Interacting with Computers Advance Access published April 5, 2016

doi:10.1093/iwc/iww004

Semi-aural Interfaces: Investigating Voice-controlled Aural Flows

Romisa Rohani Ghahari[1], Jennifer George-Palilonis[1],
Hossain Gahangir[2], Lindsay N. Kaser[1] and Davide Bolchini[1]*

[1]*School of Informatics and Computing, Indiana University, Indianapolis, IN, USA*
[2]*School of Engineering and Technology, Purdue University, Indianapolis, IN, USA*
**Corresponding author: dbolchin@iupui.edu*

To support mobile, eyes-free web browsing, users can listen to 'playlists' of web content—*aural flows*. Interacting with aural flows, however, requires users to select interface buttons, tethering visual attention to the mobile device even when it is unsafe (e.g. while walking). This research extends the interaction with aural flows through simulated voice commands as a way to reduce visual interaction. This paper presents the findings of a study with 20 participants who browsed aural flows either through a visual interface only or by augmenting it with voice commands. Results suggest that using voice commands reduced the time spent looking at the device by half but yielded similar system usability and cognitive effort ratings as using buttons. Overall, the low-cognitive effort engendered by aural flows, regardless of the interaction modality, allowed participants to do more non-instructed (e.g. looking at the surrounding environment) than instructed activities (e.g. focusing on the user interface).

RESEARCH HIGHLIGHTS

- We explore a vocabulary of simulated voice commands to control aural flows.
- We empirically compare two modalities to control aural flows: using buttons vs. voice + buttons.
- Voice command users spent 50% less time looking at the device than button-only users.
- Walking speed, system usability and cognitive effort are similar in both conditions.
- In the voice + button condition, participants use significantly more voice commands than buttons.
- Across conditions, aural flows engender more non-instructed than instructed activity.

Keywords: user studies, sound-based input/output, mobile computing, multimodal interfaces, information architecture

Editorial Board Member: Rod McCall

Received 1 April 2015; revised 17 February 2016; accepted 24 February 2016

1. INTRODUCTION

Mobile devices pervade our lives, but they also consume our attention while using them, even when it is distracting or unsafe. For example, browsing mobile news while on the go is a major task (BII Report, 2015), but continuous attention to the device to consume content can easily cause unwanted distraction from the surroundings in a variety of contexts (e.g. walking) (Anhalt *et al.*, 2001) and increase the risk of accidents (Stavrinos *et al.*, 2011).

One way to reduce visual interaction with the device is to leverage the aural channel (Yang *et al.*, 2012), i.e. in which text-to-speech (TTS) technology reads content to users or audio is supplied to users in lieu of text-based information that requires focused attention on reading. This class of user interfaces (that we call 'semi-aural') has mainly an auditory output complemented by a certain degree of visual representation. Previous work introduced new techniques to dynamically linearize existing websites and generate audio 'playlists', called aural flows, optimized for eyes-free aural browsing (Ghahari and Bolchini, 2011). Aural flows can be used to stream audio content from web sources, thus, enabling users to partially focus on the surrounding

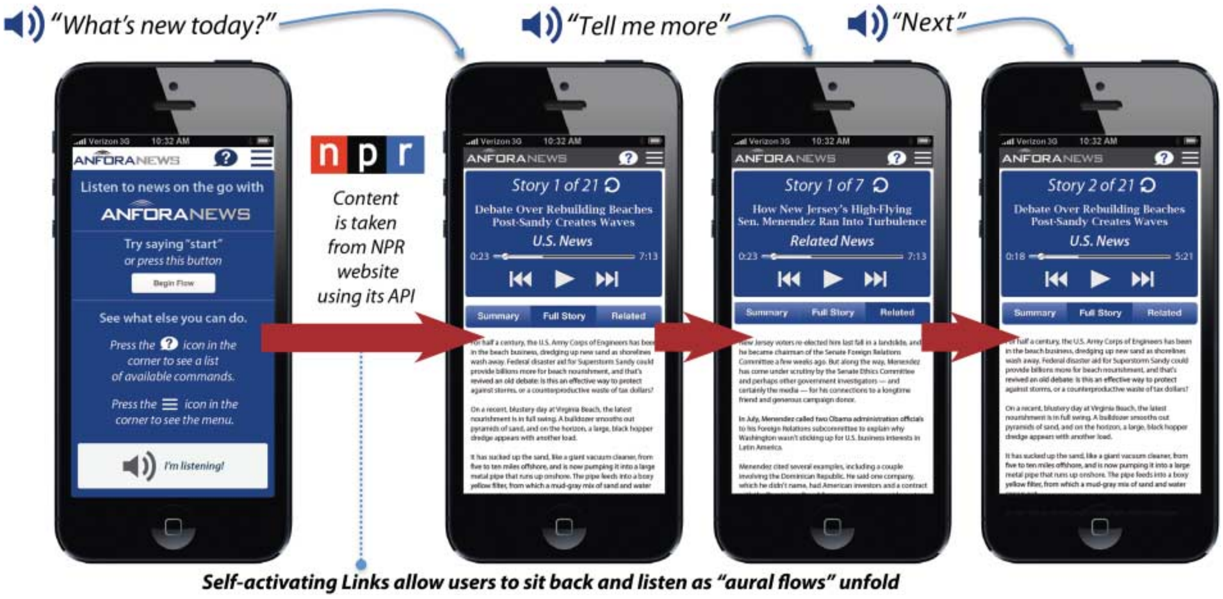
"What's new today?"
"Tell me more"
"Next"
Verizon 3G 10:32 AM
ANFORANEWS
Listen to news on the go with
ANFORANEWS
Try saying "start"
or press this button
Begin Flow
See what else you can do.
Press the ? icon in the corner to see a list of available commands.
Press the ☰ icon in the corner to see the menu.
I'm listening!
npr
Content is taken from NPR website using its API
Story 1 of 21
Debate Over Rebuilding Beaches Post-Sandy Creates Waves
U.S. News
0:23
7:13
Summary
Full Story
Related
Story 1 of 7
How New Jersey's High-Flying Sen. Menendez Ran Into Turbulence
Related News
Story 2 of 21
0:18
5:21
Self-activating Links allow users to sit back and listen as "aural flows" unfold

Figure 1. Voice commands combined with aural flows and buttons enable users to reduce visual interaction with web content. http://www.youtube.com/watch?v=a_9bgcZfpY4.

environment—instead of on the device—while engaged in a primary task, such as walking or jogging.

However, interacting with aural flows using existing input mechanisms, such as touch and gesture (Rohani Ghahari *et al.*, 2013), still forces users to pay focused attention to the screen. One way to significantly reduce visual interaction during mobile browsing is to support a rich semantic vocabulary of voice controls. This paper explores the use of simulated voice recognition to control aural flows while on the go. The rationale for exploring voice commands lies in their potential to relieve users from visual interaction with the screen while providing conversational voice controls to issue instructions while browsing an information-rich website.

To investigate the benefits and limits of voice-controlled aural flows, this paper reports a study in which participants ($n = 20$) experienced a custom-designed mobile news application rendering aural flows from npr.org in a walking environment. In the 'button' condition, users listened to and controlled aural flows through a visual user interface by selecting labels on the screen (Fig. 1). In the 'voice' condition, users controlled aural flows either with buttons or by employing an equivalent set of voice commands. In both conditions, participants walked for 15 min along the same path while listening and controlling aural flows. Aural flows were fully implemented on a working mobile system (Bolchini and Ghahari, 2013) while the Wizard-of-Oz approach leveraged a separate control device to promptly respond to participants' voice commands on a mobile phone. We chose the Wizard-of-Oz approach to quickly iterate the proof of concept and features for using voice navigation over aural flows.

We measured the time participants spent in visual interaction with the device, speed of walking, system usability and cognitive effort for each condition. Additionally, we measured frequency of using voice vs. button commands in the 'voice' condition. We also calculated number of times participants were involved in different types of activities during each task. In short, findings show that voice commands significantly reduced the amount of time that participants were required to look at the device while experiencing aural flows, but also yielded similar walking speed, system usability and cognitive effort ratings compared with the button condition. In the voice condition, participants significantly used more voice commands than button commands. Moreover, participants significantly engaged in more non-instructed activities (e.g. looking at the posters on the wall) than instructed activities (e.g. using voice or button commands) in both the conditions. All users enjoyed the directness of voice commands in combination with the visual interface, but some found voicing instructions socially uncomfortable.

Overall, this paper makes two contributions:

(i) Extends the interaction with aural flows through voice commands. These could potentially reduce visual interaction for users who browse the aural flows prepared from a content-intensive website while on the go.

(ii) Presents the findings from a controlled study to examine the amount of time that is required for visual interaction with the device, cognitive effort, and the usability of button- versus voice-controlled aural flows.

In the remainder of the paper, we introduce previous work on aural flows, explain how this research is different from existing voice user interfaces, and explain why we selected the Wizard-of-Oz technique as our design and evaluation methodology. We then present our hypothesis and study design, followed by qualitative and quantitative results, and discuss the implications of our work to advance the design of semi-aural interfaces.

2. RELATED WORK

2.1. Aural flows for mobile browsing

Previous work (Ghahari and Bolchini, 2011) introduced a semi-interactive aural paradigm—ANFORA—that enables users to listen to content-rich websites and interact with them infrequently to minimize distractions. ANFORA is based on the notion of aural flows, a 'design-driven, concatenated sequence of pages that can be listened to with minimal interaction required. Aural flow allows users to automate browsing tasks on top of web information architectures by creating a playlist that is based on the content in which they are most interested' (Ghahari and Bolchini, 2011). This approach is different from current applications such as Capti narrator (Borodin, 2014) and VoiceDream (2015), because aural flows enable users to directly select and browse an entire category of content (e.g. *US News*, *World News* etc.). Capti narrator (Borodin, 2014) and VoiceDream (2015), however, require users to select and add each piece of content or a web page individually to their playlist.

Aural flows were also previously implemented on news domain websites and named *ANFORA News* (Rohani Ghahari *et al.*, 2013). Previous iterations of *ANFORA News* used either touch or gesture commands as the primary modes of interaction with aural flows. Some systems, like EarPod (Zhao *et al.*, 2007) and Bazel-Tap (Serrano *et al.*, 2013) enable users to perform gestures without looking at the device or their hands. Unlike those systems, interacting with aural flows using touch and gesture (Rohani Ghahari *et al.*, 2013) still forces users to pay focused attention to the screen, which may be distracting or dangerous in certain situations. Thus, this study explores other semantic interaction modalities to support eyes-free experiences with *ANFORA News*. In the following sections, researchers review different literature focused on voice-based user interfaces that inspired the approach in this research and provided the foundation for voice-based controls of aural flows.

2.2. Voice-based user interfaces

During the past few years, several studies have investigated the importance of voice commands as an interaction medium. For example, the Dynamic Aural Web Navigation (DAWN) system translates HTML pages into VoiceXML pages (Gupta *et al.*, 2005). DAWN presents a small set of global voice commands for moving across documents, such as *skip* and *back*. Web-based Interactive Radio Environment (WIRE) is an in-car voice browser designed to be used safely by a driver while in transit (Goose and Djennane, 2002). Similarly, VoxBoox translates HTML books into VoiceXML (Jain and Gupta, 2007) pages that are enhanced with voice commands during document translation to improve the browsing experience and offer additional navigation controls. Voice commands such as *skip*, *back*, *start* and *pause* are also available. Finally, Apple's Siri (Apple Siri, 2015) enables people to use voice commands and ask the 'personal assistant' to do things for them, such as check the weather, schedule a meeting or set an alarm.

Voice recognition systems (e.g. Apple's Siri and Android's Google Voice) have improved dramatically in recent years, but several fundamental limitations that are recognized by researchers may lead to negative or unexpected results. For example, noisy environments (Sawhney and Schmandt, 2000), incorrect or incomplete sentences, and accents may all cause errors in a system's ability to recognize a voice command (Song *et al.*, 2012; Tang *et al.*, 2013). Although this paper does not focus on engineering voice-based interactive systems, we realize that developing robust speech-recognition systems still remains a really hard problem to date.

2.3. A design method for voice commands

Several studies have implemented the Wizard-of-Oz approach for studying voice command systems (Ashok *et al.*, 2014; Bernsen and Dybkjaer, 2001; Narayanan and Potamianos, 2002; Sinha *et al.*, 2002). In the Wizard-of-Oz approach (Dahlbäck *et al.*, 1993; Green and Wei-Haas, 1985), subjects are told that they are interacting with a computer system though they are not. Instead a human operator, the wizard, mediates the interaction. For example, SUEDE (Klemmer *et al.*, 2000; Sinha *et al.*, 2002) is an informal prototyping tool used to map and quickly test natural language interactions. SUEDE adapts the Wizard-of-Oz approach to test natural language dialogues using two modes: design mode and test mode. Design mode allows designers to map interaction flows and record voices in order to act as both computer and user. Test mode converts the dialogue sequence to a browser-based interface for the 'wizard' to use while performing the test. Along the same line, Salber and Coutaz (1993) demonstrated how the Wizard-of-Oz technique can be extended to analyse the multimodal interfaces. Oviatt *et al.* (1992) designed a rapid semi-automatic simulation method (Wizard-of-Oz approach) to compare pen and voice as an interaction modality. Likewise, another study (Vo and Wood, 1996) used the Wizard-of-Oz technique to test how users use a system in order to build a multimodal interface, which is using speech and pen as an input. Similarly, the Wizard-of-Oz technique was found to be beneficial for simulating a speech recognition system and is recommended for similar experiments in the future (Tsimhoni *et al.*, 2004).

A recent study (Ashok *et al.*, 2014) also used the Wizard-of-Oz approach to evaluate voice-enabled web browsing for visually impaired users. These studies support the notion that the Wizard-of-Oz approach is a possible method for the rapid design of voice commands.

2.4. Guidelines for effective voice commands

Researchers have also introduced guidelines for designing vocabularies for voice commands that users can easily memorize and recall. For example, one study suggests that designers should use only a few short and aurally distinct words because voice interaction is less accurate than mouse clicking (Christian *et al.*, 2000). Another study mentions that applications using small vocabularies and predefined commands can significantly reduce error rates and improve recognition accuracy (Feng and Sears, 2009). It is important to avoid multiple commands that sound alike, which leads to errors and confusion from both the user and the system. Also, a dialogue should effectively leverage a user's vocabulary, making interaction with the system natural. Additionally, Bradford (1995) suggests that a short-command vocabulary is easier to discern and retain in short-term memory. These guidelines informed the design of the vocabulary for voice commands to control the aural flows in the mobile setting.

2.5. Measuring distraction due to interaction with mobile devices

Because of the many factors composing a walking environment, this activity requires users to integrate multiple inputs and constantly attend to multiple stimuli. Specifically, prior work has acknowledged that interacting with mobile devices while walking needs a high degree of both visual (Bragdon *et al.*, 2011; Lemmelä *et al.*, 2008) and cognitive attention (Lemmelä *et al.*, 2008). Complexity of interaction plays a role in causing the cognitive distraction (Young *et al.*, 2007), while interaction mode and the nature of the secondary task affect the visual distraction (Young *et al.*, 2007). Visual distraction is measured by the number of glances and the duration of glances (Metz and Krueger, 2010), and cognitive distraction is measured through cognitive load. As shown in Table 1, cognitive load is measured directly using NASA-TLX (Task Load Index) questionnaire (Hart and Staveland, 1988) or indirectly using cognitive load theory (CLT) (Sweller, 1988). Sweller introduced different types of cognitive load such as Intrinsic Cognitive Load (ICL), Extraneous Cognitive Load (ECL) and Germane Cognitive Load (GCL). ICL (Sweller and Chandler, 1994) is the integral level of difficulty related to the task. ECL (Chandler and Sweller, 1991) is engendered by the approach in which information is presented to the subject as a part of the system design. GCL (Sweller *et al.*, 1998) is the load devoted to the processing, construction and automation of system operations related to subject's prior experiences. Measuring these three different types of cognitive load is important to understand how the interaction modality while navigating aural flows can affect cognitive effort. Moreover, understanding and measuring different types of distractions that may occur while walking and interacting with mobile devices facilitates a better experimental setup in terms of adopting the right questionnaires and data collection method.

Table 1. This research used two types of cognitive measurement: direct and indirect.

Direct measurement	Indirect measurement
NASA TLX Questionnaire	Cognitive Load Theory (CLT) = Intrinsic Cognitive Load (ICL), Extraneous Cognitive Load (ECL) and Germane Cognitive Load (GCL)

3. LINKLESS NAVIGATION OVER AURAL FLOWS

The ability to control aural flows using voice commands unleashes a 'linkless' interaction paradigm, in which users need not select interface link labels on specific pages and, instead, can activate a limited set of dialogic commands at any time.

3.1. Design methodology

In order to manifest the concept of linkless navigation, researchers first established 'full flow' as the default setting for the user experience. Full flow enables users to listen to the summaries and full versions of each news story (Fig. 2). Full flow also allows users to skip a story or go back and re-listen to a story. In addition, users also have the option to listen to related news stories for any given story.

Second, researchers defined the aural 'navigation vocabulary' to be used when moving within complex information architectures and interacting with aural flows (Fig. 2). This small and simple vocabulary of commands was inspired by common primitives identified in conceptual navigation models (Bolchini and Paolini, 2006; Bradford, 1995; Feng and Sears, 2009; Garzotto *et al.*, 1993). An aural navigation vocabulary was developed by matching new aural commands with each of the possible navigation strategies for the website. For example, a user could navigate from one news story to the next by saying 'next'. The design process for developing the final set of commands involved a team of seven designers who explored the commands and simulated the user experience through the Wizard-of-Oz technique. Although the Wizard-of-Oz approach was used, the voice commands were kept short and simple because researchers wanted users to exert less cognitive effort to enact the commands (Bradford, 1995). Table 2 lists the voice commands (and the corresponding semantics) that were

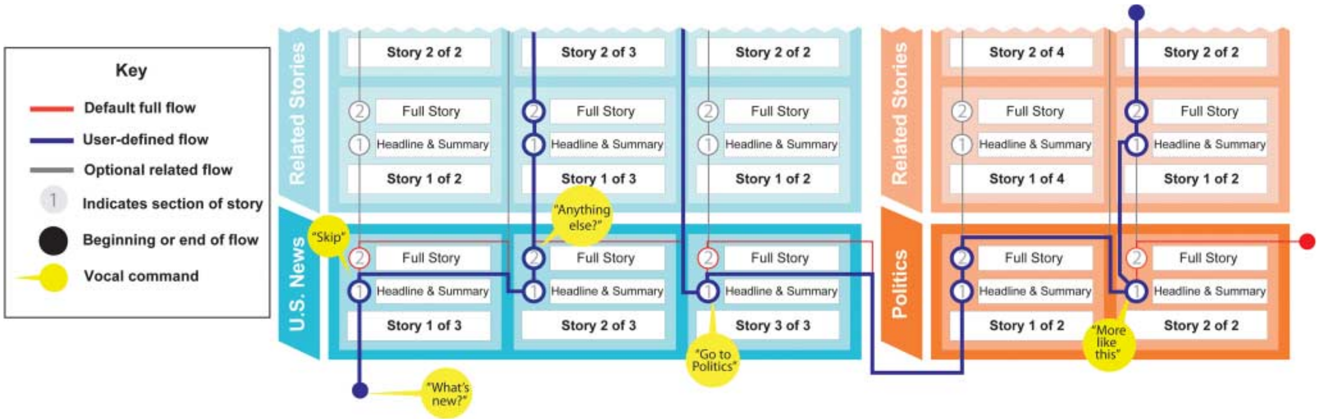

Figure 2. Semi-aural, linkless navigation strategy on ANFORA News: Architecture of aural flow types augmented by voice commands. Patent pending (Bolchini and Ghahari, 2013).

Table 2. The vocabulary of the voice commands to control the aural flows.

Voice commands	System action on aural flows
U.S., world, politics, sports, health, science, economy, or technology	Select U.S., world, politics, sports, health, science, economy, or technology news category
Start, what's new, recent news	Starts playlist of news
Restart	Restart playlist of news
Rewind	Previous section in news story
Forward	Next section in news story
Back, previous	Previous news story
Skip, next	Next news story
More, tell me more, anything else, related, like this	Related news stories
Home	Return to home page
Pause, stop, play	Click on the button to pause, resume or play

iteratively developed using this Wizard-of-Oz approach. For some of the semantics, researchers provided a few options in regard to the voice commands in order to determine which commands would be used the most.

The following basic sources were used to design our set of voice commands:

(i) The voice commands were partially inspired by the elements used to control a music player (e.g. next, skip, back, previous, pause, stop and play).

(ii) Other commands were borrowed from traditional mechanisms used to control linear media (e.g. rewind, forward, restart and start).

(iii) Another set of commands that researchers introduced was specific to the nature of aural flows (e.g. category name, what's new, recent news, home, more, tell me more, like this and anything else).

3.2. Manifesting designs in *Linkless* ANFORA

In order to explore and evaluate the implications of the proposed navigation vocabulary for users browsing complex information architectures, researchers leveraged and improved on *ANFORA News* with *Linkless* ANFORA, which supports voice control over aural flows. In *Linkless* ANFORA, the aural flows were generated in real-time from existing news source (NPR website) and read aloud to users using a TTS service (www.ispeech.org). In order to demonstrate the navigation vocabularies used for dissemination and testing, two versions of *Linkless* ANFORA have been instantiated in this research, one with button commands and one with both voice and button commands. Although the aural flows were fully implemented, the Wizard-of-Oz approach was used to control the participants' device when they used any of the voice commands.[1] Hence, one researcher manually activated the commands voiced by the user through a control console. The Wizard-of-Oz approach is a very common testing strategy for early designs of complex interfaces that need quick iterations of features that would normally require lengthy implementation processes (Dahlbäck *et al.*, 1993). For the purpose of this study, initially, the researchers conducted two rounds using the Wizard-of-Oz approach with seven designers. These designers explored and developed a set of voice commands for the evaluation study. In the evaluation

[1] The *Linkless* ANFORA prototypes are available at: Button condition: *Linkless* ANFORA 'Button'; Voice + Button condition: *Linkless* ANFORA 'Voice and Button'; Control console to manually activate voice commands: Control Console.

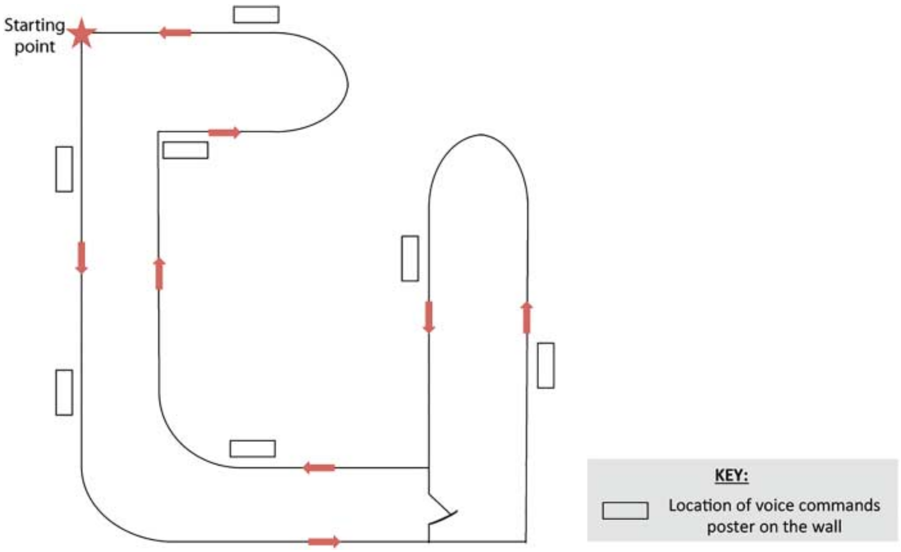

Figure 3. The path layout used in the experiment was 54.4-m long with four sharp turns, two slight turns and two U-turns.

study, however, the researchers did not use the Wizard-of-Oz approach to do a complete exploratory evaluation of the voice commands. This decision was made because it would have been difficult for the researchers to execute a random command and guess what the participants meant in a controlled evaluation study.

4. EVALUATION HYPOTHESES

Based on the principles of linkless navigation as applied to an aural website scenario, the research question (RQ) and hypotheses are as follows:

RQ: When navigating aural flows while on the go, does a set of voice commands reduce a user's visual interaction with the device and improve the user experience compared with clicking buttons in order to navigate through content?

(i) H1: Using voice commands, instead of button commands, requires less visual interaction with the device. (Although, by definition, using voice commands is expected to reduce the visual interaction, there are other factors that could come into play. For example, users might look at the screen while using voice commands because they are not yet familiar with the interaction modality or to check to see if the system did what they asked it to do.)

(ii) H2: Users will find voice commands easier to use than button commands. (Although the voice commands are expected to be a more natural form of input, both voice and button commands could cause cognitive distractions.)

(iii) H3: Users will find voice commands more enjoyable than button commands.

5. STUDY DESIGN

In order to test the hypotheses, a controlled evaluation study with 20 users was conducted. This study adopted a within-subjects design in order to maximize internal validity. What follows describes the physical set up, the detailed study design and procedures used in the study.

5.1. Physical setup

The evaluation study was conducted in an indoor navigation environment that included one large room connected to the main entrance corridor via another hallway (Fig. 3). This study established a 54.4-m long path that users walked while executing the aural browsing tasks. The path was marked on the floor using tape and included four sharp turns, two slight turns and two *U*-turns. Different static objects, such as tables and chairs, were placed along the route to simulate a real-world scenario in which an individual must safely recognize and navigate around obstacles. The participants were led through the path before they started with their tasks. The researchers limited the distractions to the available artifacts on the wall, such as posters or papers with the list of voice commands. In order to effectively compare the experience of using voice commands to button commands, this study controlled for the condition of a noisy environment. The researchers did not expect that the potential degradation of performance that might occur in a noisy setting would affect any particular problem; rather, they expected a reduction in accuracy, which would improve as the voice recognition system advanced. Additionally, the lists of voice commands were printed on an A4 size paper and placed on all the walls around the path (Fig. 3). The lists of voice commands were comfortably

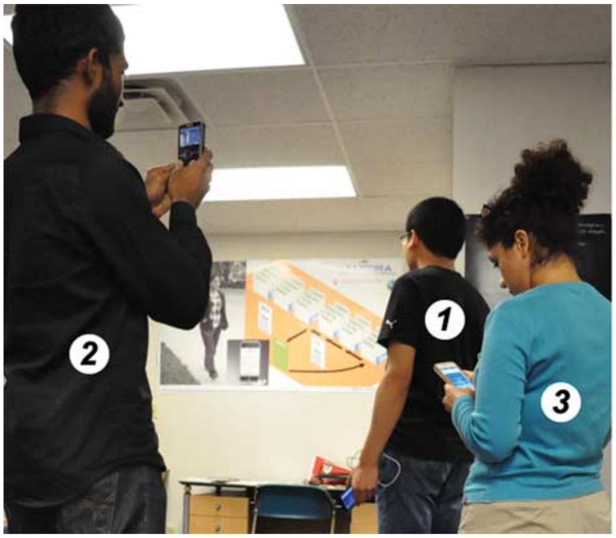

Figure 4. Experimental setup: (1) participant listens to aural flows on *Linkless* ANFORA. (2) Researcher video records the session. (3) Researcher controls the flow and interaction.

readable from a distance of 190 cm. Therefore, the users could refer to these lists at any time in order to isolate the command learnability factor of the study.

A distant side observer used a video camera to record the users' sessions and visual engagements with the application (Fig. 4). A video recorder was used for two reasons. First, the researchers did not want to add new distractions to the experiment by making people walk around with a head-mounted eye-tracking device (HED). Moreover, the condition of using a HED while walking is not externally valid. Second, the recorded video allowed the researchers to conduct post-test analysis and capture all other user activities (e.g. looking at the posters or the list of voice commands on the wall) during each task.

The participants were encouraged to listen to the TTS content using Apple headphones and interact with the application using buttons or voice commands. They were instructed to hold the phone in one of their hands with their arms down while listening to the TTS content and hold the phone up when they used the buttons to interact with it (Fig. 5). When the participants used a voice command, they had to click the button on the Apple Headphones Remote Button to simulate the real-world voice command activation. As the researcher had to walk behind the participants to hear their voice commands, the participants were made aware that the researcher was manually activating the voice commands through a control console.

5.2. Experimental conditions and study variables

The independent variable was the style of navigation over the aural flows, which varied on two levels: (i) button- or (ii) voice-plus-button commands. The researchers did not include a voice-only condition on the basis that current interfaces, such as Apple's Siri and Android's Google Voice, typically provide voice commands as only one of the *possible* modalities, and almost never employ only one interaction modality to interact. Having multiple modalities for interaction is likely to accommodate a range of individual user preferences. The dependent variables were as follows:

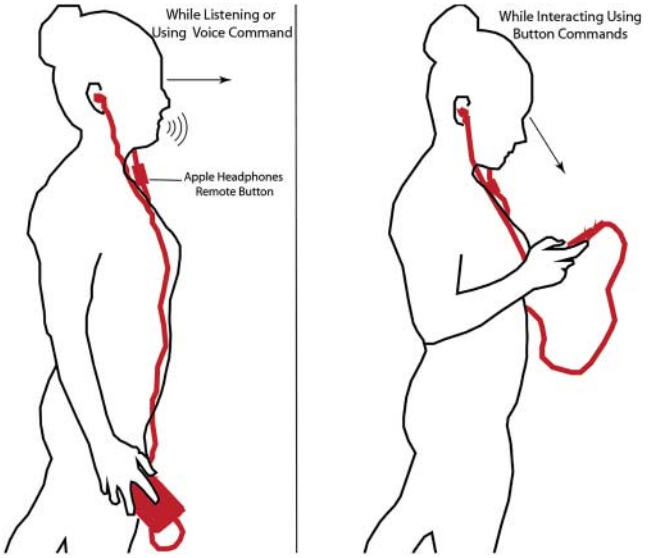
While Listening or Using Voice Command
Apple Headphones Remote Button
While Interacting Using Button Commands

Figure 5. (Left) Participant is holding the phone in her hand with her arms down while listening to the aural flows. (Right) Participant is holding the phone up when she uses the buttons to interact with the aural flows.

(i) *Interaction time (IT):* the overall time that the users were interacting with the interface *regardless of the modality* (voice or button).
(ii) *Visual interaction time (VIT):* the time that the users spent listening to the aural flows while looking at or touching the interface.
(iii) *Speed of walking*: the speed at which the participants walked while listening to the aural flows calculated by the total distance walked during a 15-min task.
(iv) *Frequency of using voice commands:* the number of times each voice command was used.
(v) *Instructed activities:* the number of activities performed by the users as instructed in the task, such as interacting via button/voice commands.
(vi) *Non-instructed activities:* the number of activities performed by the users in addition to what was instructed in the task, such as looking at and/or reading text on the interface.
(vii) *System usability:* the usability of the system as measured by the system usability scale (SUS) score (Brooke, 1996).

(viii) *Cognitive load:* the perceived mental demand of the task, as measured by the NASA-TLX (Hart and Staveland, 1988). Another strategy used to measure cognitive load is adding up the ICL, ECL and GCL scores. These scores are calculated indirectly through some of the questions in the SUS (Brooke, 1996).

The main purpose of using voice commands was to provide the users with a more eyes-free navigation experience. Thus, the researchers measured the visual interaction time in order to understand whether using voice commands required the users to look at the interface less than when they used only the button commands. In addition, visual interaction time and cognitive load were selected in order to measure visual and cognitive distraction, respectively.

5.3. Participants

Twenty participants from a large Midwestern University (10 males, 10 females) were recruited for this study. The participants ranged in age from 19 to 49 ($M = 27$; SD = 8.14) and were native English speakers and frequent news consumers. All of the participants had experience with touchscreen mobile devices and none had hearing impairments. None of the participants had prior experience with *Linkless* ANFORA. The participants each received a $20 Amazon gift card for their 90 min of participation.

5.4. Procedure

Each participant engaged in a session that consisted of three parts executed in this order: (i) training; (ii) two-stage task session, including the use of *Linkless* ANFORA in one of the two conditions, followed by usability and cognitive load surveys and (iii) a post-task interview.

5.4.1. *Training*

The participants attended a 30-min training session, during which they were introduced to *Linkless* ANFORA and briefed about the voice and button commands. In order to make sure that all of the participants could reach a common threshold of familiarity with *Linkless* ANFORA, each participant executed simple navigation tasks using different versions of *Linkless* ANFORA.

5.4.2. *Task sessions and post-task surveys*

The participants engaged in two stages of tests. The first stage used the buttons (B) as the control condition. The second stage used voice-plus-button commands (VB) as an experimental condition (hereafter to be referred to as 'voice' condition). The order in which participants engaged in each style of navigation was systematically counterbalanced across all of the participants in order to minimize the learning effect. Overall, each participant executed two tasks (Fig. 6):

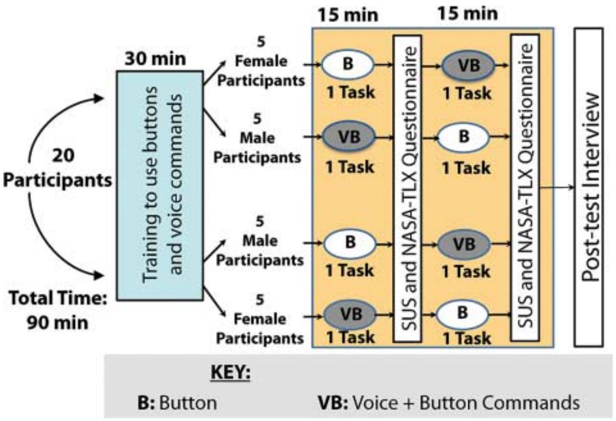

Figure 6. Within-subject design for the comparative evaluation of the different interaction modes.

(a) One task (15 min) for the button condition and
(b) One task (15 min) for the voice condition.

The structure of each task was the same across the different conditions. The only difference was the category of news stories covered. For example, the voice task was as follows.

In this version, you may navigate using either the voice or button commands. You have 15 min to use *Linkless* ANFORA. Please browse at least eight news stories during this time period and change the category to any other category at least once. Try not to listen to the category of news to which you have already listened.

The task for each condition was designed to be 15 min long because it was a good compromise between the depth and breadth of aural flows exploration and the fatigue caused by walking and listening to content. Overall, the researchers controlled for the task time (15 min), modality of interaction and continuous interaction. Within the constraint of time and modality of interaction, the researchers let the participants browse the aural flows freely in order to explore the content.

In a natural setting, users would be likely to employ several modalities at once. The combination of interaction techniques in one condition—voice and button—was used to preserve external validity. Moreover, the researchers' intentions were not to completely replace the existing button interaction techniques. Rather, they sought to provide users with more flexibility and additional options for navigating a semi-aural interface with natural and efficient aural navigation flows.

Finally, after each task, the participants rated the system's usability as well as their cognitive load using the SUS questionnaire (Brooke, 1996) and NASA-TLX questionnaire (Hart and Staveland, 1988), respectively.

5.4.3. *Post-task interview*

After the two-stage task sessions and usability and cognitive load questionnaires, the participants answered interview questions related to both conditions. The purpose of the

Table 3. Example of how the questions from the SUS were mapped to specific types of cognitive load.

Different types of cognitive load	Questions selected from the SUS
Intrinsic Cognitive Load (ICL)	Q2. I found this application unnecessarily complex. Q3. I thought this application was easy-to-use.
Extraneous Cognitive Load (ECL)	Q5. I found the various functions in this application well-integrated. Q6. I thought that too much inconsistency existed in this application.
Germane Cognitive Load (GCL)	Q4. I think that I would need assistance to be able to use this application. Q10. I needed to learn a lot of things before I could get going with this application.

interview was to understand how the participants described their experience using *Linkless* ANFORA with different modalities; which modality of interaction they preferred to use in the voice condition and why; what they liked best or least about *Linkless* ANFORA; whether they listened to the news while walking and adequately monitored their surroundings; whether the orientation cues were clear to the participants; and in what other context would the participants prefer to use *Linkless* ANFORA.

6. ANALYSIS

For the quantitative data, *repeated measure t-tests* were used in order to analyse the efficiency and effectiveness of the linkless navigation strategy as well as the effect of the interaction style. Researchers used the interaction style (i.e. button vs. voice commands) as the within-subject factor. Several outcome variables (i.e. IT, VIT, walking speed, frequency of using voice commands, instructed activities, non-instructed activities, system usability and cognitive load) were compared.

Two researchers watched the recorded videos in order to measure both the IT and VIT in order to maximize the reliability of the measurements. Walking speed, instructed vs. non-instructed activities and frequency of using voice commands were also measured by watching the recorded videos. System usability was reported using the SUS questionnaire and perceived cognitive load was calculated using the NASA-TLX.

During the analysis, however, researchers connected the questions from SUS to specific types of cognitive load (see Table 1) that they wanted to capture. We choose to utilize the SUS in this manner because cognitive load is an important variable. Hence, in order to increase the reliability of the results, researchers measured cognitive load both directly and indirectly. Table 3 shows an example of how the SUS questions were mapped to different types of cognitive load. For the qualitative analysis of the interviews, researchers transcribed each of the interviews, extracted the recurrent themes and grouped the comments by type. The emerging issues highlighted user preference for the interaction paradigms and the difficulties faced while using the voice and button commands.

7. RESULTS

7.1. Interaction times with aural flows

Figure 7a shows that the IT with the interface in the voice condition ($M = 84.5$ s, SE $= 9.93$) decreased compared with the time within the button condition ($M = 114.4$ s, SE $= 15.66$) ($t(19) = 1.835$, $p = 0.082$). However, this difference was not found to be statistically significant. In the voice condition, on average, participants spent 55.1 s out of 84.5 s interacting with the device using the buttons (Fig. 7a) and 29.4 s out of 84.5 s interacting with the device using the voice commands. On average, the participants spent 18 s looking at the voice commands posters on the wall. This activity was essential in regard to the users being able to interact with the voice commands, but it was not included in the interaction time measurement.

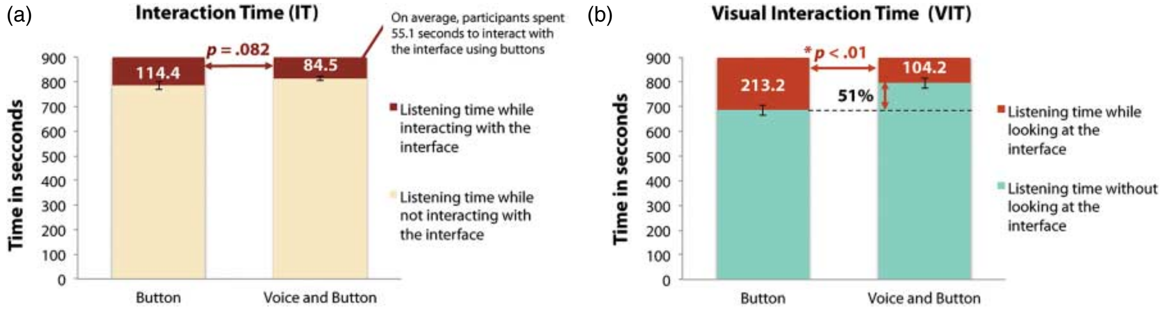

Figure 7. The voice commands (a) reduced the IT with respect to using buttons (with no statistical significance present), while the voice commands (b) also reduced the VIT with respect to using buttons (with statistical significance present).

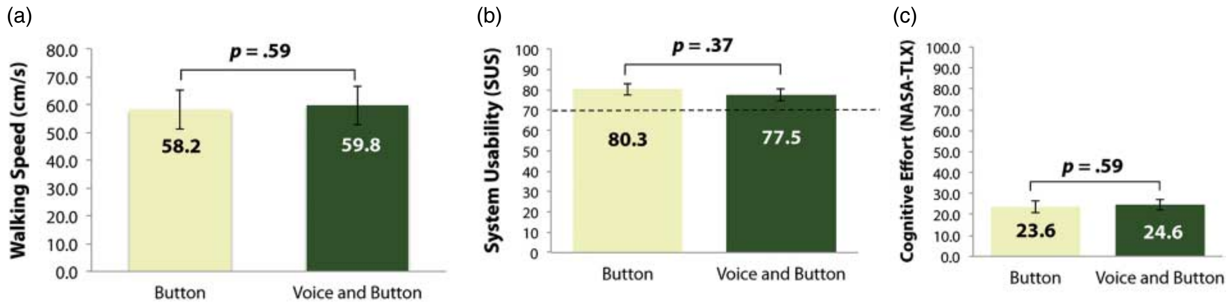

Figure 8. From left to right: no significant difference was found between the conditions for (a) the speed of walking, (b) system usability and (c) cognitive effort.

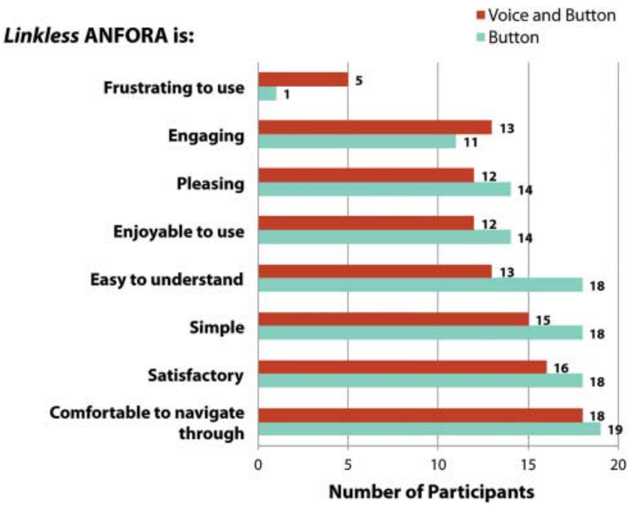

Figure 9. The participants who responded strongly agree/agree on every aspect of *Linkless* ANFORA experience.

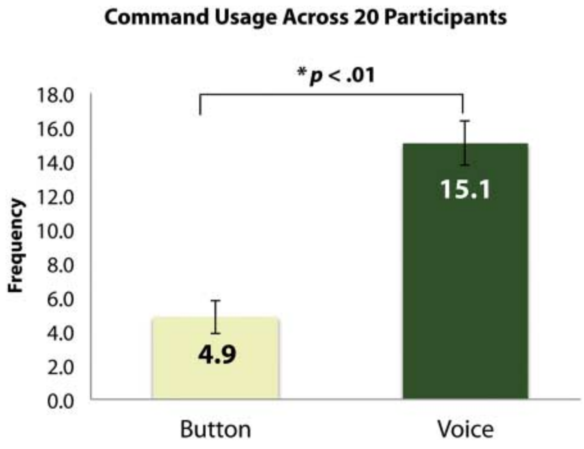

Figure 10. The participants used significantly more voice commands than button commands.

Two researchers measured the VIT. Based on the first researcher's measurements (Fig. 7b), the users spent 51% less time visually interacting with the interface in the voice condition ($M = 104.2$ s, SE $= 20.32$) than they did in the button condition ($M = 213.2$ s, SE $= 20.73$) ($t(19) = 4.289$, $p < 0.01$), which resulted in a statistically significant difference. Based on the second researcher's data, the users spent 40% less time visually interacting with the interface in the voice condition ($M = 121.0$ s, SE $= 22.65$) than they did in the button condition ($M = 202.4$ s, SE $= 19.36$) ($t(19) = 3.693$, $p < 0.01$), which is also a statistically significant difference. The inter-rater reliability correlations for the VIT by the two researchers were $r(19) = 0.057$, $p < 0.01$.

7.2. Walking speed, system usability and cognitive effort

The participants' walking speeds while listening to the aural flows appears to be similar in the button ($M = 58.2$, SE $=$ 7.03) and voice conditions ($M = 59.8$, SE $= 6.94$) ($t(19) = 0.536$, $p = 0.59$) (Fig. 8a). Based on the SUS questionnaire, the system's usability appears to be similar in the button ($M = 80.3$, SE $= 2.75$) and voice conditions ($M = 77.5$, SE $= 2.91$) ($t(19) = 0.921$, $p = 0.37$) (Fig. 8b) as well. Based on additional user experience questions, in general, the participants reported that controlling the aural flows was slightly more comfortable, enjoyable, satisfactory, pleasing, simple and easy to understand in the button condition than in the voice condition (Fig. 9). However, the participants found that their experience of using the voice commands was more *engaging* than using the buttons. *Engaging* was presented to the participants and measured as a polar opposite in the semantic differential scale to *boring*.

The users' cognitive efforts—as based on the NASA-TLX questionnaire—in the two interaction conditions are compared in Fig. 8c. The button condition ($M = 23.6$, SE $= 2.82$) yielded a similar cognitive effort as the voice condition ($M =$ 24.6, SE $= 2.74$) ($t(19) = 0.550$, $p = 0.59$). The users' cognitive efforts were also calculated indirectly using some of the questions in the SUS (Table 3). The results showed that cognitive load (indirectly calculated using SUS) was significantly correlated with cognitive load (directly calculated using the NASA-TLX) in both the button ($r(19) = 0.491$, $p < 0.05$) and voice conditions ($r(19) = 0.632$, $p < 0.01$).

7.3. Voice command usage

In the voice condition, the frequency of using the voice commands ($M = 15.1$, SE = 1.28) was significantly higher than the frequency of using the button commands ($M = 4.9$, SE = 0.97) ($t(19) = 5.293$, $p < 0.01$) (Fig. 10). The average amount of time spent using the voice commands was 14.7 s. The three sets of commands used most often were as follows: (i) the 'next/skip' command was used significantly more than all of the other commands (used 155 times; an average of eight times per participant; STD = 4.46); (ii) the category selection commands such as technology, world and health, were used the next most often (used 45 times; an average of two times per participant; STD = 1.92) and (iii) the 'forward' command was used to move from a story summary to a full version of the same story (used 41 times; an average of two times per participants; STD = 1.85). The 'anything else' and 'like this' commands were never used.

The results show that the participants used 'next' (124 times) more than the 'skip' command (19 times) to go to the next story and 'back' (four times) more than the 'previous' command (two times) to go back to the previous story. The participants used 'related' (nine times) more often than 'more' (five times) and 'tell me more' (two times) to go to a related story. They also used 'recent news' (five times) more than 'what's new' (two times) and 'start' (once) to begin listening to the aural flows playlist.

Additionally, the results show that one participant said, 'reverse' instead of 'back' or 'previous' and 'skip next' instead of 'skip' or 'next'. Another participant used 'related link' instead of 'related' and 11 participants said 'summary' for 'rewind' and 'full story' for 'forward'.

7.4. Instructed vs. non-instructed activities

In the voice condition, the participants performed significantly more non-instructed ($M = 26.7$, SE = 3.18) than instructed activities ($M = 19.9$, SE = 1.20) ($t(19) = 2.281$, $p < 0.05$) (Fig. 11). Examples of instructed activities were the use of voice or button commands to interact with the interface. Researchers also observed that the users looked at the list of voice commands or other artifacts available on the walls and glanced/read the news on the mobile interface, all of which are considered to be non-instructed activities. The participants either stopped to read the list of voice commands on the wall or glanced at it by turning their heads without stopping.

Similarly, in the button condition, the participants executed significantly more non-instructed ($M = 23.4$, SE = 3.07) than instructed activities ($M = 11.0$, SE = 1.42) ($t(19) = 3.701$, $p < 0.01$). Taken together, these sets of results show that the participants performed more non-instructed than instructed activities regardless of the modality condition.

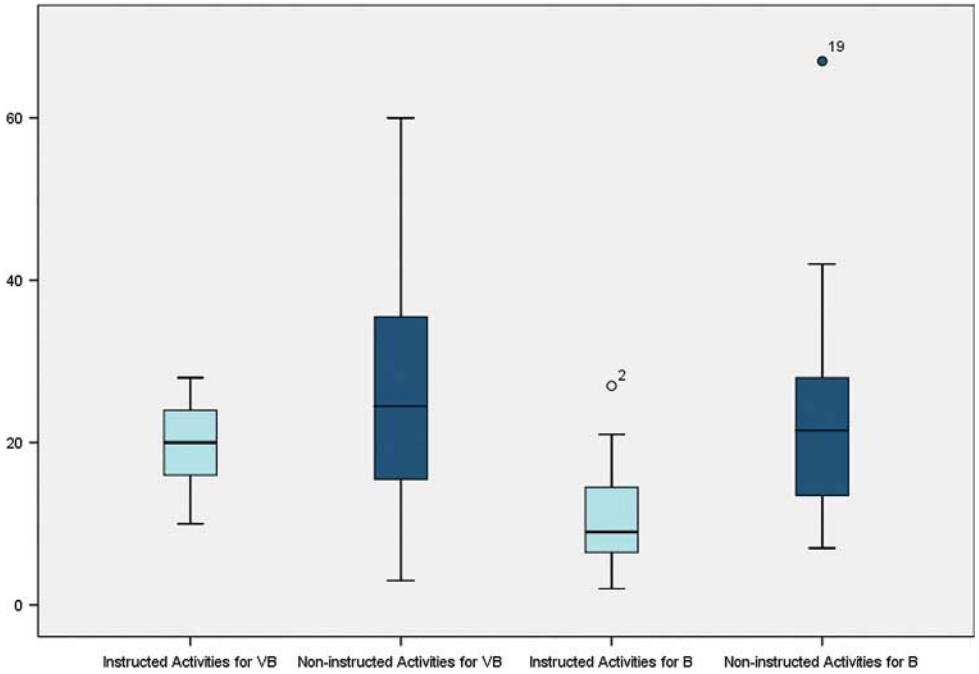

Figure 11. The participants performed significantly more non-instructed than instructed activities in both the voice and button conditions.

7.5. Interview results

7.5.1. *Self-reported experience*

The interviews confirmed the users' general satisfaction with *Linkless* ANFORA as all 20 participants reported that it was easy-to-use and convenient. In particular, three users said that they liked the wide range of categories and content taken from NPR. For example, one participant (P18) noted, 'I liked that you guys used NPR. I liked that there was lots of different news categories. It wasn't just world news. I usually like the special interest, health and science, so I liked that it had those categories available'.

Flexibility. Four of the participants reported that they liked the flexibility associated with not having to look at the screen. Furthermore, two participants reported that they liked moving from one category to another by using the voice commands. One user (P6) noted, 'I was able to walk and not get distracted. I did not have to stop walking in order to press buttons on the screen and I felt safer because I was aware of my surroundings'. Another user (P13) said, 'I enjoyed the flexibility of not looking at the screen and being able to control the news category you liked to listen to'.

Orientation. Fifteen users reported that they did not feel lost (in terms of where they were in the news content) while listening to the news story and felt that the orientation of information was good. Likewise, all of the participants recognized when a news story started or ended. One user (P12) noted, 'I did not get lost, but if I did, I could have looked at the phone to know where I was'. Another user (P18) said, 'I did not get lost in what category I was in or what story I was listening to'.

7.5.2. *Multitasking*

Eighteen of the 20 participants said that they could adequately monitor their surroundings while listening to the news. However, one participant had to stop walking while using the buttons and was not able to monitor his surroundings. One (P10) said, 'I wonder how different [my experience will be] when I am walking in a crowded area'. Three participants mentioned that the walking path was the same in both conditions and that there were not many obstacles, making it easy to monitor their surroundings.

7.5.3. *Combining the visual and voice commands*

The participants were asked whether they preferred to use the voice commands, button commands, or a combination of both types in order to interact with *Linkless* ANFORA. All of the participants used the voice commands, but three noted that they would prefer the button commands. They did not like the voice commands for three reasons. First, it was odd to speak aloud while alone in a public setting. Second, they had prior negative experiences with the use of voice commands, particularly when it came to voice recognition interfaces. For example, they had to speak the voice commands several times until the system recognized it. Third, the participants had to learn and memorize commands that were named differently than they were on the interface, which could be time-consuming. For example, the voice command to move to a full story while in the summary is 'forward' instead of 'full story' and the command to go back to a story summary is 'rewind' instead of 'summary'. The difference between the 'forward' and 'next' commands was also confusing because 'next' would go to the next story, while 'forward' would go to the full story within the same story.

Other participants, however, reported that they liked using the voice commands. Five of the participants noted that they did not have to stop walking to look down at the screen. Instead, they could do other things while using the voice commands, such as monitor their surroundings and look at posters on the walls. According to one participant (P6), 'I felt safer because I was aware of my surroundings'. Another participant (P14) said, 'The voice commands were quicker compared to the buttons'. One user (P9) noted, 'It was easy to go from category to category just by speaking into it without going back to the home screen, so it was convenient. It was just all on the fly'.

Seventeen of the 20 participants mentioned that they preferred to use a combination of the voice and button commands, but they had a variety of reasons. For example, one participant (P14) said, 'If voice does not work, I can still benefit from the buttons'. In other words, the buttons can be used as a backup navigation method if the voice commands are not working properly. Having button commands as a backup navigation method is a significant concept, as tone and tenor of voice, as well as voice quality and accents vary among individuals, making voice commands potentially less precise than button commands. The other main reason that the participants cited for preferring a combination of the voice and button commands relates to the contexts in which *Linkless* ANFORA might be used. For example, one user (P3) noted, 'I would use the voice, but, if I'm leaving class, I would click on a story and go walking from there and then use the voice'. Another user (P8) said, 'If I am at a noisy place, like a subway, I would use the button. If I am walking in a quiet place, I would use the voice. I think it depends on the environment'. A third participant (P15) reported, 'If you come to talk to somebody, you would want to pause it with your finger, but if you are just walking around, you could just tell it what to do and do it'. Another participant (P19) noted, 'Like if I were crossing a busy street or riding my bike, I would definitely prefer to use the voice than the button'. Finally, another participant (P3) said, 'If I were sitting somewhere, like a coffee shop or something, I might use the button because I'm not moving, but, if I'm walking, then I would use the voice'.

7.5.4. *Other contexts for voice-controlled aural flows*

The participants suggested other contexts in which *Linkless* ANFORA could be useful. Three participants noted they would use *Linkless* ANFORA while driving, when their eyes and hands are busy. One participant (P5) noted, ‘This app is more appropriate for a driving context than only a walking context because, while walking or sitting down, I prefer to read it, which is faster than just listening to the content’. Another participant (P18) said, ‘If I was driving, probably, I would use the voice commands because I did not have to look at my phone screen’. Several other potential contexts of use included: while on the way to work/class, outside a classroom, while sitting in a coffee shop, on the bus, while exercising, while riding a bike and while working around the house.

7.5.5. *Limitations and improvements suggested by the users*

The users also provided suggestions on how to optimize the usability of *Linkless* ANFORA.

Repetition of the orientation information. Seven of the participants were frustrated with the repetition of the orientation information. For example, each time a new story began, *Linkless* ANFORA included audio that reported the story number, category and news headline. Two of the users said that the story number was of little interest. One participant (P8) added, ‘If I was listening to a research paper, maybe it would be necessary, but not for a news story’.

Confusing category transition. Additionally, four participants said that the transition between two categories of news was not clear. One participant (P4) said, ‘I guess I didn’t understand when it switched from one category to another and I was like, oh wait, I’m not in Science anymore. I’m in Economy or whatever it was’. Two users wanted some indication of when a story was finished, such as audio stating ‘end of story’.

8. DISCUSSION

8.1. Voice commands and eyes-free browsing

This study provides some empirical support to H1: using voice commands, instead of button commands, requires less visual interaction with the device. On average, compared with the button condition, the voice condition saved about 40–51% of the time in visual interaction with the device. Therefore, combining voice commands with aural flows and buttons reduced visual interaction with the screen when compared with using button commands with aural flows. Likewise, this result validates the primary value of extending the interaction with aural flows through voice commands.

In the voice condition, researchers also observed that the participants looked at the screen not only when they used the buttons, but also, when they used voice commands for different reasons. For example, users were not yet familiar with the interaction modality or they checked to see if the system did what they asked it to do.

This study also confirms the findings from another recent study (Brumby *et al.*, 2011) on the use of mobile devices during secondary tasks. This study indicated that, although audio-based interfaces are slower to use, they are less distracting than visual interfaces. However, an important question is still unanswered: to what extent do combinations of aural flows with voice commands support eyes-free browsing while driving a car? Some of the participants noted that they would prefer to use *Linkless* ANFORA while driving. Furthermore, a recent study (Strayer *et al.*, 2013) reported that using *speech-to-text* systems for sending and receiving text or email messages in the car is risky because too many and continuous voice interactions can also cause higher levels of cognitive distraction.

8.2. Similar system usability, users’ cognitive efforts and walking speed

Both the button and voice conditions yielded a similar system usability and cognitive effort. Therefore, H2 was not confirmed. This similarity in the two conditions is, most probably, because aural flows already improve system usability and reduce cognitive effort so significantly—with respect to visually interacting with content-intensive websites on a mobile device—that merely changing the interaction style has no additional effect. Figure8b shows that the system usability for the button and voice conditions is 80.3 and 77.5, respectively, which is close to an excellent rating (Bangor *et al.*, 2009). Cognitive effort for both the button and voice conditions is 23.6 and 24.6, respectively, which is a low cognitive effort score (Fig. 8c). Overall, the results show that aural flows yield a very good user experience in both the button and voice conditions. Additionally, the low-cognitive effort engendered by aural flows regardless of the interaction modality allowed the participants to do more non-instructed than instructed activities. This finding is because the users spent 13 and 9% of the time interacting with the aural flows (i.e. instructed activities) in the button and voice conditions, respectively (Fig. 7a), and engaged in non-instructed activities during the remaining time. For example, the participants looked at the posters on the wall or glanced at the mobile visual interface, which were not instructed to them as part of the task. This result is mainly relevant for multitasking experiences while on the go because attention to the mobile device and the risk of having an accident are minimized.

Similarly, the participants’ walking speeds were similar in both the button and voice conditions. This result shows that the interaction modality did not have an effect on their walking speeds. As discussed previously, the voice commands significantly reduced the amount of time necessary to interact visually with the device. However, participants’ walking speeds show that not focusing on the device does not necessary make the users walk faster. This finding could be because the participants had to walk the same path in an indoor environment repeatedly. Figure 8a shows that the walking speeds for the button and voice conditions were 58.2 and 59.8, respectively, which is far below the average walking speed for adults (140 cm/s) in the age range of 20–30 years old (Bohannon, 1997). This finding could be because the participants had 15 min for the task and were not in a rush to finish the path or reach a particular destination. Researchers realize that the participants walked in an environment where there were no dynamic obstacles and the static obstacles were always present in the same position. Therefore, it is difficult to reach an ultimate conclusion about the real effects of distracted walking because of the nature of the environment.

8.3. Experience with voice commands

The analysis of the recorded videos revealed that the participants used the voice commands significantly more than the button commands to interact with the aural flows. However, the participants' answers to the interview questions revealed that 85% of them chose a combination of both the voice and button commands by which to interact with the aural flows. One of the reasons was because some of the users reported poor previous experiences with voice commands. The main reason for their criticism was related to their perception that the tone and tenor of their voices, as well as voice quality and individual accents, affects the systems' abilities to understand them.

8.3.1. Contradictory user experiences with navigation modalities

A few possible reasons exist as to why the user experience was slightly less favourable in the voice condition than in the button condition (Fig. 9). The Wizard-of-Oz technique introduced a longer pause between actions for when a voice command is used compared with when a button is clicked. Additionally, it may be difficult for users to quickly learn the voice commands and differentiate them from one another (e.g. 'next' and 'forward'). For example, in response to the statement, 'I found this application [voice condition] very cumbersome/awkward to use', a participant rated the application as a five on a scale of one to seven (one = strongly disagree, seven = strongly agree). This same participant also rated 'I needed to learn a lot of things before I could get going with this application [voice condition]' with a 7.

One participant reported that using the button commands was less satisfactory and less enjoyable, but also simple, easy to understand and engaging. This discrepancy between user experience attributes could exist because, although the button interface is easy-to-use, the user had to stop walking to click the button. Three of the participants reported that using the voice commands was more frustrating than the button commands, but that the voice commands were simple, pleasing and enjoyable. The reason for this apparent contradiction could be because the user was frustrated with the repetition of orientation information, although the interface was easy-to-use (see interview results, Section 7.5.5).

Our participants rated their user experiences slightly less favourably for the voice condition than for the button condition. However, they enjoyed using the voice commands slightly more than the button commands. One possible reason for this finding is that users do not have to look at the screen to interact with the device and can, instead, enjoy listening to the news while navigating with the voice commands.

8.4. Consistency between the aural and visual interfaces

This study reinforces the importance of the principle of 'consistency' between the voice commands and the written labels on the buttons. For example, the *Linkless* ANFORA interface includes two buttons, 'summary' and 'full story', but users must say 'rewind' and 'forward' to move between summaries and full stories. The design included very simple playlist-like commands (e.g. 'forward' and 'rewind'), which were applicable to the playlist metaphor. On the other hand, to control the visual condition, researchers used a tab structure that includes 'summary' and 'full story', which represents different sections of the news (i.e. world news vs. local news). At times, the users said that 'summary' or 'full story' instead of 'rewind' and 'forward'. Users reported that the labels on the buttons were not consistent with the voice commands, which caused confusion. While the common principle of consistency (Nielsen and Molich, 1990) usually applies to visual interfaces, studying semi-aural interfaces suggests the importance of examining issues related to *cross-modal consistency* (Evans and Treisman, 2010; Spence, 2011). For example, how consistent do aural and visual interfaces need to be? Does the consistency contribute to having natural interactions with the semi-aural interfaces?

8.5. Limitations of the study

One limitation of the experimental design is that the users had to walk in a controlled lab environment in order to avoid putting them in danger. Additionally, the simplicity of the walking path and not having natural distractors in the environment could have affected the cognitive load measurements and the ecological validity of the experiment. The interview findings suggest that additional studies in which participants are put in new scenarios might be valuable in the future. The second limitation is that the users had to walk the same path with the presence of static obstacles and not dynamic obstacles for both conditions. Familiarization to the path, however, is partially lessened by the counterbalancing of two conditions.

The third limitation is that the participants had to learn the voice commands and the *Linkless* ANFORA interface in a short period. Therefore, they were provided with lists of voice commands on all of the walls surrounding the path in the event that they could not remember them. Thus, learnability was factored out of the cognitive load measurement. The fourth limitation is that the voice commands were not fully implemented in the system. Instead, the Wizard-of-Oz approach was used in order to simulate voice interaction. The decision to use the Wizard-of-Oz approach was made in order to minimize the chances that many different speech patterns and/or accents would result in a high number of system errors, which would interfere with our ability to effectively measure the linkless user experience.

The fifth limitation is that researchers did not accurately capture whether the participants preferred buttons for certain types of interactions (e.g. changing the news story or the news category), although they did observe patterns of preferences while recording the participants' videos. For example, to go to the next or previous news story, sometimes the participants preferred the buttons. However, in order to change the news category, the participants preferred the voice commands instead of going through the menu selection using the buttons. The sixth limitation is that the participants were not restricted to listening to a certain number of news stories, but were simply told to have a minimum of eight news stories. Therefore, all participants did not have the equal number of interactions with aural flows, which might have affected on some of the outcome variables.

9. CONCLUSIONS AND FUTURE WORK

This study is the first study to demonstrate the properties of aural flows in the context of how to interact with them. Aural and semi-aural interfaces have the potential to augment the users' abilities to navigate any mobile applications more safely and with fewer visual distractions from their surroundings. This work compared navigating aural flows using buttons vs. voice plus buttons. The results suggest that voice commands in combination with aural flows and buttons reduce visual interaction time with the device by one-half compared with using button commands in combination with aural flows while walking. The results of the two conditions were also similar in terms of walking speed, system usability and cognitive effort. Overall, the low cognitive effort engendered by aural flows regardless of the interaction modality allowed the participants to do more non-instructed than instructed activities. We must consider that a noiseless environment and no errors in voice recognition were included as assumptions to reach the above conclusion. Hence, the ecological validity of the study is limited. In future studies, we will add errors in the Wizard-of-Oz approach to better simulate a more realistic scenario. Moreover, we will look into how users' familiarity with and trusting the application will have an effect on the visual interaction while using voice commands.

Several of our participants suggested that they would like to use *Linkless* ANFORA while driving a car. A recent study (Strayer *et al.*, 2013) suggested that using *speech-to-text* systems in the car is risky because too many voice interactions still tax our attention bandwidth. Researchers suggest that by using a small vocabulary of voice commands (Feng and Sears, 2009), which are short and easy to remember (Bradford, 1995), the cognitive effort required to use an interactive system is still minimal and would not distract too much users from effectively monitoring their environment. Based on our findings, we argue that this situation applies to *Linkless* ANFORA as well. In our current work, we are pursuing ways by which to use aural flows to mitigate the distraction by reducing both the visual and vocal interactions in a driving scenario.

FUNDING

This research is based upon work supported by the National Science Foundation under Grant Number IIS-1018054. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research is also supported by the Indiana University-Purdue University Indianapolis (IUPUI) IUPUI Office of the Vice-Chancellor for Research (OVCR) under the Funding Opportunities for Research Commercialization and Economic Success (FORCES).

ACKNOWLEDGEMENTS

We thank all study participants. We also thank Dr Mark Pfaff, Dr Steven Voida, Dr Michael Justiss, Prathik Gadde and Dr Tao Yang for their comments to improve the manuscript. We also thank Debaleena Chattopadhyay for taking pictures of the experimental set up.

REFERENCES

Android Google Voice. (2015) https://play.google.com/store/apps/details?id=com.google.android.apps.googlevoice&hl=en (accessed July 1, 2015).

Anhalt, J., Smailagic, A., Siewiorek, D.P., Gemperle, F., Salber, D., Weber, S., Beck, J. and Jennings, J. (2001) Toward context-aware computing: experiences and lessons. IEE Intell. Syst., 16, 38–46.

Apple Siri. http://www.apple.com/iphone/features/siri.html (accessed July 1, 2015).

Ashok, V., Borodin, Y., Stoyanchev, S., Puzis, Y. and Ramakrishnan, I.V. (2014) Wizard-of-Oz Evaluation of Speech-driven Web Browsing Interface for People with Vision Impairments. Proc. 11th Web for All Conf., p. 12. ACM.

Bangor, A., Kortum, P. and Miller, J. (2009) Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud., 4, 114–123.

Bernsen, N.O. and Dybkjaer, L. (2001) Exploring natural interaction in the car. CLASS Workshop on NIR, 2(1).

BII Report. http://www.businessinsider.com/bii-report-how-content-is-being-consumed-on-mobile-devices-2012-9#ixzz2dwO5aawP (accessed July 1, 2015).

Bohannon, R.W. (1997) Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. Age Ageing, 26, 15–19.

Bolchini, D. and Ghahari, R.R. (2013) U.S. Patent Application 14/024,612.

Bolchini, D. and Paolini, P. (2006) Interactive dialogue model: a design technique for multichannel applications. IEEE Trans. Multimedia, 8, 529–541.

Borodin, Y. et al. (2014) Listen to Everything You Want to Read with Capti Narrator. Proc. 11th Web for All Conf., p. 33. ACM.

Bradford, J.H. (1995) The human factors of speech-based interfaces: a research agenda. SIGCHI Bull., 27, 61–67.

Bragdon, A., Nelson, E., Li, Y. and Hinckley, K. (2011) Experimental Analysis of Touch-screen Gesture Designs in Mobile Environments. Proc. SIGCHI Conf. Human Factors in Computing Systems, pp. 403–412. ACM.

Brooke, J. (1996) SUS-A quick and dirty usability scale. Usability Eval. Ind., 189, 194.

Brumby, D.P., Davies, S.C., Janssen, C.P. and Grace, J.J. (2011) Fast or Safe? How Performance Objectives Determine Modality Output Choices While Interacting on the Move. Proc. CHI, pp. 473–482. ACM.

Chandler, P. and Sweller, J. (1991) Cognitive load theory and the format of instruction. Cogn. Instr., 8, 293–332.

Christian, K., Kules, B., Shneiderman, B. and Youssef, A. (2000) A Comparison of Voice Controlled and Mouse Controlled Web Browsing. Proc. ASSETS, pp. 72–79. ACM.

Dahlbäck, N., Jönsson, A. and Ahrenberg, L. (1993) Wizard of Oz studies—why and how. Knowl.-Based Syst., 6, 258–266.

Evans, K.K. and Treisman, A. (2010) Natural cross-modal mappings between visual and auditory features. J. Vis., 10, 6.

Feng, J. and Sears, A. (2009) Speech Input to Support Universal Access. The Universal Access Handbook, pp. 1–12. CRC Press.

Garzotto, F., Paolini, P. and Schwabe, D. (1993) HDM—a model-based approach to hypertext application design. TOIS, 11, 1–26.

Ghahari, R.R. and Bolchini, D. (2011) ANFORA: Investigating Aural Navigation Flows on Rich Architectures. 2011 13th IEEE Int. Symp. Web Systems Evolution (WSE), pp. 27–32. IEEE.

Goose, S. and Djennane, S. (2002) WIRE3: driving around the information super-highway. Pers. Ubiquitous Comput., 6, 164–175.

Green, P. and Wei-Haas, L. (1985) The Rapid Development of User Interfaces: Experience with the Wizard of Oz Method. Proc. Human Factors and Ergonomics Society Annual Meeting, Vol. 29, pp. 470–474. SAGE Publications.

Gupta, G., Raman, S.S., Nichols, M., Reddy, H. and Annamalai, N. (2005) DAWN: Dynamic Aural Web Navigation. Proc. HCI, Las Vegas.

Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX: results of empirical and theoretical research. Hum. Mental Workload, 1, 139–183.

How Content Is Being Consumed On Mobile Devices. http://www.businessinsider.com/bii-report-how-content-is-being-consumed-on-mobile-devices-2012–9#ixzz2dwO5aawP (accessed July 1, 2015).

Jain, A. and Gupta, G. (2007) VoxBoox: A System for Automatic Generation of Interactive Talking Books. Proc. UAHCI, pp. 329–338. Springer, Berlin, Heidelberg.

Klemmer, S.R., Sinha, A.K., Chen, J., Landay, J.A., Aboobaker, N. and Wang, A. (2000) Suede: A Wizard of Oz Prototyping Tool for Speech User Interfaces. Proc. 13th Annual ACM Symp. User Interface Software and Technology, pp. 1–10. ACM.

Lemmelä, S., Vetek, A., Mäkelä, K. and Trendafilov, D. (2008) Designing and Evaluating Multimodal Interaction for Mobile Contexts. Proc. 10th Int. Conf. Multimodal Interfaces, pp. 265–272. ACM.

Metz, B. and Krueger, H.P. (2010) Measuring visual distraction in driving: the potential of head movement analysis. Intell. Transp. Syst., IET, 4, 289–297.

Narayanan, S. and Potamianos, A. (2002) Creating conversational interfaces for children. IEEE Trans. Speech Audio Process., 10, 65–78.

Nielsen, J. and Molich, R. (1990) Heuristic Evaluation of User Interfaces. Proc. CHI, pp. 249–256. ACM.

Oviatt, S.L., Cohen, P.R., Fong, M. and Frank, M. (1992) A Rapid Semi-automatic Simulation Technique for Investigating Interactive Speech and Handwriting. ICSLP.

Rohani, Ghahari and Bolchini, D. (2013) Mobile web browsing with aural flows: an exploratory study. Int. J. Hum.—Comput. Interact., 29, 717–742.

Salber, D. and Coutaz, J. (1993) Applying the Wizard of Oz Technique to the Study of Multimodal Systems. Hum.—Comput. Interact., pp. 219–230. Springer, Berlin, Heidelberg.

Sawhney, N. and Schmandt, C. (2000) Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. TOCHI, 7, 353–383.

Serrano, M., Lecolinet, E. and Guiard, Y. (2013) Bezel-Tap Gestures: Quick Activation of Commands from Sleep Mode on Tablets. Proc. CHI, pp. 3027–3036. ACM.

Sinha, A.K., Klemmer, S.R. and Landay, J.A. (2002) Embarking on spoken-language NL interface design. IJST, 5, 159–169.

Song, M.G., Tariquzzaman, M., Kim, J.Y., Hwang, S.T. and Choi, S.H. (2012) A robust and real-time visual speech recognition for smartphone application. IJICIC, 8, 2837–2853.

Spence, C. (2011) Crossmodal correspondences: a tutorial review. Attention Percept. Psychophys., 73, 971–995.

Stavrinos, D., Byington, K.W. and Schwebel, D.C. (2011) Distracted walking: cell phones increase injury risk for college pedestrians. J. Saf. Res., 42, 101–107.

Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J., Medeiros-Ward, N. and Biondi, F. (2013) Measuring cognitive distraction in the automobile.

Sweller, J. (1988) Cognitive load during problem solving: effects on learning. Cogn. Sci., 12, 257–285.

Sweller, J. and Chandler, P. (1994) Why some material is difficult to learn. Cogn. Instr., 12, 185–233.

Sweller, J., Van Merrienboer, J.J. and Paas, F.G. (1998) Cognitive architecture and instructional design. Educ. Psychol. Rev., 10, 251–296.

Tang, Y., Wang, D., Bai, J., Zhu, X. and Li, M. (2013) Information distance between what I said and what it heard. CACM, 56, 70–77.

Tsimhoni, O., Smith, D. and Green, P. (2004) Address entry while driving: speech recognition versus a touch-screen keyboard. Hum. Factors, 46, 600–610.

Vo, M.T. and Wood, C. (1996) Building an Application Framework for Speech and Pen Input Integration in Multimodal Learning Interfaces. Proceedings of 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996. ICASSP-96., Vol. 6, pp. 3545–3548. IEEE.

VoiceDream. http://www.voicedream.com/ (accessed July 1, 2015).

Yang, T., Ferati, M., Liu, Y., Rohani Ghahari, R. and Bolchini, D. (2012) Aural Browsing On-the-Go: Listening-Based Back Navigation in Large Web Architectures. Proc. CHI, pp. 277–286.

Young, K., Regan, M. and Hammer, M. (2007) Driver distraction: a review of the literature. Distracted Driving. Sydney, NSW: Australasian College of Road Safety, pp. 379–405.

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R. and Baudisch, P. (2007) Earpod: Eyes-free Menu Selection using Touch Input and Reactive Audio Feedback. Proc. CHI, pp. 1395–1404.