(See Table 3b)
	Study 1	Study 2			
Aim	Evaluate grid menu and list menu usability in touchscreen IVIS	Evaluate grid menu and list menu usability in touchscreen IVIS	Evaluate icon size and location in one-handed mobile touchscreen use	Evaluate icon size and location in one-handed mobile touchscreen use	
Study design	Within-subjects 2 × 3	Within-subjects 2 × 3	Within-subjects 3 × 25	Within-subjects 3 × 25	
Devices considered	3M M2256PW (22″) display	3M M2256PW (22″) display	HP iPAQ rz1717 (3.5″)	HP iPAQ rz1717 (3.5″)	
Activity	NHTSA (2013) driving task	NHTSA (2013) driving task	Target acquisition task	Target acquisition task	
Participants	12; at least 2 years or 20,000 km driving experience	11	40: 24 had no experience using mobile devices	30: 20 had not used touchscreen handheld devices	
Age	(22—34)	"Same demographics" as study 1	22.6 ± 2.5 (18—28)	23.1 ± 2.5 (18—28)	

Table 1b (continued)

	Kujala and Salvucci, 2015		Park and Han, 2010a	Park and Han, 2010b	Sesto et al., 2012
Independent variables	Menu type (2), Menu size (3)	Menu type (2), Menu size (3)	Icon size (3); icon location (25)	Icon size (3); icon location (25)	
Dependent variables	Glance measures	Glance measures	Success rate; input offset	First transition time; task completion time; number of errors; subjective pressing convenience	
Statistical approach	Repeated-measures ANOVA	Repeated-measures ANOVA	One-way ANOVA; repeated-measures ANOVA	Two-way ANOVA	

Table 1c
Articles identified under RQ1 relating to interaction design (inputs).

	Bjorneseth et al., 2012	Chang et al., 2014	Cockburn et al., 2012		Gao and Sun, 2015
Affiliation of Primary Author	Rolls-Royce Marine AS, Ålesund, Norway	(See Table 1a)	University of Canterbury, Christchurch, New Zealand		(See Table 1a)
			Study 1	Study 2	
Aim	Compare single-touch, gesture inputs in maritime touch systems		Evaluate effect of input type on tap/drag gestures	Evaluate effect of input type on radial drag gestures	
Study design	Within-subjects 2 × 2		Within-subjects 3x2x7×2	Within-subjects 3 × 7 × 4	
Devices considered	Dell tablet computer; Rolls Royce DP system interface		HP TouchSmart 600 (23″); mouse; stylus-based tablet	HP TouchSmart 600 (23″); mouse; stylus-based tablet	
Activity	Input commands using different touch interaction styles		Target acquisition task	Radial dragging task	
Participants	19		18: all computer users; infrequent stylus and touchscreen use	18: all computer users; infrequent stylus and touchscreen use	
Age	23 (20–32)		31.33 ± 9.54 (21–59)	31.33 ± 9.54 (21–59)	
Independent variables	Simulator motion (stationary, moving); touch interaction style (one-finger menu-based, two-finger gesture-based)		Interaction type (tap, drag); input method (mouse, finger, stylus); index of difficulty (7); target direction (left, right)	Input method (mouse, finger, stylus); number of items (7); target direction (NE, SE, NW, SW)	
Dependent variables	task completion time; error rate		Error rate; target acquisition time	Error rate; target acquisition time; movement distance	
Statistical approach	Repeated measures ANOVA		3 × 7 repeated measures ANOVA	3 × 4 × 6 repeated measures ANOVA	

	Kujala, 2013	Kwon et al., 2011	Mihajlov et al., 2015	Vatavu et al., 2015
Affiliation of Primary Author	University of Jyväskylä, Jyväskylä, Finland	Pohang University of Science and Technology, Pohang, Republic of Korea	Ss. Cyril and Methodius University, Macedonia	University Stefan cel Mare of Suceava, Suceava, Romania
Aim	Evaluate scrolling methods on touchscreen IVIS	Evaluate control gain during drag on small touchscreen devices	Determine gesture learnability among older adults	Evaluate effect of age on touchscreen performance in children
Study design	Within-subjects (3 levels)	Within-subjects 5 × 2	Within-subjects (3 levels)	Mixed 3 × 2 × 4
Devices considered	iPhone 3G S (3.5″)	iPaq-hx2490b (3.5″)	SMART Table 230i (27.5″)	Samsung Galaxy Ace S5830 phone (3.5″); Acer Iconia Tab A500 tablet (10.1″)
Activity	Road driving (simulated) – maintain target speed while completing secondary task	Target acquisition task	Two drag tasks; one rotate task	Tap, double tap, drag and drop, double drag and drop tasks
Participants	24; at least 2 years driving experience	30: no experience with tap-and-drag gesture	17; described as "technology-naive"; none had used computer or smartphone	89 children between ages 3 and 6
Age	25 ± 3.6 (19–34)	23.8 ± 2.0	67.32 (54–82)	Younger than 4: 3.61 ± 0.23 4–5: 4.53 ± 0.26 Older than 5: 5.68 ± 0.45
Independent variables	Touchscreen scrolling method (3)	Gain amount (5), gain direction (push background, push viewport)	Task (drag learning, drag implementing, rotate learning)	Age group (3); device (phone, tablet); task (tap, double tap, drag + drop, double drag + drop)
Dependent variables	Glance measures; lane excursion; speed maintenance; secondary task error rate; NASA-TLX; subjective measures	Task completion time; number of touches; total navigation distance; subjective measures	Gesture learning time; gestures recorded; task time	Completion rate; task time; offset; path accuracy; motor/visual/spatial metrics
Statistical approach	One-way repeated-measures ANOVA	Within-subject ANOVA	Friedman's ANOVA, Mann-Whitney test	Repeated-measures ANOVA

rejection was enabled; this generally did not occur, suggesting that improvements in performance due to palm rejection were not due to users deliberately using their palms for support during tasks.

3.2.2. Interface

Thirteen studies among eleven papers addressed interface design with respect to touchscreen HFE concerns. Five studies discuss icon layout and arrangement in a motor vehicle IVIS,

Table 1d
Articles identified under RQ1 relating to interaction design (outputs).

	Han and Kim, 2015	Hwangbo et al., 2013	Parikh and Esposito, 2012	
Affiliation of Primary Author	Korea Advanced Institute of Science Technology, Daejeon, Korea	Yonsei University, Seoul, Korea	United States Naval Academy, Annapolis, MD	
		Study 2	Study 1	Study 2
Aim	Evaluate tactile feedback in virtual QWERTY keyboards	(See Table 3a)	Evaluate negative feedback in touch interface performance	Evaluate negative feedback in touch interface performance
Study design	Within-subjects 3 × 2		Within-subjects (2 levels)	Within-subjects (2 levels)
Devices considered	Nokia Lumia 520 phone; Microsoft Surface Pro RT tablet		iPhone 4	iPhone 4
Activity	Typing task		Keypad number entry	Typing task
Participants	12; skilled typists with touchscreen experience		12: averaging 36.6 touchscreen uses per day	12: averaging 35.2 touchscreen uses per day
Age	24.42 ± 1.68 (22–27)		20–37	21–38
Independent variables	Feedback type (Visual, visual + auditory, visual + haptic); device (Lumia, Surface)		Feedback type (positive, negative)	Feedback type (positive audio + negative tactile, positive audio only)
Dependent variables	Keystrokes per character; words per minute; corrected error rate; uncorrected error rate; subjective evaluation		Accuracy; keystrokes per character; keystrokes per second; NASA-TLX; subjective measures	Accuracy; keystrokes per character; keystrokes per second; NASA-TLX; subjective measures
Statistical approach	Repeated-measures ANOVA		t-test	t-test

	Pitts et al., 2012a	Pitts et al., 2012b	Williamson et al., 2011
Affiliation of Primary Author	University of Warwick, Coventry, UK	University of Warwick, Coventry, UK	Queensland University of Technology, Queensland, Australia
Aim	Evaluate visual and haptic feedback in a touchscreen IVIS	Evaluate visual, audio, haptic feedback in a touchscreen IVIS	Evaluate visual and audio feedback in a touchscreen IVIS
Study design	Within-subjects 3 × 2	Within-subjects (4 levels)	Within-subjects (4 levels)
Devices considered	8.4″ TouchSense haptic touchscreen	8.4″ TouchSense haptic touchscreen	Unspecified touchscreen
Activity	Road driving (simulated) – follow lead vehicle in traffic, while completing secondary tasks	Road driving (simulated) – lane-change test, while completing secondary task	Road driving (simulated) – lane-change test, while completing secondary task
Participants	36; at least 1 year driving experience	48; at least 1 year driving experience	30; licensed for average 7.9 years, 9 drivers with probationary license
Age	38.1 ± 12.6 (20–65)	Precise demographics not provided	25.7 ± 7.3 (20–54)
Independent variables	Visual feedback delay(immediate, delayed, none); Feedback mode (visual, visual + haptic)	Feedback mode (visual, visual + auditory, visual + tactile, visual + tactile + auditory)	Feedback level (none, auditory, visual, auditory + visual)
Dependent variables	Glance measures; s.d. of lane position; speed variation; headway variation; subjective measures	Mean deviation from path; mean task completion time; subjective measures	Driving measures; Mean task time; NASA-RTLX; SUS
Statistical approach	Two-way within-subjects ANOVA	Within-subjects ANOVA	Repeated-measures ANOVA

evaluating the effect of the layout on driving performance and secondary task performance. Of the remaining eight studies, four are variations on target acquisition tasks evaluating icon sizes and layouts, two examine virtual keyboard/keypads, one examines multi-page information layouts, and one explores the effect of depth cues and icon design on icon touching behavior.

The five motor vehicle studies (Irune and Burnett, 2007; Kim et al., 2014a,b; Kujala and Salvucci, 2015) evaluated target size, target spacing, menu size, and menu arrangement in touchscreen IVIS interfaces on driving performance and secondary task performance; these studies are evaluated in full under RQ2.

The target-acquisition studies (Hwangbo et al., 2013; Im et al., 2015; Park and Han, 2010a, 2010b) examined icon size, icon ratios, icon spacing, and search field size in terms of task speed, accuracy, input offset (i.e., how close to the center of the icon did the user touch), and subjective measures of preference, in a handheld device setting (3.5″ or 4.3″ screens). Park and Han's studies (2010a;

2010b) evaluated how a one-handed (right-hand) thumb grip affected target tapping behavior and preference. Due to the similarity of the studies we present their findings together. In the one-handed grip, users had the greatest accuracy at the left edge of the screen and the worst at the lower right; targets mid-screen were fastest; targets of at least 7 mm yielded the best performance. Input offsets tended to be below the target on the screen, and above the target for the lower part of the screen; larger icons had lesser offsets. Mid-screen targets were perceived as most convenient. Im et al. (2015) also evaluated target offsets, examining the two-handed grip, and found that users typically touched below and to the right of the target; they also evaluated icon array sizes in a search-and-tap task, finding that arrays with 20 or fewer icons yielded the lower error rates and faster task times. Users preferred an icon width-to-height ratio of at least 0.7, while the best performance was achieved at a ratio of 0.9. Hwangbo et al. (2013) also evaluated icon sizes and spacing, focusing on elderly participants

Table 2a
Articles identified under RQ2 with a road vehicle setting.

	Burnett et al., 2013	Burnett et al., 2011	Hagiwara et al., 2014	Harvey et al., 2011	Irune and Burnett, 2007
Affiliation of Primary Author	(see Table 1a)	University of Nottingham, Nottingham, UK	Hokkaido University, Hokkaido, Japan	University of Southampton, UK	(See Table 1b)
Aim		Compare touchscreen, touchpad, and rotary IVIS inputs	Compare task loading effects in use of a touchscreen IVIS	Compare touchscreen and rotary IVIS inputs	
Study design		Within-subjects 3 × 7	Within-subjects 3 × 2 × 2 (plus non-task baseline)	Within-subjects (3 levels)	
Devices considered		Unspecified 12″ display; touchpad; rotary controller	Century LCD-8000 (8″)	Unspecified 7″ display; rotary controller	
Activity		Follow lead vehicle in traffic, while performing secondary tasks	Follow lead vehicle, while performing secondary tasks	Driving in traffic, while completing secondary tasks	
Participants		18; driving experience 2—24 years	16; active, licensed drivers with min. 5000 km driving per year	20; all with at least 1 year driving experience	
Age		30 (21—47)	(20—50)	25 ± 2.8 (21—33)	
Independent variables		Task (7); interface type (3)	Secondary task (3), task duration (2); screen position (high, low)	IVIS interface type (control condition, rotary controller, touchscreen)	
Dependent variables		Subjective measures; task time; glance count; standard deviation of lane position; standard deviation of speed	Glance measures; tap measures; vehicle motion vs driver behavior; mean speed; mean headway; mean lateral distance; mean s.d. of speed; mean s.d. of headway, mean s.d. of lateral distance	Mean speed; centerline crossings; percentage visual attention on road; task time; task error; SUS	
Statistical approach		two-way and one-way repeated measures ANOVA	Three way, within-subjects ANOVA	Friedman's ANOVA; Wilcoxon's test	

	Jin et al., 2015	Kim et al., 2014b	Kujala, 2013	Kujala and Salvucci, 2015	Neurauter et al., 2012	Pitts et al., 2012a
Affiliation of Primary Author	Jilin University, JiLin, China	(See Table 1b)	(See Table 1c)	(See Table 1b)	Virginia Polytechnic Institute and State University, Blacksburg, VA	(See Table 1d)
Aim	Evaluate the effect of secondary tasks on driving safety				Evaluate speech-based and touchscreen-based IVIS interfaces	
Study design	Within-subjects (5 levels)				Mixed 2 × 2	
Devices considered	iPad 4; unspecified touchscreen telephone; dashboard radio				Unspecified touchscreen smartphone; aftermarket handsfree voice input	
Activity	Drive in heavy traffic while completing secondary tasks				Maintain a constant speed, while performing secondary tasks	
Participants	40; held drivers' license for at least 2 years				24; described as "drivers"	
Age	32.5				(18—30) and (45—55)	
Independent variables	Secondary task (5)				age (between-subjects) (2); interface (touchscreen, voice input)	
Dependent variables	Fixation measures; saccade measures; blink measures; derived "driving safety" metric				Task success rate; Longitudinal control measures; lateral control measures; glance measures	
Statistical approach	ANOVA				ANOVA with mixed factorial design	

	Pitts et al., 2012b	Rydstrom et al., 2012	Williamson et al., 2011	Xian and Jin, 2015	Young et al., 2014
Affiliation of Primary Author	(See Table 1d)	Luleå University of Technology, Luleå, Sweden	(See Table 1d)	Shandong Jiaotong University, Jinan, Shandong, China	Monash Injury Research Institute, Monash University, Melbourne, Australia
Aim		Compare touchscreen and rotary controller IVIS input methods		Evaluate effect of practice on touchscreen IVIS performance	Compare touchscreen and keypad phone use while driving
Study design		Within-subjects (2 levels)		Within-subjects 3 × 2	Mixed 2 × 3
Devices considered		Deltaco touchscreen (7″); rotary controller		iPad 4	Participant's touchscreen phone; participant's numeric keypad phone
Activity		Navigate winding road with opposing traffic, while completing secondary tasks		Drive with and without traffic, while completing secondary tasks	Drive in traffic while completing secondary phone-based tasks
Participants		25; possessing a valid driver's license for at least 5 years		18; licensed at least 2 years	24; licensed at least 1 year

Table 2a (*continued*)

	Pitts et al., 2012b	Rydstrom et al., 2012	Williamson et al., 2011	Xian and Jin, 2015	Young et al., 2014
Age		45.7 ± 12.5 (27—69)		25.32 (20—31)	33.4 ± 9.9 (25—50)
Independent variables		Input method (touchscreen, rotary controller)		Traffic condition (nontraffic, traffic); task (none, "easy", "difficult")	Phone type (between-subjects) (touchscreen, keypad); task (none, read text, read and write text)
Dependent variables		Task time; task error rate; glance metrics; s.d. lane position; steering wheel reversal rate		Glance measures (including region-of-interest classification of glances); task performance; driving performance	Driving performance measures; glance measures; task performance; NASA-RTLX
Statistical approach		One-way repeated-measures ANOVA		Repeated-measures ANOVA	Two-way (2 × 3) ANOVA

(65—79). They found that larger target sizes (up to 12 mm) and increased target spacing (up to 3 mm) yielded better and faster performance while also increasing subjective suitability; at the 12 mm target size the benefit of icon spacing was lost.

Two studies evaluated virtual keying interactions. Kim et al. (2014a,b) evaluated the effect of virtual keyboard key size on typing force, muscle activity, ergonomics, typing speed and accuracy, and user preference, when a virtual keyboard was presented in lieu of a physical one. Conducting the study with experienced typists (40wpm or better typing speed) as participants, they found that virtual keyboards with keys smaller than 16 × 16 mm yielded reduced objective and subjective typing performance, as well as increased wrist and shoulder extension, when compared to larger key sizes up to 21 × 21 mm. Sesto et al. (2012) compared evaluated the effect of virtual keypad key size and spacing on number-entry tasks while seated and standing, for both motor-disabled and

typical users. All users were found to have improved task times with larger keys; miss/error rates improved for key sizes up to 25 mm, with additional (somewhat lesser) improvements for sizes up to 30 mm for motor-disabled users.

The remaining studies investigate finer points of interface design for touchscreen interfaces. Garcia-Lopez et al. (2015) compared three content distribution architectures, paging (i.e., nested linked pages), internal links, and vertical scrolling, to evaluate which distribution method is best for information search tasks in a handheld touchscreen setting. Participants performed best with vertical scrolling, and rated it as highly efficient for the study task; however, participants rated paging as most appropriate for general web content. Huang and Chen (2010) evaluated depth cues and visual cues in touchscreen icons in a desktop computing setting, determining the preferred depth cues for novice and expert users and evaluating the effect of visual cues on input offset. They

Table 2b
Articles identified under RQ2 in an extraneous movement setting.

	Baldus and Patterson, 2008	Bjorneseth et al., 2012	Goode et al., 2012	Lin et al., 2010	Yau et al., 2008
Affiliation of Primary Author	John Deere Product Engineering Center, Waterloo, IA, USA	(See Table 1c)	Monash University, Melbourne, Australia	Chung Yuan Christian University, Chung Li, Taiwan	National Tsing Hua University
Aim	Compare mouse, touchpad, touchscreen in moving tractor		Examine motion in use of touch-screen battle management system	Compare mouse, trackball touchscreen in vibration setting	Evaluate the effect of motion on four different input devices
Study design	Within-subjects 2x3x2x4x4		Within-subjects 2 × 2	Within-subjects 3x3x4x4x4	Within-subjects 4 × 5
Devices considered	Intelliworxx Voice Table tablet; mouse		Unspecified (10″)	Advantech FPM-3175 (17″); mouse; trackball	Unspecified 17″
Activity	Target acquisition tasks		Operate battle management system under different road conditions	Target acquisition tasks	Target acquisition task under various movement conditions
Participants	18		22	10	12; experienced with mouse but unfamiliar with trackball and touchscreen
Age	37.8 (21—58)		26.05 ± 7.77 (20—56)	(26—31)	26.5 ± 2.5 (22—30)
Independent variables	Tractor state (moving, stopped); pointing device (3); target size (2); shape (4); color (4)		Road type (finished, unfinished); trial ("drive number") (2)	Vibration level (3); device (3); target angle (4); target size (4); target distance (4)	Motion type (static, heave, roll, pitch, random); input device (touchscreen, trackball 1, trackball 2, trackball 3)
Dependent variables	throughput, tasks per second, subjective responses		Task completion time; accuracy; visual detection; gaze time; NASA TLX; heart rate; heart rate variability; perceived usability	Movement time (MT); error rate; scatter	MT; error rate; subjective usability and preference measures
Statistical approach	Multivariate general linear model		MANOVA, ANOVA, *t*-test	ANOVA	Two-way repeated-measures ANOVA

Table 2c
Articles identified under RQ2 in a medical setting.

	Hammon et al., 2014	Herrick et al., 2013	Pal et al., 2013	
			Study 1	Study 2
Affiliation of Primary Author	University Hospital Erlangen, Erlangen, Bavaria, Germany	Johns Hopkins University School of Medicine	University College London Hospitals NHS Foundation Trust, London, UK	
Aim	Evaluate device sanitary wrapping for usability and microbial resistance	Evaluate input methods in patient self-assessment systems	Compare contamination of keypad and touchscreen phones	Compare contamination of keypad and touchscreen phones
Study design	Cross-sectional study	Cross-sectional study	Cross-sectional study	Cross-sectional study
Devices considered	iPad 3	Elo TouchSystems model ESY15A2 (15″); laptop computer with stylus input; tablet with stylus input	Touchscreen mobile phones as carried by participating users	Touchscreen mobile phones as carried by participating users
Activity	Patient information entry on tablet	Patient information entry on device	Evaluate contamination of phone surfaces	Evaluation of phone surface bacterial contamination
Participants	100: 50 male, 50 female	841	67 phones carried by clinical staff	173 phones carried by clinical staff
Age	Men: 51.6 ± 16.3Women: 46.7 ± 21.4	41 (18–95)	n/a	n/a
Independent variables	Region of interest (3); tested surface (unused tablet, used tablet, used protective bag)	Interface (touchscreen tablet, stylus tablet, stylus computer)	Phone interface type (2)	Phone interface type (2)
Dependent variables	Colony Forming Units (CFU); subjective user satisfaction measures	Time per click; error rate; various demographic evaluations and subjective usability assessments	Colony forming units (CFU) per square centimeter	Colony forming units (CFU) per square centimeter
Statistical approach	Mann-Whitney U-test	Chi-square; Kruskal-Wallis rank test	Fisher's Exact test	Fisher's Exact test

found that novices preferred a depth cue of about 10% of icon size while experts preferred a depth of about 7%; the presence of a color-contrasting visual cue on an icon tended to compel users to touch on the visual cue region rather than on the center of the icon.

3.2.3. Interaction — inputs

Eight articles, reporting 12 studies, evaluated the interaction design of touchscreen interfaces with respect to inputs. Nine studies evaluated user performance across various single- and multitouch gestures, and one study (Kwon et al., 2011) evaluated the effect of control gain (McCormik, 1976) on gesture scrolling. The final two studies investigated menu-based vs. gesture-based touch interfaces in a maritime environment (Bjorneseth et al., 2012) and scrolling techniques in an IVIS setting while driving (Kujala, 2013), and are discussed under RQ2.

The comparative studies of single- and multitouch gestures investigated tapping, double-tapping, dragging, zooming, and rotating. Cockburn et al. (2012) reported two studies comparing single-touch gesture performance among mouse-based, finger-based, and stylus-based inputs on a desktop-size device. They found that finger-based input was fastest but most error-prone for tapping, and stylus-based input was fastest for dragging and radial dragging. The remaining studies (Chang et al., 2014; Gao and Sun, 2015; Mihajlov et al., 2015; Vatavu et al., 2015) evaluated gesture performance among different user populations. Chang et al. (2014) evaluated the use of gestures by children (11–14), adults (20–28), and elderly adults (65–84) across multiple screen sizes, finding that rotation gestures were particularly challenging on larger screens (23″ and 42″), especially for children, and that elderly users had difficulty dragging and zooming on smaller screens (4.3″ and 10″). In general, the performance of young adults was highest while that of elderly adults was lowest. Gao and Sun (2015) compared older adult (52–81) and younger adult (19–24) performance for four gestures in a desktop setting, finding that older adults took longer to drag but found it easier than young users, and that younger users preferred multitouch gestures (zoom and rotate) even though the multitouch yielded objectively worse performance than button-

based zoom and rotate controls. Mihajlov et al. (2015) evaluated the intuitive learnability of drag and rotate gestures among older (54–82), "technology-naïve" (p. 1) adults in a smartphone setting, finding that dragging was readily learned by participants, although intermittent dragging was frequent; and that fewer than half of participants discovered the correct gesture for rotation. Vatavu et al. (2015) evaluated the gesture skills of children aged 3–6 on phones and tablets, and discovered that tapping speed and accuracy improved with age, but that young children's performance was sufficiently different from adults' performance to warrant special design considerations.

Kwon et al. (2011) evaluated control gain and scrolling interaction in dragging, among users without touchscreen experience. They found that push-viewport style scrolling resulted in fewer touches needed and faster task completion time as compared to push-background, and found that a gain of 5 was perceived as most preferred and easiest, while also having shorter task times and fewer touches required. A gain of 3 had slightly inferior performance to 5, but a higher perceived level of control.

3.2.4. Interaction — output

Seven studies, reported in six articles, discussed interaction design with respect to system feedback in response to inputs. All seven studies evaluated tactile, auditory, and/or visual feedback. Three studies (Pitts et al., 2012a; Pitts, Skrypchuk et al., 2012b; Williamson et al., 2011) report studies of IVIS feedback during driving with secondary tasks; these are discussed in detail under RQ2.

Four studies (Han and Kim, 2015; Hwangbo et al., 2013; Parikh and Esposito, 2012) evaluated the effect of multimodal feedback on touchscreen performance. Hwangbo et al. (2013) evaluated tactile and audio feedback among elderly adults (65–79) who were not smartphone owners, in the context of a target acquisition task on a smartphone. They found that visual + audio + tactile feedback resulted in the strongest objective performance, and that tactile-only feedback was perceived to be less usable and less satisfying than other feedback styles. In contrast, Han and Kim (2015)

Table 2d
Articles identified under RQ2 in all other settings, including aircraft, robotics, cycling, and e-readers.

Setting	Aircraft		Robotics	
	Eichinger and Kellerer, 2014	Stanton et al., 2013	Dünser et al., 2014	Eliav et al., 2011
Affiliation of Primary Author	Technische Universität München, Garching, Germany	University of Southampton, Southampton, UK	CSIRO, Hobart, Australia	Ben-Gurion University of the Negev, Beer-Sheva, Israel
Aim	Compare touchscreen, trackball inputs in jet cockpit setting	Compare four input devices in airplane Flight Management System	Compare touchscreen, mouse, eye-tracking for robotic control	Compare displays and control interfaces for robot operation
Study design	Within-subjects $2 \times 3 \times 2$	Within-subjects $4 \times 2 \times 2$	Within-subjects $4 \times 2 \times 3 \times 3$	Mixed 2×2
Devices considered	Unspecified touchscreen input; trackball	Unspecified 15″ touchscreen; trackball; rotary controller; touchpad	Acer T232L (23″); mouse; dwell-based eye tracker; click-based eye tracker	Unspecified "touch screen PC"; hand gesture input
Activity	Target acquisition and visual search tasks (no flight simulation task)	Target acquisition and visual search tasks (no flight simulation task)	Select origin and destination for pick-and-place task	Give movement commands to robot using different control paradigms
Participants	11; all with average flight experience	20; all with touchscreen experience	18	20
Age	41 ± 9.1	$31 \pm 9.7 (21{-}57)$	40 (22—56)	(19—31)
Independent variables	Task (3), task difficulty (2); input device (2)	Input method (touchscreen, trackball, rotary controller, touchpad); auditory feedback (present, absent); task (2)	Workload (2); interface (4), target size (3), target distance (3)	Display (between-subjects) (2); controller (touchscreen, hand gesture)
Dependent variables	Task time; NASA TLX	Task time; error rate; NASA-TLX; SUS; hand discomfort; body discomfort	Movement time, NASA TLX, subjective usability assessments	Implicit performance (e.g., task time); explicit performance (e.g., situational awareness questionnaire); Subjective measures
Statistical approach	Three-way repeated measures ANOVA	Factorial repeated-measures ANOVA; Friedman's ANOVA	Repeated measures ANOVA	Two-way ANOVA with repeated measures

Setting	Cycling	E-Reader
	De Waard et al., 2014	Siegenthaler et al., 2012
Affiliation of Primary Author	University of Groningen, Groningen, The Netherlands	Swiss Distance University of Applied Sciences, Brig, Switzerland
Aim	Evaluate touchscreen device use while cycling	Evaluate effects of touchscreen interface on e-reader usability
Study design	Within-subjects (8 levels)	Within-subjects (3 levels)
Devices considered	Sony Ericsson Xperia X8 phone; numeric keypad cell phone	Sony PRS-600 e-reader (6″); Apple iPad (9.7″); Sony PRS-505 e-reader
Activity	Perform various mobile phone secondary tasks while cycling	Typical e-reader tasks, e.g., navigation and bookmarking
Participants	24: 7 men, 17 women	12; no previous e-reader experience
Age	Men: 24.3 ± 4.3 (20—31) Women: 21.5 ± 1.8 (19—25) (Overall mean 22 ± 2.96)	(20—26)
Independent variables	Cycling condition (8); of particular interest are conditions TT (texting on touchscreen) and TC (texting on conventional phone)	Reading device (PRS-505, PRS-600, iPad)
Dependent variables	Speed; lateral control; reaction time; visual perception; subjective mental workload; subjective experienced risk	Task completion rate; Task time; subjective usability measures
Statistical approach	Hotelling's T	Repeated-measures ANOVA

evaluated experienced users in a virtual typing task on both phones and tablets, finding that visual + audio + tactile feedback did not yield significantly improved performance over either visual + audio or visual + tactile feedback separately. Visual + tactile feedback, however, prompted increased error correction behavior. Finally, Parikh and Esposito (2012) presented two studies evaluating the effect of negative tactile feedback, that is, feedback that is triggered on error, on key-entry tasks on a smartphone. A comparison of positive and negative tactile feedback found the two modes to be statistically indistinguishable; a follow-up study found that multimodal feedback (positive audio feedback and negative tactile feedback) improved typing efficiency by communicating early typing errors effectively.

3.3. Research question 2

Eighteen articles addressed RQ2. We identified five categories of

settings evaluated in these articles:

- *Road vehicle*: driving a car (typically in a simulator setting), while completing secondary non-driving tasks in an in-vehicle information system (IVIS) or on a mobile phone.
- *Extraneous movement*: any of several vehicle settings where vibrational or other undesired movement interference is present.
- *Aircraft*: any study with the setting of an aircraft cockpit.
- *Robotics*: any study where robotic teleoperation or supervisory control is the study task.
- *Miscellaneous*: All other settings (each with no more than one study).

3.3.1. Road vehicle

Sixteen articles evaluated touchscreens in a road vehicle setting,

Table 3a
Articles identified under RQ3 with an age-specific study population or research dimension.

	Anthony et al., 2014	Chang et al., 2014	Chung et al., 2010	Elton and Nicolle, 2013	Gao and Sun, 2015
Affiliation of Primary Author	University of Maryland Baltimore County, Baltimore, MD, USA	(See Table 1a)	Pohang University of Science and Technology, Pohang, South Korea	Brighton University, Brighton, UK	(See Table 1a)
Aim	Compare child and adult touchscreen interaction behavior		Compare physical, touch keypads between young and older adults	Evaluate temperature effects on older adults' manual dexterity	
Study design	Mixed $4 \times 4 \times 13 \times 2$		Mixed $2 \times 2 \times 3$	Within-subjects (2 levels)	
Devices considered	Samsung Google Nexus S (4″)		Daewon Computer dwcom17 (capacitive) (17″); physical numeric keypad	iPod Touch; keypad telephone; stylus-based PDA	
Activity	Target acquisition task		Keypad number entry while standing	Text and numeric entry	
Participants	44 children, 30 adults		12 younger adults, 12 older adults	31 older adults; 6 participants had arthritis	
Age	children: 12.2 ± 2.4 (6—17) adults: 23.7 ± 4.0 (18—33)		Younger: 25.6 ± 3.2 Older: 69.0 ± 3.4	70 (65—81)	
Independent variables	Age group (7—10, 11—13, 14—17, 18+); target sizes (4), locations (13), button padding (padded, unpadded)		Age group (2); keypad type (touchscreen, physical), task (3)	Temperature (19—24 °C; 5 °C); input methods (touchscreen, keypad, stylus)	
Dependent variables	Miss rate; presence of holdover		Mean entry time per unit stroke (METU) (i.e., normalized task time); error rate; subjective ease of use	Task time	
Statistical approach	repeated-measures ANOVA; one-way ANOVA		ANOVA	Paired t-test	

	Hwangbo et al., 2013		Jochems et al., 2013	McLaughlin et al., 2009	Mihajlov et al., 2015
Affiliation of Primary Author	Yonsei University, Seoul, Korea Study 1	Study 2	RWTH Aachen University, Aachen, Germany	North Carolina State University, NC	(See Table 1c)
Aim	Evaluate mobile touchscreen input layouts for elderly users	Evaluate mobile touchscreen feedback for elderly users	Compare various computer input methods among adult age groups	Examine effect of match or mismatch between task and input type	
Study design	Within-subjects 3×2	Within-subjects (4 levels)	Mixed 3x3x3x3x8	Mixed $2 \times 2 \times 4$	
Devices considered	Unspecified Android smartphone (4.3″)	Unspecified Android smartphone (4.3″)	Elo TouchSystems 1715 (resistive) (17″); wireless mouse; eye-gaze control	DataLux LMV10 (resistive) (10.4″); rotary controller	
Activity	Target acquisition task	Target acquisition task	Target acquisition task	Simultaneous video-game task and "entertainment system simulator"' task	
Participants	22; none were smartphone owners	22; none were smartphone owners	90: 30 in each age group	24 younger adults; 24 older adults	
Age	70.55 ± 3.713 (65—79)	70.55 ± 3.713 (65—79)	Group I: 26.7 ± 4.04 (20—38) Group II: 50.33 ± 6.36 (40—59) Group III: 65.47 ± 4.08 (60—72)	Younger: 20.1 ± 1.3 (18—25) Older: 65.2 ± 3.0 (60—70)	
Independent variables	Target size (3), target spacing (2)	Feedback type (none, auditory, tactile, auditory + tactile)	Age group (3); input device (touchscreen, mouse, eye gaze); target width; target distance; target angle	Age group (2); input method (touch, rotary controller) (between-subjects); attentional split between tasks (4)	
Dependent variables	Task time; error rate; Subjective measures based on ISO 9241	Task time; error rate; Subjective measures based on ISO 9241	Task time; task difficulty (ZEIS scale) (Pitrella and Käppler, 1988)	Simulator task time; simulator task accuracy; video game task accuracy	
Statistical approach	Repeated-measures ANOVA	ANOVA	Mixed design ANOVA	three-way repeated-measures ANOVA	

	Mihajlov et al., 2015	Neurauter et al., 2012	Smith and Chaparro, 2015	Sonderegger et al., 2016	Vatavu et al., 2015	Xiong and Muraki, 2016
Affiliation of Primary Author	(See Table 1c)	(See Table 2a)	Wichita State University, Wichita, Kansas	University of Fribourg, Fribourg, Switzerland	(See Table 1c)	Kyushu University, Fukuoka City, Japan
Aim			Evaluate smartphone text input methods for younger, older adults	Evaluate the effect of age in user research outcomes		Compare young adult and elderly thumb reach in handheld device use
Study design			Mixed 2×5	Between-subjects 2×2		Mixed 2×2

Table 3a (*continued*)

	Mihajlov et al., 2015	Neurauter et al., 2012	Smith and Chaparro, 2015	Sonderegger et al., 2016	Vatavu et al., 2015	Xiong and Muraki, 2016
Devices considered			Motorola Droid 4 phone touchscreen; Motorola Droid 4 physical keyboard	HTC Touch Pro 2 (3.6″) touchscreen; HTC Touch Pro 2 physical keyboard		iPhone 4; Galaxy S4
Activity			Smartphone text entry task	Data entry task; menu selection task		Sweep thumb across phone to determine coverage area
Participants			25 younger adults; 25 older adults	30 younger adults; 30 older adults		24 young adults; 24 elderly adults; all participants right-handed
Age			Younger adults: 24.4 ± 5.6 (18—35) Older adults: 68.8 ± 7.4 (60—84)	Younger adults: 23.0 ± 2.39 (19—29) Older adults: 58.13 ± 5.43 (52—73)		Young adults: 23.6 ± 1.8 Elderly: 67.5 ± 3.7
Independent variables			Age group (2); text input method (physical QWERTY, on-screen QWERTY, "Swype" style tracing, handwriting recognition, voice input)	Age group (2); text input method (physical QWERTY, on-screen QWERTY) (between-subjects)		Age group (2); device (2)
Dependent variables			Adjusted WPM; error rates; subjective usability measures	PSSUQ; PANAS; NASA-TLX; task performance		Thumb coverage area; center of gravity of thumb coverage area
Statistical approach			Two-way ANOVA	Two-factorial analysis of covariance		Two-way repeated-measures ANOVA

Table 3b
Articles identified under RQ3 with other population-specific focuses, including disability and handedness.

Population	Disability			Handedness
	Chourasia et al., 2013	Irwin and Sesto, 2012	Sesto et al., 2012	Kang and Shin, 2014
Affiliation of Primary Author	University of Wisconsin-Madison, Madison, WI	University of Wisconsin-Madison, Madison, Wisconsin, USA	University of Wisconsin-Madison, Madison, WI	Ulsan National Institute of Science and Technology, Ulsan, Korea
Aim	Examine posture effects on motor-disabled touchscreen use	Measure touch force, dwell time of motor-disabled users	Evaluate button size and spacing for motor-control disabled users	Examine effects of handedness on touchscreen use
Study design	Mixed 2x2x5×2	Between-subjects (3 levels)	Mixed 3 × 5 × 2	Between-subjects (3 levels)
Devices considered	Elo TouchSystems 1537L (resistive) (15″)	Elo TouchSystems 1537L (resistive) (15″)	Elo TouchSystems 1537L (resistive) (15″)	LG IPS236V 23″
Activity	Keypad number entry on touchscreen while sitting and standing	Reciprocal tapping task	Keypad number entry on touchscreen while seated	Touching and scrolling during web browsing
Participants	15 with motor-control disabilities ("+MCD"), 15 without ("-MCD")	22 with fine motor disability, 15 with gross motor disability, 15 non-disabled	23 with fine motor-control disabilities, 14 with gross motor-control disabilities; 15 non-disabled	21: 7 left-handed, 7 ambidextrous, 7 right-handed
Age	+MCD: 52.3 ± 12.2 -MCD: 48.9 ± 11.5	Fine motor: 53.45 ± 11.37 Gross motor: 43.53 ± 11.98 Non-disabled: 46.1 ± 12.3	Fine motor-control disabled: 52.8 ± 11.5 Gross motor-control disabled: 44.6 ± 11.7 Non-disabled: 46.1 ± 12.3	LH: 20.9 ± 1.9; AMB: 19.4 ± 1.1; RH: 20.3 ± 2.2
Independent variables	Presence of motor control disability (2); posture (standing, seated); button size (5); button spacing (2)	Motor disability level (3)	Presence and type of motor control disability (3); button size (5); button spacing (2)	Handedness (left, ambidextrous, right)
Dependent variables	Miss; error; task time; force; impulse; NASA-TLX; user preference	Performance; peak force; dwell time; impulse	Miss (touch not on a button); error (touch on wrong button); task time; user preference	Subjective discomfort rating; hand preference; touch area preference
Statistical approach	Mixed-design repeated measures analysis	ANOVA	Mixed-design repeated measures analysis	Two-way MANOVA

representing 18 studies. Two studies (Hagiwara et al., 2014; Neurauter et al., 2012) were conducted on test tracks, while the remainder were performed in driving simulators of varying fidelity. Simulator use and simulator fidelity is only a minor concern, as the validity of simulator driving in comparison to test-track driving has

been validated in an HFE context (Eichinger and Kellerer, 2014). In addition, six of the 18 studies were performed in right-hand drive settings. All studies evaluated users as they drove while engaging in secondary tasks; three articles evaluated the use of handheld devices, while the rest investigated only the use of IVISes or other

Table 4
Summary of key findings.

Research question	Subcategory	Findings and recommendations
RQ1	Hardware/driver	• **Capacitive** input yields higher usability than **resistive** • **Optical** input yields higher satisfaction and perceived performance than **SAW** • **SAW** should be favored over **optical** where mis-touches are not acceptable • **Rotation** gestures yield highest performance on **small screens** • **Dragging** gestures should be **avoided** on **small screens** • **Palm rejection technology** may yield improved ergonomic outcomes
	Interface design	• Enlarged icons and targets help minimize effects of input offset • Visual cues on icons assist with targeting tasks • 19 × 19 mm suggested key size for soft QWERTY keyboards • 12 mm icon height and 0.9 mm width-height ratio recommended for icons in handheld devices • Icon depth cues equal to 10% of target size suggested • Primary controls at top or center of screen for one-hand devices • Vertical scrolling or paging recommended in handheld settings
	Interaction design	• **Multitouch gestures** should have button-based alternatives • **Multitouch gestures** may require instruction for older and younger users • **Viewport scrolling** with a **gain between 3 and 5** recommended • **Audio plus visual and/or tactile** feedback should be provided • **Distinct positive and negative feedback** for key-input should be considered
RQ2	Automotive — hardware selection	• **Capacitive input** should be selected over **resistive input** for IVISes • Touchscreen phones may impair driving more than **keypad phones** • **Voice input** is recommended over other IVIS inputs, including touchscreens • Touchscreen input yields better overall performance than **rotary controller** or **touchpad input** for most tasks
	Automotive — touchscreen IVIS interface design	• Touchscreen IVISes may yield unacceptable safety compromises • Grid menus preferred over list menus and should not exceed 9 items • Swipe-to-scroll and button scrolling should be offered in parallel • Visual + audio + tactile feedback should be provided
	Extraneous movement	• Touchscreen interfaces yield **high error rates** under heavy movement • **User familiarity** does not mitigate movement-induced error • **Gesture-based inputs** do not mitigate movement-induced error • **Touchscreens** should be considered where **low movement time** is a priority • **Trackball input** should be considered where **accuracy** is paramount
	Medical settings	• Touchscreens help to **minimize microbial contamination**, and are compatible with disposable, **sanitary coverings** • Touchscreens have high usability as **patient information devices**
RQ3	Age — children	• **Input interaction performance** improves as children grow older • Expect **holdover touches** in children interacting with touch interfaces • **Multitouch gestures** can be difficult for children to execute
	Age — young adults	• Young adults find **multitouch gestures** highly satisfying • **Very small targets** (as small as 6.5 mm) can be acceptable for this population • **Virtual numeric keypads** are an acceptable alternative to physical keypads • **Virtual keyboards** yield high speed but low accuracy and high effort
	Age — older adults	• Touchscreen skill **degrades more slowly with age** than mousing skill • **Divided-input tasks** with a tapping component are suited to touchscreens • Touchscreen inputs see **reduced cold impairment** compared to other inputs • **Gestures** are challenging for older adults to learn • **Auditory + tactile feedback** is recommended for older adults
	Motor-control disabilities	• **Overall performance** is lower among + MCD users than general population • **Interface changes** that improve performance for the general population will likely improve performance for + MCD users as well • MCD + users display longer **holdover touches** than the general population
	Handedness	• Left-handers and ambidextrous users **tap** with the **most convenient hand** • Left- and right-handers perform **gestures** with the **dominant hand**
	General	• Users seem to **strike harder** than necessary to activate touchscreens

fixed devices. Seven studies either compared two touchscreen input devices, or compared a touchscreen input to an alternative (keypad, rotary controller, or touchpad); one study engaged participants in touchscreen tasks with different task loads to evaluate the effect on driving performance; the remaining studies evaluated the implications of IVIS interface and interaction design.

Studies used a variety of measures to evaluate driving and secondary task performance. Secondary task performance was typically reported in terms of task time and error. Key driving performance factors included longitudinal measures (distance from lead vehicle, standard deviation of speed), lateral measures (standard deviation of lane diversion), and eye glance behavior. There is no single standard that separates safe driving behavior from dangerous, but various thresholds based on glance time (i.e., the time that eyes are dedicated to a secondary task rather than the

driving task) have been proposed: presence of 1.6-s (Wierwille and Tijerina, 1998) or 2-s (Zwahlen et al., 1988; Klauer et al., 2006) or longer glances, or total glance time of 4 or more seconds for a single task (Ito and Miki, 1997); NHTSA standards for driver distraction indicate all glances less than 2 s, and total per-task glance time less than 12 s (NHTSA, 2013). In general, articles self-reported any glance duration thresholds they identified; we note significant findings relative to these self-reported measures. In addition to driving and task performance, many studies reported subjective measures of preference, satisfaction, workload, etc.

Three studies (Burnett et al., 2011; Harvey et al., 2011; Rydstrom et al., 2012) compared touchscreen IVIS input to rotary controller input; in addition, Burnett et al. (2011) also included a comparison to gesture-based touchpad input. This study found that the touchscreen and touchpad were superior to the rotary controller in

terms of driving and secondary task performance, with the touchscreen in particular being better suited to discrete selection tasks. Harvey et al. (2011) likewise found touchscreen input to be superior to rotary controller input in terms of vehicle control, secondary task times, and perceived usability. However, Rydstrom et al. (2012) found that while touchscreen input was superior for discrete tasks, rotary input was better for continuous or fine-tuning tasks, in terms of task time, glance counts, and user preference. They suggested that no single paradigm yields universally superior performance. A further study (Neurauter et al., 2012) compared touchscreen text input with voice input using an aftermarket smartphone voice recognition kit; text entry on the touchscreen demonstrated reduced lateral and longitudinal control, increased eyes-off-road time and glance count (including glances above the 2-s threshold), greater subjective mental demand, greater subjective frustration, and reduced situation awareness as compared to voice entry for the same tasks.

Two additional studies (Jin et al., 2015; Young et al., 2014) compare task and driving performance across distinct devices. Jin et al. (2015) use a novel driving safety metric to evaluate the relative safety of tasks with different levels of visual load; they find in particular that touchscreen tasks on an iPad and a touchscreen phone had a significantly greater percentage of interactions rated "low" or "very low", as compared to baseline driving, in-car and via-Bluetooth conversation, and radio tuning. Young et al. (2014) evaluated drivers while performing texting tasks on their own touchscreen or keypad phones, and found that keypad phone users encountered greater variation in longitudinal control than touchscreen users.

One study (Hagiwara et al., 2014) presented users with varying secondary touchscreen IVIS tasks designed to load different cognitive and functional capacities, to determine the effect of different task loads on driving performance. They engaged participants in a read-aloud task (visual + cognitive load), a search-and-tap touchscreen task (visual + manual + cognitive load), and a cyclic tapping touchscreen task (manual load only). The search-and-tap task, which is most similar to a typical touchscreen IVIS interaction, led to highly variable headway distance and lateral vehicle control, though headway distance was also inconsistent with the read-aloud task. Interestingly, position of the touchscreen (i.e., positioned high or low) did not have a significant effect on touchscreen performance during the search-and-tap task. Another study (Xian and Jin, 2015) evaluated the effect of easy and difficult e-mail tasks on baseline driving performance; both the easy and difficult tasks led to significant lateral deviation, with the difficult task in particular leading to a high number of eyes-off-road glances exceeding the 2-s threshold. This study also compared secondary tasks in different levels of traffic: task completion times were 1.5–1.9 times longer under traffic as compared to the non-traffic condition.

The remaining studies evaluated IVIS hardware, interface and interaction design. Burnett et al. (2013) compared capacitive and resistive IVIS touch displays, as discussed above under RQ1. Five studies (Irune and Burnett, 2007; Kim et al., 2014a,b; Kujala and Salvucci, 2015) evaluated touch interfaces in IVIS design, including target size, target spacing, menu size, and menu arrangement. Irune and Burnett (2007) reported two studies that evaluated control cluster size and arrangement, control size, and control spacing; they found that higher total control count increased task time, visual demand, and perceived difficulty, identifying clusters of 16 controls or more as resulting in total glance times above 4 s; tests of cluster arrangements indicated that increased vertical cluster size in particular yields longer task time and longer total glance time, while increased horizontal cluster size has only a lesser effect. The two studies in Kujala and Salvucci

(2015) also investigate cluster arrangement, evaluating list and grid structures of different sizes ranging from 2 elements to 12; in testing a search-and-tap task with scrolling across multiple screens, they found that all menu structures they tested would fail to meet NHTSA guidelines for glance behavior for both total glance time and maximum glance duration, but 6-item lists and 9-item grids yielded optimal behavior. In evaluating the effect of item count on driving behavior, grid menus were found to be less sensitive to increased item count than list menus. This comports with Irune and Burnett's findings, as grid menus will grow in length more slowly than list menus. Kim et al. (2014a,b) evaluated the effect of icon size on driver performance; better driving performance measures and subjective measures were seen as icon sizes increased to 17.5 mm. Additional reductions in perceived workload were seen at larger button sizes.

Four studies investigated interaction design in touchscreen interface. One study (Kujala, 2013) compared different methods of scrolling through an IVIS display (button, swipe-per-page, kinetic scrolling); button scrolling was most preferred by users, but swipe scrolling demonstrated the least degradation in driving performance measures. Kinetic scrolling in particular increased perceived visual load and workload, as well as degrading lane-keeping vigilance. The remaining three studies (Pitts et al., 2012a, 2012b; Williamson et al., 2011) evaluated multimodal (visual, audio, tactile) feedback in IVIS use. Pitts et al. (2012a) investigated whether the presence of tactile feedback could alleviate visual load (in the form of delayed visual feedback), finding that in general, immediate haptic feedback effectively offset delay or absence in visual feedback in terms of performance and user satisfaction, with haptic feedback reducing glance count and total glance time. Pitts et al. (2012b) and Williamson et al. (2011) both evaluated combinations of visual, audio, and tactile feedback. Williamson et al. (2011) examined combinations of visual and audio feedback, and failed to find a significant effect of feedback type on driving or task performance; likewise Pitts et al. (2012b) did not find a significant effect on driving or task performance of adding tactile or auditory feedback to visual feedback. However, in both studies, subjective preference improved with additional feedback modes: thus, Williamson et al. (2011) found the strongest preference for visual + audio feedback and Pitts et al. (2012b) found the strongest preference for visual + audio + haptic feedback.

3.3.2. Extraneous movement

Five studies are considered to be in the extraneous movement category: Baldus and Patterson (2008) evaluated touchscreen and mouse in an offroad agricultural tractor; Bjorneseth et al. (2012) compared menu-based and gesture-based inputs in a maritime simulator; Goode et al. (2012) evaluated the use of a touchscreen battle management system (BMS) in a military vehicle on finished and unfinished roads; Lin et al. (2010) compared touchscreen, mouse, and trackball input under induced vibration conditions; and Yau et al. (2008) compared touchscreen and three trackball designs in a motion simulator designed to emulate sea vessel motion. Although these studies differ in setting, all the settings have the common property of incursions of unwanted external motion, more severe than would be present in a road vehicle, interfering with task performance. For this reason, they are grouped together. The studies measured various performance and satisfaction metrics, including throughput, task time, error rate, glance measures, and subjective measures of usability, workload, and preference.

Baldus and Patterson (2008), Lin et al. (2010), and Yau et al. (2008) all evaluate typical target acquisition tasks in an unwanted movement environment. Baldus and Patterson compared mouse and touchscreen use in a moving and stopped tractor; mouse and touchscreen had similar overall effectiveness, but mouse was more

effective for small targets while touchscreen was more effective for larger targets. Users, however, expressed a preference for mouse. Lin et al. compared mouse, touchscreen, and trackball at different levels of vibration; the touchscreen had the lowest movement time (MT) of all devices at all vibration conditions, but had a scatter level that exceeded the 4% rate recommended in Zhai et al. (2004). The authors recommend the mouse in a vibration setting as it has a shorter MT than a trackball while maintaining an acceptable scatter rate. Yau et al. evaluated the use of touchscreens and three models of trackball under regular and random sea-like motion, finding that touchscreens were subjectively and objectively fastest but had the lowest accuracy; touchscreens also resulted in higher reported fatigue. Trackballs saw additional benefit due to the availability of dynamic gain, which increased accuracy as users slowed on approaching the target. Bjorneseth et al. (2012) and Goode et al. (2012) each investigated tasks specific to their setting. Bjorneseth et al. compared menu-based and touchscreen-based maritime dynamic positioning interface. The participants selected were relatively novice maritime students, to minimize familiarity bias. Contrary to the hypothesis that gestures would be easier to execute successfully than point-and-tap actions under simulated vessel movement, gestures did not reduce task time or error rate, or improve situational awareness; nor did gesture input demonstrate resistance to movement compared to the point-and-tap interface. Goode et al. evaluated the effect of road surface and repeated trials on the use of a touchscreen BMS in a moving military vehicle; traveling on an unsealed road significantly impaired accuracy of BMS tasks and yielded a higher reported workload, while repeated trials did little to improve task performance.

3.3.3. Medical

Three papers reported four studies evaluating the use of touchscreens in medical settings. All four studies were cross-sectional in nature, conducted *in situ* and uncontrolled in a clinical setting. One study (Herrick et al., 2013) evaluated touchscreen versus stylus kiosks as patient self-administered interview systems, while three studies (Hammon et al., 2014; Pal et al., 2013) evaluated touchscreen devices for bacterial contamination. Herrick et al. (2013) compared a touchscreen, tablet-based patient information system to a stylus-based tablet and a stylus-based laptop. The study evaluated 841 users conducting a self-administered assessment, and measured time per click, error rate, and subjective usability assessments; the touchscreen input demonstrated reduced time-per-click and a lower error rate than the stylus-based methods. Hammon et al. (2014) and Pal et al. (2013) examined the capacity of touchscreen devices to limit microbial contamination, measuring colony-forming units (CFUs) as a measure of contamination. In addition, Hammon et al. (2014) tracked user satisfaction. Hammon et al. (2014) evaluated whether wrapping a touchscreen patient information system in a disposable bag could effectively prevent contamination of the touchscreen. Microbial contamination of the touchscreen was effectively limited by use of the plastic bag without impairing reported functionality or user experience of the device. Pal et al. (2013) reported two studies at two different clinical locations, measuring the contamination of personal and on-call phones belonging to medical staff. CFU counts on keypad phones were 8.5 and 3.7 times higher than touchscreen phones.

3.3.4. Aircraft

Two studies (Eichinger and Kellerer, 2014; Stanton et al., 2013) compared touchscreen inputs to alternatives (trackball, rotary controller, touchpad) in a simulated flight deck setting. Tasks for both studies were target acquisition and visual search tasks; dependent measures included task time, error rate, subjective workload, subjective usability, and subjective discomfort.

Eichinger and Kellerer (2014) compared touchscreen and trackball inputs in a simulated military jet cockpit, with 11 participants reported to have average flight experience. Although the setting for the experiment was a simulated cockpit, participants were not engaged in simulated flight while conducting the study tasks. The study found that the touchscreen interface yielded better performance and lower reported workload than the trackball, with the relative difference between touchscreen and trackball performance increasing with task difficulty. Stanton et al. (2013) compared touchscreen, trackball, rotary controller, and touchpad inputs in a simulated commercial jet cockpit, with 20 non-pilot participants. The study did not include a simulated flight task. Touchscreens produced shorter task times than the other input methods and low error rates; participants reported highest workloads for the touchscreen but also the highest usability. (It is unclear how to reconcile the low reported touchscreen workload of Eichinger and Kellerer with the high reported workload in Stanton et al.)

3.3.5. Robotics

Two studies evaluated touchscreens in the context of robotic teleoperation. Dünser et al. (2014) compared touchscreen, mouse, and two forms of eye-gaze control for a pick-and-place task, and Eliav et al. (2011) compared touchscreen inputs and hand gestures for remote control of a robot. Performance was evaluated through implicit and explicit performance measures (movement time, task time, error rates, situational awareness) and subjective measures of workload and usability.

Dünser et al. (2014) compared touchscreen, mouse, eye-tracking with gaze clicking, and eye-tracking with explicit clicking in a simulated robotic pick-and-place task. The study found that touchscreen interaction was faster than the other input methods, but yielded a higher workload. Eliav et al. (2011) evaluated supervisory control of a remote robot through either touchscreen inputs or hand gestures. Touchscreen performance measures and perceived usability were higher than gesture control, and yielded faster task completion.

3.3.6. Miscellaneous

Two studies had contexts that were not a good fit for any other category, but are included here for completeness. De Waard et al. (2014) compared the use of keypad phones and touchscreen phones while riding a bicycle on an open course, evaluating speed maintenance, lateral control, visual perception, subjective workload, and subjective experienced risk. Touchscreen texting led to greater road incursions and reduced peripheral visual perception as compared to keypad phones, but without an increased subjective sense of risk or workload. Siegenthaler et al. (2012) compared key-operated and touchscreen e-reader devices, evaluating task completion rate, task time, and subjective usability. Subjects gave touchscreen e-readers higher subjective scores on design, ease of navigation, and handiness as compared to key-operated models.

3.4. Research question 3

Twenty-six articles addressed RQ3. These articles are summarized in Table 4 below. We identified three population categories evaluated in these articles:

- *Age*: articles comparing or focused on specific age groups.
- *Disability*: articles focused on participants with a disability, or comparing disabled with typically-abled users.
- *Handedness*: articles comparing participants with different handedness.

3.4.1. Age

Fourteen articles, representing eighteen studies, evaluated the use of touchscreens by different age groups. 14 studies included age group as an independent variable. Of these, 11 studies compared adult age groups, with 8 studies comparing young adult and elderly populations with mean ages 20—26 and 64—71 respectively; one of the remaining adult studies had three age groups; one study had a slightly lower mean for the older/elderly group, and one compared young adults to middle-aged adults. The remaining three studies included children in the study population: one compared 3, 4, and 5-year olds; one compared three child groups to young adults; and one compared children (11—14) to young and elderly adults. The remaining four studies limited study population to a specific age group of interest; all four of these studies limited their study populations to "older" or "elderly" adults, with three of the studies limited to participants 65 and older and the last to users 54 and older.

Eleven studies evaluated target acquisition tasks, or variations suitable for testing touchscreen gestures; varying target size, distance, direction, and other factors affecting targeting difficulty, as well as environmental variables such as alternative input methods, touchscreen type, screen size, and inclination angle. Dependent variables in these studies include task time, error rate, performance value, and subjective measures of usability, workload, and preference. Four studies evaluated key-entry tasks, evaluating keyboard type, temperature, and text input method; dependent measures for these studies included task performance measures such as typing speed and error rate, subjective usability measures, subjective affect measures, and subjective workload measures. One study evaluated a dual-workload condition intended to evaluate mismatches between task type and input type, with input method and attentional split as independent variables and task time and accuracy as dependent variables; one study offered a road driving task with secondary tasks, comparing touchscreen and voice inputs and evaluating task success rate, longitudinal vehicle control, lateral vehicle control and glance measures; and a final study evaluated thumb reach in a one-hand grip for two different devices, measuring thumb coverage area.

Of the studies evaluating target acquisition-type tasks, five studies among four papers (Anthony et al., 2014; Gao and Sun, 2015; Hwangbo et al., 2013; Jochems et al., 2013) evaluated simple (i.e., point and tap) target acquisition tasks. Anthony et al. (2014) compared the performance of children (7—10, 11—13, 14—17) and adults (18—33) at a simple target acquisition task on a handheld device, varying target sizes, locations, and edge padding, evaluating miss rate, as well as holdover, which occurs when a user lingers on touching the target. The authors found that holdover was present in a small but significant percentage of touches, especially among 7—13 year olds. Children had poor success rates with targets smaller than 0.375″, though adults maintained accuracy on targets as small as 0.25". Edge padding led to a significant increase in error rate, especially among children. Gao and Sun (2015) compared 40 older adults (52—81) and 40 younger adults (19—24) in a target acquisition task on a desktop-sized screen, evaluating button spacing, button size, touchscreen technology, and screen inclination angle, and found that older adults' completion times and error rates were worse than younger users; notably, younger users improved in both measures as button size increased, while older users stopped seeing performance gains as buttons increased beyond 10.6 × 6.0 mm (for speed) and 15.9 × 9.0 mm (for accuracy). Additionally, older users reported the greatest perceived difficulty, but highest satisfaction, in using the smallest button size. Hwangbo et al. (2013) reported two studies on a handheld device that evaluated target acquisition performance of elderly users (65—79) given different target size, target spacing, and feedback types. They found

that larger button sizes and greater button spacing yielded faster completion times, lower error rates, and greater subjective suitability scores, though performance gains were not seen at the largest button sizes; auditory + tactile feedback yielded the strongest improvement to objective performance; participants found tactile-only feedback less usable and satisfying than other feedback modes, as well as yielding higher reported workload. Jochems et al. (2013) evaluated three age groups of adults (20—38, 40—59, 60—72) in a target acquisition task on a desktop-sized interface, comparing touchscreen input with eye-gaze and mouse input. They found that touchscreen performance was strongest for all three groups; while overall execution time increased with age when considering all devices, this effect was not seen for the touchscreen interface when considered separately.

Seven studies reported in 4 papers (Chang et al., 2014; Gao and Sun, 2015; Mihajlov et al., 2015; Vatavu et al., 2015) evaluated the execution of various gestures (drag, scale, rotate) among age groups. All these studies used variations on target acquisition tasks appropriate to gesture interactions to evaluate performance. Chang et al. (2014) evaluated the gesture performance (drag, scale, rotate) of participants in three age groups (children 11—14; adults 20—28; elderly 65—84) while using handheld, tablet, desktop, and large format touchscreen displays, evaluating performance in terms of performance value. Four studies reported in Gao and Sun (2015) evaluated 40 older adults (52—81) and 40 younger adults (19—24), with separate studies investigating tapping, dragging, rotating, and zooming. Independent variables included touch technology used (optical vs. SAW), error rate, task time, and subjective measures of ease of use and satisfaction. Mihajlov et al. (2015) evaluated the learnability of drag and rotate gestures on a large format display, among older adults (54—82) with limited technology experience, recording gesture learning time and task time. Vatavu et al. (2015) examined the performance of tap, double tap, drag-and-drop, and double drag-and-drop gestures among children aged 3—6, measuring completion rate, task time, offset, and accuracy. The salient outcomes of all these studies regarding age-related differences in gesture performance are covered under RQ1; readers are referred to Section 3.2 for details.

Four studies (Chung et al., 2010; Elton and Nicolle, 2013; Smith and Chaparro, 2015; Sonderegger et al., 2016) focus on touchscreen key-entry tasks among various age groups. Chung et al. (2010) compared number entry in physical and touchscreen keypads between younger adults (25.6 ± 3.2 years) and older adults (69.0 ± 3.4 years), evaluating performance of three numeric entry tasks (4 digits, 4 digits hidden, 11 digits) in terms of mean entry time per unit (i.e., normalized task time), error rate, and subjective ease of use. Keypad type did not affect the performance of younger users, but older users had faster performance on touchscreen keypads than on physical keypads. All users considered the touchscreen keypad to be easier to use than the physical keypad. Elton and Nicolle (2013) compared the effect of cold (5 °C) and ambient (19—24 °C) temperatures on the task performance of older users (65—81) in performing common manual tasks. Although the tasks were not directly comparable, three tasks related to data entry in technology settings: physical numeric keypad entry on a mobile phone; text entry on a soft QWERTY keyboard on a touchscreen device; and equation entry on a stylus-based PDA. Task performance was evaluated in terms of task completion time. While cold temperatures increased the task time for physical keypad entry by 6%, neither the stylus nor the touchscreen saw a significant change in performance under the cold condition. Smith and Chaparro (2015) compared the performance of younger (18—35) and older (60—84) adults in text entry tasks on handheld devices across five text input methods: physical QWERTY keyboard, virtual QWERTY keyboard, SHARK-style word-gesture input (e.g., Zhai and

Kristensson, 2012) on a virtual QWERTY keyboard, handwriting recognition, and voice recognition. Dependent measures included adjusted words-per-minute, error rates, and subjective measures of usability. Older adults were found to be significantly slower than young adults in all conditions except voice input, and committed more errors in all screen-based input modes; while younger adults rated physical QWERTY and voice highest for usability, older adults rated only voice most highly. Sonderegger et al. (2016) compared the typing and menu navigation performance of younger (19–29) and older (52–73) adults, evaluating physical and virtual QWERTY keyboards as a between-subjects factor. Dependent measures included the Post-Study System Usability Questionnaire (Lewis, 1995), the Positive and Negative Affect Schedule (Krohne et al., 1996), and NASA-TLX (Hart and Staveland, 1988) to measure subjective measures of usability, affect, and workload, respectively, and evaluated task performance through typing speed, number of clicks, and number of errors. Both sample populations were slower with touchscreen typing than QWERTY typing. Younger users' speed for both touchscreen and physical keyboards was higher than older users', though young adults suffered from a greater relative loss of speed when switching from physical to touchscreen keyboards. Although no statistical difference in error rate was found, an evaluation of speed-accuracy tradeoff indicated that younger users were more likely to favor speed over accuracy, and older users the opposite. Older users demonstrated increased negative affect in a touchscreen context versus the physical keyboard. However, although younger adults rated the touchscreen as having the higher workload and less efficient perceived performance than the keyboard, older users felt the opposite, indicating that the touchscreen felt more efficient and less taxing.

Three further studies investigated dual-task performance with a touchscreen component, text input driving, and thumb coverage area, each comparing performance between two age groups. McLaughlin et al. (2009) examined how the correspondence between input device and task affects task performance between younger (18–25) and older (60–70) adults, in a setting with divided attention. Divided attention was implemented by instructing users to conduct a video game task with a joystick controller in their left hand while using a simulated entertainment system with their right hand. Additional independent variables included interface type (either rotary controller or touchscreen, a between-subjects variable) and attention split between tasks (20/80, 50/50, 80/20, 100/0); dependent variables included task time and task accuracy in the simulated entertainment system and task accuracy in the video game. Matched and mismatched tasks were modeled by different tasks in the entertainment system emulator, some of which corresponded more strongly to touch selection, and others of which were more effectively modeled by rotary selection. Thus, performance in input-matched and input-mismatched tasks was also treated as a dependent variable. The authors found that older adults' performance was degraded more severely by divided attention than younger adults' performance was, and further, older adults' performance was reduced more greatly than younger adults' performance when there was a mismatch between task and input type. However, under divided attention conditions, older adults gained a considerable relative performance boost when input type matches task in comparison to younger adults. In other words, a match between task and input type can help to mitigate the effects of divided attention for older users. Neurauter et al. (2012) evaluated the driving performance and secondary task performance of users entering text on a handheld device while driving, evaluating touch and voice inputs and comparing the performance of younger (18–30) and middle-aged (45–55) adults. They found that older adults' performance with voice input was comparable to that of younger adults, but considerably worse with touch input;

subjective ratings for mental demand and frustration were likewise higher for older adults than for young adults when using touchscreens, but similar when using voice input. Xiong and Muraki (2016) evaluated the thumb reach of young (23.6 ± 1.8) and elderly (67.5 ± 3.7) adults for two different handheld touchscreen devices, measuring the area and center of the thumb reach envelope. They found that elderly participants' thumb coverage area was approximately equal to younger participants, though the centers of the thumb envelope for older users were generally somewhat to the left of and above the centers of the younger users' thumb areas.

3.4.2. Disability

Three studies evaluated users with disabilities; all three compared the performance of users with motor-control disabilities to those with typical motor performance. The studies differed in task (two tested physical versus virtual keypad number entry while the third evaluated performance in a reciprocal tapping task) and in the precise separation of participants into study groups (two studies distinguished fine- and gross-motor control disorders while the third simply separated on presence/absence of motor control disability). All three studies evaluated task performance (task time, miss rate), while two evaluated the force and impulse used to tap the touchscreen; additional dependent measures included dwell time, NASA-TLX, and user preference. All three studies used the same desktop-sized touchscreen hardware for their studies (a 15" Elo TouchSystems 1537L resistive touchscreen).

Two studies evaluated motor-control disabled individuals' use of keypads. Chourasia et al. (2013) evaluated how posture (sitting vs. standing) affects task performance and touch force in touchscreen use between motor-control disabled and typically abled users, in a keypad number entry task; Sesto et al. (2012) compared the performance of users with no, fine, and gross motor-control disorders in using virtual keypads while seated only. Independent variables common to both studies included miss rate, error rate, task time, and user preference; in addition, Chourasia et al. (2013) evaluated NASA-TLX workloads, touch force, and touch impulse. In Chourasia et al. (2013), posture, button size, and button spacing served as independent variables. Standing led to higher force and impulse levels for both sample populations, though differences in posture became negligible at button sizes of 20 mm or greater. Although users without motor control disabilities had higher overall performance than disabled users, the lack of interaction effects suggest that similar design considerations may be valid for both populations. In Sesto et al. (2012), independent variables included button size and button spacing. Here improvements were seen in miss and error rates for all users up to 25 mm button size, and disabled individuals saw lesser additional improvements up to 30 mm button size. Task times improved slightly with increased button size as well. Users from all populations used a force much higher than the minimum needed to register a touch on the touchscreen.

A final study evaluated motor-control disabilities in a more generic target acquisition task. Irwin and Sesto (2012) evaluated typically-abled users, users with fine motor-control disorders, and users with gross motor-control disorders in a reciprocal tapping task. Dependent variables were performance, peak dwell time, peak force, and impulse. Motor-control disabled individuals had longer dwell times and higher impulse than typically-abled users, and those with gross motor-control disorders had higher impulse than expected given their dwell times; as in Sesto et al. (2012), all participants used a force considerably higher than the minimum level of force needed to register a touch on the touchscreen.

It should be noted that all three of these studies shared authors, and appeared to share study apparatus as well. Given the

demographics of the sample populations, it seems likely that all three studies may be reporting findings from a series of studies conducted on the same or similar study populations; this seems especially likely for Chourasia et al. (2013) and Sesto et al. (2012). In addition, a number of papers by the same set of authors were excluded from this review due to duplicate reporting. Although this by no means invalidates the findings of these studies, it seems prudent to temper the weight given to these studies, as the findings that the papers share may not actually represent replicated results but rather different views on what may ultimately be the same (or very similar) study data.

3.4.3. Handedness

One paper evaluated the influence of handedness on preferred touch area and bodily discomfort in touchscreen use. Kang and Shin (2014) compared left-handed, ambidextrous, and right-handed users (as determined by the Edinburgh Handedness Inventory, Oldfield, 1971) in touching and scrolling tasks on a desktop-sized touchscreen while seated. Handedness was the only independent variable, with subjective discomfort ratings, hand preference, and touch area preference as dependent variables. They had participants browse the web, and tracked all touch and scroll actions, along with the screen area in which the action took place (dividing the screen into 9 regions), as well as the hand used for each gesture. They found that bodily discomfort levels increased as duration of use increased for all groups, and that both right-handed and left-handed users preferred the dominant hand for scrolling tasks. For tapping tasks, however, both left-handed and ambidextrous users used whichever hand was most convenient at the time. All groups preferred to tap at the center-middle region of the screen and scroll in the lower-middle of the screen.

4. Discussion

One of the purposes of this review is to inform the development of systems with respect to whether and how touchscreens should be implemented, given the system context. To that end, we identify the following findings from our inquiries into the research questions.

4.1. Findings drawn from RQ1

In our review, we identified research comparing touchscreen hardware and driver technology, evaluating interface design elements, examining user capabilities with respect to input interactions, and determining the effect of visual, audio, and tactile feedback on user performance and satisfaction in a touchscreen setting. We found that each of these dimensions can affect performance, user satisfaction, and safety when touchscreens are in use. Our findings are summarized below, with implementation recommendations intended to maximize performance, safety, and satisfaction (note that additional safety findings with respect to touchscreen technology are addressed under RQ2, Section 4.2).

With respect to hardware selection, we identify recommendations on touch technology, size, and palm rejection technology. Capacitive screens demonstrate a clear advantage over resistive in terms of performance and user satisfaction, especially in an automotive context (Burnett et al., 2013), and presumably in similar contexts where responsiveness to touch inputs has safety implications. However, care should be taken to consider whether users may be wearing gloves, or may otherwise be unable to trigger capacitive technologies (e.g., where users may rely on prosthetic limbs). Optical screens provide a strong subjective experience at the cost of slightly impaired performance (Gao and Sun, 2015), so they may be considered over SAW screens, except where mis-

touches must be minimized for safety or performance reasons. For cases where gesture-based interactions may be encountered across multiple screen sizes, rotation-heavy interfaces should be restricted to smaller screen sizes, while dragging-focused interfaces should be avoided on small screens, especially where the user population may include elderly users. Finally, if palm-rejection technology is available, it may lead to improved ergonomics and task performance (Camilleri et al., 2013). Although Camilleri et al. did not find users explicitly taking advantage of palm-rejection technology to use the screen itself for support, it may be that this behavior is learned over time, yielding improved ergonomic outcomes.

For interfaces used on touchscreen devices to be successful, they must be designed to optimize for the touchscreen setting. For example, input offset is a recognized issue when icons must be touched in a touchscreen interface (Im et al., 2015; Park and Han, 2010a); Park and Han attribute this in part to the interaction between apparent thumb position and contact point between thumb and screen. In addition, we hypothesize that in some cases users may be striving to maintain a visual understanding of the icon image while targeting it; thus the user attempts to touch the icon with her or his thumb without fully occluding it. Regardless of the cause, target offset should be considered in the design of icons and other targets; in particular, larger targets present an immediate way to reduce offset effects. In addition, the presence of a visual cue on the icon (Huang and Chen, 2010) may help to drive the visual aspects of targeting.

A number of dimensions of interface design have implications on inclusive design; by designing as to accommodate users with diverse needs and capabilities we can design touchscreen interfaces that reach the widest possible audience. The following recommendations are made with this in mind. In virtual QWERTY keyboards, keys should be made at least 16×16 mm for most users, but when possible 19×19 mm yields a more inclusive design (Kim et al., 2014a). In a handheld setting, icons of at least 12 mm (Hwangbo et al., 2013) and an icon width-to-height ratio of 0.9 (Im et al., 2015) will help to accommodate older and novice users; in general, icons with a depth cue of approximately 10% of target size (Huang and Chen, 2010) will help to signal interactivity to users. We recognize that depth cues in particular are not fashionable as design elements at the moment; presumably some other interactivity cue (e.g., button border) of similar size would be similarly effective. Primary controls should be designed to be easily reachable, especially considering that handheld devices are often held in a one-handed grip; to that end primary controls should be kept towards the top edge or center of the screen to maximize availability to thumb-grip users (Park and Han, 2012a; 2012b). When visual-search of data may be needed, presenting that data in a vertical scrolling format may be appropriate for a handheld device, though multiple pages should be considered depending on the nature of the information being presented (Garcia-Lopez et al., 2015).

Interaction design likewise provides an opportunity for designing for inclusivity, with respect to both reading user input and providing feedback to those inputs. Furthermore, carefully considered input design can help to bridge the gulf of execution (for input design), which can be especially critical for novice or technology-averse users; and well-designed feedback can help to conquer the gulf of evaluation, which likewise can help to ameliorate the anxieties of novices, while also providing signal reinforcement that can help to offset noisy or attention-divided environments. When designing for very young children or the elderly, designers must compensate appropriately for age-related gesture performance limitations (Vatavu et al., 2015; Chang et al., 2014). As a particular example, given the enhanced performance

of button-based alternatives to multitouch gestures, as well as the fact that button-based approaches are preferred by elderly users, a reasonable design approach would be to provide both means of executing zoom and rotate functions. Indeed, gestures, especially multitouch gestures such as rotating, may not be sufficiently intuitively learnable, especially for technology-naïve and/or older audiences (Mihajlov et al., 2015), so for interfaces that rely on such gestures that may include audiences with limited technology experience, training or contextual instruction in executing gestures should be provided. For interfaces that rely heavily on drag-to-scroll, the interface can be made more efficient and appealing to inexperienced users by using viewport scrolling with a gain level of 3–5 (Kwon et al., 2011). To provide feedback on user input, multiple feedback channels are well-advised. Providing at least one alternative to visual input can counteract the effects of visual load (Pitts et al., 2012a); even if additional channels do not enhance performance, they can provide a more satisfying experience (Pitts et al., 2012b; Hwangbo et al., 2013; Williamson et al., 2011). In particular, visual + audio + tactile feedback should be considered for touch inputs to maximize the inclusivity of the design; likewise visual + tactile feedback without audio should be avoided particularly if the audience includes older users (Hwangbo et al., 2013). Finally, for key-entry tasks, multimodal (positive and negative) feedback with separate messages assigned to auditory and tactile responses may help to reduce error rates.

4.2. Findings from RQ2

In our review, we sought research into the use of touchscreens in different task-oriented settings. We identified considerable research in the use of touchscreen IVISes and touchscreen phones during road driving, as well as a number of studies across several settings (battlefield, maritime, agricultural, general unwanted movement) that had a common factor of extraneous movement potentially interfering with the input task. Touchscreens were evaluated in medical settings in terms of microbial contamination and in terms of user experience for self-directed patient data input. Additional studies focused on robotic teleoperation, aircraft flight decks, bicycle use, and e-readers. These studies indicated numerous implications of touchscreen use to performance, safety, and satisfaction in both primary and secondary tasks. Our findings are summarized below.

Numerous studies evaluated the use of a touchscreen IVIS in a road vehicle while driving in comparison with other input methods, with additional studies exploring the use of touchscreen handheld devices in the same setting. Naturally, these studies were all focused on the effects of touchscreen use on the primary road-driving task, with a particular eye to safety; secondary task performance was also a key element of interest. We provide the following recommendations on touchscreens in a road driving setting. With respect to driving safety, several studies found that IVIS tasks of any sort, regardless of input method, lead to attentional or driving behavior that could be considered unsafe (Hagiwara et al., 2014; Jin et al., 2015; Young et al., 2014); nevertheless, as trends in automobile design and driver behavior tend towards the use of IVISes or other devices, any recommendations that can mitigate unsafe driving behavior are warranted. In particular, capacitive touchscreens are more responsive than resistive touchscreens in IVIS tasks, leading to a reduction in long (2 + second) glances off the road (Burnett et al., 2013); touchscreen input seems to yield safer driving behavior than rotary controller input (Harvey et al., 2011); and touchscreen phones seem to degrade driving performance less during texting than keypad phones (Young et al., 2014). However, for secondary tasks where voice input is a possibility, the selection of voice input may be

universally safer and more efficient than touchscreen input, especially for older users (Hagiwara et al., 2014; Neurauter et al., 2012); this finding is supported by the stronger performance of voice input over touch input outside the driving setting as well, once again especially for older users (Smith and Chaparro, 2015). Although driving performance is critical to safety in the on-road setting, secondary task performance is also important, not least because more efficient secondary tasks clearly reduce the attentional divide away from driving. Findings for input selection for secondary task performance, presumably for this reason, align with the findings for driving performance. In particular, capacitive IVIS touchscreens lead to faster and more satisfying secondary task performance than resistive touchscreens (Burnett et al., 2013); touchscreens lead to lower overall task times than rotary controllers or touchpads (Burnett et al., 2011; Harvey et al., 2011), though there is some evidence that rotary controllers may yield better performance for fine-tuning tasks (Rydstrom et al., 2012). Voice input, once again, seems to be a reasonable alternative to execute secondary tasks efficiently and in a satisfying manner, especially for older users (Hagiwara et al., 2014; Neurauter et al., 2012); when considered with the improved driving performance it affords, therefore, voice input should be implemented as an option when possible for maximum safety.

Additional studies investigated the effect of touchscreen IVIS interface and interaction design on driving performance and safety, as well as secondary task performance and user satisfaction. Once again, several studies found that regardless of design, touchscreen IVISes may well result in unacceptable safety compromises (Kujala and Salvucci, 2015; Xian and Jin, 2015), but the effect can be reduced through interface and interaction design considerations. In particular, the information display on an IVIS has a significant effect on driving performance, therefore touchscreen IVIS information displays should be brief – no more than 6 items in a list or 9 items in a grid (Kujala and Salvucci, 2015); since vertical scanning is more detrimental to safety measures than horizontal (Irune and Burnett, 2007), grid layouts may generally be preferable. Touch targets in the IVIS interface should be made as large as possible to improve both secondary task performance and subjective usability (Irune and Burnett, 2007; Kim et al., 2014a,b), at least 17.5 mm in size. For scrolling between information screens, swipe-to-scroll should be considered (Kujala, 2013) due to the fact that it limits degradation of driving performance; however, the fact that users most prefer button scrolling in this context, and that gesture-based interactions may be difficult for novices to learn (Mihajlov et al., 2015), button scrolling may be a reasonable alternative if performance degradation can otherwise be mitigated. Auditory and/or tactile feedback should be provided in response to touch inputs to reduce visual load (Pitts et al., 2012a) and to enhance subjective usability; the effect is greatest when visual, audio, and haptic feedback are combined (Pitts et al., 2012b; Williamson et al., 2011).

Several studies investigated the use of touchscreens in extraneous movement settings, and their findings can help to inform the design of interfaces in these settings. In particular, touchscreen interfaces are prone to high error rates in settings with unexpected or undesired movement (Goode et al., 2012; Lin et al., 2010; Yau et al., 2008), and neither practice nor the use of gesture input seems to mitigate this effect (Baldus and Patterson, 2008; Goode et al., 2012). In addition, touchscreens lead to higher fatigue/lower comfort than alternate input methods such as mouse and trackball (Baldus and Patterson, 2008; Yau et al., 2008). However, touchscreens had the fastest movement time (Lin et al., 2010; Yau et al., 2008), therefore in cases where fast movement time is a high priority and can come at the cost of accuracy, touchscreens are recommended. If accuracy is paramount, mouse or trackball should be considered instead, especially as these devices can take

advantage of dynamic gain.

Touchscreen interfaces show great promise in medical settings from a safety perspective, where microbial contamination is a serious concern. Not only do touchscreens seem to hold onto less contamination by virtue of their very design (Pal et al., 2013), they can also be used in a disposable wrapping that will further limit contamination (Hammon et al., 2014) without impairing their perceived usability as patient information devices. Their use in patient assessment shows promise as well, especially over stylus-based mechanisms (Herrick et al., 2013). As the spread of microbial contamination is a keen concern in clinical settings, especially where many patients may be interacting with the same device or where doctors may be using devices in the presence of many patients, the value of a contamination-resistant device that remains usable even when enclosed in a sanitary wrap cannot be overstated. Therefore, touchscreens are highly and unambiguously recommended for use in medical settings or other situations where contamination may be a serious concern.

Findings for aircraft use, robotics, e-readers, and bicycling are less concrete, due to the small number of studies in each area. Nevertheless, the findings may be informative. In aircraft flight-deck settings, touchscreens demonstrated better performance and perceived usability than other common input devices like trackballs, touchpads, and rotary controllers (Eichinger and Kellerer, 2014; Stanton et al., 2013), though the validity of these findings may be limited by the fact that the studies did not evaluate these inputs during real or simulated flight; relative workload of input devices needs to be investigated further as well. In robotic control, touchscreen inputs seem more effective than mouse or eye-gaze for pick-and-place tasks (Dünser et al., 2014) and as compared to hand gestures for more general teleoperation (Eliav et al., 2011); workload for touchscreen was found to be lower than hand gestures but higher than eye gaze and mouse input. Touchscreen texting while bicycling leads to higher lateral deviation and reduced situation awareness (De Waard et al., 2014), however these findings seem to somewhat contradict the findings of Young et al. (2014) where touchscreen texting led to reduced lateral deviation as compared to keypad texting in a road driving task. Granted, these settings are quite different, but it is not unreasonable to suspect that degradation of performance between the two vehicle-piloting tasks would occur in a similar fashion; more investigation is needed. Finally, touchscreen e-readers are seen as more convenient and usable than key-operated e-readers (Siegenthaler et al., 2012).

4.3. Findings from RQ3

In exploring research that investigated touchscreen use across different populations, we identified three key population groups that were evaluated: age groups (roughly divided into children, young adults, and older adults), motor-control disability severity, and handedness. We found that each of these groups interact with touchscreens differently, with implications for performance, user satisfaction, and safety. Our findings are summarized below, with implementation recommendations.

We begin with a review of the age-based findings, which we address separately for children, younger adults, and older adults. A key finding is that children's motor control is limited and still evolving, and this has implications for targeting tasks, gesture use, and dwell time. Age-based differences in aptitude at touchscreen use, in particular dragging and tapping performance, can be seen in children as young as 3–6 (Vatavu et al., 2015); children as old as 11–13 may linger on the touchscreen longer than expected (Anthony et al., 2014), leading to unexpected system responses. In addition, children in general have particular trouble with

multitouch gestures such as rotation (Vatavu et al., 2015), especially on larger screens (Chang et al., 2014). Due to these differences, care must be taken in developing interfaces where children are among the user base: the interfaces should be resistant to holdover touches, should provide larger targets (at least 0.375"/9.5 mm for a small device; Anthony et al., 2014), and should minimize the complexity of gestures needed to interact with the system.

Although young adults may in some ways be considered the reference population in the age-based studies, certain behaviors and needs were identified that were specific to this population group. For example, although single-touch, button-based input is often more efficient and less error-prone than multitouch gestures, young adults derive satisfaction from multitouch input (Chang et al., 2014), despite the reduced performance of these gestures. Young adults also have the greatest success with small targets: they were able to effectively interact with targets as small as 0.25" (6.5 mm), though performance was seen to increase with button size. For key-entry tasks, young adults were equally capable with virtual and physical keypads (Chung et al., 2010), but young users saw considerably reduced performance in QWERTY input on touchscreens as compared to physical keyboards, perhaps because they favored speed over accuracy; additionally, young adults evaluated touchscreen keyboards as being more taxing and less efficient than physical keyboards (Sonderegger et al., 2016).

Older adults and elderly users demonstrated an array of unique needs and preferences, as age brings with it not only reduced motor skill (and potentially cognitive skill) but also a distance from new technology. Thus the needs of older adults may serve as a good benchmark for inclusive design. To this end, the selection of input method is critical. Touchscreens seem to provide some particular benefits for older adults: compared to mice, as users age, their performance in using touchscreens degrades at a slower relative rate (Jochems et al., 2013). When attention is divided, older adults' performance suffers more severely when input method and task are mismatched, however, a good match between task and input mode can help to offset the effects of divided attention (McLaughlin et al., 2009), thus, in divided attention scenarios for older users, touchscreens may be a good choice for systems effectively controlled through tapping. Older users are faster on touchscreen keypads than physical ones, and consider touchscreen keypads easier to use (Chung et al., 2010), further, elderly performance on a touchscreen in cold environments resists impairment when compared to physical keypads (Elton and Nicolle, 2013). The benefits of touchscreen keypads for older users do not extend to QWERTY keyboards, though they find them subjectively more efficient and less taxing than physical keyboards (Sonderegger et al., 2016); a more viable alternative for older users may be voice input (Smith and Chaparro, 2015; Neurauter et al., 2012). When touchscreen systems are being designed with older users in mind, the capabilities of these users with respect to input and interaction design should be considered. Older adults tend to see greater success with large targets in tapping tasks, though improvement tapers off above a certain size (15.9 × 9.0 mm; Gao and Sun, 2015). Gao and Sun also found that younger adults had the best performance in both tap and gesture based tasks, followed by children, followed by older adults, so designs to maximize performance for older adults should largely encompass the needs of all age groups. It should be noted, however, that older users unfamiliar with gestures may need instruction to learn them effectively, especially multitouch gestures (Mihajlov et al., 2015). As far as feedback goes, older adults found tactile feedback on its own to be less usable and satisfying than auditory + tactile feedback; further, auditory + tactile feedback had improved performance over either alone (Hwangbo et al., 2013). In handheld settings, it should be remembered that grip posture and thumb coverage positions for

one-handed grips differ between older users and younger users (Xiong and Muraki, 2016). Finally, one finding of note regarding user satisfaction is that older adults often ascribe high satisfaction to tasks with a high subjective or objective difficulty level (Gao and Sun, 2015); this may be more valuable in entertainment settings than in the development of interfaces for critical tasks, however.

There were few studies of other populations, though the three studies on motor-control disabled individuals and the study on handedness did identify some key points relevant to the performance and usability of touch interfaces for these populations. Users with motor disabilities have a specific array of needs that were identified in our review. Although overall performance of motor-disabled users is lower than typically-abled users, both populations respond similarly to variations in interface design (Chourasia et al., 2013), suggesting that findings regarding normally-abled users may also apply to motor-disabled users. In addition, motor-disabled users demonstrate increased dwell time on touchscreen touches (Irwin and Sesto, 2012), similar to the holdover effect seen in younger children (Anthony et al., 2014); similar design principles to account for this effect may apply equally to both audiences. Finally, our review identified the following information on user preference in individuals with different handedness. Ambidextrous and left-handed users will readily tap with either hand; scrolling will generally be executed by the dominant hand (Kang and Shin, 2014).

One final finding is that all users seem to hit touchscreens harder than needed to activate them (Sesto et al., 2012; Irwin and Sesto, 2012). This may have implications for ergonomics and fatigue of long-term touchscreen use.

4.4. Opportunities for further research

4.4.1. Research question 1

There are a number of opportunities for meaningful research in the area of touchscreen implementation that remain unexplored. Burnett et al. (2013) and Gao and Sun (2015) demonstrate that the selection of touch input technology in particular can have a profound influence on screen usability, both in objective and subjective terms. A study to meaningfully compare a wide variety of touch technologies to each other directly, perhaps under different task conditions, would provide a clearer picture of the strengths and weaknesses of each technology and better inform selection. The ability to successfully perform gestures seems at least partly related to the touch technology used, which not only suggests that results from the current studies on gestures may not be directly comparable if studies have different input technologies, but also suggests that gesture performance may be a key evaluative performance of a comprehensive study of touch input methods. Gestures themselves are also worthy of more studies, once again perhaps more comprehensively. Several studies indicate that age plays a strong role in the ease of learning and executing gestures (Chang et al., 2014; Gao and Sun, 2015; Mihajlov et al., 2015; Vatavu et al., 2015), but again, variations in study design make it difficult to directly compare findings; a more comprehensive study across multiple age ranges may be valuable. Research into multimodal feedback, at least, has had relatively consistent findings across studies, but there is an opportunity to extend how audio, visual, and tactile feedback are combined (e.g., can the channels be used simultaneously to convey different information about the nature of the touch input, and if so, what is the bandwidth of each channel?). In addition, limited research has been done on the use of feedback modes with more complex gestures such as drag, rotate, and zoom; it remains to be seen how audio and haptic feedback can inform these common touchscreen interactions. Finally, a notable omission from the research evaluated here is the capability of multitouch

inputs to support multiple simultaneous users. The ability to support a large number of discrete touches is a key feature of many touch input technologies, and an obvious use case is to support multiple users on the same display; the capability of touchscreen inputs to support multi-user collaboration is a worthy avenue of exploration.

4.4.2. Research question 2

Despite the extensiveness of research into some task settings (such as touchscreen IVIS use), there is opportunity for more research. Although touchscreen use while driving is well-studied, since attentional divide between primary and secondary tasks has such grave implications for safety, more research in this area is welcome. As we have acknowledged, it is likely that trends in vehicle design will continue to include IVIS and handheld devices, and the complexity of such interactions seems likely only to increase. Therefore, any findings that can lead to more efficient task completion and safer driving performance are of value, especially as touchscreen secondary tasks have been demonstrated to degrade driving performance to an unreasonable degree. Voice input, in particular, seems promising as a means of reducing secondary task glance time and improving driving performance, so more research is warranted here. Additionally, a more thorough investigation into the relative strengths and weaknesses of touchscreens vs. rotary controllers may provide clarity into when and how these inputs should be chosen. For touchscreen IVISes, any efficiencies that can be gained from improved interface or interaction design of touchscreen inputs would likely improve safety as well. One key area not represented in the extant studies is the effect of touchscreen IVISes on new drivers (e.g., those with no more than 1 year of solo driving experience), as it is known that this audience presents especial safety risks. Research in understanding how IVIS design affects the performance of these drivers, and how they can be implemented to maximize safety for this audience, would be valuable.

Further investigation into the use of touchscreens across different task settings may be conducted along a number of different avenues. Research into touchscreen use in extraneous movement settings demonstrates consistent findings; however, it is not clear that mouse or trackball input is appropriate in the presence of extreme movement either, given the mechanics of their designs. There is an opportunity for novel means of input, specifically designed for settings with a high degree of extraneous movement, to be explored. The findings of the research identified in medical settings is extremely promising, especially considering the concerns in that setting relating to microbial contamination; research into how to maximize the sanitary qualities of touchscreens in this setting would be very valuable. Although touchscreen inputs in aircraft seem promising based on the studies seen here, more practical evaluations of touchscreen cockpit interfaces (i.e., with simulated or real-world flight task) are needed for validation; further, the relative workload of different input devices needs to be investigated further. In the bicycling setting, further investigation into the use of handheld devices may be warranted as well; the effects of secondary tasks and attention splitting have a great degree of safety relevance, but the setting is sufficiently different from automobiles that it cannot be assumed that findings in automobiles will transfer. Finally, there are an endless variety of different settings where touchscreen interfaces seem wholly unexplored. For example, the use of touchscreens in industrial control apparatus may be worthy of investigation, as these settings often combine unwanted movement, high background noise, and safety-critical decisionmaking; it is worth investigating how touchscreens, with their relative lack of tactile affordances and positive action, may affect safety and performance

in such a setting. In addition, the ubiquity of touchscreens in household appliances and home automation is worthy of study: although the use of touchscreens for such purposes seems to be part of the current zeitgeist of today's consumer industrial design culture, whether touchscreen inputs are appropriate for these uses has not been formally explored.

4.4.3. Research question 3

In our review of research question 3, we found a large number of studies investigating the performance preference of users of different ages, as well as a number of studies on motor-control disorders and a single study on handedness. While these studies identified several important findings, there were significant opportunities present for additional research into each population, as well as a number of populations not adequately investigated. Across all studies, one dimension that was not well-researched was the safety implications of touchscreen use for these populations, outside a limited number of studies regarding older users and driving safety. More investigation into the ergonomics of touchscreen use for specific populations, including fatigue, repetitive stress, and degradation in performance over extended use sessions, would be warranted to understand the implications in short-term ergonomic performance and long-term ergonomic safety issues. Context-specific studies relative to specific user populations (e.g., touchscreen interfaces for consumer medical devices for the elderly or those with specific medical conditions such as diabetes) is needed to better understand how touchscreen interfaces might affect the safety of users among those populations, for example in health-critical medical interventions. The studies on motor-control disabilities yielded important findings, but questions linger regarding the applicability of those studies due to the apparent overlap between study populations; certainly this is a population where more research is warranted. Finally, there may be opportunities for the use of touch interfaces in training or treating mentally disabled and learning disabled individuals, since the direct interaction that touchscreens afford may help to mitigate the challenges of mapping input to task for users with impaired cognition.

4.5. Limitations of this review

This review is limited by the following factors. First, the systematic review process may have failed to discover or incorrectly excluded relevant articles. Second, our definitions of "implementation," "task setting," and "population" may not represent the full scope of variables underlying these concepts. Third, specific research areas and research types excluded, such as case studies, research on handwriting recognition, conference papers, performance modeling, etc., may provide valuable additional insight into the research questions. Fourth, studies relevant to selection for or against touchscreens, but not specific to touchscreen implementation, task settings, or populations, has been excluded; for example, general ergonomic research into touchscreens may inform the selection of touchscreens purely based on the body posture of the user. Fifth, generalizability of the findings in these studies cannot be guaranteed, especially given the relatively small n of many of the studies. Finally, this is a systematic review, not a meta-analysis; several of the covered research areas, in particular studies on touchscreen use and age, and studies on touchscreen IVISes in automobiles, are sufficiently rich where a meta-analysis approach may provide more comprehensive and meaningful findings.

5. Conclusions

We exhaustively reviewed recent literature on the HFE performance of touchscreens across touchscreen implementations, across task settings, and across populations. We defined the HFE concerns of interest broadly as task performance, user satisfaction, and safety. With respect to implementations, we focused on how touchscreen technology, interfaces, and interaction design affect these HFE concerns; with respect to task settings, we considered environments where specific environmental factors or situational tasks might affect interface use or where interface use might affect situational tasks; and with respect to populations, we considered any and all properties of touchscreen users that are intrinsic and generally immutable through the users' actions.

We find that each of these dimensions has the potential to interact with the selection of touchscreen inputs to affect the identified HFE concerns. For implementation, the identified dimensions were touchscreen hardware selection, touchscreen driver capabilities, design of the touchscreen interface, inputs the interaction design was capable of supporting, and feedback provided by the system in response to interaction. We identified the following categories of task settings: on-road motor vehicle driving with touchscreen inputs used to execute secondary tasks; settings with unwanted extraneous motion that might interfere with the execution of computing tasks; aircraft flight decks; robotic teleoperation; bicycling while executing secondary tasks; and e-reader use. Three population groups were evaluated in the context of touchscreen use: age group (where children, young adults, and older/elderly adults were considered), motor control disabilities (where typically-abled users, users with fine motor-control disabilities, and users with gross motor-control disabilities were considered), and handedness (where left-handed, ambidextrous, and right-handed subjects were considered).

The considerations relevant to each dimension are highly varied and context-dependent. Therefore we identified key findings across all studies and made recommendations for each context separately. It should be noted that while we identified evidence for each recommendation in the evaluated studies, these recommendations must be considered preliminary in nature, as a number of them are not (or only weakly) verified among other studies. Nevertheless, the findings and recommendations provide a set of guidelines for evaluating computing scenarios to determine whether touchscreen input may be appropriate, and to create more effective, more satisfying, and safer touchscreen implementations in any context. In addition, we identified a number of opportunities for further research into touchscreen technology from the contexts of implementation, task setting, and user populations.

It is clear that decision to develop a system around a touchscreen interface must be made with care if safety, performance, and user satisfaction are to be prioritized. It is hoped that this review effectively summarizes the major issues concerning touchscreen implementations in context, so that systems designers can make an informed decision about the human factors and ergonomics implications of touchscreen integration into their designs, and so that HFE researchers can conduct effective and meaningful research on touchscreens in the areas most in need of attention.

Foundation.

References

Anthony, L., Brown, Q., Tate, B., Nias, J., Brewer, R., Irwin, G., 2014. Designing smarter touch-based interfaces for educational contexts. Personal. Ubiquitous Comput. 18 (6), 1471—1483.

Apple, 2007, January 9. Apple Reinvents the Phone with IPhone [Press Release]. Retrieved from. http://www.apple.com/pr/library/2007/01/09Apple-Reinvents-the-Phone-with-iPhone.html. on 15 November 2016.

Baldus, T., Patterson, P., 2008. Usability of pointing devices for office applications in a moving off- road environment. Appl. Ergon. 39 (6), 671—677.

Beck, F., Stumpe, B., 1973, May 24. Two Devices for Operator Interaction in the Central Control of the New CERN Accelerator [Memorandum]. CERN, Geneva.

Bjorneseth, F.B., Dunlop, M.D., Hornecker, E., 2012. Assessing the effectiveness of direct gesture interaction for a safety critical maritime application. Int. J. Hum. Comput. Stud. 70 (10), 729—745.

Burnett, G., Lawson, G., Millen, L., Pickering, C., 2011. Designing touchpad user-interfaces for vehicles: which tasks are most suitable? Behav. Inf. Technol. 30 (3), 403—414.

Burnett, G.E., Large, D.R., Lawson, G., De-Kremer, S., Skrypchuk, L., 2013. A comparison of resistive and capacitive touchscreens for use within vehicles. Adv. Transp. Stud. 31, 5—16.

Camilleri, M.J., Malige, A., Fujimoto, J., Rempel, D.M., 2013. Touch displays: the effects of palm rejection technology on productivity, comfort, biomechanics and positioning. Ergonomics 56 (12), 1850—1862.

Chang, H., Tsai, T., Chang, Y., Chang, Y., 2014. Touch panel usability of elderly and children. Comput. Hum. Behav. 37, 258—269.

Chourasia, A.O., Wiegmann, D.A., Chen, K.B., Irwin, C.B., Sesto, M.E., 2013. Effect of sitting or standing on touch screen performance and touch characteristics. Hum. Factors 55 (4), 789—802.

Chung, K., Shin, G., 2015. Effect of elastic touchscreen and input devices with different softness on user task performance and subjective satisfaction. Int. J. Hum. Comput. Stud. 83, 12—26.

Chung, M.K., Kim, D., Na, S., Lee, D., 2010. Usability evaluation of numeric entry tasks on keypad type and age. Int. J. Ind. Ergon. 40 (1), 97—105.

Cockburn, A., Ahlstrom, D., Gutwin, C., 2012. Understanding performance in touch selections: tap, drag and radial pointing drag with finger, stylus and mouse. Int. J. Human-Computer Stud. 70 (3), 218—233.

De Waard, D., Lewis-Evans, B., Jelijs, B., Tucha, O., Brookhuis, K., 2014. The effects of operating a touch screen smartphone and other common activities performed while bicycling on cycling behaviour. Transp. Res. Part F Traffic Psychol. Behav. 22, 196—206.

Dünser, A., Lochner, M., Engelke, U., Rozado Fernández, D., 2014. Visual and manual control for human-robot teleoperation. IEEE Comput. Gr. Appl. 35 (3), 22—32. http://dx.doi.org/10.1109/MCG.2015.4.

Eichinger, A., Kellerer, J., 2014. Between laboratory and simulator: a cognitive approach to evaluating cockpit interfaces by manipulating informatory context. Cognition, Technol. Work 16 (3), 417—427.

Eliav, A., Lavie, T., Parmet, Y., Stern, H., Edan, Y., 2011. Advanced methods for displays and remote control of robots. Appl. Ergon. 42 (6), 820—829.

Elton, E., Nicolle, C., 2013. Designing inclusive products for everyday environments: the effects of everyday cold temperatures on older adults' dexterity. Univers. Access Inf. Soc. 12 (3), 247—261.

Fitts, P.M., 1992. The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychology:General 121 (3), 262—269.

Gao, Q., Sun, Q., 2015. Examining the usability of touch screen gestures for older and younger adults. Hum. Factors 57 (5), 835—863.

Garcia-Lopez, E., Garcia-Cabot, A., de-Marcos, L., 2015. An experiment with content distribution methods in touchscreen mobile devices. Appl. Ergon. 50, 79—86.

Goode, N., Lenne, M.G., Salmon, P., 2012. The impact of on-road motion on BMS touch screen device operation. Ergonomics 55 (9), 986—996.

Hagiwara, T., Sakakima, R., Kamada, T., Suzuki, Y., 2014. Effect of different distractions on driving performance for drivers using a touch screen. Transp. Res. Rec. 2434, 18—25.

Hammon, M., Kunz, B., Dinzl, V., Kammerer, F.J., Schwab, S.A., Bogdan, C., Uder, M., Schlechtweg, P.M., 2014. Practicability of hygienic wrapping of touchscreen operated mobile devices in a clinical setting. Plos One 9 (9), e106445.

Han, B., Kim, K., 2015. Typing performance evaluation with multimodal soft keyboard completely integrated in commercial mobile devices. J. Multimodal User Interfaces 9 (3), 173—181.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX: results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Hum. Ment. workload. Elsevier, Amsterdam, pp. 139—183.

Harvey, C., Stanton, N.A., Pickering, C.A., McDonald, M., Zheng, P., 2011. To twist or poke? A method for identifying usability issues with the rotary controller and touch screen for control of in-vehicle information systems. Ergonomics 54 (7), 609—625.

Herrick, D.B., Nakhasi, A., Nelson, B., Rice, S., Abbott, P.A., Saber Tehrani, A.S., Rothman, R.E., Lehmann, H.P., Newman-Toker, D.E., 2013. Usability characteristics of self-administered computer-assisted interviewing in the emergency department. Appl. Clin. Inf. 4 (2), 276—292.

Hewlett-Packard Development Company, 2012. HP-150 Touchscreen Personal Computer with HP 9121 Dual Drives, 1983. In HP History: Virtual Museum [Online exhibit]. Retrieved from. http://www.hp.com/hpinfo/abouthp/histnfacts/museum/personalsystems/0031/. on 15 November 2016.

Huang, H., Chen, L., 2010. Enhancing human-computer interaction and feedback in touchscreen icon. World Acad. Sci. Eng. Technol. 65, 428—433.

Hwangbo, H., Yoon, S.H., Jin, B.S., Han, Y.S., Ji, Y.G., 2013. A study of pointing performance of elderly users on smartphones. Int. J. Human-Computer Interact. 29 (9), 604—618.

Im, Y., Kim, T., Jung, E.S., 2015. Investigation of icon design and touchable area for effective smart phone controls. Hum. Factors Ergon. Manuf. 25 (2), 251—267.

Irune, A., Burnett, G.E., 2007. Locating in Car Controls: Predicting the Effects of Varying Design Layout. Advances in Transportation Studies, pp. 25—32.

Irwin, C.B., Sesto, M.E., 2012. Performance and touch characteristics of disabled and non-disabled participants during a reciprocal tapping task using touch screen technology. Appl. Ergon. 43 (6), 1038—1043.

Ito, T., Miki, Y., 1997. Japan's Safety Guidelines for In-vehicle Display Systems. Proceedings from the its Congress in Berlin: 1997 its Technology Review: 4th World Congress on its, Berlin, Germany. Institute of Transportation Engineers, Washington, D.C.

Jin, L., Xian, H., Niu, Q., Bie, J., 2015. Research on safety evaluation model for in-vehicle secondary task driving. Accid. Anal. Prev. 81, 243—250.

Jochems, N., Vetter, S., Schlick, C., 2013. A comparative study of information input devices for aging computer users. Behav. Inf. Technol. 32 (9), 902—919.

Johnson, E.A., 1965. Touch display - a novel input/output device for computers. Electron. Lett. 1 (8), 219—220.

Kang, H., Shin, G., 2014. Hand usage pattern and upper body discomfort of desktop touchscreen users. Ergonomics 57 (9), 1397—1404.

Kim, J.H., Aulck, L., Thamsuwan, O., Bartha, M.C., Johnson, P.W., 2014a. The effect of key size of touch screen virtual keyboards on productivity, usability, and typing biomechanics. Hum. Factors 56 (7), 1235—1248.

Kim, H., Kwon, S., Heo, J., Lee, H., Chung, M.K., 2014b. The effect of touch-key size on the usability of in-vehicle information systems and driving safety during simulated driving. Appl. Ergon. 45 (3), 379—388. http://dx.doi.org/10.1016/j.apergo.2013.05.006.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The Impact of Driver Inattention on Near-crash/crash Risk: an Analysis Using the 100-car Naturalistic Driving Study Data (NHTSA publication no. HS-810 594).

Krohne, H.W., Egloff, B., Kohlmann, C.W., Tausch, A., 1996. Investigations with a German version of the positive and negative affect schedule (PANAS). Diagnostica 42 (2), 139—156.

Kujala, T., 2013. Browsing the information highway while driving: three in-vehicle touch screen scrolling methods and driver distraction. Pers. Ubiquitous Comput. 17 (5), 815—823.

Kujala, T., Salvucci, D.D., 2015. Modeling visual sampling on in-car displays: the challenge of predicting safety-critical lapses of control. Int. J. Hum. Comput. Stud. 79, 66—78.

Kwon, S., Choi, E., Chung, M.K., 2011. Effect of control-to-display gain and movement direction of information spaces on the usability of navigation on small touch-screen interfaces using tap-n-drag. Int. J. Ind. Ergon. 41 (3), 322—330.

Leahy, M., Hix, D., 1990. Effect of touch screen target location on user accuracy. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 34 (4), 370—374.

Lewis, J.R., 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Human-Computer Interact. 7 (1), 57—78.

Lin, C.J., Liu, C.N., Chao, C.J., Chen, H.J., 2010. The performance of computer input devices in a vibration environment. Ergonomics 53 (4), 478—490.

McCormik, E.J., 1976. Human Factors in Engineering and Design. McGraw Hill, New York.

McLaughlin, A.C., Rogers, W.A., Fisk, A.D., 2009. Using direct and indirect input devices: attention demands and age-related differences. ACM Trans. Computer-Human Interact. 16 (1).

Mihajlov, M., Law, E.L., Springett, M., 2015. Intuitive learnability of touch gestures for technology-naive older adults. Interact. Comput. 27 (3), 344—356.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA group, 2009. Preferred reporting items for systematic reviews and metaanalyses: The PRISMA statement. PLoS Med. 6 (6).

National Highway Traffic Safety Administration, 2013. Visual-manual NHTSA Driver Distraction Guidelines for In-vehicle Electronic Devices (NHTSA publication No. NHTSA-2010-0053). U. S. Government Printing Office, Washington, DC.

Neurauter, M.L., Hankey, J.M., Schalk, T.B., Wallace, G., 2012. Outbound texting comparison of speech-based approach and handheld touch-screen equivalent. Transp. Res. Rec. 2321, 23—30.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97—113.

Pal, P., Roy, A., Moore, G., Muzslay, M., Lee, E., Alder, S., Wilson, P., Powles, T., Wilson, P., Kelly, J., 2013. Keypad mobile phones are associated with a significant increased risk of microbial contamination compared to touch screen phones. J. Infect. Prev. 14 (2), 65—68.

Parikh, S.P., Esposito, J.M., 2012. Negative feedback for small capacitive touchscreen interfaces: a usability study for data entry tasks. IEEE Trans. 5 (1), 39—47.

Park, Y.S., Han, S.H., 2010a. One-handed thumb interaction of mobile devices from the input accuracy perspective. Int. J. Ind. Ergon. 40 (6), 746—756. http://dx.doi.org/10.1016/j.ergon.2010.08.001.

Park, Y.S., Han, S.H., 2010b. Touch key design for one-handed thumb interaction with a mobile phone: effects of touch key size and touch key location. Int. J. Ind. Ergon. 40 (1), 68—76. http://dx.doi.org/10.1016/j.ergon.2009.08.002.

Pitrella, F.D., Käppler, W.D., 1988. Identification and Evaluation of Scale Design Principles in the Development of the Sequential Judgment, Extended Range Scale. Bericht Nr. 80, Institut für Anthropothechnik.

Pitts, M.J., Burnett, G., Skrypchuk, L., Wellings, T., Attridge, A., Williams, M.A., 2012a. Visual-haptic feedback interaction in automotive touchscreens. Displays 33 (1), 7—16.

Pitts, M.J., Skrypchuk, L., Wellings, T., Attridge, A., Williams, M.A., 2012b. Evaluating user response to in-car haptic feedback touchscreens using the lane change test. Adv. Human-Computing Interact. 2012, 1—13.

Rydstrom, A., Brostrom, R., Bengtsson, P., 2012. A comparison of two contemporary types of in-car multifunctional interfaces. Appl. Ergon. 43 (3), 507—514. http://dx.doi.org/10.1016/j.apergo.2011.08.004.

Sesto, M.E., Irwin, C.B., Chen, K.B., Chourasia, A.O., Wiegmann, D.A., 2012. Effect of touch screen button size and spacing on touch characteristics of users with and without disabilities. Hum. Factors 54 (3), 425—436.

Shin, G., Zhu, X., 2011. User discomfort, work posture and muscle activity while using a touchscreen in a desktop PC setting. Ergonomics 54 (8), 733—734.

Siegenthaler, E., Bochud, Y., Wurtz, P., Schmid, L., Bergamin, P., 2012. The effects of touch screen technology on the usability of e-reading devices. J. Usability Stud. 7 (3), 94—104.

Siek, K.A., Rogers, Y., Connelly, K.H., 2005. Fat finger worries: how older and younger users physically interact with PDAs. In: Human-Computer Inter-action—INTERACT, vol. 2005. Springer Berlin, Heidelberg, pp. 267—280.

Smith, A.L., Chaparro, B.S., 2015. Smartphone text input method performance, us-ability, and preference with younger and older adults. Hum. Factors 57 (6), 1015—1028.

Sonderegger, A., Schmutz, S., Sauer, J., 2016. The influence of age in usability testing. Appl. Ergon. 52, 291—300.

Stanton, N.A., Harvey, C., Plant, K.L., Bolton, L., 2013. To twist, roll, stroke or poke? A study of input devices for menu navigation in the cockpit. Ergonomics 56 (4), 590—611.

Vatavu, R., Cramariuc, G., Schipor, D.M., 2015. Touch interaction for children aged 3 to 6 years: experimental findings and relationship to motor skills. Int. J. Hum. Comput. Stud. 74, 54—76.

Wickens, C.D., Lee, J.D., Liu, Y., Gordon Becker, S.E., 2004. An Introduction to Human Factors Engineering. Pearson Prentice Hall, Upper Saddle River, NJ.

Wierwille, W.W., Tijerina, L., 1998. Modeling the relationship between driver in-vehicle demands and accident occurrence. In: Gale, A.G., Brown, I.D., Haslegrave, C.M., Taylor, S.P. (Eds.), Vision in Vehicles VI. Elsevier, Amsterdam, pp. 233—243.

Williamson, A.R., Young, K.L., Navarro, J., Lenne, M.G., 2011. Music selection using a touch screen interface: effect of auditory and visual feedback on driving and usability. Int. J. Veh. Des. 57 (4), 391—404.

Xian, H., Jin, L., 2015. The effects of using in-vehicle computer on driver eye movements and driving performance. Adv. Mech. Eng. 7 (2) http://dx.doi.org/10.1155/2014/981908.

Xiong, J., Muraki, S., 2016. Effects of age, thumb length and screen size on thumb movement coverage on smartphone touchscreens. Int. J. Ind. Ergon. 53, 140—148. http://dx.doi.org/10.1016/j.ergon.2015.11.004.

Yau, Y., Chao, C., Hwang, S., 2008. Effects of input device and motion type on a cursor-positioning task. Percept. Mot. Ski. 106 (1), 76—90. http://dx.doi.org/10.2466/PMS.106.1.76-90.

Young, J.K., Trudeau, M., Odell, D., Marinelli, K., Dennerlein, J.T., 2012. Touch-screen tablet user configurations and case-supported tilt affect head and neck flexion angles. WORK A J. Prev. Assess. Rehabil. 41 (1), 81—91.

Young, K.L., Rudin-Brown, C., Patten, C., Ceci, R., Lenne, M.G., 2014. Effects of phone type on driving and eye glance behaviour while text-messaging. Saf. Sci. 68, 47—54.

Zhai, S., Kristensson, P.O., 2012. The word-gesture keyboard: reimagining keyboard interaction. Commun. ACM 55 (9), 91—101.

Zhai, S., Kong, J., Ren, X., 2004. Speed-accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. Int. J. Human-Computer Stud. 61, 823—856.

Zwahlen, H.T., Adams, C.C., DeBald, D.P., 1988. Safety aspects of CRT touch panel controls in automobiles. In: Freeman, M.H., Haslegrave, C.M., Smith, P., Taylor, S.P., Gale, A.G. (Eds.), Vision in Vehicles II. Elsevier, Amsterdam, pp. 335—344.